Federal Ministry
of Food
and Agriculture

Position Paper
Data management in agriculture

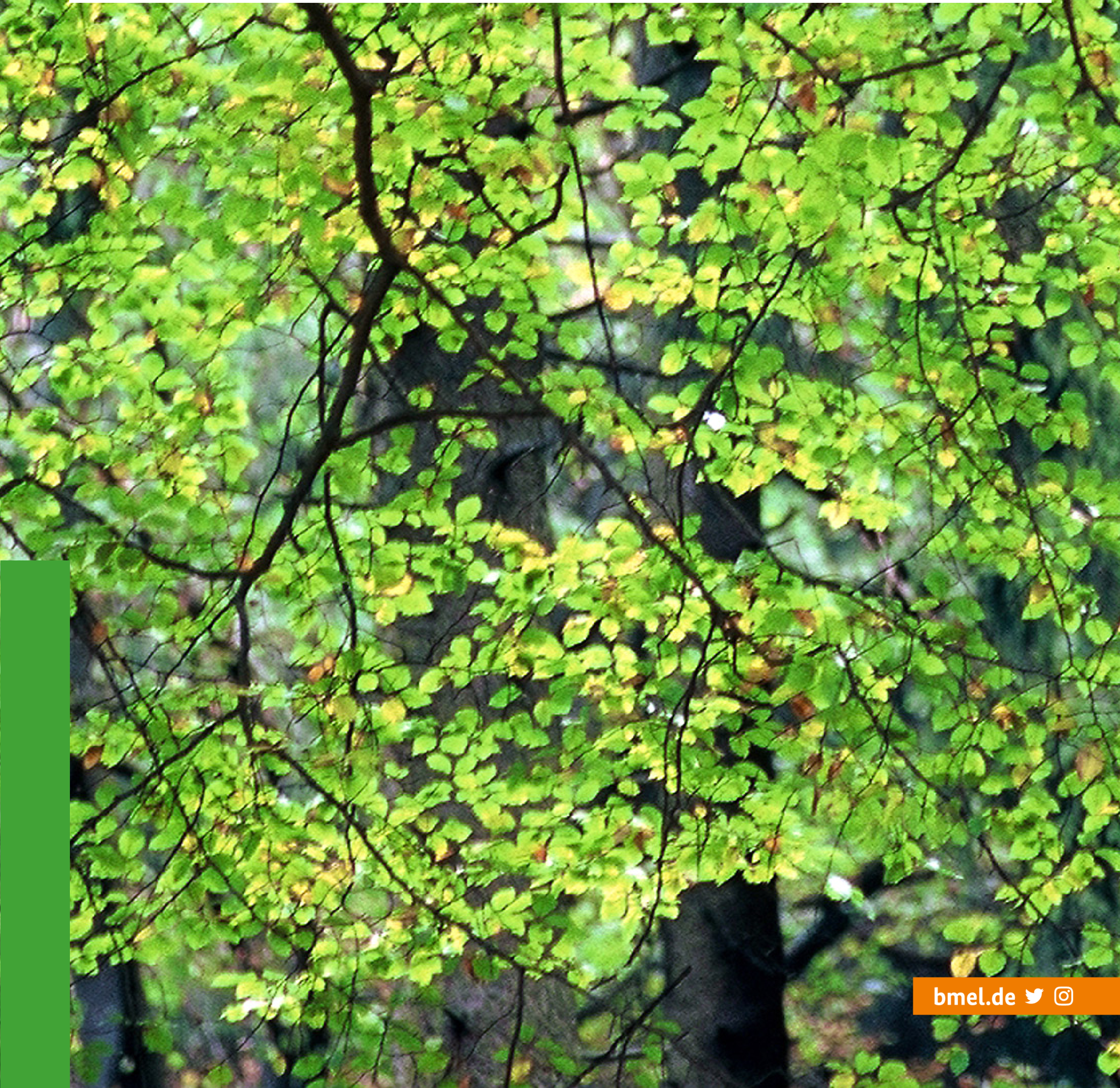

Table of Contents

1 Initial situation

Agricultural work and business processes are increasingly influenced by digital technologies. The status of digitalisation in the agricultural sector is heterogeneous, some farms already have a very high level of digitisation, others are just starting to consider introducing digital technologies [SG18]. The basic gains from digitisation are expected to come from efficiency improvement, quality management, meeting regulation requirements, but also convenience and worker satisfaction. In some cases, the benefits of digital technologies for farms are not directly visible, in particular when weighing up necessary upfront investment and running costs. Often, the benefits depend on the farm size. In principle, different digital solutions offer potential benefits for holdings of every type and size.

The high level of heterogeneity of the available solutions with respect to the data management of digital solutions is considered a major challenge. This applies to both software and apps used by public administrations (s. [BDF+20], chapter 5.2) and private-sector systems [s. [BDF+20], chapter 3.4 and 5.2.). In principle, heterogeneity is a manageable issue, provided that the public sector, but also the private sector economy establish the required conditions. Practitioners currently identify the following challenges:

- → Barriers to the reusability of data: Data from one application are needed by other bodies, but are not accessible to them. Standardisation notwithstanding, lock-in effects can be observed.
- → New dependencies: The growing significance of accurate data for agricultural production activities turns data availability and the reliable functioning of the globally connected IT systems - ranging from mobile phone connections and internet systems to applications - into an increasingly vital prerequisite. Often, there is no awareness of the complexity and potential risks of disruption. Since agricultural production is ultimately part of the vital critical infrastructure, system-wide resilience against possible disruptions deserves particular attention.
- → Complex cross-references: With respect to the entire set of data of a farm (that is currently often stored at different places) there are strong cross-references between the data and complex relations between entities, which can hardly be understood by final users without IT skills.
- → Insufficient interoperability of standards: different domains supplying data to farmers each have their own standards that are not interoperable, e.g. Geodata (OGC, INSPIRE), machines (ISO11783) and weather (diverse APIs and formats are available).
- → Insufficient interoperability of systems: Not all systems are interoperable (apart from ISOBUS, only very few official and/or de-facto standards exist, see [BDF+20], Chapter 5.2.3.3). Figure 1 provides a schematic overview of the current state of the practice. The interoperability of systems is a multi-faceted quality characteristic that cannot be achieved by individual concepts, but only by a set of different, yet aligned principles and concepts at different levels of abstraction.

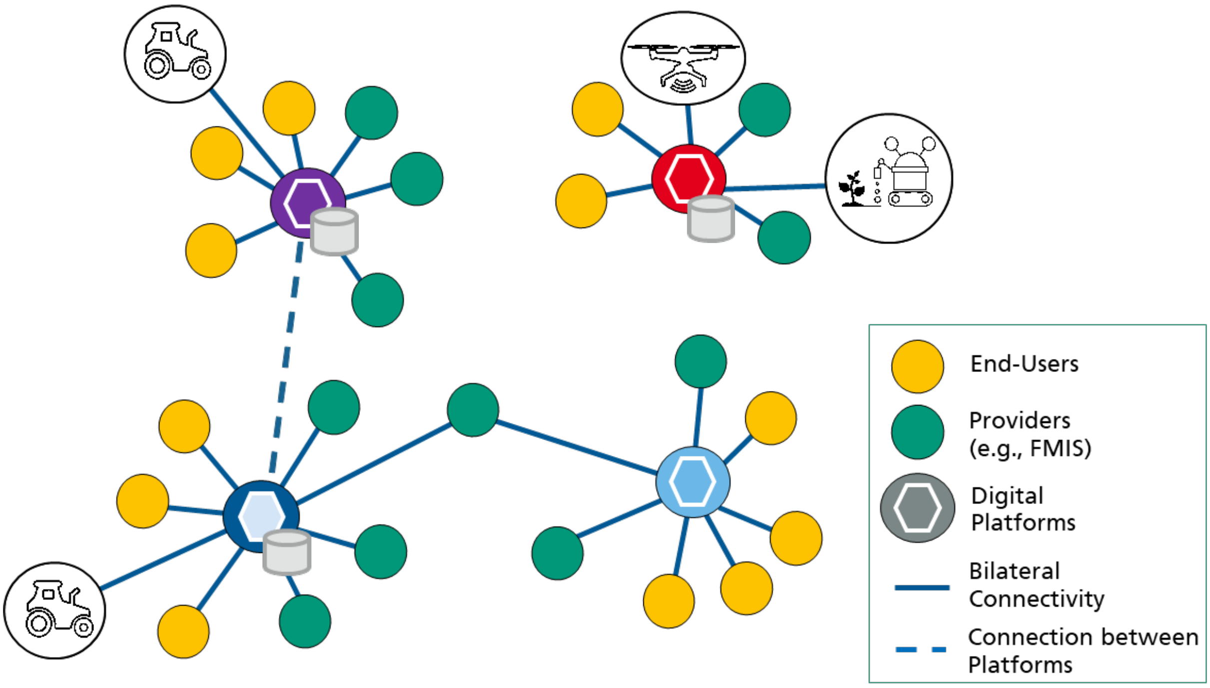

Figure 1: Exemplary presentation of the digital domain ecosystem 'agriculture' with different independent digital ecosystems, three of which are connected with each other (illustration modelled based on [BDF+20]).

Consultation on necessary technical concepts and the standardisation work associated with it requires much effort and should not be underestimated. In some cases, diverse trade-offs prevent interoperability, in particular when people sense that they no longer have sovereign rights over their data or whenever IP issues need to be resolved. Missing or even deliberately impeded interoperability can be understood as a tool to retain customers ("lock-in"); but it leads to higher development and application costs and slows down the possibilities to react for all stakeholders.

Meanwhile there is a large number of (parallel) initiatives and projects, such as the BMEL's digital trial fields, diverse EU projects (i.a. ATLAS and DEMETER), the BMEL's AI projects, the Gaia-X projects of the Federal Ministry of Economics (Agri-Gaia and NaLamKI) and more projects (see [BDF+20], Chapter 3 and Annex A), which will generate more data-intensive and data-processing systems. Thereby more data volumes are opened up, and there is a risk of growing heterogeneity if the systems continue to be fragmented and different paradigms to be used (see also [BDF+20], Chapter 5.2.3.4). Consultations are held between initiatives, projects and trial fields, yet frequently on a bilateral level; the framework of consultations and the possibility for aligning / codetermining concepts is limited.

The heterogeneity described and the lacking interoperability as well as missing similar/identical concepts are disadvantageous for many target groups. Ultimately, agricultural enterprises are, however, affected by the effects of this fragmentation. As opposed to this, better interoperability can also result in more benefits for farms.

2 Target group/objective

This position paper addresses developers and users of data-processing solutions in the agricultural value chain, including upstream and downstream sectors, such as producers of agricultural technologies and farm software, public administrations, and policy-makers. This also includes the provision of relevant basic data, for example data of thepublic sector, data with respect to in-process data processing in primary production and the further processing of food and bio-resources.

The aim of this position paper is

→ to achieve a comprehensive understanding of the requirements related to agricultural data management insofar as they are a result of the cooperative, cross-company and cross-organisation character of work processes.
→ to illustrate the principles of farm data management that need to be observed to secure flexible interoperability that is independent of individual applications and manufacturers.
→ to identify specific examples and recommendations for action for the design of solutions and systems. This specifically includes ideas for designing the effective access to data under official control.
→ to highlight potentials and general conditions of digital technologies and to formulate proposals for an effective basic and further training of the agricultural stakeholders involved.

The position paper summarises the positions of the stakeholders in the field of data management of digital trial fields in Germany, yet it does not reflect individual positions in every case.

3 Positions and technology options

The uniformly understood syntactic, content-related and conceptual modelling of the data to be exchanged as well as seamless connectivity of connected systems (technology and software) are the basis of every comprehensive data management. The internet with its established protocols is to be used as a technical basis and enable seamless networking of many actors and participants in agricultural data spaces.

In order to facilitate flexible, functioning, application- and manufacturer-independent data management in data spaces, different requirements and aspects need to be taken into account. In our view, this includes in particular:

→ Data transfer on the basis of established protocols
→ Simple and automated data access by FAIR principles and Semantic Web technologies
→ Locatability of information/data and services via registry/directory services
→ Consistent use of technologies to ensure the purpose limitation of data use as well as specific/general data sovereignty
→ Use of modern, efficient concepts such as digital twins
→ Resilience of digital infrastructures
→ Further training/knowledge transfer
→ to illustrate the principles of farm data management that need to be observed to secure flexible interoperability that is independent of individual applications and manufacturers.

3.1 Data transfer

The organisation of the data transfer between sensors in the field (this possibly includes agricultural machinery in use), and of all further data processing in the back office requires suitable transfer protocols. In view of the potential variety of interconnected sensors and receivers, the Message Queuing Telemetry Transport protocol MQTT , which has been tried and tested in the industry 4.0 environment, has proven to be particularly suitable. MQTT is advertised as standard for the internet of things by the Organization for

the Advancement of Structured Information Standards (OASIS) . The protocol does not require too much bandwidth and provides quality-of-service mechanisms such as reliable delivery via confirmations of receipt. This makes it particularly suitable for applications in less powerful networks with intermittent disconnections, as can occur in agriculture.

3.2 FAIR and Semantic Web

The modelling of transmitted contents needs to take into account the needs of an open universe of connected and/or interested communication partners. It should be possible for new partners (i.e. their computers) to connect to the data exchange at any time without a high a priori configuration effort. This requires the machine-readable modelling of the more complex data formats as well as the formal grounding of the respective exchanged data values.

As early as the 1970s, methods were developed in computer science with which data can be represented in such a way that computers can independently infer types, functions and conceptual links from the data. Instead of restrictively agreeing on data structures in advance, a descriptive approach is taken: Data as well as their interrelationships are comprehensively described by means of logical statements, and the descriptions are made as accessible as the data itself. The fact that the data can be interpreted automatically supports interoperability. In a global context with distributed data sources of any content, two basic requirements are needed:

→ globally unique identifiers: All "things" that play a role in the data exchange context - i.e. objects but also elements of classification systems, properties of objects, etc. - must be given globally unique identifiers, not only within their domain or own environment, but across domains (i.e. e.g. an ISO11783 data dictionary identifier (DDI) must never consist of the same character string as an IACS crop code).

→ A data model into which all available data structures of existing systems, but also of systems to be newly developed, can be transferred with manageable effort, at least at interface level. It should be upgradable to include future contents, too. Technologically, this can be achieved by graph-oriented data models or data models that are based on statements.

Standards of the Internet Engineering Task Force IETF and the World-Wide Web Consortium W3C within the context of the Semantic Web such as URIs, RDF and SPARQL are available for technical implementation. This ensures that all data can be queried and integrated with each other in a uniform manner. The modelling of JSON-LD objects based on RDFS data format descriptions, which offers a universally readable, flexible formalisation, is an example of a pragmatic solution. A shared understanding of exchanged data requires shared, ontological foundations, in simple cases, in the shape of a binding vocabulary that can be used uniformly by all stakeholders. When using more expressive ontology languages, powerful reasoning becomes possible; examples include the agricultural ontologies based on SKOS or OWL provided by the FAO.

[1] https://mqtt.org/

[2] https://www.oasis-open.org/

In the context of the initiative taken by the European Open Science Cloud, the FAIR principles (https://www.go-fair.org/fair-principles/) were formulated with respect to the desired properties of data:

- → Findable: The first step towards reusing data is to make it locatable at all. Machine-readable meta data is essential in this matter.
- → Accessible: It must be transparent for users how they can access data. This requires open communication protocols which, however, may also include authentication and autorisation.
- → Interoperable: It must be possible to integrate data into other data in analysis, storage and processing. This requires appropriate representations and vocabularies.
- → Reusable: For data to be reusable, it must be described in its relevant contents (machine-readable).

The FAIR principles implicitly include the above requirements for identification and data models. To achieve the desired properties, they also propose specific mechanisms for implementation. Unequivocal identification, description of data and services and the use of open data exchange protocols play a central role in this respect.

In total, the above principles and approaches contribute to the realisation of linked open data solutions, which promise a broad usability, in particular for data that is publicly available.

3.3 Locatability of data and services

For participants in the agricultural ecosystem to get an overview of which data and services are actually made available by various providers, they can be recorded and described via so-called registration services (see IDS and Gaia-X). If this overview is not offered at a central location, but in a decentralised way from different locations, end users have a problem elsewhere:

- → How do you get an overview of the registration services?
- → Are their data and services described in the same or at least in a similar way?

Therefore a centrally operated registration service for the agricultural ecosystem would be expedient.

3.4 Technologies for purpose limitation and data sovereignty

According to our understanding, data sovereignty in agriculture comprises the legal legitimation and the organisational and technical possibilities to use the data from one's own operational context. They include the following:

Data use only with consent

- → Transparency as to how the data is handled and the possibility to use data in different systems. The consent includes several aspects: who is allowed to use the data, for how long, in what context and for which purposes. Transparency means that the handling of the data can be traced back and that the consequences of the consent to use the data are understandable.
- → The digital continuity of data and the possibility to use them flexibly and repeatedly. There are technologies that can enable data sovereignty, for instance XACML-based technologies and technologies for data use control in IDS.

[3] https://www.ietf.org/

[4] https://www.w3.org/

[5] http://www.fao.org/agrovoc/

[6] https://www.w3.org/2004/02/skos/

[7] https://www.w3.org/TR/owl2-overview/

3.5 Digital twins for the agricultural sector

The concept of the "digital twin", for instance of the field, the animal or the process chain under consideration promises even greater performance.
As opposed to pure modelling, the digital twin also takes the dynamic behaviour of the considered reality into account. For example: If a specific field is modelled, the size, location and planned sowing could be recorded from 1 March as date of documentation. If there are data queries to this model in July, it will still provide the unmodified data as documented originally. However, a digital twin of the same field would behave differently: Based on the information collected in March and with further modelled knowledge, e.g. about regular plant growth, a query in July, for example, would provide the current stand by density and height. Hence the digital twin combines aspects of modelling and simulation to also enable predictive statements [Fra20].
The implementation of this idea, which originally emerged in industry 4.0 factories, in a farming environment, offers diverse prospects for development and automation which are currently under intensive research.

3.6 Resilience of digital infrastructures

Since agriculture is a fundamental element of the critical infrastructure (KRITIS) for food, the use of relevant operational data also in a crisis situation should be possible even when there is no internet connection, for instance based on the “offline first principle“. Due to the somewhat tight time slots for harvesting processes, short-term disruptions in operational processes can have strong negative impacts on harvests. In animal husbandry, even short-term disruptions in technical installations can seriously threaten animal welfare. In systems which in the past showed a high level of stability, resilience is therefore of utmost importance. This was underlined not least by the vulnerability paradox [BMI09]. As opposed to a mere online application, a hybrid use of online and offline applications should be preferred. For the realisation of such a resilient architecture, edge computing as the next stage in the field of network-based data processing increasingly comes to the fore. If this paradigm is used to design the management of the edge devices in a decentralised manner that also works offline, it also strengthens the resilience of the sector vis-à-vis infrastructure failures and makes a contribution to preventing the failure of digital systems in agriculture along the lines of ‘Resilient Smart Farming’ (RSF).

3.7 Further training/knowledge transfer

Knowledge transfer via the use of digital applications in agricultural practice will have a major impact on its acceptance. Farming practitioners need open access to necessary and compiled information.
The requirements to be met by the adoption of digital technologies include higher levels of acceptance, increasing the benefits at and above farm level and the targeted transfer into practice. In particular, this relates to the use of existing digital applications and technologies. Therefore, the focus of the knowledge transfer is to demonstrate the palpable benefits of data management for farms as well as for upstream and downstream sectors. The approach is intended to further qualify and inform both "beginners" and "advanced" users.

[8] https://internationaldataspaces.org/

[9] https://www.bmwi.de/Redaktion/DE/Dossier/gaia-x.html

4 Recommendations

4.1 Transparency

The increasing intense interconnectedness between all agricultural stakeholders, but also between human beings and machines, imperatively requires improved approaches to creating transparency. Data use and data security require arrangements that are easy to understand and legally secure. To realistically assess modern decision support systems and AI, it is necessary to get an understanding of the features and limits of the algorithms used. The following points are of major importance in this context:

→ Developers of innovative and especially data-based solutions should disclose information on performance and limits.
  The agricultural environment is highly variable. Diverse “smart” solutions perform analysis tasks, offer decision-making support or carry out tasks fully automatically, while being dependent on diverse boundary conditions, environmental factors, sensor data, etc. Responsible users need to be enabled to assess the quality of the system performance in the specific case and rectify matters when the performance limits of the system are reached. This requires a fundamental understanding of how the systems work and what the effects of the different influencing factors are.

→ Agreements on data use by industry and research need to be legally certain, but also comprehensible.
  An increasing number of data of all agricultural holdings, either on the holding itself, on machines and systems, or on the managed assets are stored and processed by the provider. In addition to a legally secure contractual basis, getting understandable instructions is just as important for the users. At any time, it should be transparent for the user which data are stored and used for which specific purposes. Where necessary, it should also be possible to exercise control in a simple way over the purposes for which data may be used. Private sector mechanisms can be used that are similar to those recommended in [BDF+20] for public schemes. If the goal of comprehensible data use agreements is not achieved in the foreseeable future, consideration could be given to creating uniform guidelines on how these must be structured in the future.
  On the other hand, it is just as important to create an awareness among users that in the age of innovative business models, providers have a legitimate interest in storing and also using data so that new value-added services can also be created for the benefit of users. For the best possible acceptance, the interests and objectives of all parties should be communicated as clearly as possible. The revocable consent of the user is the basis of a trustworthy data exchange.

→ To maintain agricultural production, resilient digital infrastructures are to be established as critical infrastructure (KRITIS).
  A higher level of digitalisation means that the availability of data must be guaranteed in order to keep processes running, or that functionalities can be provided even when subsystems (networks, systems, ...) fail. The concepts for building resilient infrastructures should be further promoted and implemented in practice.

4.2 FAIR provision of data

Both *the public sector* and *private-sector enterprises* should provide data of general interest on the basis of the FAIR principles.
The relevant methods and technologies are available. Depending on the framework conditions, existing infrastructures that already follow the FAIR principles can also be used and data can be fed into them (e.g. community-driven approaches such as Wikidata[1], platforms in the field of bioinformatics[2], catalogue and provision systems for public data[3] or research data management[4]). The FAIR principles deliberately do not demand that data be made available exclusively as open data - on the contrary, it is even noted that access protocols should also include mechanisms for the responsible handling of data worthy of protection. In other areas, acceptable models for partnerships and data sharing between public and private sector actors have been found, and considerable progress and innovation has been achieved (e.g. in pharmaceutical research and development).

4.3 Innovation

In recent years, digital innovations, that is novelties and improvements, have contributed more data, information, knowledge and – as a result – knowledge-based action. The skills that are built this way are a basic condition for follow-up developments.
It is imperative to make a clear distinction between developers and users. Developers are looking for new applications for a new product/service, and users are looking for ways to solve real problems. Both approaches can contribute to improvement, yet the limits are blurred. It is evident that both sides can drive innovation. The issue is how comprehensive the scan/search of both sides for innovative ideas is. Active exchange between research, development and users through partnerships and association work should be encouraged and promoted at all levels.

→ Society (policy-makers) should promote more infrastructure projects and basic technologies that are of social benefit, but whose Return on Investment is not immediately obvious (semantics, digital twins, sovereignty).
→ *Industry & science* need to put more emphasis on the existing ecosystem and existing solutions and initiatives (cf. Chapter 3). (Voluntary) commitment not to constantly invent new things, but to use existing solutions.
→ Active exchange between researchers, developers and users through partnerships and association work should be encouraged and promoted at all levels. (Voluntary) commitment of economic operators to be actively engaged.

4.4 Further training/knowledge transfer

It will be necessary to develop comprehensive training manuals and material for initial and further training that open up the opportunities and limits of current developments for users-practitioners. The presentation of topical tools and solutions of producers is just one path. What's even more important is the transfer and provision of

---

1 https://www.wikidata.org
2 https://www.embl.org/services-facilities/
3 https://www.wikidata.org
4 https://www.embl.org/services-facilities/
5 https://www.govdata.de/
6 https://www.publisso.de/

- → basic mechanisms (possibilities and limitations) for the application of digital solutions and AI. These include properties of fundamental basic data and solid foundations for classifying and assessing data-based solutions. The result is the management and application of specific products. A manufacturer-specific application doctrine should not take centre stage.
- → Understanding of algorithmic thinking and knowledge about statistical correlations.
- → The users of digital tools need to get an understanding of the technologies. To this end, basic principles of data collection, statistical correlations and typical sources of errors or mistakes must be included in initial and further training, as well as standard assessments of digital applications, e.g. diagnostic suitability (sensitivity/specificity, error rates: false positive, false negative, etc.).
- → Materials of educational and instructional value (training manuals, worksheets, work assignments, explanatory videos, etc.).
  They should be pooled and published by a neutral body. In this context, information materials of public sources (universities, vocational and technical schools) and the private sector play an important part. For all sizes of holdings to benefit from digital applications, private contractors, machinery pools and consultants also need to be considered in the knowledge transfer. To provide users easy access to solutions for their enterprise, independent experts should explain, assess and convey real examples. The initial threshold of the enterprises for digitisation must be taken into account (high or low level of digitisation). Practical digital practical data (model farms) should also demonstrate the possibilities and benefits of digital applications.
- → A shared terminology in the field of digital applications.
  This should be developed by the centres of vocational and university education and training. The development of a uniform terminology is important for the knowledge transfer and should be available to all target groups. This is the only way to make sure that identical content is conveyed. We would recommended to establish and extend a central reference guide (http://Farmwissen.de).
- → Clear and easily understandable facts on data sovereignty and data security.
  The purpose is to make users familiar with legal foundations, data-based business models and the possibilities of data sovereignty in knowledge transfer. Both operating data and data of external labourers should be included.

This makes it possible to assess the performance and the use of digital solutions and AI. In the face of the rapid development dynamics, close cooperation between researchers, solution developers, initial users and active teachers is indispensable. Transparent, innovative and well-demonstrated applications increase user acceptance.

5 Afterword

The Competence Network on Digital Technologies in the Agricultural Sector was commissioned by Federal Minister Klöckner to strengthen digitalisation in agriculture in Germany and established in 2019. The Competence Network consists of the spokespersons of the 14 digital trial fields in the agricultural sector and other important stakeholders from administration, academia, industry, associations and the federal government's research institutions. One key task is to promote the transfer of knowledge to agricultural practice, to the upstream and downstream sectors, to the general public and to policy-makers. The Working Group on Data Management was established in spring 2020 and comprised members of the Competence Network and external experts.

The following people have participated in the working group:

Dr. Ansgar Bernardi, Dr. Fabian Billenkamp, Dr.-Ing. Hermann Buitkamp, Prof. Dr. Jörg Dörr, Daniel Eberz-Eder, Dr. Antje Fiebig, Prof. Dr. Hans W. Griepentrog, Anton Huber, Dr. Hartwig Kübler, Dr. Martin Kunisch, Daniel Martini, Kirstin Ohlendorf, Bernd Rauch, Prof. Dr. Yves Reckleben, Nikolaus Staemmler

This position paper is a summary of the meetings of the working group and does not in each case reflect the position of each individual participant.

→

6 References

[BDF+20] BARTELS, Nedo; DOERR, Joerg; FEHRMANN, Jens, et al. „Machbarkeitsstudie zu staatlichen digitalen Datenplattformen für die Landwirtschaft (Feasibility study on governmental digital learning platforms for the agricultural sector)“, 2020, https://www.bmel.de/DE/themen/digitalisierung/datenplattform-machbarkeitsstudie.html; last retrieval on 11.4.2021

[BMI09] BMI, „Nationale Strategie zum Schutz Kritischer Infrastrukturen - KRITIS-Strategie (National Strategy for Critical Infrastructure Protection - CIP Strategy)“, 2009,https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/bevoelkerungsschutz/kritis.html, p. 8, last access on 16.07.2021

[Fra20] Digitale Zwillinge für die Landwirtschaft (Digital twins for the agricultural sector), https://www.iese.fraunhofer.de/blog/digitale-zwillinge-fuer-die-landwirtschaft/, last access on 18.5.2021

[SG18] SCHLEICHER, Sebastian; GANDORFER, Markus, “Digitalisierung in der Landwirtschaft: Eine Analyse der Akzeptanzhemmnisse (Digitalisation in agriculture: an analysis of acceptance barriers”, 2018, Presentations of the 38th Annual Conference of the Association for informatics in the agri-food-forestry sector (GIL-Jahrestagung) in Kiel. Ruckelshausen et al. (eds.)