

Practice Abstract #42

Transforming large quantities of diverse agricultural data for Findability, Accessibility, Interoperability and Reusability

The aim of the EUREKA project (<https://www.h2020eureka.eu/>) was to test the feasibility of building an open-source knowledge reservoir (the **EU FarmBook**) for making the results of EU-funded research and innovation projects in food, agriculture and forestry more widely accessible (<https://eufarmbook.eu/>). Many of the lessons learnt during the project will be of interest to the developers of other similar digital platforms.

An important aspect of creating such an agricultural knowledge platform is gathering data from numerous sources and transforming it to become more FAIR (Findable, Accessible, Interoperable and Reusable). In order to achieve this goal, data should be stored together with metadata that attaches semantic meaning to the objects. The recommended way to achieve this is to make use of a Resource Description Framework.

To increase interoperability, widely used agricultural ontologies such as AGROVOC (<https://www.fao.org/agrovoc/>) should be used to attach common understanding to objects. Additionally, for non-domain specific information, common ontologies (e.g. www.schema.org) exist which allow the general properties of data and other knowledge objects to be conveyed.

Rather than developing custom solutions, storage of such data should involve existing digital repository solutions such as DSpace to greatly reduce the amount of manual development work.

In order to ingest the data into such a repository objects will need to be extracted from various locations, transformed into an understandable format and loaded into the system. Although most of such a workflow can be automated, some manual effort by experts should still be expected in order to map unfamiliar concepts and formats to an understandable format.

