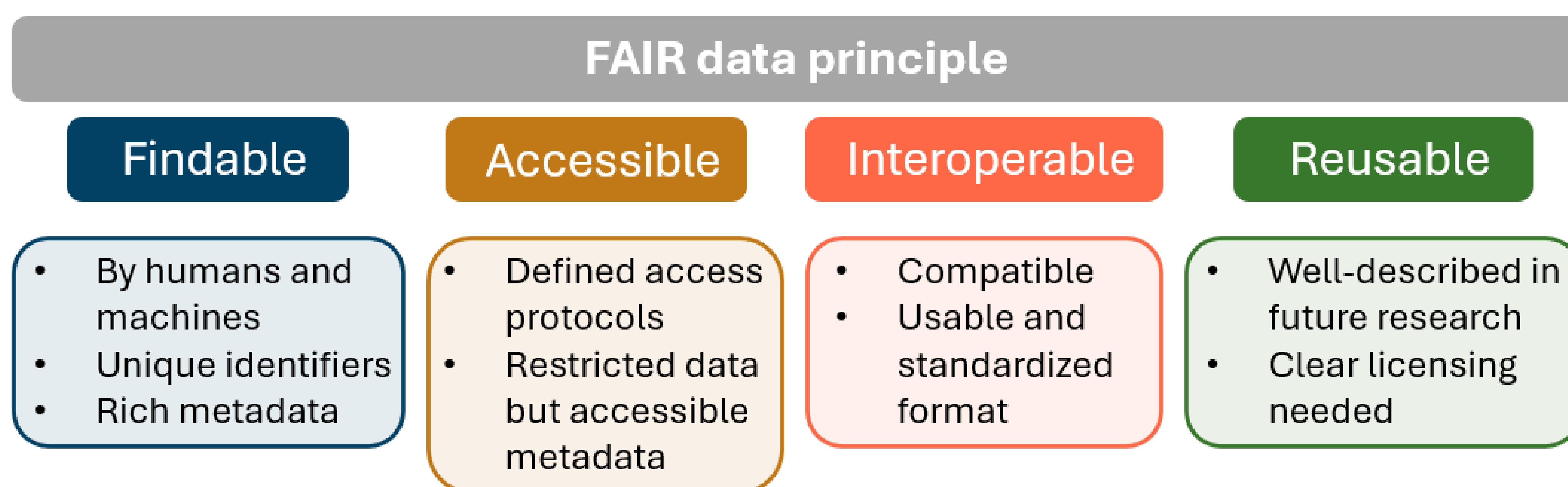


GUIDELINES

How to make biosecurity assessment data FAIR?

Data on the implementation of biosecurity measures on livestock farms are collected worldwide. However, if stakeholders have no insight into what biosecurity data is collected, where, how, when, and by whom, there is a risk of duplicating assessments and collecting the same data repeatedly, which is inefficient. To tackle this, collected data should become findable, accessible, interoperable and reusable (FAIR) as much as possible. But how do you make biosecurity assessment data more FAIR? These guidelines will show some basic principles and handles to start with. Before going there, let's first introduce the FAIR principle!

The FAIR data principles were developed to enable and enhance the reuse of data in 2016. These principles emerged as a response to the growing volume of data being generated and the need for a standardized approach to ensure the long-term value, utility and integrity of data. The FAIR principles are a guide to assist in how to make data reusable by others, humans and machines, they do not suggest any specific technology or standard themselves.



Findable

It should be possible for others to find the data. This does not necessarily mean the data itself should be findable but at least the metadata (documentation, descriptive information about the data) should be. These metadata should be as rich and detailed as possible. Findability of data can be enhanced by providing the metadata or data itself with a unique and persistent identifier such as a Digital Object Identifier (DOI).

Accessible

The data should be available to others. When the data itself are not publicly available it should at least be clear how the data can be accessed, including any necessary authentication or authorization procedures. Even when restrictions exist, metadata should always remain accessible, providing potential users with information on how to gain access to the data.

Interoperable

It should be possible to integrate the data into other research. Therefore, data must be formatted in ways that enable compatibility with different platforms, tools, and research disciplines. Furthermore, it should use standardized languages, vocabularies, and ontologies to describe the data, making it easier for both humans and machines to understand and combine datasets.

Reusable

Data should be accompanied by clear licenses detailing usage rights and restrictions, ensuring that it can be reused in different contexts. Reusability enhances the longevity and impact of datasets, maximizing their potential to contribute to new discoveries and innovations.

Make data findable

Step 1: Describe the data with rich metadata (detailed information about the data) including:

Standard general metadata

- Title of the data
- Creator of the data
- Persistent identifier link
- Version history if needed
- Audience to whom these data might be useful
- Information about the availability of the data
- Information about the license: how to cite them
- The file format of the data
- Description of the data

Biosecurity compliance specific metadata

- Geographic location of data collection
- Year or timeframe of data collection
- Species and production system (type of farms) in which the data is collected
- Reason for data collection
- Language of data collection

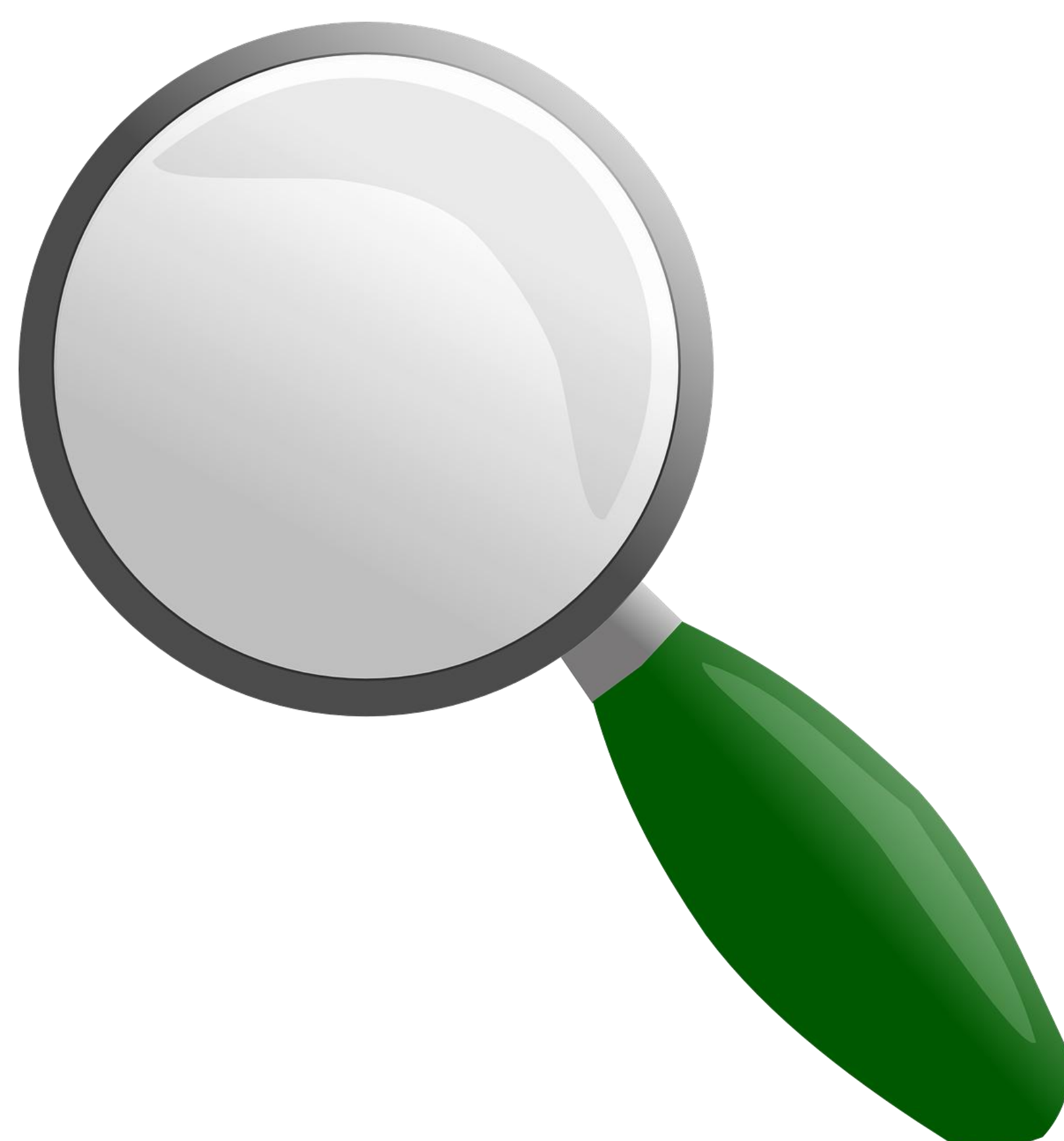
=> These metadata can be given separately or can be mentioned in one section such as 'Description of the data'

Step 2: Deposit the (meta)data in an online trusted repository to assign a persistent identifier

Making sure the (meta)data are referred to with a persistent identifier ensures that the (meta)data can always be found even when the web address changes and it is unique for these specific data. This persistent identifier can be used as reference when others are reusing and citing the data. Furthermore, it can be placed on the website of the institution collecting the data as a unique link where more information can be retrieved.

Step 3: Deposit the protocol used for data collection in an online trusted repository

Many institutions collect data about the implementation of biosecurity measures on farms. Some are doing a complete farm assessment, others are only interested in one aspect of biosecurity. Often each institution develops its own protocol/survey to collect this information and by sharing the protocols others know exactly which data are collected. If interested in the same information, the institution can be contacted for a collaboration instead of reinventing the wheel and duplicating the work. In case the protocol changes due to new insights or changing priorities, sharing this new protocol is recommended. Depositing the protocol can be done separately, giving the protocol a persistent identifier and metadata or it can be uploaded together with the data as a supporting document to understand and interpret the data.



Make data accessible

Step 1: Make metadata accessible

Metadata should always be accessible, even if the data itself are not. By sharing the metadata people can take action to retrieve more information preventing already the duplication of work and data collection.

In case the data are NOT immediately accessible

Step 2: Provide clear instruction on how the data can be retrieved

- State who should be contacted to receive access to the data
- State what protocol will be followed to receive access to the data e.g., which identification/information of the third party is needed, what type of agreement should be signed, etc. Then the interested party can immediately prepare those documents.
- State how the data sharing will be done, which will depend on the format and the size of the data. Emailing the data upon request is an option with limitations such as size and safety. Other specialized data sharing platforms can be considered such as Cloud Files for which identification steps can be incorporated.

Step 3: Prepare the data to be shared

Although the data are not accessible by others immediately, it is important to collect, store and identify the data with sharing in mind. This means formats are used that can be read by anyone (interoperable) and that information can be provided on how others should cite the data once shared (reusable).

In case the data are immediately accessible

Step 2: Do not make personal data accessible

Personal data including but not limited to names of farmers, exact locations of farm, contact details etc., cannot become public unless the person has given written consent to do so.

Step 3: Share the data

Make the data available for the public in a format that can be read/used by anyone (interoperable) and provide information on the license of the data and how to cite (reusable). There are two main ways of making the data directly available:

- Via the website of the institution or company collecting the data
- Via an online trusted repository which is the preferred way of data sharing as this provides the data with a persistent identifier and supports the description with metadata (findable).



Make data interoperable

Step 1: Give the data a standardized name

Always put the same type of information in the file name

Step 2: Supplement the data with ReadMe files

ReadMe files contain all necessary information that is needed for others to understand and interpret the data. For example, if calculations are present behind the provided data, describe them in detail in this ReadMe file. Make standardized ReadMe files which can be used multiple times and easily adapted when necessary. Furthermore, provide the ReadMe files in a format that is accessible by the public, for which no specific software is needed.

Step 3: Use standardized vocabulary and terminology

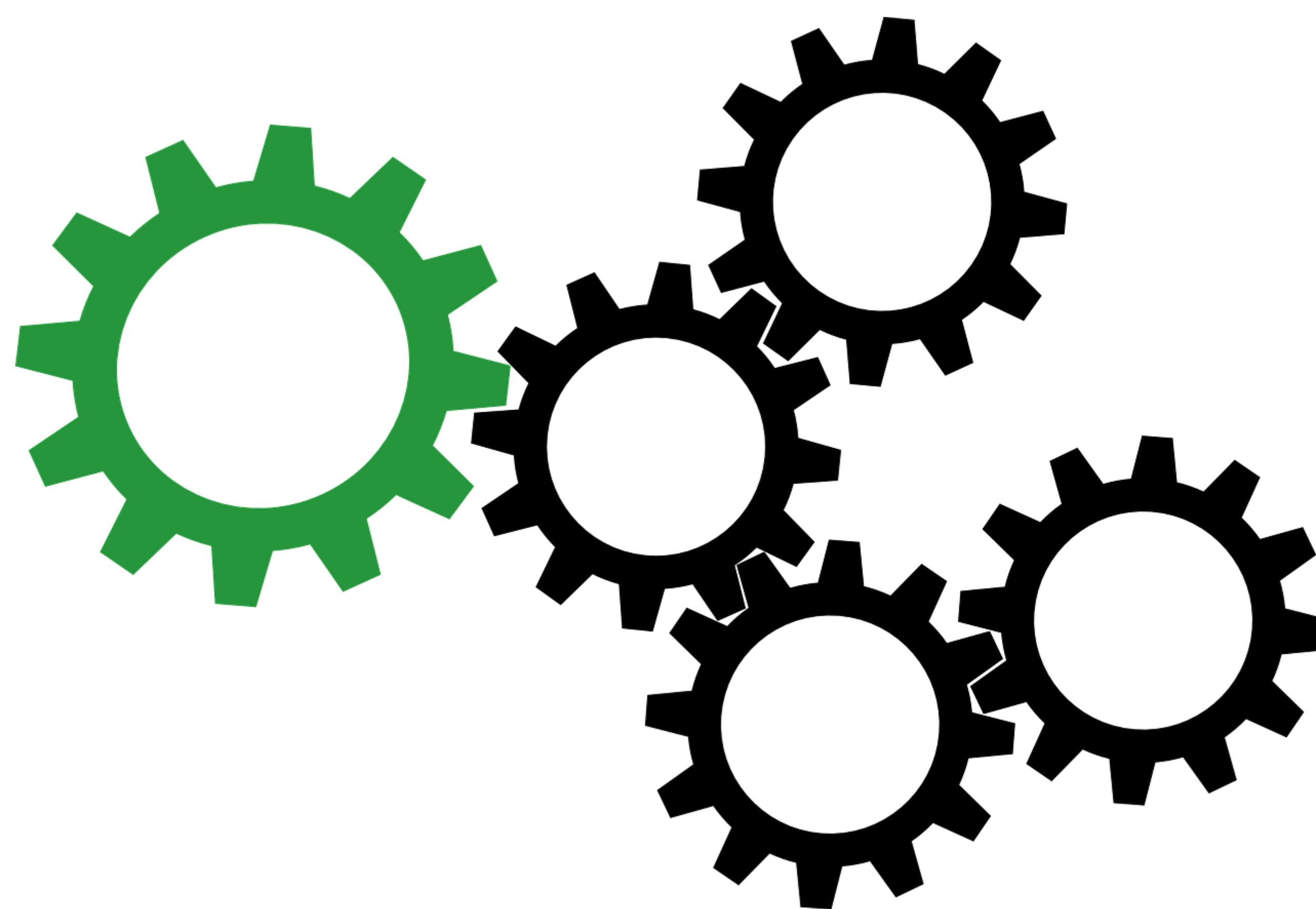
As data are often collected across borders it is important to use standardized wording. Some examples of internationally accepted norms:

- Time: yyyy-mm-dd ([ISO 8601](#))
- Country codes ([ISO 1366](#))

No standardized vocabularies exist specifically for the field of animal biosecurity. However, the World Organization for Animal Health developed a glossary to understand the Terrestrial Code and includes words in the field of biosecurity ([WOAH Glossary](#)). It is recommended to always include a code book providing detailed information about the terminology used.

Step 4: Provide or deposit data in an accessible file format

When sharing data, others should not only be able to read the data but also to integrate it into their own research meaning the data should be extractable. Data about the implementation of biosecurity measures on farms are often collected by completing surveys or checklists. These data can be transferred into Excel files for example which are easily readable and make it possible to extract the data. In case specific software is needed to make the data readable and usable by others, this should be mentioned clearly and preferably a link provided to the needed software in the metadata description.



Make data reusable

Step 1: Clearly state the license that is attributed to the (meta)data

The license that is attributed to the (meta)data describes how others can use these (meta)data. The most commonly used licenses are the Creative Commons (CC) licenses ([CC license types](#)).

The FAIR principle states that metadata should be licensed under CC-0, which means metadata belong to the public domain. The metadata can be used, shared, etc. without any conditions. The FAIR principle stimulates to share data under the CC-BY license meaning that others can distribute, adapt, continue working on the data and even use them for commercial purposes as long as credit is given to the creator. Variations exist on the CC-BY license such as NC (only non-commercial use) or ND (no derivations or adaptations are permitted). Important to note is that once a license is attributed to data, it cannot be changed. Even if the license is changed afterwards, when people receive data with a certain license that license stays valid for their use. The type of license attributed to the data should be mentioned clearly in the metadata.

