# Protection and privacy of data in cooperative agricultural processes - the challenges of the future

**Franz Kraatz [a,*], Frank Nordemann [a], Ralf Tönjes [a]**

[a] Faculty of Engineering and Computer Science, University of Applied Sciences Osnabrück, Osnabrück, Germany

* Corresponding author. Email: f.kraatz@hs-osnabrueck.de

## Abstract

In agriculture, the growing usage of sensors, smart mobile machinery and information systems results in high volumes of data. The data differs in accuracy, frequency, volume, type and, most importantly, owner of the information. However, cooperative processes and big data analyses require access to comprehensive amounts of data for successful agricultural operation and reasoning. In some processes instructed contractors even gather data belonging to other owners and use it for machinery operation optimisation and accounting (e.g. yield in maize harvest). Today's approach of data handling has a high potential to conflict with European and national regulations for data protection and privacy. This article presents a concept for continuous data protection and privacy in cooperative agricultural processes. The concept aims at ensuring data sovereignty for the owner while making as much data usable for process operation and big data research at the same time. Briefly explained, owners pick a collection of data and create usage licenses for other players. The licenses specify time-limited and / or position-bound access to the data collection. Privacy environments in soft- and / or hardware protect access rights on end user devices, data share hubs and machinery devices such as agricultural terminals. In addition to access right configurations, digital signatures prevent data manipulation when cooperative players capture data during processes. So-called signature boxes represent certificated soft- or hardware components, which are located close at data sources (e.g. as hardware attached to sensors on mobile machinery) and bind the data captured with digital signatures.

**Keywords:** privacy, data security / encryption, digital certificates / signatures, cooperative agricultural processes

## 1. Introduction

During an agricultural process mobile machinery records a growing amount of data using the company independent ISOBUS. With data such as crop yield (de facto output rate), the agricultural processes can be optimised. For example, application maps with subareas can be generated by a precision farming provider to do fertilising work more efficiently. The data is also used to document for compliance with regulations. The amount of manure and the crop yield are combined to a nitrogen balance, which is analysed and observed by authorities [4].

The transfer of recorded data is mostly done directly from the agriculture machinery to the farmer's office (e.g. by USB-drives). However, more and more data is transferred by or managed in central agricultural data hubs. Via data hubs, a farmer can control and realise data sharing with other process players. But the transfer and handling of important and sensitive data is often done with little or without any data security mechanisms by now. Furthermore, all process players (data hub, agricultural contractor, agricultural service provider) may have access to the data. Avoidance of unauthorised data access or data manipulations are difficult to prevent.

To implement a secure data handling for all process players a distributed and player overlapping digital right management is missing. The right management system must be able to handle agricultural environments with interrupted or no communication. Ownership and access rights of data needs to be well-defined at all times and locations. Manipulation of data must be prevented from data capturing on field till the end of data transfer at the destination. Only then process data can be used for a distinct proof of compliance with regulations. After these requirements are implemented as a company independent and standardised data handling in agricultural processes, overall data analyses over all players are possible. Comprehensive data analyses and reasoning will be used to research plant production / crop production, resulting in optimisations for managing and realising agricultural processes.

## 2. State of the Art

There are several technical mechanisms to implement data security and privacy. The similarity between all mechanisms is the usage of cryptography for realisation. The main mechanisms are described in the following short summary of the state of the art.

### 2.1. Encryption mechanisms

The basis for a secure data exchange is set by crypto graphical mechanisms such as encryption methods. These encryption methods are differed in two main techniques. With the symmetric-key algorithm the communicating parties share the same digital key for encryption, known as a shared secret [2]. This shared secret has to be distributed to all parties prior to a secure data exchange, shown on the left in Figure 1. Contrarily to this, the public-key algorithm uses pairs of keys for every communication participant [5]. One part of this pair of key is the public key, which is published to the community, shown on the right in Figure 1. The other part is the private key. This key stays in the possession of the key owner and is kept secretly. In this algorithm the public key is used to encrypt the data and only the owner of the private key can decrypt the data with his private key. The public-key algorithm doesn't have to share a secret for beginning a secure information exchange. But the processing speed of the public-key algorithm is much slower compared to the symmetric-key algorithm.

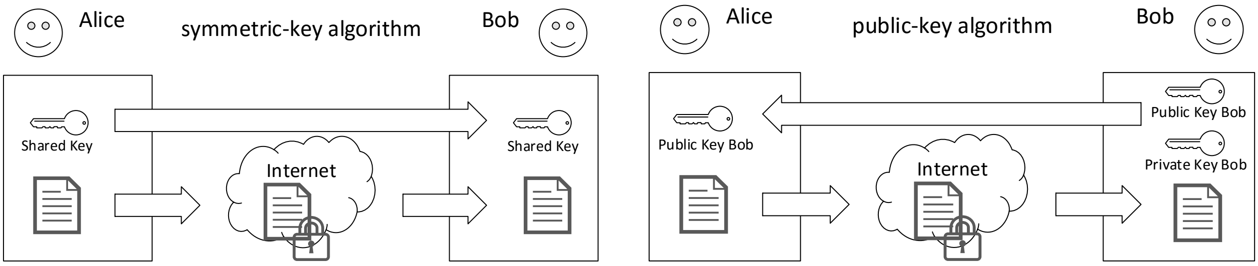

Figure 1: Symmetric- and public-key algorithm

For secure data exchange at public communication channels a hybrid version of these two techniques is used primarily [17], shown on the left of Figure 2. The public-key algorithm is only used to commit the shared secret of the symmetric-key algorithm at the communication establishment. After the establishment, the secure data exchange is realised with symmetric-key algorithm. In cases where it is not possible or exaggerated to use the public-key algorithm for key exchange, the Diffie-Hellman key exchange is used [6]. This method, shown on the right in Figure 2, doesn't transfer the complete key between the two communication partners. Only the calculated parameters (A, B) based on the shared parameter (G) and the secret (a, b) are transferred. With the secret (a, b) and the parameter (A, B) form the other side, both communication partners can calculate the same secret key. After this, the secret key can be used for secure data exchange.

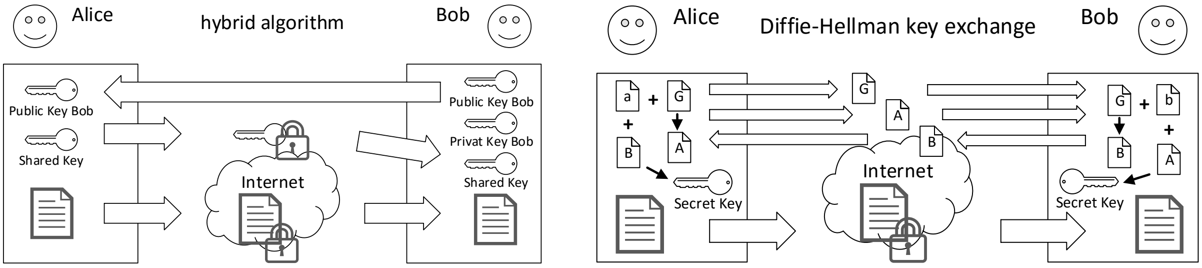

Figure 2: Hybrid algorithm and Diffie-Hellman key exchange

The pair of keys of the public-key algorithm can also be used to proof the ownership of data. With the private key a digital signature of data can be generated and the public key can be used to verify the ownership [10]. The only weak point of public-key algorithm and digital signature is the public key. The community has to trust that the public key really belongs to the assumed participant. To handle this problem a Public-Key-Infrastructure (PKI) is needed [1]. With this infrastructure a trustable organisation signs the public key and confirms the authenticity, shown in Figure 3. The user of the public key only has to trust this signing organisation. To lower the number of needed trustworthy organisations the PKI is built with a hierarchy structure. The public key user trusts the root organisation, but the public key is signed by an organisation some layers under the root organisation. To verify the ownership, the user can go up through the structure and reaches at the end the trusted root organisation.

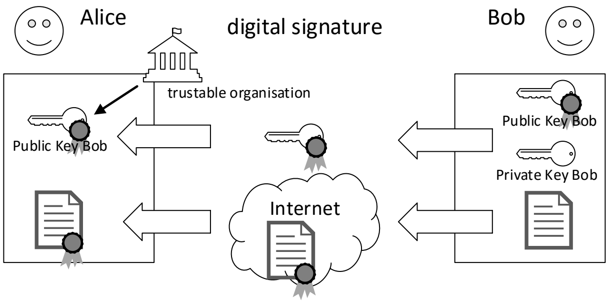

Figure 3: Digital signature

### 2.2. Watermarking

Digital watermarking is a technique used to label / brand content in files [3]. Usually, watermarks are not noticed by users reading, listening or viewing the branded content (e.g. books, music, movies). As an example, this is done by adding / manipulating minor frequencies in audio files. Watermarks are ideal to label files given to different users / groups. Every user / group gets differently marked files. In case of a data leak, the origin of the leaked file can be identified up to user- and / or group-level. This helps identifying the leak location (traceability) and limiting the users / groups / systems involved in the leak. One or more techniques for digital watermarking may be combined.

### 2.3. Access right management

Creating copies of digital data is fairly easy. Files may be accessed, duplicated, transferred, used and distributed without knowledge of the owner. Hence, mechanisms for access control are required.

Most filesystems provide file access control based on users and groups managed by system administrators [12] [13]. Typical rights given are read, read-write and execute access to files. Users are only allowed to access and manipulate files they were given access to. While this is sufficient for many computer systems, it does not prevent users to duplicate and transfer files they may have (temporary) access to. In addition, such systems mostly reflect data management inside one organisation / company. Typical filesystems are not appropriate for handling data in cooperative processes. Many different systems of various process players cooperate with each other. Different techniques and administrators of cooperating companies are involved; security leaks may occur at many levels. In addition, companies / administrators need to trust each other. In the case of data leaks, responsible systems and persons may be hard to be identified.

In media industry, controlling of digital data access for different users / parties is realised by using digital rights management (DRM) [18]. DRM assigns access rights to digital content such as music, movies and books based on users and groups. A user subscribes to a service and is granted access to corresponding content. However, DRM requires special hard- and / or software to show and play the media files. The DRM-environment checks if user access is still granted and prevents duplication of data. In other DRM-systems, data access is controlled by issued digital licenses, valid for a certain period of time. The license will be checked by certified media players such as Windows Media Player. DRM systems allow copyright holders to grant access to media for a certain period of time without worrying about unauthorised data duplication. Some systems allow to play media without being online (e.g. Spotify's Offline-Mode for 30 days [16]). This is possible by fetching the current date / time from official servers every time a client is online. In Offline-Mode, access to music files will be restricted after 30 days until the user connects to an official server again and still has a valid subscription. A problem in using existing systems in the area of agriculture is that most systems are designed for usage in specific domains. Due to the diversity of users and systems, a general and flexible approach is desirable for use in agriculture.

Additional features to control access to digital files are provided by Attributed Based Encryption (ABE) [11]. Whether or not access is granted depends on the fulfilment of attributes, which are defined by the data owner. Attributes represent different characteristics of the user, system, component and other technical devices that want to access a file. However, attributes may also represent environmental aspects such as a geographical location and / or a time that need to match for data access.

## 3. Problem statement

Many agricultural processes require cooperation between different players to be realised economically and ecologically efficient. Since cost for agricultural machinery as well as legal regulations will continue to increase in the future, less farmers will be able to handle complex processes such as precision farming and harvesting fields on their own.

In practice, today's cooperative processes often show critical issues in terms of privacy and security. The following Figure 4 lists four main issues in a typical agricultural scenario including the players farmer, precision farming provider, authority and agricultural contractor.

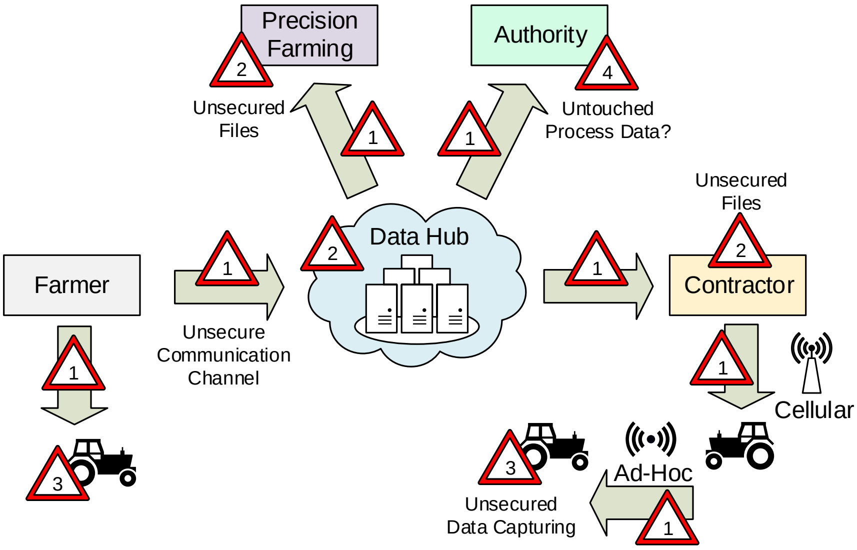

Figure 4: Privacy and security issues of cooperative agricultural processes

First of all, a lot of data is shared between players using unsecured communication channels (Figure 4, sign no. 1). For example, a farmer may send yield maps of his land to a precision farming provider by e-mail. After processing the map, the precision farming provider may return an application map to the farmer, also using e-mail. Communication via e-mail is not secured in any way, data can be captured along the way from farmer to precision farming provider and back. Integrating a centralised data hub for sharing of digital files requires the data hub provider to support secure communication channels for data transfers.

Additionally, offline data exchange by USB drives and other media is also potentially unsecure. USB drives may be stolen or lost, even after the data was already processed by the precision farming provider. Figure 4 shows many more positions where data loss can occur in agricultural processes.

Another potential risk for privacy and security is working with unsecured files. Usually, every player receives process / task data and imports it in their systems and machinery for process work. Since the data is not secured and digital right management is not implement across system borders, data duplication is possible without knowledge of the owner or the following processing players. This is presented in Figure 4, sign no. 2.

Beside mobile communication networks using EDGE, UMTS, HSPA and LTE, other communication techniques are applied for data exchange in agricultural scenarios. A growing number of devices connects directly with each other, without any need for communication infrastructure [15]. These so-called mobile ad-hoc networks (MANET) enable agricultural machinery and user devices to connect dynamically and exchange data whenever they meet on street, field, farm or silo. However, establishing a secured, encrypted connection with an element of the agricultural process is often not carefully discussed and implemented. An essential step in guaranteeing privacy and security is the authentication and authorisation of possible communication partners. Since there is no need for a central infrastructure in ad-hoc networks, there is also no central authentication and authorisation. A person interested in process data may fake its identity to act as a communication partner and to capture process data. Hence, decentralised mechanisms to identify possible partners and to verify their data access rights is crucial. Many scenarios are missing such mechanisms. In addition, the ad-hoc communication channel between authenticated partners needs to be secured. Otherwise, data can be captured by other players or intruders (Figure 4, sign no. 3).

Furthermore, data can be manipulated by different players during an agricultural process. Players have not guarantee that logged and accounted data really reflects what happened during process operation. This provides possibilities for fraud between players. Data manipulation may also be used for faking adherence to legal requirements. For example, data can be manipulated to testify that the amount of liquid manure brought out reflects the legal maximum. However, an authority would not be able to detect higher amounts of liquid manure that have been applied on field in reality (Figure 4, sign no. 4).

The problem statement presented above leads to a number of requirements for cooperative agricultural processes with respect to privacy and security aspects. The requirements are:

- A comprehensive mechanism to define data access rights by owners.
- A distributed data right management system with sophisticated access right control mechanisms based on process characteristic (ABE).
- Authorisation: Reliable identification of communication partners.
- Authentication: Reliable verification of data access rights of communication partners. An instance of trust for all players / participants is required.
- Instance of trust on every device (systems, machinery and user devices at all cooperative players)
  - to control access rights of digital data (only provide data that is needed for the subprocess) and
  - to sign process data during runtime (protection against data manipulation at or after process runtime).
- Whenever access right controlling by trusted instance is not possible: a mechanism for tractability in case of data leaks (Watermarking).
- Anonymisation of data collections for optimising analyses and reasoning by external service providers (e.g. data sharing for big data analyses).

The requirements need to be implemented in an open and reasonable way for all cooperative players of agricultural processes. A possible solution requires flexibility and adaptability to be appropriate for many different agricultural scenarios.

## 4. Utilisation of privacy and security techniques in cooperative agricultural processes

The state of the art presented in chapter 2 includes different techniques for establishing privacy and security. The techniques originate from use-cases other than agriculture. Most technological approaches can be adapted somehow to different use-cases. The complexity consists of designing and configuring an open, consistent and dependable concept for their use in cooperative agricultural processes. This chapter explains how and with what technology the requirements of chapter 3 can be fulfilled. The following chapter 5 will present how the utilisation of technologies results in privacy and security for cooperative agricultural processes.

### 4.1. Data licensing, data encryption and data watermarking

At the beginning, the data owner specifies who should be able to access what data in a cooperative process. Usually the data owner is represented by a famer who wants a specific agricultural task done. Therefore, he prepares relevant process data and starts by issuing digital data access licenses for the different cooperative players of the process. The data access rights for the data collection are written to the license before the data owner signs the license with his private key. As part of the access rights definition, ABE is used to allow comprehensive access rights based on process characteristics (e.g. location and time of process data access). Licenses are the basis for distributed data access along the cooperative process.

Prior to sending data with access licenses to the process players, it will be encrypted using the public / private mechanism. The data is no longer accessible by unauthorised players and is protected against unsecure communication channels in Internet, mobile communication networks and mobile ad-hoc networks. Intruders may capture encrypted data, but without private keys they are not capable of decrypting it. Data encryption using public keys of trusted instances (RM-Box, explained in the following subchapter 4.2) is preferred. Alternatively, public keys of other cooperative players or their systems and machinery respectively can be applied for encryption.

In some situations, cooperative players need full access to original files / data. An example is the processing of data by a precision farming provider. The provider calculates manure application maps with his own tools based on data formatted as yield maps. Here, usage of trusted instances for access right protection is impossible. Digital watermarking can be an appropriate mechanism to prevent data sharing of cooperative player with unwanted parties. Watermarked data is traceable back up to the originally issued player, which helps to identify non-trustworthy players or players with security leaks in their systems and machinery.

### 4.2. Access right management for distributed data access

Ensuring access rights across systems of various players is challenging (cf. chapter 2.3). A possible solution is to add a trusted component on every system that is part of the agricultural process. This component acts as a black-box to systems and is responsible for managing the data access rights. The black-box represents a privacy environment and specifies interfaces which allow other system components to ask for process data. Depending on the current system and the licensed data, the black-box will grant or deny data access. Therefore, all data is encrypted with the public key of the RM-Boxes (cf. subchapter 4.1). The black-box is named after its main purpose, Right-Management-Box (RM-Box). It may be realised as software to be installed on user devices and computer systems. Alternatively, a RM-Box can be designed as a piece of hardware connectable to systems and machinery of players.

Due to the fact that data access rights are defined for certain time frame, the RM-Box is able to check for updated licenses online. This enables a data owner to extend data access time frames or other characteristics of the process after initial license release. Since not all mobile machinery of an agricultural process is connected continuously to the Internet, a RM-Box supports an Offline-Mode. Data can still be accessed during process runtime on remote locations without mobile communications coverage.

### 4.3. Protection against data manipulation

Data manipulation happens for many reasons. However, fraud against other players or authorities is the major one in agriculture. As a protection against manipulation data should be signed directly at process time by a trusted instance. This trusted instance can be represented by a Signature-Box (Sig-Box). With the help of certifications for Sig-Boxes (e.g. by an accepted institute or authority) manipulation freeness can be achieved. Digital signatures created by Sig-Boxes ensure that captured process data reflects reality, no applied manure amount or crop yield has been modified. Sig- and RM-Boxes can be implement as a single piece of soft- or hardware.

### 4.4. Infrastructure for digital key management

An essential aspect of privacy and security is provided by data encryption using public keys. Encrypted data is only readable by intended recipients, other parties end up with numbers of meaningless bits. However, there is an important factor to take care of: you need to be sure that public keys used for encryption really belong to the intended recipients. Intruders could easily read data if they manage to implant their own public keys into the encryption mechanism.

This results in the need for a reliable public key infrastructure, which is organised by a trusted institute. Every player who wants to be part of a cooperative process brings his public key to the institute. The institute checks the identity of the player and signs the public key with the institutes private key, allowing everyone else to verify the public key of the player using the resulting digital certificate. Furthermore, other trusted organisations may sign public keys of players, systems or machinery. This builds up a chain of certificates, led by the initial trusted institute. Dissemination and updating of public keys / certificates can be realised by public key servers of trusted organisations.

### 4.5. Anonymised access to process data collections

Data owners should only provide needed data to cooperative players in their own interest. However, some situations may require sharing and collecting data from many players. For example, the ongoing research in the area of big data technologies may allow optimisations for plant production and process planning. Big data analyses and reasoning require high amounts of agricultural data from many different players. Performing analyses / reasoning without violations of privacy concerns can be done by anonymising data before integration into data collections. The position of a field can be altered from GPS-like positions to a defined region, making the field no longer linkable to a specific farmer. The results of ongoing analyses may help the farmer to optimise his operation, e.g. the crop yield by adapting the process of fertilisation during a season.

## 5. Solving data security and privacy issues in cooperative agricultural processes

The big picture of solving data security and privacy issues in cooperative processes is presented in this chapter. In regard to Figure 4 typical issues of today’s agricultural process realisations and possible solutions using technologies from chapter 4 are discussed.

Figure 5 summarises issues and solutions of cooperative processes. The first issue of unsecured communication channels along the cooperative process is treated by requiring encrypted communication channels. This can be easily realised by applying well-known encryption techniques for data transfers (e.g. HTTPS [7], TLS [9], SFTP [8]). As an alternative, a virtual private network (VPN) can be built between communication partners. With technologies such as OpenVPN a secured virtual tunnel connects communication endpoints [14]. Communication channel encryption provides solid protection against data capturing along a communication path (so-called Man-in-the-Middle attacks) for mobile communication networks, mobile ad-hoc networks and Internet connections in general.

The second issue showed by Figure 5 is handled by distributed right management using RM-Boxes on systems and machinery of cooperative process players. Data owners issue digital licenses, describing data access rights for players, systems and machinery. The access rights may include process attributes and characteristics like geographical location and time of data access. The right management after data dissemination is controlled by RM-Boxes. Unauthorised data access or data sharing is no longer possible. Data access is restricted for the time of process realisation. RM-Boxes can be implemented as pieces of soft- or hardware.

Furthermore, data is encrypted using the public / private key principle. The data owner encrypts data collections using the public key of RM-Boxes. This ensures that only RM-Boxes can control data access, even when using central data hubs for data dissemination. If necessary, public keys of players, their systems and machinery can be used additionally. It is important to keep in mind that this softens exclusive controlling of access rights by the RM-Boxes: duplication and sharing of unencrypted data on systems of cooperative players cannot be prevented. Digital watermarking should be used to trace back data in case of data leaks. However, data can never be read by cooperative players it was not encrypted for. This is certainly important in case of using data hubs and dissemination chains (e.g. from farmer to precision farming provider to contractor to authority).

Encryption of data requires a reliable distribution of public keys. A public key infrastructure featuring a trusted institute / organisation is responsible for checking the identify of cooperative players. If successful, the institute approves the public key of a cooperative player by issuing a digital certificate. Every player is able to authenticate a public key for a specific player using the certificate.

Handling of issues 3 and 4 requires Sig-Boxes distributed across the cooperative process. Operation of Sig-Boxes need to be approved and certificated by a trusted institute or authority. Placed on systems and machinery, Sig-Boxes can sign captured process data right at process runtime to prevent data manipulation. Since Sig-Boxes are certified, cooperative players as well as authorities can trust signed process data, e.g. for crop yield or applied manure.

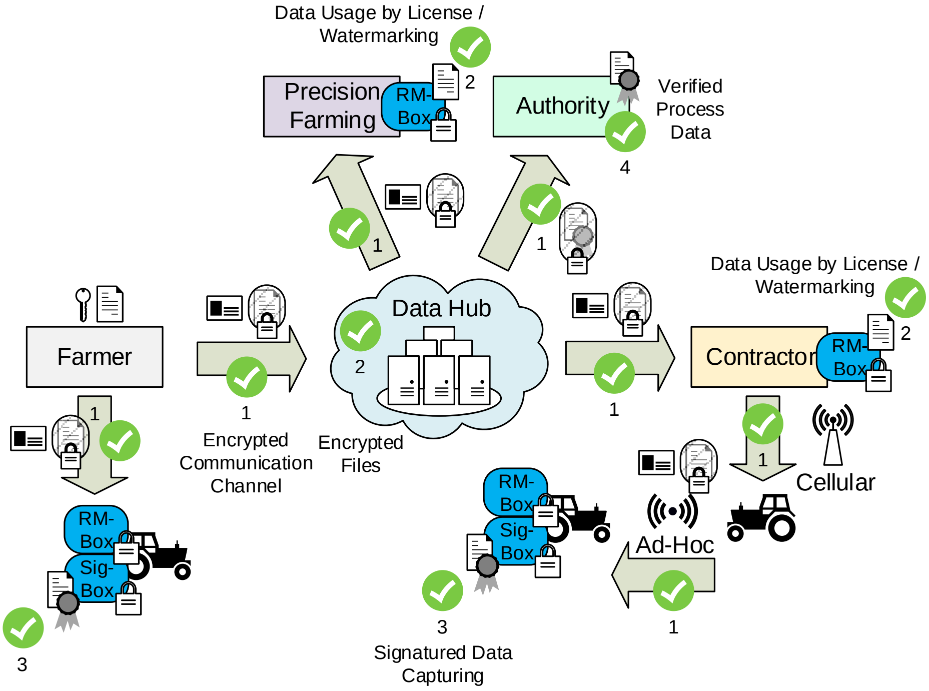

Figure 5: Solution for privacy and security issues of cooperative agricultural processes

## 6. Outlook

The paper presents typical privacy and security issues in cooperative agricultural processes. First of all, this aims at providing awareness and a better understanding of privacy and security concerns when cooperating with other players. Secondly, the discussed issues intend to create a common view on the main challenges faced in cooperative agricultural processes. Finally, a concept for solving privacy and security issues is presented and is meant to act as a basis for discussions to solve issues in a sophisticated and reliable manor for all cooperative players.

Distributed data access management is challenging due to the different systems and machinery of cooperative players. Access right definitions based on digital licenses issued by data owners can be one aspect of a solution. Attributed based encryption (ABE) allows comprehensive definitions of access rights. An essential part is the management of access rights by RightManagement-Boxes (RM-Boxes), which prevent unauthorised data duplication and data sharing. Digital watermarking may help to identify data leaks. Encryption of data is another important factor. Using of the public / private key mechanism and a reliable public key infrastructure results in unreadable data for unauthorised players and intruders. This is also applicable for central instances such as data sharing hubs.

A possible solution for data manipulation is represented by Signature-Boxes (Sig-Boxes). Certified by a trusted institute, Sig-Boxes sign process data directly at process runtime. Cooperative players and authorities may trust signed process data. Reasoning and analyses of agricultural data with big data methods can be done using anonymisation of data.

Planning, realisation and documentation of cooperative processes will continue to evolve in the area of agriculture. Privacy and security aspects need to gain focus when sharing digital process data. An acceptable solution requires comprehensive analyses and discussions including all agricultural players. The presented approach may be seen as a proposal and will be further elaborated and prototyped in the future.

## References

[1] Adams, C., Lloyd, S., 2003: *Understanding PKI – Concepts, Standards, and Deployment Considerations*. Boston: Addison-Wesley.

[2] Delfs, H., Knebl, H., 2007: *Introduction to Cryptography – Principles and Applications*. Heidelberg: Springer-Verlag Berlin.

[3] Digital Watermarking Alliance, 2016. http://www.digitalwatermarkingalliance.org Accessed May 31, 2016.

[4] European Environment Agency, 2010: Agriculture: nitrogen balance. http://www.eea.europa.eu/data-and-maps/indicators/agriculture-nitrogen-balance Accessed May 31, 2016.

[5] IEEE P1363, 2008: Standard Specifications for Public-Key Cryptography. http://grouper.ieee.org/groups/1363/ Accessed May 31, 2016.

[6] IETF, 1999: Diffie-Hellman Key Agreement Method. https://tools.ietf.org/html/rfc2631 Accessed May 31, 2016.

[7] IEFT, 2000: HTTP over TLS. https://tools.ietf.org/html/rfc2818 Accessed May 31, 2016.

[8] IETF, 2006: SSH File Transfer Protocol. https://tools.ietf.org/html/draft-ietf-secsh-filexfer-13 Accessed May 31, 2016.

[9] IEFT, 2008: The Transport Layer Security (TLS) Protocol. https://tools.ietf.org/html/rfc5246 Accessed May 31, 2016.

[10] Katz, J., 2010: *Digital Signatures*. New York: Springer.

[11] Lewko, A., Waters, B., 2011: *Decentralizing Attribute-Based Encryption*. Heidelberg: Springer-Verlag Berlin, P. 568-588.

[12] Linux Kernel Documentation, 2016: Ext4 Filesystem. https://www.kernel.org/doc/Documentation/filesystems/ext4.txt Accessed May 31, 2016.

[13] Microsoft, 2003: NTFS Technical Reference. https://msdn.microsoft.com/de-de/library/cc758691.aspx Accessed May 31, 2016.

[14] OpenVPN Technologies, 2016: OpenVPN Website. https://openvpn.net Accessed May 31, 2016.

[15] Schiller, J., 2003: *Mobile Communications*. Edinburgh: Addison-Wesley.

[16] Spotify, 2016: Listen offline. https://support.spotify.com/au/article/Listen-offline/ Accessed May 31, 2016.

[17] Stinson, D.R., 2006: *Cryptography – Theory and Practice*. Boca Raton: CRC Press.

[18] Zeng, W., Yu, H., Lin, C.-Y., 2006: *Multimedia Security Technologies for Digital Rights Management*. San Diego: Elsevier.