Growing Smart Farming Services

How to get the best out of farming data

Susanne Braun, Marcus Trapp, Claudia Nass, Matthias Gerbershagen, Stefan Schweitzer, Rodrigo Falcão, Fraunhofer IESE, Kaiserslautern
Markus Schweitzer, Nikolaus Wire, John Deere European Technology Innovation Center, Kaiserslautern

Abstract
Many enterprises in the agricultural sector can access a lot of data, but often (still) fail to treat this data as an asset. Even though farming data has great potential for the creation of new business models, most manufacturers of agricultural equipment do not know which data they possess as it is spread across different data silos of complex and historically grown application landscapes. In order to come up with innovative data-driven smart farming services, seamless collaboration of cross-functional teams staffed with different roles like software engineers, data scientists, business experts, and end users is crucial for the successful realization of smart farming services.
In a joint research cooperation, John Deere and Fraunhofer IESE have developed an agile engineering method for the value-driven creation of smart farming services. In this paper, we present our method, which is based on different design thinking techniques used to identify concrete data-driven user stories that are of true value to farmers. We also present a data science notation that we developed to facilitate the targeted analysis of the data-science-significant requirements of data-driven user stories and to foster interdisciplinary collaboration and knowledge dissemination in our cross-functional team. We also demonstrate the Insights Collaboration (short: ICSpace) app, which supports the application of the notation.

Introduction
With the creation of new business models based on insights generated by data-driven services, data is becoming an important business asset. This is mostly addressed from a purely technological standpoint, using bottom-up approaches for the creation of all-encompassing data catalogs (e.g., [1]) or canonical data models. The result is an overwhelming collection of data. Even if enterprises have a perfect overview of their data assets, they still have problems

leveraging them because it is hard to come up with innovative data-driven services by merely looking at a large data catalog.

An outstanding data-driven service delivers insights to the users that are novel to them and of added value regarding some business concern. Therefore, the users' domain and problem space needs to be very well understood. Data analytics and machine learning models must cover all relevant real-world aspects. Besides that, technical feasibility requires a number of "data-science-significant requirements" to be met. We use this term to summarize any requirements that need to be fulfilled so that the data science model backing the data-driven service can be implemented and run in production in such a way that it delivers results with sufficient quality. Examples are: Which data is required? With which quality, in which volume or amounts, with which resolution, frequency, and currency are data items required so that the model can deliver results with sufficient quality and accuracy? Which software-architecture-significant requirements need to be derived so that the required data processing rates and volumes can be realized? Which legal and data privacy regulations are associated with datasets, and is analysis of the datasets, as required by the data science model, permitted?

The bottom line from the research cooperation between John Deere ETIC and Fraunhofer IESE on the development of smart farming services over the course of more than three years is: The individual contributions making up a data-driven service developed by different experts such as data scientists, software engineers, and domain experts are much more interrelated than in traditional software engineering projects and hence require much closer collaboration and coordination. Our goal is to provide optimal technological and methodological support for this challenging cross-disciplinary collaboration. This motivated us to develop an agile engineering method and a custom app supporting the method, both called ICSpace.

We evaluated the ICSpace app and method in a project, the Plant protection Application Management (PAM) project. PAM aims to simplify the agrochemical application process, by calculating buffer zones for protected areas and integrating them into machine-readable intelligent application maps. The required data has a broad variety of sources and frequencies ranging from application machinery via plant protection products to terrestrial elements adjacent to the field. Considering the wind (with live weather data) could enable buffer zones to adapt to the working conditions. This could allow an increase of the applicable area while at the same time improving spray drift mitigation [2]. The ICSpace app and method helped to clarify whether all required data is available while operating in the field, and whether it is available at an adequate resolution and update rate.

A Graphical DSL for the Analysis of Data Science Significant Requirements

At the core of the ICSpace app and method is a graphical DSL (domain-specific language) that has been designed to facilitate the targeted and joint analysis of the data-science-significant requirements. We intentionally designed the language to be very simple and easy to understand, even for non-technical users. The language consists of only four graphical elements: Business Question, Dataset, Processing Step, and Data Source (Fig. 1). Note that data sources may have different color codes, which will be explained in more detail below. As depicted in Fig. 1, elements are organized as trees, where the root element always has to be a business question.

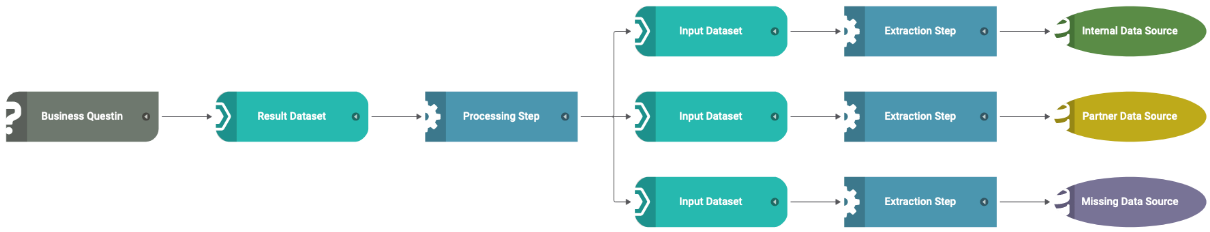

Fig. 1: Valid ICSpace tree with minimal set of elements

As the goal is to gain new insights based on data analytics regarding the business question, the envisioned result is represented by a dataset. Thus, the child of a business question is always a dataset element. Dataset elements are coherent collections or aggregates of data usually representing some real-world concepts or entities. Processing steps process and transform datasets, producing new datasets as a result. Data extraction or data cleaning steps, calculations or transformations of data, as well as statistical or machine learning models are examples of processing steps. Datasets and processing steps are alternating elements within the tree, but the leaves have to be data source elements. The latter is an important design element of our DSL: we wanted it to be completely obvious (even visually) that all datasets required for the realization of the service are available and accessible in some internal or partner data source. If a required dataset is not available, it must be connected to a data source marked as “unavailable”. In the ICSpace app, data source elements can represent databases, but also services delivering data such as REST services. Data sources can be classified as internal, external (e.g., at the site of a partner or a public site), or unavailable. Depending on this classification, the color style of the graphical element is adapted accordingly (see Fig. 1). For each element, additional metadata can be included. Users can add descriptions in prose to all elements. Also, any number of open issues can be added. Elements with unresolved open issues will be rendered with an orange indicator to make it easier to keep track of them. Depending on the element type, users can add specific metadata. Processing steps can be enhanced with alternative approaches and constraints, like legal regulations or company policies, with which the computation is compliant. Furthermore, users can document known limitations of the procedure if, e.g., a certain business case is not supported or the procedure is not applicable to a specific user group. Also, any relevant documentation can be linked to the processing step. In case the processing step contains a REST API call, the method and request URL can be documented. Details on the data model, sample data, and the origin of the data can be provided for dataset elements. For data sources, details about the database schema and information on how to access the data can be included. If a data source represents a data service provider, relevant access information can be captured also, as can a link to the API documentation.

Additionally, the ICSpace app provides the possibility to specify alternatives at the subtree level. A dataset can be calculated, e.g., by two (or more) alternative processing steps depending on different data sources. The currently chosen alternative can be visualized (see Fig. 2). Having different algorithms producing the same data set but with different levels of accuracy is an example of modeling with alternatives.

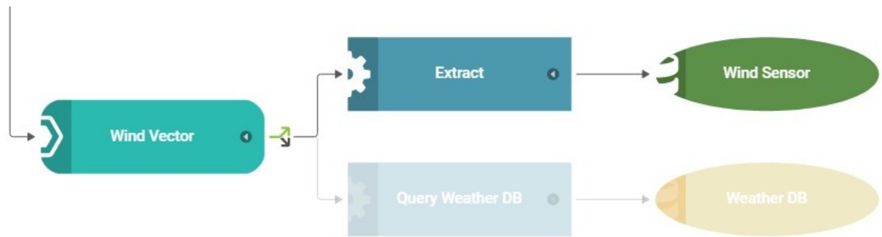

Fig. 2: Data set calculated by alternative data sources

The Deere Engineering Method

As mentioned above, the goal of ICSpace is to provide optimal technological and methodological support for this challenging cross-disciplinary collaboration. Starting from the business question, the notation facilitates the targeted and joint analysis of the data-science-significant requirements of the service. This is the common knowledge base for the team, linking together documentation and design artifacts of the data-driven service. But it also has the potential to become an organization-wide knowledge base for different teams, eventually documenting the key datasets, key data sources, and key data science models, delivering the key insights of the enterprise.

At John Deere, the engineering method is divided into two phases, which is directly supported by the app (see Fig. 3). The first phase is called “Project Exploration” and starts with a marketing representative and a domain expert jointly formulating the business question and sketching out the first levels of data dependencies in the app. After that, software engineers and data scientists refine the identified data sets by adding additional datasets and processing steps, down to the elementary data sources and services. Furthermore, the data scientist adds data quality specifications and assesses whether those can be fulfilled by existing data sources or services or whether further alternatives must be explored. The project exploration phase ends in a data dependency tree, which basically shows all the datasets required to answer the business question. Open issues like poor data quality or unknown data sources are documented and ready to be investigated further. The exploration phase typically takes place as a sequence of hand-offs between the different roles.

"Insights Collaboration" is the second phase, which takes place in multiple meeting-style sessions bringing together all involved roles of the cross-functional team to the ICSpace app for discussion, further refinements, and closing of open issues. These discussions open a lot of potential for developing a realistic solution plan for the smart farming service. Feedback cycles between understanding customer needs from the marketing perspective and mapping them to their data dependencies are shortened. Overall, this interdisciplinary group builds up common knowledge and empathy between the different roles over time and strengthens empathy with the customers' needs. The result of this phase is a mostly complete view of the data dependencies for the development of the service.

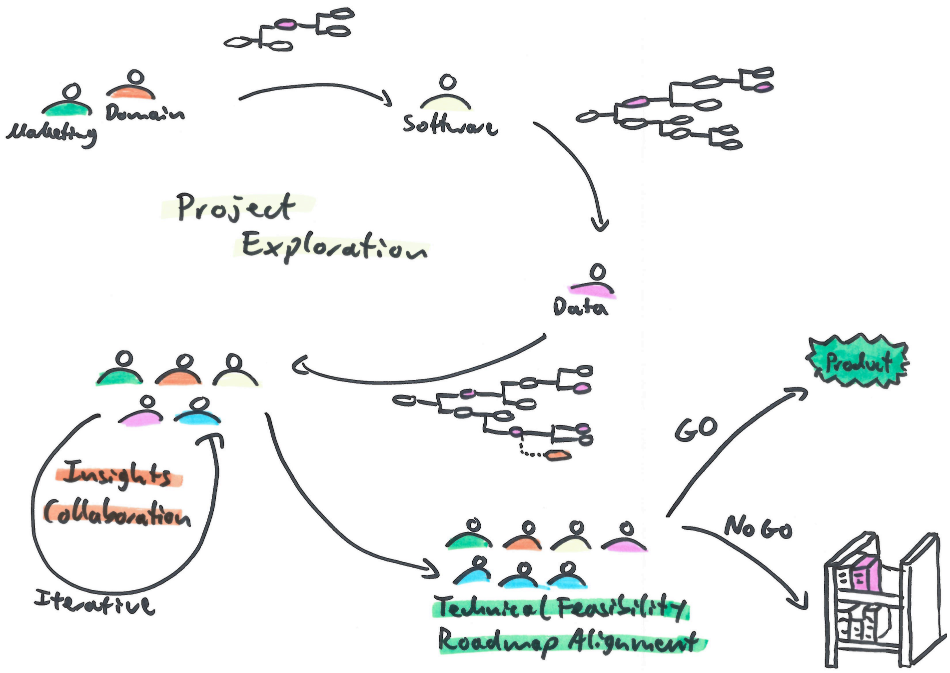

Fig. 3: The ICSpace engineering method at Deere

At the end, there is a project readiness check in analogy to readiness checks in agile software development. A formal review of the solution concept takes place and open issues with high impact are discussed. These do not necessarily kill a project but require special awareness. That is, quickly delivering a minimum viable product (MVP) and incrementally improving this MVP with customer feedback can still be a valid strategy and generate novel insights [3]. Additionally, a mid-term alignment of the roadmap with other enterprise stakeholders can take place in order to close gaps and solve the open issues for future development.

The ICSpace app supports this method in applying the DSL precisely and clearly structuring the identified data dependencies and open issues. The DSL and web-based app have been designed for minimal visual complexity so that they are understood by all kinds of stakeholders.

First Evaluation Results

We have been developing ICSpace for two and a half years. Right from the beginning, we started to use the app in internal projects to get continuous feedback for improvement. On the one hand, non-technical team members like agronomists or marketing representatives can use the app with the same proficiency as members with a technical background. On the other hand, the app greatly supports the evaluation of implementation alternatives in an easy way and can be used to transparently visualize to management if a project is too ambitious at the current point in time. Some project propositions have been put on hold given that the data dependency tree showed fundamental missing or inaccurate data. Thus, the data dependency tree greatly supports the development of mid- and long-term strategies regarding the creation or preparation of data sources. Furthermore, it supports decision-making in questions of strategic partnering with 3rd-party data providers. Most importantly, it prevented us from wasting precious capacities and supported us to focus development activities on services that were at the time technically feasible. In the ideal case those services also generate new data required to implement the next generation of smart farming services. After having received a lot of great feedback internally, we are now planning to open up our development activities and look for external partners, interested to collaborate with us on future product increments of ICSpace.

References

[1] "Collibra Catalog," [Online]. Available: https://www.collibra.com/data-governance-solutions/data-catalog/. [Accessed 6 September 2019].

[2] J. Feldhaus, Prozessoptimierung zur automatisierten Einhaltung und Dokumentation von Abdriftauflagen bei der Pflanzenschutzmittelapplikation in Flächenkulturen, Hohenheim, 2017.

[3] E. Ries, The lean startup: How today's entrepreneurs use continuous innovation to create radically successful business, USA: Crown Business, 2011.