Workshop on a Common European Agricultural Data Space in the Digital Europe Programme

(INPUT BY DEMETER)

The DEMETER vision:

The H2020 DEMETER project (https://h2020-demeter.eu/) has a vision towards the cooperation between and the possible federation of existing Farm Management Systems (FMS) thus allowing the usage of applications and data from multiple platforms and various providers. Its stated goal is to adopt advanced technologies, such as Internet of Things, Artificial Intelligence, Big Data, Decision Support, Benchmarking, Earth Observation, etc., in order to increase performance in multiple aspects of farming operations and to put these digital technologies at the service of farmers using a human-in-the-loop approach that constantly focuses on mixing human knowledge and expertise with digital information. To facilitate the federation of existing systems, platforms and data sources it focuses on interoperability as the main digital enabler, extending the coverage of interoperability across existing data, platforms, services, applications. It also adopts a modular approach to application design (aided by the aforementioned interoperability) that allows to create new apps from existing data sources, sensors, services and platforms.

To achieve this vision, DEMETER provides two indispensable contributions (among many others):

- The design and implementation of a Reference Architecture (RA) that is suitable to address the aforementioned challenges. This follows a modular approach and its key components is a Brokerage Service Environment (BSE) and the DEMETER Enabler Hub (DEH). These allow various providers to give data, sensors or other apps (e.g. semantic interoperability enablers) that can be used to form a complete system from various existing or new components.[1]
- The design and delivery of the DEMETER Agriculture Information Model (AIM) that specifies the common data models used to represent agricultural information that are handled by FMSs. This is the model that forms the basis for enabling semantic interoperability between DEMETER and existing agrifood systems and ontologies (used for example by existing FMS and sensors). It has been developed by taking concepts from the dominant existing ontologies, such as FIWARE (Agri), ADAPT, Saref4Agri, INSPIRE and FOODIE, etc., as well as from the well known AgroVoc vocabulary. This allows the semantic mapping of those ontologies and systems to AIM, which is a key factor for semantic interoperability that in turn enables collaboration between and federation of FMSs and existing platforms.[2]

[1] Details about this RA can be found in deliverable D3.1 of DEMETER.
[2] Details about AIM and the semantic interoperability support can be found in deliverable D2.1 of DEMETER.

==========

1. Is the federation of some of the Farm Management System (FMS) platforms and other data platforms feasible?

The DEMETER Agricultural Information Model (AIM) approach aims to enable federated data management for the food and agriculture sector. By offering AIM that specifies the common data models used to represent the data handled by FMSs and related applications in the agrifood domain, this is a key step towards achieving semantic interoperability, which in turn is a necessity for the federation of platforms such as FMS (see also the answer to question 2 in this document).

Thus, we envisage that it might be feasible to achieve a loosely coupled Federation of Distributed FMS, although it might need public support and participation with public platforms like (IACS) "Integrated Administration and Control Systems". In addition, the European data portal(s) (https://data.europa.eu/) provide a way to expose provided datasets (metadata and data) to re-users. From a more general perspective, the federated architectures of International Data Spaces (IDS)[3] and the more recent GAIA-X, which explicitly aims at delivering a technical solution for the data spaces envisaged by the European Data Strategy[4], have been designed to accommodate a broad range of legacy systems and platforms. In other settings, it has been proven that a federation of existing platforms is possible by following these approaches. DEMETER is committed to building on previous work like IDS. Going beyond approaches that, like IDS, focus on data exchange, DEMETER will require service and API brokerage. The latter, in a way compatible with IDS data exchange, is what GAIA-X aims at.[5]

Now, as already briefly discussed, semantic interoperability is the key technology for achieving this. Common to federated systems is the use of a common semantic model and protocol, to which each participant agreed to adhere and behave. For instance according to the Department of Defense (DoD) Architecture Framework[6], in order to federate systems architectures, there must be elements of semantic agreement so that pertinent information can be related appropriately. Ways to achieve that include adhering to a common framework that includes definitions of common data elements, semantics, and data structures, or the use of enterprise taxonomies (vocabularies) for aligning different models. DEMETER supports this via the definition of the aforementioned Agriculture Information Model (AIM), which provides a common vocabulary and underlying semantics, enabling different systems to exchange data. Each of those systems may use and produce data using different models, but they need only to implement wrappers that will allow the transformation (translation) of such data to and from the common language to expose and consume data of the federated systems. DEMETER will also provide tools that facilitate the creation of these wrappers.

[3] IDS Reference Architecture Model v3.0, 2019. https://www.internationaldataspaces.org/publications/reference-architecture-model-3-0/

[4] GAIA-X: A Pitch Towards Europe. Status Report on User Ecosystems and Requirements. June 2020. https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Publications/gaia-x-a-pitch-towards-europe.pdf

[5] GAIA-X: Technical Architecture. June 2020. https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Publications/gaia-x-technical-architecture.pdf

[6] DoD (2007) DoD Architecture Framework Version 1.5. 23 April 2007.

==========

2. Assuming that the implementation option for the Common European Agricultural Data Space for agriculture is based on a federated distributed system of existing data platforms, what is needed to implement a European data space from a technical point of view (definition of the interoperability mechanisms)?

Achieving semantic interoperability between the various platforms and systems is the key technology needed to be able to federate these systems. As discussed in more detail in DEMETER Deliverable 2.1 section 8 (Semantic Interoperability Support), *interoperability* needs to be considered at multiple levels in a distributed system:

- *organizational level*: coordinated processes in which different organizations achieve a previously agreed and mutually beneficial goal (this goal could pertain to satisfying desired policies).
- *semantic level*: precise meaning of exchanged information, which is preserved and understood by all parties, which includes the desired behaviour of assets (agents, machines, systems) and data exchanged.
- *technical level*: planning of technical Issues involved in linking computer systems and services, part of which is to achieve syntactic interoperability and to define the communication transport between assets.

The DEMETER approach focuses on bridging best practices at the organisational and technical levels with a common semantic model (DEMETER AIM) which is created by collating together and adapting terms and concepts for the best known and widely used ontologies and platforms for the agri-domain. This is the basis of the AIM data model which is turn facilitates interoperability between DEMETER and existing platforms as long as the necessary AIM compatible data wrappers are created to translate data between AIM and any existing ontology. Thus, AIM can be used to derive simplified technical implementation views which can then be explicitly mapped to existing and future data sources, and drive transformations of local data models into a common view needed by federated systems.

To sum up these key points, AIM constitutes the basis to enable the *semantic interoperability* between different systems and data models. In fact, semantic interoperability concerns functionality (e.g., provided by services) involving semantic (data) integration as well as, e.g., interoperability of services and tools made possible and driven by semantic integration.

Note that DEMETER's notion of interoperability is in line with cross-domain approaches such as the IDS Reference Architecture Model and with upcoming broader approaches such as GAIA-X. With "Agri-Gaia"[7], a preliminary idea for an AI ecosystem for the SME agricultural and food industry based on GAIA-X has already been defined. The corresponding working group within the GAIA-X community has not been active yet, but is scheduled to start in September 2020. This expert workshop thus provides a good opportunity for reviving it and establishing connections between DEMETER and Agri-Gaia.

[7] https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Artikel/UseCases/agri-gaia.html

Furthermore, the DEMETER Reference Architecture (RA) aims to facilitate the merging of different sensors, data sources (in general) and existing platforms into integrated apps, which then allows the collection, processing and usage of the data used by DEMETER enabled apps (such as those now used in the DEMETER pilots) and provides an integrated view over different and heterogeneous datasets that can support the discovery and extraction of new knowledge, as well as the decision making of farmers and other stakeholders. The various components, services and data sets (or data sources) are offered through the hub as services for hire and consumption into the final DEMETER-enabled apps, and can be composed in to a homogeneous app through the usage of the data wrapper enablers made available and developed (mostly by third parties) through use of the interoperability support tools. Remember that AIM provides the common vocabulary that is used to exchange data between different components, and to provide the integrated view over different and heteronomous data sources (semantic data integration).

==========

3. How can we reach an agreement on a set of interoperability mechanisms (avoiding locking into existing platform architectures)?

It is necessary to agree on a common data model in order to achieve interoperability. The DEMETER AIM approach is to build on a framework of interoperable data models published either by standards bodies and other coordinating activities such as EU-INSPIRE and DCAT(-AP), or that are well-known and used in particular domains. Moreover, in line with best practices and recommendations, the specification of DEMETER AIM follows a modular approach in a layered architecture, enabling among others:

1. eased interoperability with existing models by reusing available (well-scoped) models in the modules, instead of defining new terms, whenever possible,
2. easy mapping/alignment with other models, by module instead of the whole model,
3. easy extension of the domain/areas covered in AIM with additional modules,
4. easy extension of the domain model, by modifying only specific modules,
5. easy mapping to top-level/cross-domain ontologies.

This limits the scope of agreements needed to:

- domain specific terminology needed for specific data federation goals
- adoption of low-level syntactic and API level interoperation - or provision of gateways and brokers able to perform the simple task of syntactic transformation
- semantic translation specifications - understanding how to map meaning across different data elements.

Semantic translation support is by far the hardest interoperability mechanism to achieve. DEMETER is addressing this challenge by:

- Publishing formal data models and controlled vocabularies as Linked Data and including available semantic alignments and relationship information within the published views of the model.

- Implementing brokers to exercise semantic and syntactic translations using these resources.
- Generating alternative syntactic models from the common semantic model in a way which provides traceability from syntactic elements back to the underlying definitions in the semantic model that allow for meaning to be discovered.

The mechanisms for data translation are also part of broader pipelines for the generation of linked data using AIM as underlying model.

IDS and GAIA-X provide a basic architecture for platform-independent interoperability, plus the governance infrastructure and processes for fostering agreements among multiple stakeholders. Thus, agreement around an Agricultural Data Space could be reached by working within these established bodies of government, or by establishing new ones that follow the best practices established by IDS and/or GAIA-X. However, existing platforms have also to evolve to new architectures addressing new policies like CAP post 2020.

==========

4. Are the suppliers of FMS ready to share their data? And willing to federate their data platform with other suppliers?

It cannot be taken for granted that supplier of FMS are ready or willing to share their platforms or data. DEMETER provides tools and its RA would enable the consumption of these as provided services, but it does not explore fully the incentives (monetary or otherwise) that need to be given in order for provider to make these available. Afterall, modern economic theory assumes that entities and individuals are self interested, and they would therefore need to be compensated appropriately in order to share data and services. Note that data has intrinsic value and that the providers of this data or of FMS platforms have invested time and money to set up sensors and the rest of their system. Therefore, either monetary compensation or an exchange of services (be it data or access to processing software for example) needs to be offered most likely to incentivize such collaboration. Even though, DEMETER does not explore these incentives, it does provide the tools and infrastructure however so that these incentives and the related business models can be explored using DEMETER's RA, hub and tools/services.

In addition, data owners need to:

1. be enabled to make their data FAIR[8] (findable, accessible, interoperable, reusable), and
2. be given trust, by data sovereignty mechanisms such as data usage control by mechanisms that technically force data consumers to comply with the data usage policies defined by data owners. In the IDS the latter has already been implemented[9].

[8] https://www.go-fair.org/fair-principles/ – not to be confused with "Open Data"; the FAIR principles can also be applied to non-open data.

[9] Whitepaper "Usage Control in the International Data Spaces", IDSA 2019. https://www.internationaldataspaces.org/papers-studies/#positionpapers

Nevertheless, it should be noted that in order to fully realize the era of digital farming and to obtain its benefits, sharing data between different partners in the agri-food chain is required. However, this can only happen it the sharing is being conducted in a fair and transparent way. That's why a coalition of associations from the EU agri-food chain launched a joint EU Code of Conduct on agricultural data sharing in Brussels on April 2018[10]. The signatories believe that access to accurate agricultural data is vital to develop digital farming enabling farmers and cooperatives to produce more using less resources.

==========

5. Which existing platforms supported by ecosystems (at regional or national level) are already sharing data? In which sub-sectors are they sharing the data?

European data portals covering different sectors including agriculture[11] [12], European statistical information related to different sectors like Eurostat (e.g., agri indicators[13]), national open data portals (e.g., Polish open data covering different sectors including agri[14]), the Open Linked Data Cloud with data related to different sectors[15], agri specific datasets like the farm accountancy data network[16], the EU agri-food data portal[17] (including CAP indicators, markets and economics), or geospatial datasets (e.g., Sentinel[18]). One example for a national data marketplace that is based on the same advanced technology as the DEMETER platform is given by the future version of the Mobility Data Marketplace to go live in early 2022: the tender, which is currently being negotiated, demands an IDS-based architecture[19].

==========

6. Which public data sets would be of particular relevance for increasing the effectiveness of the Common European Agriculture Data Space?

Based on experience with the initial preparation and execution of the DEMETER pilots (which is currently ongoing), it looks like certain types of data sets would be pertinent. Below, we identify such data that could come from public data sets, but obviously these are not the only ones needed and data from proprietary sources might also be needed.

First, data sets regarding EO and satellite image data are definitely necessary. Some of these like Copernicus are public. More specifically, data sets and images from the National Regional IACS, FMIS, IoT Platforms, Copernicus Program Sentinel 2 and Galileo, LOD cloud public EndPoints could be used and be useful.

[10] https://t.co/cu4BRaQJqQ
[11] https://www.europeandataportal.eu/en
[12] https://data.europa.eu/euodp/en/home
[13] https://ec.europa.eu/eurostat/web/agriculture/agri-environmental-indicators
[14] https://dane.gov.pl/
[15] https://lod-cloud.net/
[16] https://ec.europa.eu/agriculture/rica/
[17] https://agridata.ec.europa.eu/extensions/DataPortal/home.html
[18] https://sentinel.esa.int/web/sentinel/sentinel-data-access
[19] https://www.evergabe-online.de/tenderdetails.html?0&id=322425

Second, agri-related statistical and economic data, e.g., regarding the efficiency of running particular farms or plots of land are used for benchmarking in certain pilots. For example, FADN (Farm Accountancy Data Network) data together with Eurostat agri indicators and CAP indicators are being used.

Third, information regarding a number of farm operations, e.g., regarding crops, crop yields, rest and disease statistics, soil maps, weather data and forecasts (e.g., from public meteo services and meteorological national agencies that publish such data). In addition, data regarding farm animals, or agricultural machinery would also be very relevant, as well as available databases on plants and seeds. Finally, data regarding the prices of agri products, as well as the costs of various consumables/resources, such as fertilizers, pesticides, water costs, energy costs, fuel prices, etc. are useful for determining the operating costs and to gauge the efficiency of agricultural operations.

==========

7. Are their experiences with taking public data sets as input to FMS, farmers' applications or agricultural

DEMETER facilitates the use of nearly any type of data source included in DEMETER-enabled apps, provided that the appropriate data wrapper to/from AIM is also available. While it is not the goal of the core of the project to look at particular data sets, but rather the ontologies and vocabularies that such data is represented by, and to facilitate interoperability with these, nonetheless several difference data sources and data sets are used by the individual DEMETER pilots.

First, geospatial and EO data are being widely used by most of the pilots. More specifically, Copernicus Sentinel 2 Images are being used by a number of pilots for agriculture with AI process addressing DSS services for smart farming, environment, sanitary control, etc. These are combined with the data from other proprietary source, e.g. images taken from privately owned drones.

Second, weather data and forecasts both from public and from private weather stations are being used as input to determine e.g. irrigation needs.

Third, DEMETER pilots develop decision support systems, which use and rely on open data like FADN, Eurostat, as well as geospatial data from Sentinel and other missions, in order to facilitate applications such as benchmarking the productivity and sustainability performance of the farms, which involves monitoring different conditions and parameters affecting such indicators, collecting the data and integrating it in a unified layer accessible by the appropriate DSS.

Editor:

Ioanna Roussaki (ICCS – NTUA)

Contributors:

Kevin Doolin (TSSG)

Raul Palma (PSNC)

Poulakidas Athanasios (INTRASOFT)

Mariano Navarro (TRAGSA)

Christoph Lange (Fraunhofer FIT)

Anja Linnemann (Fraunhofer FIT)

Óscar Miguel (Vicomtech)

Ioannis Vetsikas (ICCS)

Nuria De Lama (ATOS)