2021 IEEE 23rd Conference on Business Informatics (CBI)

Conceptualizing Data Ecosystems for Industrial Food Production

Rix Calvin
Institute for Industrial Management (FIR) at RWTH Aachen University
Aachen, Germany
calvin.rix@fir.rwth-aachen.de

Stein Hannah
German Research Center for Artificial Intelligence
Saarbrücken, Germany
hannah.stein@dfki.de

Chen Qiang
German Research Center for Artificial Intelligence
Saarbrücken, Germany
qiang.chen@dfki.de

Frank Jana
Institute for Industrial Management (FIR) at RWTH Aachen University
Aachen, Germany
jana.frank@fir.rwth-aachen.de

Maass Wolfgang
German Research Center for Artificial Intelligence
Saarbrücken, Germany
wolfgang.maass@dfki.de

***Abstract*—Industrial food production represents one of the largest industries, accounting for a share of ten percent of the world's gross domestic product. Simultaneously, it is responsible for 26 percent of global greenhouse gas emissions. Due to increasing CO2 taxes and population's call for sustainability and CO2 reduction, it is facing challenges in terms of economic profitability and stakeholder demands. These challenges could partly be overcome by participating in data ecosystems in which data are refined as data products, understood, exchanged and monetized as economic goods. Despite large amounts of data, collected parenthetically along the value chain in food production, potentials of data analytics and data ecosystems are only marginally exploited. Food production mainly focuses on traditional, product-centric business models. This work shows the conceptualization of a data ecosystem for food production, enabling data-based business models. Therefore, resources, actors, roles and underlying relationships of future ecosystem are analyzed. Building on these, corresponding architectural and analytical artifacts that support data ecosystem exploitation are presented. A food production data ecosystem is exemplified by applying data analytics to compressor data, which reveals high potentials for CO2 reduction.**

***Index Terms*—data ecosystem, digital business models, ecosystem design, value stream mapping, industrial food production**

I. Introduction

With a share of ten percent of the world's gross domestic product, food production represents one of the largest industries nowadays [1]. In 2019, agriculture as one part of food production process employed approx. 27 percent of world population [2]. Therefore food production plays an important role both socially and economically. Though food production emits 26 percent of global greenhouse gas emissions [3]. Against the background of United Nations' sustainable development goals, action is required for food production. Goal 13 describes taking urgent actions in order to combat climate change and its impacts, as it is changing countries and disrupting economies. Goal nine aims towards sustainable solutions in economics and environment through innovation and industries. With Paris Agreement, 196 countries agreed to limit global warming to 1.5 degrees Celsius [4]. Several countries therefore raise taxes on CO2 emissions, which also increases costs in food production. In addition, population is calling for industries to reduce CO2 emissions. Reducing them in food production would on one hand lead to partly high investment costs for modern, climate friendly production machines and transportation means. On the other hand, if no efforts for CO2 reduction would be taken, the results would be (1) image loss and in consequence decreasing sales figures, (2) high costs due to CO2 taxes and (3) global issues, as climate targets become harder to achieve.

First steps towards a more sustainable food production could be taken by food industry through participating in a data ecosystem (DE). These can be defined as "set of networks composed by autonomous actors that directly or indirectly consume, produce or provide data and other related resources (e.g., software, services and infrastructure)" [5]. Within a DE, data are understood, exchanged and monetized as an economic good. Furthermore, these ecosystems can offer services for data analytics and in consequence generation of higher quality data products.

In food production, big data are collected parenthetically along the whole value chain [6]. Despite the high potential, data are seldom used beyond local process optimization in food production. Using, analyzing and trading raw or aggregated data within an ecosystem and combining data across companies enables to increase productivity and returns for companies [7]. In consequence, new data-driven business models evolve through data usage and analytics. Still, many companies are reluctant to DEs, as knowledge on potential value, data scientists or related skilled staff and trust in platform

This work is part-funded by the German Federal Minsitry for Economic Affairs and Energy (BMWi) under grant number 01MT19003B within the funding program "Smart Data Economy" (project: "EVAREST: Development and utilization of data products in the food industry via smart services).

2378-1971/21/$31.00 ©2021 IEEE
DOI 10.1109/CBI52690.2021.00031

security mechanisms are often missing. Within this work, the following research question will be addressed: *Which design elements are necessary for data ecosystem conceptualization?*

Therefore we outline the conceptualization of a food production DE by complementing the business ecosystem perspective with underlying technical concepts for ecosystem implementation. For this purpose, we adapt the methodology by Oliveira and Lóscio [5] and performed a case study combined of several workshops together with business model, data analytics and food production experts. Thereby we identified resources, roles, actors and relationships in order to conceptualize a DE in food production. A data ecosystem is based on these four "constructs" and defined as follows: "a set of networks composed of autonomous actors that directly or indirectly consume, produce, or provide data and other related resources (e.g., software, services, and infrastructure). Each stakeholder assumes one or more roles and is connected through relationships with other stakeholders, such that collaboration and competition among stakeholders promotes self-regulation of the data ecosystem". Hereby it differs from the basic concepts of Business Ecosystems (BE) and Digital Business Ecosystems (DBE). According to Moore [8] a BE is a economic community of interacting organisations and individuals who work cooperatively and competitively to support new products, satisfy customer needs, and eventually incorporate the next round of innovations. Advances in digital technology have led to the development of new collaborative organizational networks, giving rise to DBE. DBE is "a socio-technical environment of individuals, organisations and digital technologies with collaborative and competitive relationships to co-create value through shared digital platforms" [9]. Since the mere use of shared digital platforms not yet defined a DE, it concretes this concept.

After the introductory section, we present related work on business and technical DE design and conception. Next, we present the methodology for capturing resources, roles, actors with underlying relationships, enriched with approaches for deriving elemental technical concepts for ecosystem conception. Results of our case study will be described in chapter four, based on the example of CO2 reduction in food production ecosystem building on compressor data. Thereby, we present design elements and technical ecosystem features including added value for ecosystem participants. We conclude the paper by summarizing our results, review limitations and give an outlook on future work.

II. Related Work

There are three main aspects to overcome major hurdles in DE design: (1) considering the fear of participants in exposing valuable or sensitive data [10]; (2) constructing an ecosystem where participants can do more (added-value) with available data without the need of seeing all data [11]; (3) showing that all actors can benefit from ecosystem participation [10]. In this section, we review literature on DE design and conception, along with underlying technical platform aspects.

A. Data Ecosystem Design

Since DEs attracted a lot of attention, there has been an increase in recent publications. Nevertheless, research in this area is still in its infancy.

Tsai et al. [12] have derived the roles and responsibilities in a DBE based on a literature review. The proposal is generic in nature and therefore not focused on the manufacturing industry or DE. Further, the authors validated the roles, actors and the responsibilities in the context of preventive healthcare, which differs from the sector focus in this work. Akzan et al. [13] address the key aspects and perspectives of DEs in context of value co-creation, drivers, success factors, and in particular role-specific actors. Although certain approaches can be adopted for this work, the main focus is on industrial manufacturing industry. Furthermore, technical platform functions are disregarded. Cao et al. [14] present a dynamic cloud-based marketplace of near-real-time human sensing data (MARSA) where various stakeholders can sell and buy near-real-time data. MARSA is designed for environments where information technology (IT) infrastructures are not well developed, but the need for collecting and selling near-real-time data is high and was tested as part of a traffic information system. The work offers valuable insights into the design of a DE, but deviates from this papers' objective with regard to the object domain. Using a case study of six large companies, Otto et al. [15] demonstrate the importance of situational nature of master data for its value as a strategic resource. They identify data management, data use for information production, storage and maintenance as four phases of a data life cycle. As the authors do not consider the design of DEs, the basics of data processing were observed for developing data-based use cases in this work. Lange, Stahl and Vossen [16] focus on investigating and differentiating data marketplaces. The authors examine characteristics of data as an information good and explore the DE as a market and possible business models, characterized by specific pricing models. The paper provides a comprehensive overview of data marketplace types and data products to be traded. However, to the best of the authors' knowledge there exist no design approaches for DEs in the context of industrial food production. Russo and Albert [17] present four different levels of data monetization and point out the difference between internal and external use of data. Furthermore, they provide a DE in manufacturing as well as essential players and their business relationships. Still, they describe DEs at high level, so no detailed description of the business models or technical functionalities are listed. Savastano, Amendola and D'Ascenzo [18] examine the impact of a new DE on the value chain of a leading company in food industry. They describe opportunities and threats of ecosystem implementation. The content listed by the authors could thus be used as basis for analyzing the environment of food industry and the influence of digitization. The paper by Yu et al. [19] presents a big DE that focuses on implementing predictive maintenance with real industrial big data collected directly from large global manufacturing facilities. The authors address technical specifics of data ingestion from different manufacturing plants within real-time environment. Furthermore, a detailed insight is given on challenges and solutions, arising in context of data management and analytics with predictive maintenance. However, ecosystem design aspects in a cross-industry context remain unconsidered. Furthermore, the object area refers specifically to plant engineering.

B. Architectures for Data Ecosystems

In general, decentralized and centralized platform architecture approaches can be differentiated in DE design. Relevant approaches for this work are presented in the following.

The DE "International Data Spaces" aims to enable exchange of data between data providers and users. To secure this data exchange, a broker role, an App Store operator role and a certification authority are introduced [20]. Furthermore, implementing legal restrictions for data exchange, operationalized through electronic or smart contracts and supervision of the exchange and contract creation through a broker role could reduce fear of data misuse when exposing it within the ecosystem [11]. The concept of GAIA-X targets modular, distributed solutions, enabling data processing and hosting separately. Anonymized data can be brought together by networked structure where they are required for analytics. This allows sensitive data to remain in place, i.e., with the provider, while other data can be exchanged for processing and analysis [21]. From the IoT industry to the financial sector, data protection, data security and data traceability are decisive factors in data trading [22]. This has limited the conventional cloud platform with Software as a Service (SaaS) to the use of data from individual nodes to the central server. In addition, real-time computing plays a dominant role in many aspects of the industrial field. With the increase in the number of IoT devices, latency has become a bottleneck preventing the development of potential smart services for applications [23]. A decentralized smart service platform can solve this challenge, allowing edge devices to create valuable data products through smart service applications in a local security environment. It ensures that smart services run on the low latency network where data are protected in isolated environment, and provides peer-to-peer communication and modularized smart services.

Khalyly et al. [24] compared the existing microservices-based IoT platforms using a weighted score method to establish the relevant criteria. The result of the paper has concluded that the main problem with the IoT platform is that it requires the context of awareness and interoperability, as well as the challenge of the IoT platform in terms of data security, privacy and Devops methods. Wei and Li [25] present that the current two data trading modes, hosting mode and aggregation mode, cannot transparently track data ownership and usage. They suggested and verified the smart contract with blockchain and smart service as a solution to keep the data in the process of the data. Data acquisition in the real world is difficult due to data protection, traceability of data usage, data security and legal requirements (GDPR), especially data from multiple organizations [26]. This results in many isolated data sources that are not used together to create new valuable data products.

The literature review revealed that even if the current state of research provides building blocks to design DEs, there is no work that takes a systematic approach to design DE in the food industry. To the best of the authors' knowledge, this research gap holds not only for the food industry, but also for the entire manufacturing sector. In addition, there is no holistic approach of an underlying technical platform for DE that treats data security, data trading and automatized application of microservices to data of numerous market participants without the participants seeing data of other ecosystem participants. Furthermore, the combination of centralized and decentralized approaches in DEs was only marginally studied so far.

III. Methodology

The aspect of added value and especially value capture in and through DEs remains mostly unanswered [27]. Therefore, we aim to design a DE with clear added value and value proposals, e.g., by deriving monetizing strategies for data products and describing an exemplary DE use case, adding value for all participants. To get there, we adopt an ecosystem design methodology by Oliveira and Lóscio [5]. In the following, we present required design elements and the procedure applied in order to conceptualize a DE for food production.

A. Design Elements

Four basic model elements are necessary for DE design [5]: (1) Resources, (2) Roles, (3) Actors, and (4) Relationships. A DE is based on these four "constructs" as follows: "a set of networks composed of autonomous actors that directly or indirectly consume, produce, or provide data and other related resources (e.g., software, services, and infrastructure). Each stakeholder assumes one or more roles and is connected through relationships with other stakeholders, such that collaboration and competition among stakeholders promotes self-regulation of the data ecosystem" [5]. The elements are defined as follows:

- **Resources:** Resources include all products, services or capabilities that are produced, provided or consumed by actors. In DEs, resources range from data to software and services to infrastructure. Software mostly represents applications that are used to process or provide data. Resources can be exchanged individually or in combination. In addition, the resources usually comply with the predefined platform standards and are use-restricted by licenses. Finally, all resources can be evaluated with respect to various quality metrics.
- **Roles**: Roles represent different functions that an actor can take in a DE. These are associated with tasks and activities. Many different roles can be identified in a DE, which are characterized differently for each individual DE. Typically, at least data consumer (user) and producer (provider) are listed as consistent roles in each DE. However, there are many other roles that can be responsible for different tasks and activities. Actors can take on different

roles at the same time, and it is also common for the responsibilities of roles to overlap.

- **Actors**: An actor is an autonomous entity such as a company, an institution, or an individual that has one or more specific roles in a DE. Actors that are bound to a role must have all the capability to fulfill the obligations that exist through a specific role. Existing business needs motivate actors and each of them has different expectations from the DE. Actors commit to abide by the rules of the data platform because they want to be active in the ecosystem through the platform via different business relationships.
- **Relationships**: Relationships represent the interactions between actors in the DE. These are based on a common business interest and are related to the role each actor plays in the ecosystem. Basically, data or other types of resources are exchanged by the actors through transactions. If the transactions form an innovative and new added value, the goal is to implement data-based business models for the actors through the business relationship. Thus, these represent the value streams in a DE.

B. Procedure

The procedure for the design of the DE is based on a four-step approach (cf. fig. 1) by Oliveira and Lóscio [5]. We adopted this procedure within a funded research project in terms of a case study, consisting of a workshop period from January to December 2020. Work was conducted together with experts, researchers and practitioners in food production, data science and business model development.

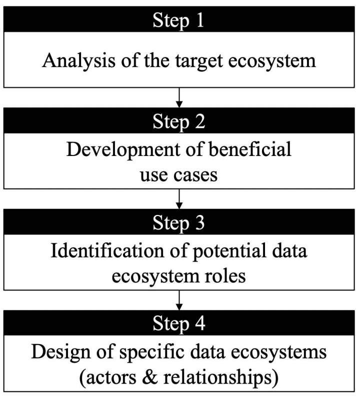

Fig. 1. Procedure of the Data Ecosystem Design Process

In the first step, the analysis of the targeted ecosystem takes place, in which primarily the area of investigation is delimited and discussed. For this purpose, the existing ecosystem of a confectionery producer was analyzed. Essential actors as well as their expertise and services were presented transparently and their respective interests were derived. The relationships along the food production value chain were also analyzed in order to identify order- and process-related data flows and to derive the best possible potential for data-based services. Furthermore, possible target corridors of existing and potential ecosystem actors were defined in order to specify a strategy for the targeted DE.

Subsequently, beneficial use cases (step 2) were developed. To this end, possible data monetization strategies were first examined and specific data products were derived that can be traded as basis of the business models in the DE. Then, taking into account the existing ecosystem requirements from step 1, value-added use cases for the future DE were developed. Since the targeted DE is characterized by a network-like structure and thus non-linear business relationships with multiple entry and exit points of actors, user stories were developed that highlighted the added values and requirements of the stakeholders in the context of the use case. Furthermore, the required data streams and services of each stakeholder for the implementation of the use cases were defined. The development of the use cases was based on iterative expert group discussions between the authors and the project consortium, in which ideas were developed on the one hand and prioritization took place through repeated exclusion of use cases.

This step is followed by the identification of potential roles of the DE (step 3). Following Oliveira and Loscio [5], the basic functions of the DE with associated tasks and activities were derived. For this purpose, existing roles from relevant literature were analyzed and critically evaluated with regard to the targeted DE. In addition, further roles were defined for tasks that provide essential added value in at least one of the identified use cases and were not covered by the existing ensemble. The results were developed iteratively and verified through expert group discussions with food industry representatives from the project consortium.

The next step (step 4) represents the concrete design of the individual data-based value creation systems that are combined in order to form a holistic DE. For this purpose, the respective actors and their characteristics (roles, resources, expectations and capabilities) as well as their relationships with each other must be designed. In order to obtain the most efficient but also concise mapping of the value network, a new system was developed for the design of the business relationships. In addition to the processed data products traded by the actors, cash flows and metadata as well as the processing of local raw data are to be marked. Furthermore, a distinction is made between conventional and data-based services, which can be used optimally based on analytical procedures or can also generate significantly higher added value for the customer. The design followed the premise of value co-creation [28] to create collaborative value systems in which cross-company and cross-industry data and service transfer creates greater value for the overall system and each individual actor benefits from this increase in value.

The upcoming section describes the results of methodology application.

IV. Results

In the following the achieved results are described. We first present monetizing strategies for data products, representing potential added value for market participants. Next, we describe generic roles that can be taken by actors in the DE. Finally, we validate the building blocks through a real life use case which we developed with a chocolate producer. The use case focuses on the reduction of CO2 emissions in the context of chocolate production, through the targeted use of compressor data.

A. Monetizing Strategies for Data Products

Various approaches in literature describe and categorize different monetizing strategies for data products. Tempich [29] defines three different types: ”Data as a Service”, i.e. data are made available by providers and used to generate direct revenue (data * price = revenue). ”Data as Insights” where data is used to improve product marketing and achieve higher economic results. In this case data themselves are opaque within customer interaction. ”Data-enhanced Products”, data enrich physical or virtual products. Hereby, increasing revenue of the enhanced physical product corresponds to revenue generated by data. Wixom and Ross [30] distinguish monetization strategies for data products as follows: ”Selling data”, ”optimize existing products or service” or ”improving internal processes”. Liozu and Ulaga [31] add “new business models and revenue streams” to these types. Laney [32] defines data products based on economic value they capture for businesses: ”direct exchange with goods, services or monetary resources”, ”use to increase income, or reduce risks and expenses”. Until now, definitions of monetization strategies for data products and corresponding business models pursued with them are inconsistent. Hence, a typification was carried out within the ecosystem design development process. The typification enables to identify different generic data product types and assign them to three overarching monetization options: data enhanced-products or services, data products as insights and data products as performance. These are presented in the following.

Data-enhanced products or services enable various special features through data connectivity compared to conventional products or services. In addition to positive economic factors, such as the possibility of charging higher prices or implementing services more cost-effectively, data connectivity enables innovations for the pricing model. The complementary bundle of services can be billed on usage basis and even supply and demand models are feasible. Furthermore, generated data can be used to achieve additional value through bartering. Here, data are provided free of charge in the sense of a barter transaction for the receipt of added value. Moreover, data-based features allow customers to select performance levels with little effort, even for hardware products. There are countless implementation possibilities for food industry. For example, Celli Group’s smart dispensing systems [33] or Bizerba’s innovative weighing technology [34] can be applied to directly analyze consumer behavior in order to optimize inventories, reduce waste or directly measure the success of marketing campaigns for the food manufacturing companies.

Data products as insight are composed of raw data and analytical services aggregated into a data product that aims to answer business-critical questions for customers. The price of this data product is measured by potential added value that additional insights bring to a business decision. Basic criteria such as quality and relevance in business context and degree of analytical maturity plays a major role for added value. Value increases from a descriptive data product over a diagnosis to forecasts or even decision support that shows direct guidance for action in future [35]. Therefore, three data product types are distinguished herein depending on the respective aggregation level: **data product as report, data product as diagnosis** and **data product as prediction**. Since it is outrageous to clearly allocate added values for customers, value can merely be determined and estimated ex-ante on the basis of existing value attributes. The price can be paid once or on a subscription basis, e.g. for the purpose of updating data. In food production, data products as insight can add great value to a wide range of areas. For example, in addition to classic price tables, sales forecasts can provide particularly valuable data products as insight, as these are very volatile and influenced by various external factors such as weather, seasonality, constantly changing customer needs or political influences. This shows a positive impact on inventory management throughout the food production value chain, from retailers to distributors to producers and farmers. Stocks can thus be reduced and capacity limits improved, thereby reducing the loss due to over-capacity.

Data products as performance go beyond mere provision of insights and are composed by prescriptive analytics as part of a holistic solution to generate concrete benefits for customers. This requires both data analytics and industry-specific expertise. Through participatory business constellation of provider and customer, this data product type enables implementing performance-oriented pricing models and transferring them into a contractual framework. For this purpose, an interactive price determination with the customer is needed, in which the price is ultimately determined by the result. Pricing takes place ex-post to allow using the actual added values achieved and, optimally, to enter into a long-term partnership with the customer that is result-oriented and represents a win-win situation. Data generated in the process provide opportunities to continuously improve offers and transform them into innovations. In food production, for example, compressed air as a service such as AirPlan from Atlas Copco provides a value-added solution [36]. Through a system of compressor, service and analytical services, the air volume per kilowatt hour consumed can be increased and billed per outcome. This ensures sustainable production and reduces high energy losses that many food producers have to deal with.

In addition to monetization strategies, that represent an important basis for designing data-driven business models, generic roles are formulated next. These perform the necessary tasks and have the capabilities to carry the desired DE.

B. Roles

Roles taken by actors represent key drivers of DEs and provide the basis for data-based business models [5]. Several related work highlights specific roles in DEs with varying degrees of detail. Based on literature review and expertise of the ecosystem development team, generic roles for DEs were derived, representing the building blocks of DE actors.

The **data marketplace operator** acts as an intermediary between DE actors. It's main tasks include setting up and operating the platform. Enabling value-creating framework conditions through platform governance is crucial for success. The marketplace operator regulates interactions of the different actors through central decision-making authority and control mechanisms (e.g. pricing, distribution of revenues, security aspects, transaction and communication regulations) [37]. The governance strategy chosen for a platform ecosystem therefore contributes significantly to platform status and future success [37]. In addition, a central function is to set standards for data quality and minimum requirements that are offered via the platform.

Data providers provide data they generated or processed within the DE [16]. Possible data providers include actors in the value chain of the food industry as well as external data providers (e.g. weather, transport or energy).

Data product users consume raw data and data products in the DE. They may be interested in acquiring data products in order to improve business decisions [38], to increase efficiency [13] or to develop new products or services [39]. In addition, it is also possible to use the data for research purposes or to aggregate them into higher-value products that are offered in the DE [40].

The **data orchestrator** represents a central role in DE [17]. He orchestrates multiple data streams as an intermediary and holds necessary domain knowledge to provide value-added services to customers in an efficient and profitable way. In knowledge-intense B2B sectors such as mechanical and plant engineering, comprehensive knowledge of production processes and machine configurations is necessary in order to be able to offer adequate services [13].

Due to a lack of competences of data product users, data providers or data orchestrators in dealing with data and creating value-added analytic services a market for various applications offered by **smart analytics service providers** emerged. Examples are applications simplifying query generation for customers on the data marketplace or support the integration of data from volatile and heterogeneous sources [41]. Application developers can sell or rent their algorithms and applications on the data marketplace as black-box applications or as transparent algorithms to other players [41].

These applications are used by different **smart analytics service users** in the DE. On the one hand, these can be data providers who want to develop their data into valuable data products in order to achieve a higher price or to be able to use them for internal company processes in a beneficial way [42]. On the other hand, they can also be used by data orchestrators to develop or improve their service offering for customers.

The **service customer** in the sense of the DE is taken by companies that profit directly from the data product-based services without wanting to process data provided themselves. These are direct customers of data orchestrators, who use provided data and smart analytics services to optimize the customers' business processes or decision making. The role differs from data product user as they pay for added value of the data-based services, but not for use of raw or aggregated data.

Furthermore, **industrial service providers** play a role in DEs, offering a broad range of activities that can provide products, services and solutions along the entire life cycle of industrial companies [43]. In addition to maintenance and repairs, this also includes e.g. engineering, logistic services or financial and insurance services. In summary, industrial services are essential to realize most of solutions envisaged in DEs. On the basis of data analysis, these can be used more efficiently and in a more beneficial way for the customer in order to achieve a higher added value [43].

The necessary IT infrastructure such as the platform or connectivity services are provided by the **infrastructure provider**. They offer a supporting service, especially for the marketplace operator [13]. However, these roles do not have to be fundamentally separate. It is also possible for data marketplace operator to provide the infrastructure itself.

Platform security as well as data and process quality are guaranteed by a **certificator**. The corresponding certificates increase trust of customers in the platform and can thus contribute to gaining market shares [41].

The **consultant** can demonstrate potential to data providers in monetizing their data and support them in developing a data sharing strategy by assessing suitability of existing data and developing business cases. This enables data providers to enter the business or optimize it. Data consumers can be supported by consultancies in finding suitable data sources or developing their own products based on the data acquired or already available in the company [41]. Figure 2 illustrates

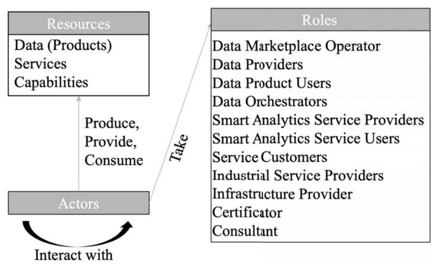

Fig. 2. Actors, Roles and Resources in Food Data Ecosystems

the connection between actors, resources and roles. Actors within the DE interact with other actors therein. In addition, each actor produces, provides or consumes different resources. Each actor can take one or more roles. The resources and roles defined are generic to food DEs, whereas actors are not generalized yet.

C. Exemplification of the Data Ecosystem

Building on the connection between actors[1], roles and resources, the following section puts them together to form a DE (cf. fig. 3). This includes the exemplary representation of actors and their relationships to each other within a real life use case. The aim is to optimize energy consumption and reduce CO_2 emissions in compressed air generation by using standard compressor data. For this purpose only use case essential actor characteristics according to Oliveira and Loscio [5] are included in the results, so irrelevant capabilities of food manufacturers are left out. Further the technical platform features that support business relationships in the targeted DE are listed and briefly described.

The actor **compressor manufacturer** takes the role of a data orchestrator, data product user as well as SAS provider and consultant. Thus, it performs various value-adding functions as a junction of the DE and is responsible for its effectiveness. The compressor manufacturer represents a solution provider for conservatively set up food manufacturers who have paid little attention to electricity consumption of industrial plants in the past. Their focus is on optimizing their raw material and supply chain practices. They are unaware of opportunities to save energy and be energy efficient. Often, available electricity data are only collected plant-wide and not further analyzed to find energy saving opportunities. Thus, a *data product as report* is offered by compressor manufacturer as part of the DE. It provides a benchmark that allows the customer to compare energy efficiency for production of compressed air with other food manufacturers. For this purpose, raw data of the respective data product customers' compressor systems are aggregated. The data are available in a standardized form, easy to generate and have an excellent comparability. These consist of the volume flow per time as well as energy consumption that must be expended for generation. The benchmark lists performance through an indicator. This indicator shows how much energy (in kilowatts) the site invests per cubic meter of compressed air produced. A low value indicates that a site is operating efficiently as it uses less energy to produce the same amount of air, while a high value indicates inefficiency. The smart analytics service to aggregate data and create the benchmark is loaded from the platform's app store into the data provider's local environment, instantiated and applied to input data to generate the data product. By processing raw data locally, they remain secure with the data provider and only the processed data product is visible to users. Furthermore, a *data product as prediction* is offered, which shows saving potentials of food manufacturer. In order to realize this, air pressure applied to the compressor is required in addition to previously listed data. Besides comparison via benchmarking, various detailed evaluations of compressed air system data are carried out to determine the potential. Efficiency losses due to leakages or machine malfunctions can be determined by analyzing coverage of compressed air demand at weekends compared to permanent load operation. Beyond that, inefficiencies in compressed air generation can be identified and evaluated by detecting irregular pressure drops when volume flow increases. Evaluation of the saving potentials provides the basis for *data product as performance*, in which a concrete optimization of the air pressure system is realized based on target measures. The compressor manufacturer has relevant domain knowledge and is able to harmonize data from different suppliers in the DE and translate it into novel value-added services. Added value is generated by the actual data-based service providing lower energy consumption per volume flow at customer and corresponding result-oriented revenue. Data-based analysis services also enable to anticipate customer needs ever better over time and thus also continuously increase customer loyalty. Continuous learning algorithms and targeted use of expertise are used to achieve a permanent increase in customer performance and thus a continuous reduction in energy consumption and CO_2 emissions. Further added value results from extensive database of benchmarking customers, which represents a sales potential, as many food manufacturers become aware of possible savings potential through the data product as performance.

Other actors in the DE are **food manufacturers**. Due to large demands for compressed air in various food production processes, they have many compressors on site, which are operated extremely energy inefficiently by many suppliers due to a lack of expertise and no necessary attention for these processes. Thus, in most cases, there is potential for savings of 30 percent on average. Food manufacturers could take the role of data provider and data product user if compressor data is only made available to obtain benchmarking via the platform. This creates transparency about their own performance and is provided free of charge. The larger the network of DE becomes and thus the chance to incorporate value-added data sources into data products, the higher the chance that the data product will continuously improve and the customer can benefit from the overall system. In addition, the role of service customer can be taken, whereby the food manufacturer benefits from optimizing energy efficiency. Since the service customer faces the challenges not only once, but recurrently, a benefit-oriented, long-term business model logic should be aimed for. Against the backdrop of continuously learning algorithms and ever new data sources that significantly improve predictive capabilities and thus prescriptive decision support, the economic potential increases with a medium- to long-term business relationship. "For sustainable success [in a DE], feedback mechanisms need to be integrated so that providers receive critical feedback on their data offerings from data users" [13]. In the role of data provider and service customer, the food manufacturer thus benefits from an opportunity to indirectly monetize data provided and improve the sustainability of production processes.

Since listed business processes can only work with the

[1]Although a wide network of food producers is envisioned with the use case in the future, we initially started with only one chocolate producer to initiate the development of the DE and pilot the infrastructure. The pilot installation serves to test the data products and the underlying smart services presented in the following to optimize them for the roll-out.

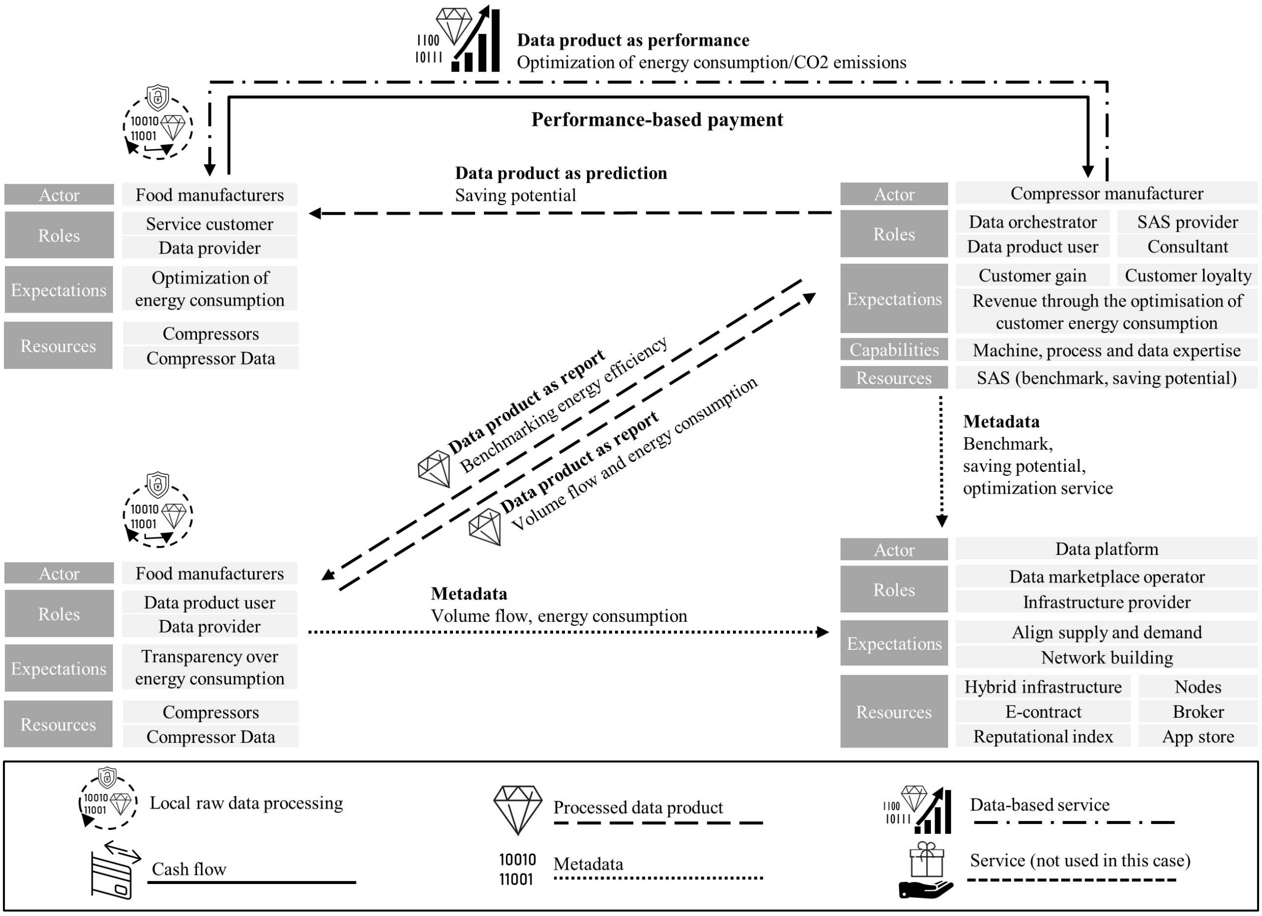

Fig. 3. Exemplary Presentation of the Data Ecosystem of Compressor Data in the Food Industry

help of a needs-oriented IT infrastructure, the underlying **data platform** of the ecosystem takes on the next actor of this use case. It takes the role of data marketplace operator and infrastructure provider and benefits from an ever larger network through which value-added data products are traded: it gains monetary added value, e.g., through transaction fees or a one-off license fee. For this reason, the data platform should provide the best possible breeding ground via IT infrastructure, its features and business framework conditions. In addition the app store, which makes smart analytics services available, the communication and information sharing of existing data products or available raw data is realized via metadata. Furthermore, solutions for data security and data protection are provided. In order to cover needs from data analytics services, an underlying technical platform, which includes decentralized smart analytics services as well as centralized smart analytics services is proposed. The platform consists of web-based technologies and microservices, which are naturally supported by browser engines and can be used across platforms and domains. In addition, a trustee agent is implemented that can not only enable smart analytics services for data aggregation from various stakeholders, but also provide a runtime engine to launch a smart service pipeline in a secure environment. Beyond that, the process of smart service and data usage has been defined in e- or smart contracts as regulation in order to track and scale data usage and the process of data analysis service. This enables easy, fast and legally binding processing within the DE. In addition, nodes are a key feature of the platform architecture, containing functions comparable to a connector. The connector consists of WebRTC technology, which enables peer-to-peer communication or broadcast communication for the platform users in the decentralized DE. At a high level, the node supports inheritance and abstraction. It can create instances with various functions that are derived from the general node. Three different instances are assigned to each node:

- The Broker acts as a virtual trustworthy agency that provides an isolated environment to launch smart services with corresponding data. In this approach, data and process of the smart service were aligned in a smart contract between data provider, service provider and broker. On one hand, data providers can retain usage of data. On the other hand, the entire process of smart services is transparent to the provider. The Broker is the central role that ensures the process of smart services and data in a secure, traceable and transparent environment. This encourages data providers and service providers from different fields to enrich the DE.

- Market participants represent either data provider or consumer. The market participant is instantiated by a decentralized smart analytics services platform, which enables edge computing. In the context of compressor data, service provider provides smart service to operate in real time and an isolated runtime environment of the data owner. In terms of collaboration with multiple data owners, market participants can sign an e-contract with the service provider and the broker. Based on the e-contract, the broker executes the smart service with the data in a specified and transparent process with regard to data traceability and data protection.
- Service provider provides the bundled package of data processing modules which can be applied on the data to create a valuable data product. Each module is based on the convention of the platform to enable the security and quality of the service.

For this use case, secure and anonymous participation in benchmark of energy consumption per volume flow is facilitated for the food manufacturer. The data provider can process raw data in its local environment by smart analytics services of the compressor manufacturer without great effort and learn about quality of its processes without risk. The smart analytics service is sent to the data provider, i.e. data does not leave its environment if provider is not choosing so. It also facilitates to find and enter a framework agreement with assured data usage rights. This benefits the compressor manufacturer, since accuracy and meaningfulness of data product are improved and a larger circle of potential customers for optimization is created. If customers want to obtain higher-quality data products, next step is to securely allocate data streams in order to derive optimization potential for each individual customer and ultimately to determine and initiate measures for optimizing energy consumption. Additional security and transparency is ensured via reputation index, which improves the number and satisfaction of realized transactions of the selected data product.

V. Conclusion & Outlook

In this work we derived a DE for industrial food production by defining actors, roles, resources and relationships for the DE. We found that the applied methodology [5] supports DE design for food production. Actors and their roles, resources, as well as relationships represent relevant elements for conceptualizing DEs. Ten typical roles were derived, serving as building blocks of the DE. These were used to develop an exemplary DE consisting of actors and underlying relationships, which develops significant added value for each actor. By defining a food production DE we overcome three hurdles: (1) fear of participants in exposing valuable or sensitive data [10], (2) constructing an ecosystem where participants are able to generate added-value with available data without the need of seeing all data of other participants [11] and (3) showing all actors can benefit from ecosystem participation [10].

To overcome (1) we defined the resource of a broker in a decentralized platform ecosystem. It represents a trustworthy, independent unit, that temporarily stores data of numerous ecosystem participants, analyzes their data and, e.g., provides an anonymous benchmark where participants can compare for example their CO2 reduction efforts through improving their compressors. In addition, the broker renders obsolete that all participants need to see raw data of other participants. The results further show, that data products represent one of the most important resources in a DE. We derived four types of data products and three interrelated monetization strategies, enabling added value for DE participants. Furthermore, the exemplification of a the DE through a compressor data use case (cf. fig. 3 shows cash and data product flows between the actors. Beneath obvious cashflows, insights generated by data products further enables optimization of compressed air generation and, in turn, reduction of CO2 and CO2 costs.

On one hand, the results aim to provide a methodological basis for the systematic design of DEs. On the other hand, by clarifying central aspects of data-driven business models and underlying information technology components, essential factors for the design are highlighted. The systematic design approach serves to improve future research and to expand existing state of knowledge regarding the design and conceptualization of DEs in the industrial food industry and beyond. As food production and industrial production have several aspects in common, e.g. quality control of pre-products, functionality and infrastructure of plants (e.g. use of compressors), this method could also be applied to the development of DEs in industrial production and manufacturing.

The research results are subject to limitations which need to be overcome by future work. First, due to the new domain of DEs, the developed use cases require frequent updating to stay relevant and to incorporate data-based services and added value. Second, since the design was carried out within one project consortium, other researchers and practitioners could derive generalized actors and further relationships that they deem significant. Therefore, we are planning an extended evaluation of the workshop results together with food producers in order to evaluate relevance and possibly extend actors and relationships. In addition, further studies are required to prove the generalizability of the results to further use cases and to expand level of formalization. Third, although the exemplary of DE shown is specified for industrial food production, further research could reuse the methodology for other sectors and an extension could be made for the entire manufacturing industry. Furthermore, the case of compressor data for CO2 reduction could be applied in industrial DEs in general. Thereby, companies will be able to support global CO2 reduction goals. In addition, a profitability analysis could be conducted in future work to derive the monetary value of data products and DEs.

REFERENCES

[1] Plunket Research. (2018) ”Food, Beverage and Grocery Overview”. [online]. Available: https://www.plunkettresearch.com/industries/food-beverage-grocery-market-research/.

[2] World Bank Group. (2021) ”Employment in Agriculture”. [Online]. Available: https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS.

[3] H. Ritchie. (2019) ”Food production is responsible for one-quarter of the world's greenhouse gas emissions”. [Online]. Available:https://ourworldindata.org/food-ghg-emissions.

[4] United Nations. (2016) ”The Paris Agreement”. [Online]. Available: https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement.

[5] M. I. S. Oliveira and B. F. Lóscio, “What is a data ecosystem?,” in *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, pp. 1–9, 2018.

[6] C. W. Yoo, S. Parameswaran, and R. Kishore, “Knowing about your food from the farm to the table: Using information systems that reduce information asymmetry and health risks in retail contexts,” *Information & Management*, vol. 52, no. 6, pp. 692–709, 2015.

[7] A. Anand, T. Coltman, and R. Sharma, “Four steps to realizing business value from digital data streams,” *MIS Quarterly Executive*, 2016.

[8] J. F. Moore, “Predators and prey: A new ecology of competition,” *Harvard Business Review*, vol. 71, no. 3, pp. 75–83, 1993.

[9] P. K. Senyo, K. Liu, and J. Effah, “Digital business ecosystem: Literature review and a framework for future research,” *International Journal of Information Management*, vol. 47, no. 4, pp. 52–64, 2019.

[10] J. Gelhaar and B. Otto, “Challenges in the emergence of data ecosystems.,” in *PACIS*, p. 175, 2020.

[11] C. Cappiello, A. Gal, M. Jarke, and J. Rehof, “Data ecosystems: Sovereign data exchange among organizations. report from dagstuhl seminar 19391,” 02 2020.

[12] C. H. Tsai and J. Zdravkovic, “A survey of roles and responsibilities in digital business ecosystems,” in *PoEM*, 2020.

[13] C. Azkan, L. Iggena, L. Meisel, M. Spiekermann, T. Korte, and B. Otto. (2020) ”DEMAND - Perspektiven der Datenwirtschaft”. Fraunhofer-Institut für Software- und Systemtechnik (ISST). [Online]. Available: https://www.demand-projekt.de/paper/DEMAND-2020.

[14] T.-D. Cao, T.-V. Pham, Q.-H. Vu, H.-L. Truong, D.-H. Le, and S. Dustdar, “Marsa,” *ACM Transactions on Internet Technology*, vol. 16, no. 3, pp. 1–21, 2016.

[15] B. Otto, “Quality and value of the data resource in large enterprises,” *Information Systems Management*, vol. 32, no. 3, pp. 234–251, 2015.

[16] J. Lange, F. Stahl, and G. Vossen, “Datenmarktplätze in verschiedenen forschungsdisziplinen: Eine übersicht,” *Informatik-Spektrum*, vol. 41, no. 3, pp. 170–180, 2018.

[17] M. Russo and M. Albert. (2018) ”How IoT Data Ecosystems Will Transform B2B Competition”. Boston Consulting Group. [Online]. Available: https://www.bcg.com/de-de/publications/2018/how-internet-of-things-iot-data-ecosystems-transform-b2b-competition.

[18] M. Savastano, C. Amendola, and F. D'Ascenzo, “How digital transformation is reshaping the manufacturing industry value chain: The new digital manufacturing ecosystem applied to a case study from the food industry,” in *Network, Smart and Open* (R. Lamboglia, A. Cardoni, R. P. Dameri, and D. Mancini, eds.), vol. 24 of *Lecture Notes in Information Systems and Organisation*, pp. 127–142, Cham: Springer International Publishing, 2018.

[19] W. Yu, T. Dillon, F. Mostafa, W. Rahayu, and Y. Liu, “A global manufacturing big data ecosystem for fault detection in predictive maintenance,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 183–192, 2020.

[20] B. Otto and M. Jarke, “Designing a multi-sided data platform: findings from the international data spaces case,” *Electronic Markets*, vol. 29, no. 4, pp. 561–580, 2019.

[21] German Federal Ministry of Economic Affairs and Energy (BMWi). ”Project GAIA-X: A Federated Data Infrastructure as the Cradle of a Vibrant European Ecosystem”, 2019.

[22] P. Vepakomma, O. Gupta, A. Dubey, and R. Raskar, “Reducing leakage in distributed deep learning for sensitive health data,” *arXiv preprint arXiv:1812.00564*, 2019.

[23] S. Shukla, M. F. Hassan, D. C. Tran, R. Akbar, I. V. Paputungan, and M. K. Khan, “Improving latency in internet-of-things and cloud computing for real-time data transmission: a systematic literature review (slr),” *Cluster Computing*, pp. 1–24, 2021.

[24] B. El Khalyly, A. Belangour, M. Banane, and A. Erraissi, “A comparative study of microservices-based iot platforms,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 11, no. 7, 2020.

[25] W. Xiong and L. Xiong, “Smart contract based data trading mode using blockchain and machine learning,” *IEEE Access*, vol. 7, pp. 102331–102344, 2019.

[26] Europian Union. (2016) ”REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation”. [Online]. Available: https://eur-lex. europa. eu/legal-content/EN/TX.

[27] A. Hein, J. Weking, M. Schreieck, M. Wiesche, M. Böhm, and H. Krcmar, “Value co-creation practices in business-to-business platform ecosystems,” *Electronic Markets*, vol. 29, no. 3, pp. 503–518, 2019.

[28] S. L. Vargo, “Service-dominant logic: Backward and forward,” in *SAGE handbook of service-dominant logic* (S. L. Vargo, R. F. Lusch, and K. Koskela-Huotari, eds.), pp. 720–739, Los Angeles and London and Melbourne: Sage, 2018.

[29] C. Tempich, “Konzeption und entwicklung von data-driven products / datenprodukten,” in *Data Science* (U. Haneke, S. Trahasch, M. Zimmer, and C. Felden, eds.), Edition TDWI, pp. 45–64, Heidelberg: dpunkt, 2018.

[30] B. Wixom and J. Ross, “How to monetize your data,” *MITSloan Management Review*, vol. 58, no. 3, 2017.

[31] S. M. Liozu and W. Ulaga, *Monetizing data: A practical roadmap for framing, pricing & selling your B2B digital offers*. Anthem, AZ: Value Innoruption Advisors Publishing, first printing ed., 2018.

[32] D. B. Laney, *Infonomics: How to monetize, manage, and measure information as an asset for competitive advantage*. New York, NY and Abington, UK: Taylor & Francis, first edition ed., 2018.

[33] PTC. (2021) ”Celli Group transformiert Nahrungsmittel- und Getränkeindustrie”. [Online]. Available: https://www.ptc.com/de/case-studies/celli.

[34] Bizerba. (2020) ”Smart Shelf Lösungen: Vernetzte Smart Shelf Lösungen mit integrierter Wägetechnologie”. [Online]. Available: https://www.bizerba.com/media/themen/security-of-supply/bizerba-insight-smart-shelf-loesungen-de.pdf.

[35] K. Amann, J. Petzold, and M. Westerkamp, “Prädiktive analytik,” in *Management und Controlling* (K. Amann, J. Petzold, and M. Westerkamp, eds.), pp. 251–256, Berlin, Heidelberg: Springer, 2020.

[36] Atlas Copco. (2021) ”AIRPlan: Druckluft erzeugen ohne Investitionskosten”. [Online]. Available: https://www.atlascopco.com/de-de/compressors/service/plans/airplan.

[37] A. Tiwana, *Platform ecosystems: Aligning architecture, governance, and strategy*. Amsterdam and Waltham MA: MK and Morgan Kaufmann, 2014.

[38] Bundesverband Digitale Wirtschaft (BVDW). (2018) ”Data Economy: Datenwertschöpfung und Qualität von Daten”. [Online]. Available: https://www.bvdw.org/der-bvdw/news/detail/artikel/data-economy-datenwertschoepfung-und-qualitaet-von-daten/.

[39] G. Ji, L. Hu, and K. H. Tan, “A study on decision-making of food supply chain based on big data,” *Journal of Systems Science and Systems Engineering*, vol. 26, no. 2, pp. 183–198, 2017.

[40] R. Clement, D. Schreiber, and P. Bossauer, *Internet-Ökonomie: Grundlagen und Fallbeispiele der digitalen und vernetzten Wirtschaft*. Berlin Heidelberg: Springer, 4., aktualisierte und überarbeitete auflage ed., 2019.

[41] F. Stahl, A. Löser, and G. Vossen, “Preismodelle für datenmarktplätze,” *Informatik-Spektrum*, vol. 38, no. 2, pp. 133–141, 2015.

[42] M. Spiekermann, “Data marketplaces: Trends and monetisation of data goods,” *Intereconomics*, vol. 54, no. 4, pp. 208–216, 2019.

[43] G. Schuh, G. Gudergan, and A. Kampker, *Management industrieller Dienstleistungen*. Berlin, Heidelberg: Springer, 2016.