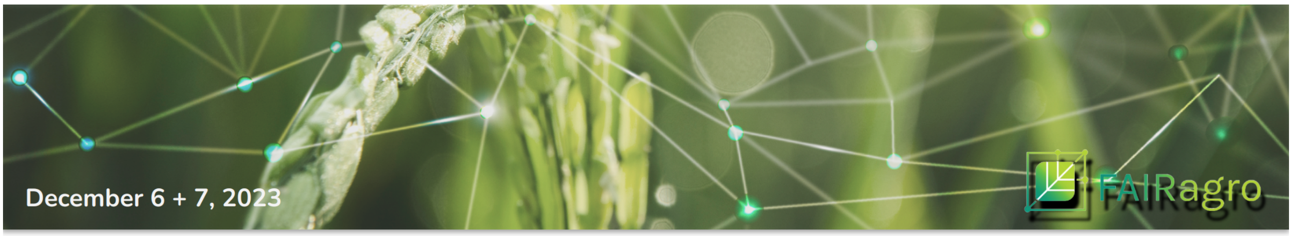

Report on the FAIRagro Workshop on Data Quality for Data Analytics in Agrosystem Science (DQ4DA)

edited by *Sven Gedicke**, *Shiyaza Risvi**, and *Jan-Henrik Haunert**

*Rheinische Friedrich-Wilhelms-Universität Bonn, Institute of Geodesy and Geoinformation, Geoinformation Group

1. INTRODUCTION

In the domain of agricultural systems, the fusion of mobile sensing, machine learning, and plant phenotyping unlocks far-reaching opportunities for sustainable crop production. With the aim of exploiting these potentials in the best possible way, a quality-assured research data management (RDM) for generating, publishing, and accessing research data is currently being developed in the initiative of FAIRagro[1]. To make such a data infrastructure usable across different disciplines, it is crucial to identify aggregation, processing, and data quality levels that are optimal for multi-modal data analytics.

Addressing this aspiration, we organized the *Workshop on Data Quality for Data Analytics in Agrosystem Science* (DQ4DA) at the University of Bonn on December 6 and 7, 2023. Our aim was to stimulate an interdisciplinary exchange on best practices to ensure data quality through documentation, standardization, and automation. This includes methods for automatically assessing data quality as well as methods for pre-processing the data in order to ensure the required data quality for subsequent data analysis.

Over the course of two half days, we gathered expertise, ideas, and needs from different perspectives of both data users and data providers. A total of 19 participants with different research focuses from various institutions throughout Germany and even from international affiliations took part in the workshop. Presentations by the participants offered the opportunity to provide insights into ongoing research with a particular focus on

(1) collected in-field data and its necessary (spatial and temporal) characteristics and quality standards and

(2) data processing as an intermediate step between data acquisition and further use (e.g., in a neural network).

One keynote speech and two impulse talks stimulated lively discussions in subsequent break-out sessions. Participants discussed commonalities in the processing and quality assurance of data across their varying domains. To allow the quality of data to be traceable and transparent, the round tables facilitated exchange to establish requirements for the documentation of data quality.

We consider the insights gained from the workshop to be particularly valuable for the development of guidelines that help data providers resolve ambiguities regarding the required quality level of the data they feed into the data infrastructure.

In the following, we will first provide a description of the workshop agenda in Section 2. Furthermore, we will share a collection of abstracts from presentations given in Section 3. Subsequently, Section 4 summarizes the main findings from the discussions that took place. A final conclusion and outlook for future research follows in Section 5.

2. ORGANIZATION AND PROGRAMME

The workshop was organized by Sven Gedicke, Shiyaza Risvi, and Jan-Henrik Haunert as part of the FAIRagro initiative. Invitations were primarily sent via the various FAIRagro channels (e.g., newsletter, mailing list), but also by inviting colleagues from related domains of agrosystem science and data management with the request to spread the word. We have encouraged the submission of short abstracts on the basis of which presentations were selected by the following committee.

- Jan-Henrik Haunert, *University of Bonn*
- Uwe Rascher, *Forschungszentrum Jülich*
- Markus Möller, *Julius Kühn-Institute in Braunschweig*
- Carsten Hoffmann, *Leibniz Centre for Agricultural Landscape Research (ZALF)*
- Juliane Fluck, *Information Centre for Life Sciences in Bonn*

[1] `https://fairagro.net/`

A total of 19 participants registered, six of whom submitted an abstract, all of which were accepted for presentation. The accepted abstracts are listed in Section 3. The workshop took place over two half days on December 6 and 7, 2023 at the University of Bonn[2]. In the following, we briefly describe the agenda of the two days.

Day 1 After a welcoming address by Jan-Henrik Haunert, in which the scope and agenda of the workshop were explained to the participants, a first abstract session with three presentations (20 minutes each) held by Jannes Uhlott, Florian Beyer, and Lucia Vedder took place. Subsequently, Shiyaza Risvi gave an impulse talk on the potential of automation in data quality assessment. Taking the talk as impetus, a break-out session followed on the topic of *Important Metrics of Data Quality for Data Processing*. A summary of the break-out session's outcomes are described in Section 4. Towards the end of the first workshop day, Prof. Dr. Thomas Döring gave an inspiring invited talk on data quality in the context of biodiversity assessment. The first day ended with an optional joint dinner in the evening.

Day 2 Day two started with the second abstract session with two presentations by Susanne Lachmuth and Stefan Kühnel. Sven Gedicke then gave an impulse talk on visualizing data quality, which introduced the second break-out session on the topic of *User Expectations for an Interactive Quality Assessment Tool*. This was followed by a third and final abstract session with the two presentations of Iris Vogeler and Johannes Timaeus. After a short closing speech by Jan-Henrik Haunert, the official part of the workshop ended and there was further opportunity for bilateral discussions.

3. ABSTRACTS FROM THE WORKSHOP

In the following, we list the six abstracts that were submitted by participants and accepted by the committee for presentation. Please note that we present the abstracts as they were submitted by the authors.

Data Fitness for Use in Practice: Determination and variation of data quality in the derivation of biodiversity indicators from satellite-based crop type classifications

Jannes Uhlott[1], *Florian Beyer*[1], and *Markus Möller*[1]

[1]Julius Kühn-Institute, Institute for Crop and Soil Science

For several years, the FAIR principles have been an established guideline to ensure sustainable and efficient research data management. Data should be organized, structured, and documented in a way that makes them findable and accessible. Additionally, data should be interoperable with other data sources and reusable for future applications. Reusability should not only cover data storage but also enable effective use of the data by other users. The data should be fit for use, meaning they should be suitable and reliable for their intended purpose. Basic criteria such as accuracy, completeness, relevance, and consistency contribute to a certain level of data quality.
The overall quality of a dataset is influenced by every phase of the data lifecycle, emphasizing the importance of documenting the entire data history. This documentation forms the basis for tracking data and evaluating their adaptability to various applications. This is necessary because the quality of a dataset significantly depends on the specific application area, such as different steps of product generation. The adaptability of data to specific purposes is referred to as Data Fitness for Use, describing the ability of data to meet the specific requirements and expectations of a particular purpose. In the derivation of biodiversity indicators from satellite-based crop type classifications, the Data Fitness for Use approach has been actively implemented. Various measures of data quality were used to assess data quality during different phases of product generation in which semantic and spatial aggregations lead to changes in data quality. This confirms the Data Fitness for Use approach, demonstrating its ability to differentiate data quality in various application areas.
Another example shows how local accuracy metrics can be used for quantifying the data fitness for use uncertainty. A prerequisite for the derivation of agricultural weather indices (AWI) is the definition of phenological time windows within which statistical operations (e.g. number of days an agricultural weather parameter such as precipitation exceeds a threshold value) are carried out. Phenological time windows are in turn based on modelled occurrence dates of phenological events, which are associated with uncertainties, and which can be used for the derivation of AWI variants. In our presentation, we introduce our Data Fitness for Use examples and demonstrating how data quality can be determined during different process steps and how it changes. Additionally, we plan to present an initial draft on how these quality changes can be mapped and documented.

Data quality along the data life cycle of an earth observation based in-season crop type classification

Florian Beyer[2], *Jannes Uhlott*[2], and *Markus Möller*[2]

[2]Julius Kühn-Institute, Institute for Crop and Soil Science

In-season national-scale crop type classification based on earth observation data requires substantial amount of data and processing capacity on the one hand. On the other hand, knowledge about data provenance, quality and usability is crucial for the data life cycle documentation. Using an in-season classification approach as example, the presentation will show the key points concerning the data life cycle under FAIR principles and data quality.

before classification: Starting with the initial data provided by third parties to feed a machine learning model, pre-processed image data from ESAs Copernicus Sentinel-2 satellites (S2) are needed. The pre-processing procedure is scientifically published and extensive metadata (machine-readable) are available. Secondly, training data originates from the Integrated Administration and Control System (IACS), the European agricultural subsidy program, to get agricultural parcels and their cultivated crops. Strict quality standards are applied to these data sets, but some of these are not always fully comprehensible.

classification: During the training of the machine learning (ML) algorithms, various quality parameters are created, which vary from classifier to classifier, and there are also standard measures to assess the quality of the classification results. During the training of a ML algorithm (here: Random Forest) the out-of-bag-error (OOB) and importance metrics (Imp) are calculated, which allows an initial quality assessment of the trained model and the individual input variables. In addition, it is common practice in such classification procedures to prepare an independent test set that is unseen during the training. It is used to calculate the accuracy measures of the classification. The over-all accuracy (OA) is a global measure to evaluate the accuracy averaged over all classes. There are also various measures to evaluate the classification accuracy of the individual classes. These are the precision or user accuracy (from the point of view of the product user) and

[2]Google Maps

the recall or producer accuracy (from the point of view of the product maker). The F1-score is a measure, which harmonizes both class-specific values.

after classification: Nowadays, classification products are often published, and sometimes the source code is also available on github. This usually happens in parallel with the publication of a scientific manuscript. In such manuscripts, some of the quality measures mentioned above are also published (usually OA, often F1-score, rarely OOB, and quality measures of the source data, here S2 and IACS). These quality measures are thus indirectly human-readable, but not directly machine-readable. Taking the FAIR principles into account, however, it would be important to publish such quality measures together with the classification products in a machine-readable extra file of the metadata (standard in online repositories) or even as tiff tags within the classification product itself.

recap: Some of the mentioned quality metrics are already standardized, some are not. How can the metrics categorized and documented in a structured and pragmatic manner? In addition, the interpretability of quality metrics depends on the applied validation schema. How can such background information be provided (in a machine-readable way)?

Assessing the quality of DNA sequencing data

Lucia Vedder[3]

[3]University of Bonn, Institute of Crop Science and Resource Conservation

Within the broad field of crop science, there are many different aspects of research, such as cultivation, fertilization, crop protection and breeding. But crop genetics also plays an important role. In general, all genetic analyses require some type of genetic data, most commonly DNA sequencing data, which is becoming increasingly available as the cost of DNA sequencing has decreased dramatically over the past decade due to next-generation sequencing (NGS) methods. However, NGS is more prone to sequencing errors than other established long-sequence methods. This raises the question of how to assess the quality of the raw sequencing data and how to find a trade-off between the rate of (possible) sequencing errors and the highest possible sequencing coverage. Going one step further in the default analysis workflow of a genetic experiment, the NGS sequencing reads will most likely be mapped to a reference genome. This allows for further downstream analyses such as SNP detection (as part of the analysis of genetic variation of different crop lines and/or their hybrids) or differential expression analysis (for example, to analyze crop responses to stress). But once again, the quality of the sequence is critical: this time, the quality of the reference genome. In addition to these data quality issues, the quality of metadata is also an important topic: What metadata is needed to adequately describe DNA sequencing data so that the sequencing effort can be reused to analyze other questions?

DQ-Kit – a tool fostering data quality to support both data authors and reusers of the BonaRes Repository

Susanne Lachmuth[4a], *Carsten Hoffmann*[4a], *Thomas Kühnert*[4a], *Stephan Lesch*[4b], *Sebastian Rick*[4b], *Viet Hoang Nguyen*[4a], *Xenia Specka*[4a], and *Nikolai Svoboda*[4a]

[4a]Leibniz Centre for Agricultural Landscape Research (ZALF)

[4b]Senckenberg - Leibniz Institution for Biodiversity and Earth System Research

Well-curated research data repositories play a vital role in facilitating the discovery, access, integration, and analysis of scientific data, maximizing research impacts, and ensuring the accuracy and reliability of data-driven technologies such as modeling and artificial intelligence. This objective aligns with the growing adoption of the FAIR data principles by funding agencies and publishers, promoting data that is Findable, Accessible, Interoperable, and Reusable. Here, we introduce and would like to discuss DQ-Kit, a web application currently in development as a tool amending the data publishing process in the BonaRes Repository. DQ-Kit is a tool for data authors (providers), data reusers, and anyone seeking to perform quality checks on their research data. DQ-Kit will offer automated guidance on elements of the data that require review and confirmation. The checks conducted by DQ-Kit will encompass four main categories. First, formal criteria such as semantic and structural consistency, atomization of data, and other formatting issues will be addressed. Second, DQ-Kit will provide a comprehensive and well-structured summary of variables, their properties, and summary statistics. Third, DQ-Kit will allow for exploration of relationships among variables, as well as temporal and spatial patterns, and patterns of missingness. Lastly, we are planning to implement data plausibility checks that flag variables containing theoretically “impossible” values and values that seem empirically implausible based on available datasets, scientific literature, and expert knowledge. Initially, this functionality may be limited to soil research data, where our team possesses the necessary expertise. However, the determination of plausibility for variables flagged as containing “implausible values” rests with the data author, who has the ultimate authority in assessing their validity. Importantly, we embrace the concept of “Fit for use” rather than a binary “acceptable vs. unacceptable” approach, focusing on the suitability of data for specific purposes while acknowledging the efforts of data providers and amplifying their impact. Eventually, the BonaRes Repository metadata will be enriched with the DQ-Kit results, enabling seamless quality control and facilitating the comparison of different datasets. In summary, these checks ensure the integrity and reliability of scientific data in support of research endeavors.

Streamlining pest and disease data to advance integrated pest management – A FAIRagro use case

Stefan Kühnel[5a], *Benno Kleinhenz*[5b], *Bettina Klocke*[5a], *Markus Möller*[5c], *Manfred Röhrig*[5d], *Ulrike Stahl*[5e], *Matthias Senft*[5e], *Jörn Strassemeyer*[5a], and *Til Feike*[5a]

[5a]Julius Kühn-Institute, Institute for Strategies and Technology Assessment

[5b]Central Institute for Decision Support Systems in Crop Protection (ZEPP)

[5c]Julius Kühn-Institute, Institute for Crop and Soil Science

[5d]Informationssystem für die integrierte Pflanzenproduktion e.V. (ISIP)

[5e]Julius Kühn-Institute, Department of Digitalisation and Artificial Intelligence

The UC 3 “Streamlining pest and disease data to advance integrated pest management” aims for improved RDM regarding pest and disease data and respective yield loss data. Crop protection aims to minimize yield losses due to pests and diseases (P&D). However, there is increasing scientific and public concern regarding the use of pesticides. Accordingly, the farm-to-fork strategy of the EU targets a 50% reduction in the use of pesticides by 2030 (EU, 2020). Integrated pest management (IPM) aims at minimizing the use of pesticides and related environmental impacts by utilizing versatile crop management options, including decision support systems (Barzman et al, 2015). IPM may thus help to minimize related trade-offs.

Despite increasing efforts from policy, science and extension regarding the promotion of IPM, its resounding success has not been achieved so far. One major reason lies in the lack of findability, standardization, accessibility and integration of IPM-related data, models and respective decision support. There are several major challenges regarding RDM of P&D data, which are mainly data from yield-loss trials, epidemiological experiments, and P&D infestation data.

1. Comparison and integration of data is challenged by differences in experimental design (e.g. regarding control treatments) and disease assessment procedures (i.e., timing, scale and sample size).
2. Information on the existence and potential accessibility of specific P&D data in Germany is insufficient.
3. Different types of models for IPM-related decision support exist, building on the above-described data. However, there is a lack of integrated decision support for plant protection that considers the potential yield loss and environmental risk of pesticide application.

The future integration of different types of models is therefore of vital importance to advance IPM-related decision support and make IPM work. Finally, the interplay of continuous crop genetic adaptation, agronomic management changes, climatic change, landscape level effects and P&D evolution is highly complex. It requires a solid database that can be utilized effectively through data integration, analysis and modeling by the research community. To overcome the above-described RDM-related limitations and challenges, UC 3 has three main objectives,

1. Develop guidelines for standardization of yield loss trials
2. Establish an inventory for and improve the accessibility of IPM-related data
3. Integration of P&D models and crop yield models.

Nitrogen Cycling in Agriculture: Approaches for Dealing with Incomplete Data, Interpolation, Water Flux Estimation and Variability

Iris Vogeler[6]

[6]University of Kiel, Department of Agronomy and Crop Science

The study of soil nitrogen (N) cycling processes and its associated N losses has been at the focus of attention in agricultural research for a very long time. Yet our understanding is still incomplete, with large amounts of N not being accounted for in N balances. Nitrogen cycling in the soil and its uptake by plants is influenced by a myriad of dynamic and. This lack of understanding is partly due to the lack of tools for measuring the myriad and linked biological, physical, and chemical processes which govern N cycling. As the soil solution is a very dynamic pool, responding to any changes in the nutrient supply/demand system, collection and analysis of it is vital for studying N cycling. Traditionally, soil solutions are obtained via suction cups, which is time consuming and costly, limiting how many samples can be taken, leading to uncertainty regarding temporal and spatial variability. Other problems include uncertainties related to the water flux, which is required for calculating N leaching loads from measured concentrations. Another problem is the interpolation of N concentrations between measurement days, which is commonly done by linear interpolation. Similar problems exist for estimating gaseous N losses, such as nitrous oxide (N2O) emissions, which are commonly measured with manual chambers.
We investigated how differences in water flux, obtained from different models (EVACROP and APSIM), affect NO3-N leaching loads. Differences of up to 9% in N leaching were estimated for a cereal system in Denmark based on different estimates of the water flux, while variabilities between replicated blocks owed differences in N leaching ranging from 17 to 35%. Soil variability and the determination of sensitive model parameters affecting soil moisture status, drainage and N leaching was investigated via APSIM modeling and a global sensitivity analysis. The microporosity and hydraulic conductivity were identified as the most sensitive parameters for a grassland system in New Zealand. The effect of interpolation of NO3-N concentrations, either based on linear interpolation between measurement days or drainage-weighted was also addressed, based on measurements from a cropping system in Northern Germany. Over a period of 2 weeks differences of 23% were obtained. Finally, a stepwise regression analysis was done for a pastoral system in New Zealand for determining environmental conditions that drive N2O emissions. The predictive ability of the regression model was variable for different sites and soil types, but the analysis indicates substantial differences in emission factors (EF) when using linear interpolation between measurements with an EF of 1%, compared to using environmental explanatory variables, with an EF of 0.7%. This highlights that further research and the development of smart tools is required for better understanding and quantifying N cycling in agricultural systems.
As a final point we introduce a sensor system that provides real-time soil data on soil solution concentrations of N (ammonium, nitrate), moisture, pH and dissolved oxygen, which will be developed under the FAMOSOS (FArm MOnitoring via Real-time SOil Sensing) project.

Metadata for agroecological research and intercropping development: a systematic mapping approach and its challenges

Johannes Timaeus[7a], *Wopke van der Werf*[7b], and *Thomas Döring*[7a]

[7a]University of Bonn, Institute of Crop Science and Resource Conservation (INRES)

[7b]Wageningen University and Research, Centre for Crop Systems Analysis

Developing cropping systems towards a stronger inclusion of agroecological principles requires experimental evidence that should also inform further strategic research development. One approach promoted in agro-ecology is intercropping, i.e. growing two species simultaneously on the same field. Intercropping has a long tradition, especially in countries with a large share of small-scale agriculture. In recent years intercropping also increased in prominence in more industrialized countries. This produced a large legacy of empirical evidence. A simple search in Web of Science for intercrop * OR "crop mixture" OR "species mixture" OR "relay crop" OR "strip crop" yielded 10 276 articles (search conducted 12.06.2023 in title/abstract/keywords). A current challenge is to structure this diverse evidence base to inform strategic intercropping research, e.g. in terms of avoiding duplication, identifying research gaps and facilitating knowledge synthesis. As a step towards these goals we apply a systematic mapping of intercropping experiments. This approach extracts meta data from experimental studies such as independent (crop species, cultivar, sowing density, spatial arrangement, fertilization), intervening (plant traits, soil parameters, pathogens, pests), and dependent variables (yields, quality, plant health, biodiversity, nutrient efficiency greenhouse gasses), as well as the experimental design. These descriptors could then be used to identify clusters of knowledge for quantitative secondary analysis or gaps for new experimental studies. Data quality is not a main focus of systematic mapping approaches since primary data are usually not analyzed. Internal study validity, however, is captured by extracting meta-data about the experimental design (number of replications/locations/seasons or randomization) and the type of experiment (controlled environments, replicated field experiment, on-farm experiments). External validity or transferability is captured by extracting the studied crop species, the climate zone or geolocation, and the type of the studied intercropping system (mixed, alternating rows, strips). In this contribution we will report our experiences with systematic mapping of intercropping research.

4. SESSION SUMMARIES

During the two half days of the workshop, there were two break-out sessions in which we discussed

(S1) important metrics of data quality for data processing and

(S2) user expectations for an interactive quality assessment tool.

During the first break-out session (S1), different priorities regarding the importance of different quality metrics dependent on the area of application became evident. However, there was a consensus that quality metrics can and should be divided into different categories. The following categories with corresponding examples of metrics were mentioned in particular.

Completeness: spatial and temporal coverage, data gaps, documentation, metadata
Plausibility: data limits, standards, outliers, randomization of the data
Resolution: spectral, temporal, spatial, number of samples, repeated measurements
Accuracy: positional accuracy, outlier, confusion matrix, cross-validation
Consistency: scaling, aggregation level, format and naming, units

Statistical summaries (e.g., standard deviations, mean and median, correlations) were mentioned as helpful aids for gaining an overview over a data set. In this context, reference is made to the existing concepts of five-number and seven-number summaries. Intuitive representations such as histograms or box plots can be used to easily convey such statistical values. Particularly for machine learning applications, measures of model-specific validation (e.g., F-Score, precision and recall, confusion matrix) were also highlighted as relevant quality metrics. In addition to the quality of the data itself, the completeness of (optional) metadata was emphasized as important. Information on environmental influences (e.g., weather, soil moisture, light conditions) is particularly valuable. Generally, the ability of data to fulfill the specific requirements and expectations of a particular purpose (fitness for use) is an implicit measure of quality.

Generally, it was suggested that quality assessment should be conducted in two modes, namely (1) intrinsic assessment of a data set and (2) comparison between data sets. While intrinsic evaluation assesses the data itself (e.g., using statistical measures) not only at the level of the entire data set but also at the level of the variables, a comparison with benchmarks and standards provides an external perspective that offers valuable insights into the relative performance and quality of the data under consideration.

It is not only important to select suitable metrics for evaluating data quality, but also to document and communicate them. It was suggested that data records that have passed through a quality control should be labeled with a corresponding *flag*. Furthermore, certain quality aspects could be documented using additional metadata, for example whether the data meets certain established standards (e.g., OGC standards of geospatial data). However, it is crucial for the social acceptance of such documentation not to be too judgmental. Data providers will not use certain quality assessment tools if their data is labeled inferior. The documentation should be transparent with regard to the metrics and algorithms used for quality assessment. If changes are made to a corresponding tool, users should be able to track the progress.

The second break-out session (S2) focused even more on a possible quality assessment tool that would be used by both data users and data providers. Many of the aspects discussed on the first day were reiterated. In general, it was pointed out that "less is more", which means that the appearance of an interface should be clear and self-explanatory. Important quality features should be presented intuitively in an initial overview and in-depth explorations should be optionally possible through further interaction. It was recommended to initially present key statistical variables using common display formats like histograms, box plots, or violin plots alongside a visualization of the raw data. To ensure that a quality assessment tool is actually used, it was noted that both data users and providers need further incentives alongside an intuitive design. People are generally impatient, which means that the algorithms used should be designed to have a fast running time. Comprehensive language models should enable text and keyword-based searches for data and glossaries and tool tips should make the interface easier to understand and use. Since different people might use different search terms, onthologies should be integrated to improve the quality of the data search.

5. CONCLUDING REMARKS

Over two half days, we brought together people from different areas of agricultural science and data management to discuss data quality for data analysis in agricultural systems science. In insightful presentations and lively exchanges in the form of break-out sessions, it became clear that there is a wide range of different quality metrics whose applicability and relevance depend heavily on the area of application and the data type. However, there was a consensus that the presentation of important statistical values together with the visualization of the raw data is meaningful to get an overview of the data. This idea is in line with another conclusion, namely that *less is more* when it comes to evaluating and presenting data quality. For a quality assessment tool, which is being currently developed within FAIRagro, this means that the focus should primarily be on simplicity, clarity, and comprehensibility. The handling and presentation of information must be intuitive. In-depth explorations of data quality should be optional and executable through self-explanatory interaction capabilities. The transparency of the automated data analysis also plays a key role. It is important to disclose and document the algorithms used and make modifications and progress traceable. Metadata is suitable for documenting certain quality aspects and standards to which a data set conforms.

Overall, the workshop demonstrated the immense value of interdisciplinary exchange for all participants. The gathering of different perspectives enables the synthesis of a comprehensive overview and avoids the restriction to one's own field of research.

A. PARTICIPANTS OF THE WORKSHOP

In the following, we provide a complete list of the participants of the workshop. We would like to thank all participants for their contribution to the success of the event; your involvement was indispensable in making it a successful endeavor.

- Shiyaza Risvi, *University of Bonn* (organizer)
- Sven Gedicke, *University of Bonn* (organizer)
- Jan-Henrik Haunert, *University of Bonn* (organizer)
- Florian Beyer, *Julius Kühn-Institute in Braunschweig*
- Prof. Dr. Thomas Döring, *University of Bonn*
- Christian Dold, *Aarhus University*
- Juliane Fluck, *Information Centre for Life Sciences in Bonn*
- Niklas Heidemann, *University of Bonn*
- Carsten Hoffmann, *Leibniz Centre for Agricultural Landscape Research (ZALF)*
- Ireneusz Kleppert, *Forschungszentrum Jülich*
- Dr. Stefan Kühnel, *Julius Kühn-Institute in Braunschweig*
- Susanne Lachmuth, *Leibniz Centre for Agricultural Landscape Research (ZALF)*
- Markus Möller, *Julius Kühn-Institute in Braunschweig*
- Ruben Remelgado, *Dresden University of Technology*
- Dr. Torsten Thalheim, *German Biomass Research Centre (DBFZ)*
- Johannes Timaeus, *University of Bonn*
- Jannes Uhlott, *Julius Kühn-Institute in Braunschweig*
- Lucia Vedder, *University of Bonn*
- Iris Vogeler, *University of Kiel*