Institutionalizing Analytic Data Sharing in SME Ecosystems – A Role-Based Perspective

Henning Baars
Chair of Information Systems I,
University of Stuttgart
henning.baars@bwi.uni-stuttgart.de

Patrick Weber
Ferdinand-Steinbeis-Institute
patrick.weber@ferdinand-steinbeis-institut.de

Ann Tank
Chair of Management Accounting
and Control, University of Stuttgart
ann.tank@bwi.uni-stuttgart.de

Abstract

There is a variety of reasons that sharing data among Small and Medium-Sized Enterprises (SMEs) carries business potential, particularly for analytical applications. But outside a few niche domains, the number of success stories for data sharing is rather modest. Based on a qualitative study and first experiences from a research project with pilot implementations, we argue that this is mainly due to a lack of an institutionalized governance structure: Founding a separate legal entity for data sharing and analysis can address core concerns regarding sharing valuable data assets. However, this requires a well-calibrated set of defined roles for the involved partners. Based on our results we propose a first concept on delineating and mapping out those roles.

1. Introduction

Across a variety of industries, Analytics is considered a cornerstone of sustainable success and a relevant building block for a digital transformation [1] [2] [3]. Analytics subsumes the systematic collection, transformation, and analysis of data as well as its use in reports and analysis models [1]. The latter also includes applications of Artificial Intelligence (AI) in general and Machine Learning (ML) in particular. Larger enterprises can often build upon mature data infrastructures and supporting organizational structures that are rooted in overarching Business Intelligence (BI), Data Warehouse (DW), and Big Data environments. Often such organizations have established BI Competence Centers (BICC), data governance structures, and efficient data pipelines [3] [4] [5].

Small and Medium-Sized Enterprises (SMEs), however, rarely bring along this bedrock for data management, and also fundamentally lack the scale, the scope, the necessary capital, and the personnel (data scientists, data engineers, etc.) needed for building it [6] [7]. In addition, while more advanced ML techniques (like deep learning) are a key competitive factor for SMEs, they are even more demanding, both

from a personnel and from a data perspective [8] [9]. The data side concerns questions of scope and volume. Regarding scope, a single, specialized SME (e.g. a machine provider) is by its very nature confined to a subset of the attributes that are needed to holistically capture a problem (e.g. under consideration of material and process data). Regarding volume, a single SME often has access to a limited number of cases and is therefore restricted when it comes to collecting a sufficiently large set of training data. The issue is aggravated by the fact that the collected data is often imbalanced. A typical example are fault detection applications where it is often easy to gather large volumes of event data, however, only a small fraction actually describe faults [10].

From this perspective, sharing analytical data among peer SMEs seems like a viable approach. A systematic literature review by Rupek (2021) indicates that there is not only a scarce body of pertinent publications but also that those that exist primarily document a few niche applications e.g. for smart farming or condition monitoring [11].

Our assumption is that this is mainly a result of the lack of an overarching governance that defines the structures, processes, and relational mechanisms [12] needed for a well-defined, fair, and transparent handling of the data to be shared and analyzed. Not only have issues of governance and trust been identified as prerequisites for data sharing in other scenarios [13], they have also been shown to be prerequisites for successful Business Analytics endeavors within a single organization [14]. It is not far-fetched to assume that its importance should become even more critical outside a single hierarchy. We address the respective research gap and propose to anchor such a governance within a separate legal entity that is responsible for coordinating resources and activities associated with sharing and analyzing data within an ecosystem of SMEs. We also derive a set of roles (*role* in the sense of a set of related and organizationally defined functions) for such an ecosystem.

HⁱCSS

Our research question is therefore: *How can a structured role catalogue be designed to enable and foster data sharing and analysis in a business ecosystem under consideration of the option to establish an institutionalized legal entity dedicated to those tasks?*

This entails three sub-questions:

1. *How can an institutionalized data sharing and analysis entity support data sharing?*
2. *What roles should such an entity assume in the context of data sharing and analysis? What further supporting roles are needed?*
3. *How are the roles distributed among the different organizations?*

The research question(s) already suggest(s) a design-oriented approach which also needs to be qualitative and explorative due to the novelty of the subject. As to our knowledge the proposed structures do not exist, we conduct two complementary series of interviews that provide insights a) into the nature of data sharing and analysis across enterprise borders and b) into the roles of an institutionalized, cooperative unit and of the surrounding ecosystems.

The result of the research is a framework for a business ecosystem with a legal data sharing and analysis entity. In its center is a role catalogue that is complemented by findings regarding its implementation. We see our results as a contribution to practice that can guide SMEs to found a data sharing ecosystem with an institutionalized entity. We also contribute to research as our framework can be used as a conceptual tool for analyzing and/or designing business ecosystems with a focus on data sharing and analysis.

The remainder of this paper is structured as follows: in the next section, we describe the underlying conceptualization. In section three we describe the state of the art of data sharing in a SME context. In section four, we give an overview of our methodology and in section five, we present our findings. In section six we describe the results of the evaluation process including workshops and a pilot implementation. The final section reflects and discusses our findings.

2. Conceptualization

The terminology in the field of Analytics has become rather blurred after several decades of tides and ebbs of hypes [15]. In all its incarnations however, Analytics requires processes that collect, transform and analyze data. The results of these processes can take a variety of forms – from a simple report to a model that is operationalized and integrated into a transactional enterprise system. For the purposes of this research,

we focus on applications that are geared towards business-oriented applications, i.e. on Business Analytics (BA) [16]. This supersedes the older term Business Intelligence (BI) that was once understood to be an integrated approach to management and decision support [17] [18] but has recently shifted to focus mainly on reporting- and dashboard-oriented solutions.

BA in general and ML/AI in particular require relevant amounts of data [19], a combination of specialized high-end statistical, IT and domain know how [20], and an up-front organizational preparation [21]. These aspects are barriers especially for the application in an SME environment [7]. A general approach to address these types of challenges is to collaborate across company and domain borders and thereby establish a business ecosystem. A business ecosystem is defined as follows: "*Companies coevolve capabilities around a new innovation: they work cooperatively and competitively to support new products, satisfy customer needs, and eventually incorporate the next round of innovations*" [22]. Business ecosystems can be found in a variety of settings. In few cases those already incorporate some degree of data sharing, e.g. in the domain of the Internet of Things (IoT) [23]. Such ecosystems are often built around a shared set of interconnected assets, e.g. machinery, transportation equipment, buildings, or IoT devices. A cooperative data use in these ecosystems can foster the development of new services that harness the otherwise untapped potential of BA. It needs to be emphasized that those cases are usually driven by larger organizations rather than SMEs, often with a leading *focal* partner in the center [22]. In 2017 Adner recommended that business ecosystems focus on a core value proposition and an explicit specification of the contribution of every partner [24]. Important success factors of a business ecosystem include a delineation of relevant activities, a thorough selection of the partners, a specification of the relationships between the partners as well as a definition of their respective roles [22]. The roles in particular form fundamental building blocks of an ecosystem as the cooperatively defined value proposition coalesces around them [25]. Ideally, they complement each other for cooperative value creation [22].

3. State of the art

A concept closely related to the one presented here is the *data spaces* approach that aims at a sovereign, trustworthy and provider-neutral data federation in order to support data ecosystems [26]. Most notable are the European GAIA-X project [27] and the embedded standards of the International Data Spaces Association (IDSA) [28]. We see those more infrastructure-

oriented solutions as complementary to our organizational one. In fact, a business ecosystem with an institutionalized entity responsible for data sharing and analysis would greatly benefit from a working data space as it would remove doubts regarding data security and provider dependence. The institutionalization can be established in a variety of forms: from a joint venture, over a virtual enterprise, up to founding a co-operative. Independent of the concrete legal form, we focus on the relevance and the roles of such an institutionalized entity as well as the roles of the other partners of the business ecosystem this entity is embedded into. A *role* in the context of business ecosystems can be defined as a set of interrelated and organizationally defined functions that a member takes responsibility for. Roles help to represent a member of a business ecosystem and its involvement in different services [29]. A single member can assume several roles in an ecosystem [30]. We consider it a core challenge that there so far is no defined catalogue of roles for data-sharing ecosystems that go beyond a technology perspective. We compile our results from two perspectives: A bottom-up, *micro perspective* that gathers relevant functions and roles from the body of research on BA within enterprises, and a top-down, *macro perspective* that is informed by the literature on business ecosystems that considers the interplay between organizations.

The **micro view**: in the realm of BI organization, there has been some early work on establishing BICCs and BI steering committees. BICCs are understood as units dedicated to data integration and analysis [31]. The sizes and types of BICCs that can be found within various organizations vary strongly. This has been explained by the different roles that a BICC can assume. One explanation that directly feeds into this research is as follows: some BICC are mainly responsible for solution development, others for operation and support, and a third group of BICC merely act as coordinators between the different Lines of Business (LoB) and IT [32, 34]. Interestingly, similar results come from the data governance and data quality (DQ) world, where competence centers are seen as integral for both data governance and data quality [34]. The picture is also mirrored in recent discussions on establishing AI centers of excellence [35]. Given the data focus of all those areas, it does not come as a surprise that there is a strong overlap among AI, BI, and DQ competence centers, and the roles assigned to those units.

Often, the main responsibility for devising an overarching governance framework (particularly concerning the portfolio of solutions, the architecture, as well as the rules and responsibilities) resides in a higher-level body like a AI/BI/DQ data steering committee that has top-management involvement [36].

From this research also comes a set of more granular roles. Noteworthy are the roles of the *data owner* (a unit organizationally responsible for a set of data entities), the *data architect*, the *data engineer,* and the *data modeler* (technically responsible for central data assets), the *data steward* (who maintains selected local data entities), the *data analyst* (as a more business-driven analyst role), the *data scientist* (as a more method-versed analyst), the *chief data officer* (a top-management role for dealing with data assets as a strategic resource), the *report and dashboard developer,* the *data integrator/ETL developer*, *the BA marketeer/evangelist* and the *BA trainer* [37] [38]. Those roles can be grouped as follows:

- Roles concerned with *supplying the solution components*, namely the infrastructure and the platform.
- *Data and BA/AI management roles* that include managerial roles responsible for overarching decisions, rules, and guidelines regarding the data assets as well as technical roles for their handling (data architects, data engineers, data modelers).
- *Support roles*, e. g. for training or internal marketing.
- Roles concerned with *development of BA solutions;* this includes dashboard and report designers, data scientists who develop relevant models, as well as roles concerned with data transformation (like the data integrator or the ETL developer).
- More *operational and local roles* concerned with the coordination of data assets across various units (including data quality management). Those need some degree of embedding into the LoB units and have gained in relevance with self-service or Low/No code BI/AI solutions.

The **macro view:** relevant results can be drawn from work on business ecosystems in the realms of IoT [32] and Cloud Computing [29]. For the IoT side, Yoo et. al (2010) specify the general roles of *object*, *network*, *service,* and *content provider* [39]. In the realm of Cloud Computing, Böhm et. al. (2010) identify *infrastructure provider, platform provider*, *application provider*, *aggregator*, *integrator*, *consultant*, and *customer* [29]. These are very similar to the roles defined in the GAIA-X data space environment [26]. The former typology defines layers of (IT) abstraction that can be expected to be present in a data sharing and analysis environment as well. The latter can most likely be directly applied given the relevance of Cloud-based BA, which is particularly high for SMEs and their limited IT infrastructure capabilities. A separate group of roles is related to cost clearing and revenue accounting. Those can be expected to be critical for shaping the overall business model of the ecosystem and its members [40]. The shape and the importance of the discussed functions and roles above

will likely shift with a step towards cross-company-border solutions and the establishment of the separate legal unit. The means to enforce rules by hierarchy are diminished here and the need for coordination is significantly increased. In addition, external players gain relevance, particularly on the infrastructure and solution side. For those reasons, the macro view in particular needs additional insights.

4. Methodology

Due to the nature of our research, we follow a design-oriented, explorative, and qualitative approach. We follow the principles of design science [41]. Our design artifact is the role-based framework for an ecosystem of organizations that found a joint legal entity that encapsulates cross-company-border data sharing and analysis activities. We collected the data in qualitative, semi-structured interviews (relevance cycle) [42, 43]. Based on our results we derived an initial design (design cycle) that we evaluated and iteratively refined in four evaluation workshops as well as ongoing pilot implementations in real-world SME settings (rigor cycle). For the *kernel theories,* we draw from the presented literature on Business Analytics and Business Ecosystems.

As there are to our knowledge no established solutions for our design, we conducted a two-pronged approach to our interviews. In the *first series of interviews*, we talked to representatives of 8 initiatives from different industries and countries, which we consider to be experts in data sharing and analysis (**Table 1**, left column). In the SME context, we could not identify cases in which IoT data had been shared across the borders of companies. That is the reason why in the first series of interviews were mainly large companies. This side of the study corresponds to the micro-view of the state of the art section. We focused on cases where IoT data was shared, as we assumed that the willingness for data sharing is higher than in cases where data is sensitive and comes with more strings attached, e.g. product design or customer data. The interview partners were mostly accessed through the Industrial Internet Consortium (IIC) based on criteria defined by the researchers. A crucial criterion was experience with pilot implementations of data-driven IoT applications.

In the *second and larger series of interviews*, we examined 14 established German business ecosystems with institutionalized entities for cooperation tasks (**Table 1**, right column). The interview partners were selected by the Baden-Württemberg cooperative association (BWGV), based on criteria defined by the researchers. The interview partners were exclusively members of this association. Despite not being in the

realm of analytics, we deemed those valuable for transferring insights for the macro-view in particular as we gained insights on the overall structure of the ecosystems, the (potential) relevance and function of an institutionalized unit, the roles in the ecosystem, aspects of governance, processes and tasks, as well as cost structures. Focusing on German cooperatives was motivated by the fact that those organizations are typically formed by SMEs in order to generate additional value by cooperating across the company borders [44].

Interview series with experts in data sharing and analysis	Interview series with existing business networks
DS1 Logistics and supply chain with object tracking and analysis	**CO1** Cooperative in the context of a large wine-maker
DS2 Streaming IoT data for analysis, fishing industry	**CO2** Cooperative in the bedding supplies specialist trade
DS3 Analytics infrastructure for device coordination; drone hospital deliveries initiative	**CO3** Cooperative in the context of bakeries
DS4 Floor planning for smart factories	**CO4** Cooperative in the context of energy distribution
DS5 Optimizing plastic injection molding machines	**CO5** Supraregional banking and finance
DS6 Port traffic management	**CO6** Cooperative in the context of open source software development
DS7 Port traffic management	**CO7** Cooperative in the context of regional banking
DS8 Data analysis for smart factories and smart logistics in retail	**CO8** Cooperative in the context of regional banking
	CO9 Cooperative in the context of regional banking
	CO10 Cooperative in the context of regional banking
	CO11 Cooperative in the context of regional banking
	CO12 Cooperative in the context of logging and forest management
	CO13 Cooperative in the context of wood wholesale
	CO14 Cooperative in the context of specialized consulting
Note: Due to the COVID-19 pandemic all interviews of the first series were conducted in a virtual setting.	

Table 1. List of the interviews

Taken together, we conducted 22 interviews, after which we observed a saturation. Each interview took on average one hour and focused on the business potential of cooperation entities, success factors, as well as the relevant roles within potential SME data sharing networks. Both interview series were based on a conceptual framework developed by three research institutions from various fields (SME Networks, Business

Analytics, Managerial Accounting) and validated by experts on cooperatives. **Figure 1** depicts the framework that grounded both the design of our interview questionnaire as well as the analysis of its results. It mostly reflects the contents of the related work part and is geared towards the three research questions.
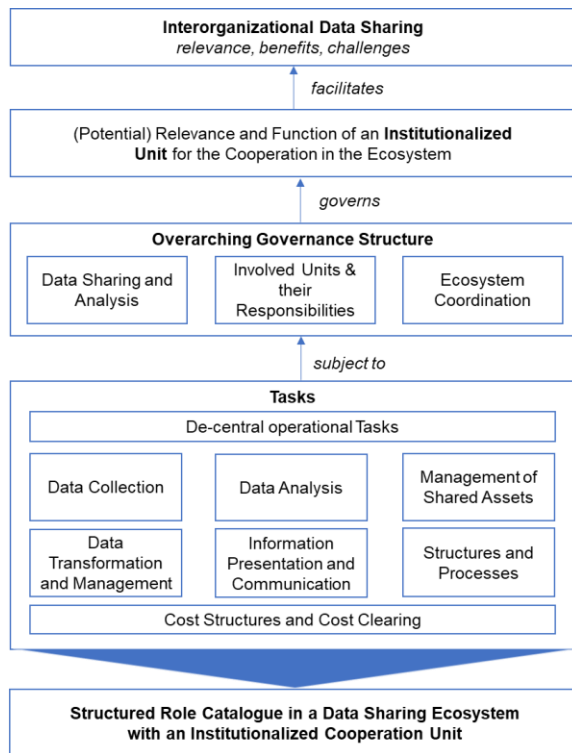

Figure 1. Conceptual framework

We transcribed the interviews, coded the answers, and then iteratively condensed them using a qualitative content analysis [45, 46]. The results were combined to a preliminary framework that includes the role-based concept and that was iteratively revised in the evaluation and prototyping phases.

5. Findings

5.1. Interview series with experts in data sharing and analysis

As a result from the sample, it is no surprise that data sharing was central to a varying degree in all eight discussed cases, be it for tracking transportation equipment (DS1), sharing data between fishers, and banks (for calculating credit lines), the World Wide Fund For Nature (WWF) and authorities to calculating the exploitation of fishing quotas in near-time (DS2), to enable drone deliveries and develop new delivery ser-

vices (DS3), to combine building information management, with machinery, and production data for optimizing shop floor designs (DS4), optimizing molding processes (DS5), monitoring and steering trimodal port traffic streams (DS6, DS7).

Finding 1: The interviews support the claim that data sharing across enterprise-borders enables new services and solutions.

While not being SME, two organizations still struggled massively with barriers especially regarding the acceptance of data sharing. In DS5 the machine provider was not willing to share machinery data, in DS6 most of the initial participants of the pilot project rejected to share any data (despite being more than willing to benefit from the data shared by others). DS6 only barely succeeded with the help of circumventing a large part of the data sharing by independently collecting traffic data with cameras and a deep learning-based object detection. This was surprising for the project management, as the benefits for all participants were clearly outlined before the project and all seemed to be committed. The drone deliveries in DS3 worked primarily because it was backed up by a government mandate (and it was unclear if and how a planned transfer to delivery applications in Europe and the US would turn out). In the fishery case DS2 the banks had a strong incentive to push the solution and could also bring convincing financial incentives, and DS1, DS8, and DS9 were also only possible because of the involvement of a large, focal player (in the case of DS7, the name "authority" was even part of the name of the enterprise). DS4 was strongly motivated by a stream of industry research funding and the involvement of several research institutions.

Finding 2: Even clear economic benefits for the participants do not guarantee the willingness to share data. Mandates and power promoters help – although including those might not be an option for SME peers.

Institutionalized cooperative units were not present in this string of interviews. However, in the ones that struggled with data sharing, the calls for such units were very vocal. DS7 was most pronounced in this regard, attributing the problems of the project to a lack of a binding set of rights, roles, and responsibilities, to problems with locating overarching tasks, the lack of a neutral entity, and unclear cost and benefit attribution. Interestingly, in the similar (albeit larger) setting of DS8, this was not the case. The BICC of the port authority could draw from experiences in reporting and DW management and could build upon existing governance structures for BI and data management.

Finding 3: An institutionalized entity could bundle overarching tasks and resources and also helps designing and enforcing a governance for data sharing.

As most cases came from technological testbeds, the technical solutions were impressive. DS2 featured a lightweight, yet highly scalable infrastructure for global data streaming. In the drone setting of DS3, a multi-layered infrastructure for drone coordination applied complex, deep-learning-based path predictions was used. DS4 applied math-heavy simulations and optimizations. DS5 proudly used state-of-the-art classification and regression algorithms. DS6 distributed image analysis. Often, larger IT enterprises took over the development and operation of the solutions (hoping to utilize them as showcases), acting both a technology promoter and as a technology provider. With the exceptions of DS1, DS7, and DS8, the organizational side (definition of roles, structures, and processes for managing those tasks) was not as mature, which to be fair was not the focus on the endeavors. DS7 suggests that the role of a BICC is needed in a data sharing scenario.

Finding 4: Solutions benefit from a dedicated technology promoter and professional technology provider.

Finding 5: It is advisable to consider the pendent to a BICC for a data sharing ecosystem (which could e.g. be located in the institutionalized unit).

Closely intertwined with the organizational side, costs and revenues were seen as a success factor for a sustainable data sharing cooperation. DS8 brought a comprehensive scheme for cost sharing and management to the table that recognized the need for some costs to be shared evenly and others to be connected to the consumption of services. Because of the bank involvement, DS2 was intrinsically tied to a scheme in which data sharing was connected to palatable financial benefits for both the fishers and the bank. DS3 was in the process of deriving respective business models.

Finding 6: Establishing transparent cost clearing and revenue accounting is a success factor for the sustainable success of a data sharing initiative.

To sum up our findings from this first series of interviews: we found a set of relevant aspects that support the benefits of an institutionalized entity and could identify a first set of roles (governance bodies, power and technology promoter, developer, data sharing and analytics competence center, accounting).

5.2. Interview series with existing business networks

In the 14 interviews of the second series, we observed that an independent entity (here: cooperative) is needed as a neutral and trustworthy intermediary. Ideally, the entity is also the power and technology promoter. The interviewees presented strong arguments to establish this partner as its own legal identity

capable of and allowed to coordinate external suppliers and/or customers. All business networks were based around a palpable value proposition. The interviewees clarified that the value generation can *only* be realized cooperatively. For example, in CO5 the cooperative is responsible for the procurement of agricultural products for all members, thus creating bargaining power. In interview CO2, the procurement, development, and financing of beds produced by the ecosystem members was the cooperative value proposition. The cooperative of CO14 focused on a joint development of surface technology.

Finding 7: The institutionalized entity is formed by the members to be able to realize a cooperative value proposition.

Finding 8: It is advisable that this institutionalized entity has its own legal identity.

Across interviews, the necessity for an explicit role specification was uttered. This includes the specification of the roles the ecosystem members take and the ones assigned to the institutionalized entity. For example, in CO6, the roles capital provider and distributor were assigned to the institutionalized entity, while other roles were divided among the partners. In interviews CO8, CO10 and CO11, the institutionalized entity also assumes the role of an investment manager, a sales partner, and a data analyst.

Finding 9: A clear definition of the roles of the institutionalized entity and the partners is important for the development of the business ecosystem.

The structure of the network also differed widely. In a part of the interviews the group of the members were very homogeneous (CO1, CO2, CO4, CO9, CO10). In the interviews CO4, CO7 and CO13 partners were rather heterogeneous. Interestingly, the ecosystems also varied strongly with respect of centralization which led us to give up on one of our initial goals to come up with a clear "best practice" structure for the roles. In CO4 and CO14 we found the most decentral structures with the institutionalized entity only realizing a handful of necessary roles. Here, the primary role of the institutionalized entity was coordinating member activities. Sometimes, the institutionalized entity took over specialized tasks for the ecosystem members, especially in interview CO7, where the cooperative designs open-source software. Another type was characterized by centrally concentrating a large set of roles. Examples of this were interviews CO1, CO2, CO3, CO5, CO7, CO8, CO9, CO10 and CO13.

Finding 10: The roles assumed by the institutionalized entity can vary between the two types: centralized and decentralized cooperation.

In the interviews CO5, CO8, CO9, CO10 and CO11, we found individual roles that were taken neither by the institutionalized entity nor by the partners. In these cases, roles were sourced to external suppliers with the institutionalized entity responsible for acquisition and coordination.

Finding 11: Individual roles can be handed over to external suppliers with the institutionalized entity being responsible for the coordination.

In total, we identified a set of core roles based on the interviews. Some of those overlap with the roles from literature. Despite the fact that this series did not focus on IT-related tasks, we still re-discovered two of the roles for *Data Sharing and Analysis* that are known from the BICC and data governance literature [34, 36, 37], namely the *data analyst* (in CO1) and the *data provider* (in CO5 and CO9).

We deem it to be plausible that in the analytics-driven ecosystems we want to establish, more of the "classical" data roles would emerge which is why we also added additional groups of roles to our list, namely: *central data roles* (esp. the chief data analyst, the data governance, BA, AI & DQ manager, the data architect, data engineer & data modeler)*, support roles* (like the BA marketeer/evangelist or the BA trainer), *application development and operation roles* (including the data scientist, or the report/dashboard designer), as well as local roles (esp. the data owner, the data steward, and the data analyst).

The roles from the business ecosystems literature were mostly found in the interviews, though. The *object provider* appeared in CO1 and CO7 (with ATM being the provided assets here). The *infrastructure provider* was mentioned in the interviews CO1, CO3, CO5, CO6, CO7, CO9 and CO12 (in all cases provider of data centers). The *application provider* was identified in interview CO12. It became clear that this role could not be derived from the institutionalized entity or the members, so this task was performed by an external partner. An application provider was also present in the ecosystem of CO5, preparing analyses for various stakeholders (e.g., banking supervisory institutions). In interview CO14, the *data integrator* was mentioned as a central point for the collection of data. In interview CO3, a *consulting* role was identified with a focus on legal questions. Further roles we found were:

- *Coordinator*: Takes over the orchestration of different partners. In interview CO8, this role coordinates the building of a new bakery. Players were an architect, a manufacturer, and an investor.
- *Member representatives*: Elected member who represent the interests of the members of the institutionalized entity, especially relevant for larger ecosystems. CO1 featured a network of 40,000 members where the member representative was deemed as critical.
- *Strategy board*: Defines the future direction of the collaborative ecosystem. This role was again found in larger ecosystems. CO7 and CO11 (banking) described a special role for the definition of the overarching strategy. This strategy was defined for several business sub-ecosystems with the aim to define a common vision. Such a meta-organization is probably not relevant for a small data sharing ecosystem.
- *Technical support*: Contact person for analytics and data sharing related advice and assistance. In the case of the interview partner CO12, this role was characterized as first level support.
- *Broker/Negotiator*: Intermediary with the right to negotiate contracts. This role was derived from interview CO5, where the brokerage of raw materials on the stock exchange was specified.
- *Distributor*: Procures and trades products from different manufacturers. The interviewees CO5, CO8, CO12 and CO13 named this role.
- *Investment manager & capital provider*: Takes over the management of investments, pays interest on the invested assets and gives loans to members. This role was identified in interviews CO2, CO5 and CO8. For example, the financing of investment goods in bakeries.
- *Insurer*: Compensates for negative consequences (for example, in the form of payment defaults) were mentioned in interview CO6 as a trade credit insurer.
- *Payment processor*: Handles the transfer or assignment of money from one party to another. The interviewee in CO6 named this role to take over the handling in terms of payment processing.
- *Protection facility*: Ensures the protection of deposits. Interviews CO3 and CO5 specified deposit protection in the banking context.
- *Sales partner*: Does the marketing of the service offering. In interviews CO6, CO7, CO8, CO9 and CO11 a commission rate is defined, which is paid for contract conclusions.
- *Testing facility*: Checks compliance with defined requirements (e.g.: according to the German cooperative law). In the interviews CO1, CO2, CO3, CO6 and CO7 this role could be identified. It is examined when the legal requirements are fulfilled.

6. Evaluation & framework development

6.1. Evaluation workshops

The evaluation was conducted with four (online) workshops and pilot implementations that prepare the actual founding of an institutionalized cooperation entity for various business ecosystems.

A *first (virtual) workshop* included representatives of an interest group for business ecosystems. The workshop focused mainly on the understandability and the conformity with legal requirements.

This led to no relevant changes to the presented role catalog and framework apart from a less academic wording. Members of the interviewed organizations from the second series of interviews participated in the *second and third (virtual) workshops*. Here we particularly addressed the inclusion of specialized IT and infrastructure providers. We consider the related roles to be of essential nature since their positioning has a variety of implications for dependencies and negotiating power and the results from the interview series were inconclusive. The participants gave a variety of (mostly power- and fairness-related) reasons to keep IT-providers external to the ecosystem. If the institutionalized entity cannot provide the IT by itself, it should source them with clear-cut delivery and performance contracts. The participants also highlighted aspects of data security and privacy and the need for IT-related partners to be certified.

The *fourth and last evaluation workshop* took place in a global online conference of the IIC. The participants strongly supported the idea of founding a dedicated entity for data sharing. The following discussions mostly revolved around the choice of the legal entity and equivalents to the German cooperative.

6.2. Pilot implementations

The two-year goal of our research is to establish the discussed entities in three ecosystems that are currently at various stages. In the most advanced one we already developed first prototypical data sharing and analysis applications that we currently transform into running pilots. The case focusses on the management of coolant in a manufacturing scenario, an area of application that is not only costly but also critical for life and limb (as the coolant can turn toxic).

The currently involved partners of the ecosystem include the supplier of the coolant, a provider of coolant management machinery, and a manufacturing company for metallic automotive parts. Analytics and AI based on data from all three companies promise to facilitate a more cost efficient, less error-prone, and safer coolant management with tangible benefits for all partners. We could already gain additional insights into the process of establishing the institutionalized entity and the distribution of roles. The SME context meant that the companies did not have the capabilities to set up a data management and analysis infrastructure (similar to the literature findings presented in the first two sections). We currently assign the roles from our framework in a stepwise fashion.

6.3. Framework – a structures role catalogue

Our findings strongly suggest that an institutionalized entity fosters acceptance for data sharing among SMEs. The entity can bundle overarching tasks and resources. We are able to identify roles that are relevant for analytics and data sharing as well as additional ones that are relevant for the value creation in business ecosystems but. From this, we derive the structured role catalogue visualized in **Figure 2**.
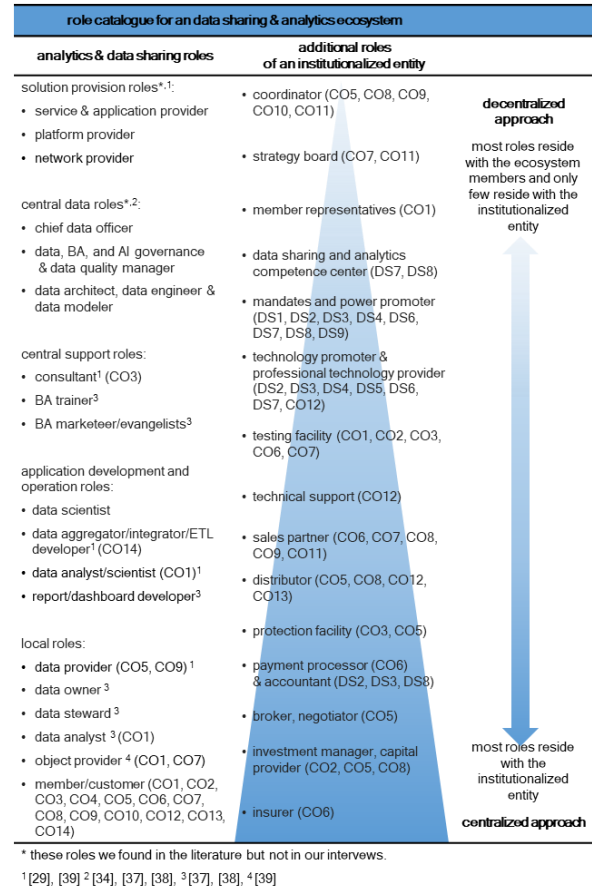

Figure 2. Structured role catalogue in business ecosystems with an institutionalized entity

The catalogue helps to identify which roles are needed in the context of joint data sharing and analytics and thus, to realize a value creation scenario. It also helps to assign these roles and thereby to determine whether to follow a rather *decentralized approach*, in which most of the roles are assigned to the members of the ecosystem, or a *centralized approach* where most roles reside with the institutionalized entity.

The framework supports partners to identify additional roles required for the realization of the value creation. In the interviews, it also became clear that for

many roles there is some leeway regarding their location in the ecosystem. The "centrality" of the ecosystem is therefore not a binary choice but a gradual one with various alternative design option. The assignment of the roles to the decentralized and centralized approach is a first classification, which has to be evaluated and extended in further research. **Figure 2** shows where we currently see the roles on the central-decentral spectrum, although those positions cannot be considered as final and require further research.

7. Discussion and outlook

Our results strongly support the assumption that an institutionalized legal entity can act as an enabler for a SME data sharing ecosystem. Besides bundling resources for data management and data analysis and exploiting economies of scale and scope, such an entity can act both as a trusted and neutral middleman in the ecosystem and as an anchor for an ecosystem-wide data governance. The related findings also positively answer our *first research question* on the relevance of an institutionalized entity.

The *second research question* led to the derivation of a preliminary role catalogue that is visualized in **Figure 2**. As for the role distribution – *third research question* – we found that there is indeed some leeway. There is a spectrum between central and decentral models that can be chosen from according to the idiosyncrasies of the business network in discussion. This is similar to the implementation of an AI/BI/DQ competence center. Unlike those, the institutionalized entity is a separate legal unit that can take over additional administrative and supportive roles like capital provider. This includes one core role that is innately tied to it: the *coordinator of the ecosystem*. The coordinator is responsible for IT resources (either by providing them directly or by sourcing them), and for the strategic alignment of the data analytics activities.

One of the remaining open questions is what legal forms are best suited for implementing the given approach. We are piloting our approach in a German environment that has the advantage of providing an established legal form for exactly our purpose: the (German) cooperative. We acknowledge the necessity to assess other legal forms and settings and thereby come to stronger generalizations of our findings.

The presented work is embedded in a wider context where we analyze further topics: aspects of cost accounting and pricing, concrete data governance, acceptance of data sharing, suited BIA architectures, BIA solutions, and process models. We consider these to have great potential and even deem it to be strategically critical for SMEs. For this reason, we encourage research of all sorts into this field.

9. References

[1] Davenport, T. H. and H. G. Harries, Competing on analytics: The new science of winning, Harvard Business Press, Boston, 2007.

[2] Krishnamoorthi, S. and S. K. Mathew, "Business analytics and business value: A comparative case study", Information & Management, Elsevier Science, 55(5), 2018, pp. 643-666.

[3] Shanks, G., N. Bekmamedova, and L. Willcocks, "Using business analytics for strategic alignment and organisational transformation", International Journal of Business Intelligence Research, IGI Global, 4(3), 2013, pp. 1-15.

[4] Dinter B., "The Maturing of a business intelligence maturity model", Americas Conference on Information Systems (AMCIS 2012), 2012, 37.

[5] Comuzzi, M. and A. Patel, "How organisations leverage big data: A maturity model", Industrial Management & Data Systems, Emerald, 116(8), 2016, pp. 1468-1492.

[6] Bianchini, M. and V. Michalkova, "Data analytics in SMEs: Trends and policies", OECD SME and Entrepreneurship Papers, No. 15, OECD, Paris, 2019.

[7] Coleman, S., R. Göb, G. Manco, A. Pievatolo, X. Tort-Martorell, and M. S. Reis, "How can SMEs benefit from big data? Challenges and a path forward", Quality and Reliability Engineering International, Wiley, 32(6), 2016, pp. 2151-2164.

[8] OECD, The Digital Transformation of SMEs, OECD Studies on SMEs and Entrepreneurship, OECD Publishing, Paris, Chapter 5: Artificial intelligence: Changing landscape for SMEs, 2021.

[9] Kaisler, S. H., J. A. Espinosa, F. Armour, and W. H. Money, "Advanced analytics – Issues and challenges in a global environment," 47th Hawaii International Conference on System Sciences (HICCS), IEEE, 2014, pp. 729-738.

[10] Fathy, Y., M. Jaber, and A. Brintrup, "Learning with imbalanced data in smart manufacturing: a comparative analysis." IEEE Access 9, IEEE, 2020, pp. 2734-2757

[11] Rupek, T., "Establishing Governance Structures for Analytics-Driven Interorganizational Data Sharing Networks – Designing a Framework Based on a Qualitative Study", Lernen, Wissen, Daten, Analysen (LWDA 2021), CEUR-WS, 2021.

[12] De Haes, S. and W. Van Grembergen, "IT governance and its mechanisms", Information Systems Control Journal, ISACA, 1, 2004, pp. 27-33.

[13] Ghosh, A. and J. Fedorowicz, "Governanance mechanisms for coordination and information sharing in supply chains: The role of trust", Americas Conference on Information Systems (AMCIS 2005), AIS, 18.

[14] Espinosa, J. A. and F. Armour, "The big data analytics gold rush: A research framework for coordination and governance," 49th Hawaii International Conference on System Sciences (HICSS), IEEE, 2016.

[15] Marjanovic, O. and B. Dinter, "25+ years of business intelligence and analytics minitrack at HICSS: a text mining analysis", 50th Hawaii International Conference on System Sciences (HICSS), IEEE, 2017.

[16] Power, D. J., C. Heavin, J. McDermott, and M. Daly, "Defining business analytics: an empirical approach", Journal of Business Analytics, Taylor & Francis, 1(1), 2018, pp. 40-53.

[17] Baars H. and H.-G. Kemper, "Management support with structured and unstructured data – An integrated business intelligence framework", Information Systems Management, Taylor & Francis, 25(2), 2008, pp. 132-148.

[18] Chen, H., R. H. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact", MIS Quarterly, AIS, 36(4), 2012, pp. 1165-1188.

[19] Chui, M., J. Manyika, M. Miremadi, N. Henke, R. Chung, P. Nel, and S. Malhotra, Notes from the AI frontier: Insights from hundreds of use cases, McKinsey Global Institute, 2018.

[20] Verma A., K. M. Yurov, P. L. Lane, and Y. V. Yurova, "An investigation of skill requirements for business and data analytics positions: A content analysis of job advertisements", Journal of Education for Business, Taylor & Francis, 94(4), 2019, pp. 243-250.

[21] Vidgen, R., S. Shaw, and D. B. Grant, "Management challenges in creating value from business analytics", European Journal of Operational Research, Elsevier 261(2), 2017, pp. 626-639

[22] Moore, J. F., "Predators and prey: A new ecology of competition", Harvard Business Review, 71(3), 1993, pp 75-86.

[23] Mazhelis, O., L. Eetu, and H. Warma, Defining an internet-of-things ecosystem, Internet of Things, Smart Spaces, and Next Generation Networking, Springer, 2012, pp. 1-14.

[24] Adner, R., "Ecosystem as structure: An actionable construct for strategy." Journal of management, SAGE, 43.1, 2017, pp. 39-58.

[25] Weber, P., S. Hiller, and H. Lasi, "Identifying business potentials within an IoT ecosystem – An explorative case study in the industrial domain," Americas Conference on Information Systems (AMCIS 2020), AIS 2020, 19.

[26] Braud, A., G. Fromentoux, B. Radier, and O. Le Grand, "The road to European digital sovereignty with Gaia-X and IDSA", IEEE Network, 35(2), 2021, pp. 4-5.

[27] GAIA-X – a federated data infrastructure for Europe - Federal Ministry for Economic Affairs and Energy of Germany, Last accessed 15th of June 2018 https://www.data-infrastructure.eu/GAIAX/

[28] International Data Spaces e. V., https://internationaldataspaces.org/, Last accessed June 15th of 2021.

[29] Böhm, M., G. Koleva, and H. Krcmar, Towards a generic value network for cloud computing, International workshop on grid economics and business models, Springer, 2010.

[30] Tian, C. H., B. K. Ray, J. Lee, R. Cao, and W. Ding, "BEAM: A framework for business ecosystem analysis and modeling", IBM Systems Journal, IBM, 47(1), 2008, pp.101-114.

[31] Miller, G. J., D. Bräutigam, and S. V. Gerlach, Business intelligence competency centers: A team approach to maximizing competitive advantage, Wiley, Hoboken, 2006.

[32] Foster, K., G. Smith, T. Ariyachandra, and M. N. Frolick, "Business intelligence competency center: Improving data and decisions", Information Systems Management, Taylor & Francis, 32(3), 2015, pp. 229-233.

[33] Unger, C., H.-G. Kemper, and A. Russland, "Business intelligence center concepts", Americas Conference on Information Systems (AMCIS 2008), AIS, 147.

[34] Otto, B. and H. Österle, Corporate data quality: Prerequisite for successful business models, Springer, Berlin and Heidelberg, 2015.

[35] Davenport, T. H. and S. Dasgupta, "How to set up an AI center of excellence", Harvard Business Review Digital Articles, 2019.

[36] Hawking P. and C. Sellitto "Business intelligence strategy: A utilities company case study", International Journal of Enterprise Information Systems, IGI Global, 11(1), 2015, pp. 1-12.

[37] Gansor, T. and A. Totok, Von der Strategie zum Business Intelligence Competency Center (BICC): Konzeption-Betrieb-Praxis, dpunkt, Heidelberg, 2015.

[38] Cupoli, P., S. Earley, and D. Henderson, DAMA – DMBOK: Data management body of knowledge, DAMA International, 2014.

[39] Yoo, Y., O. Henfridsson, and K. Lyytinen, "Research commentary—the new organizing logic of digital innovation", Information systems research, INFORMS, 21(4), 2010, pp. 724-735.

[40] Cooper R. and R. Slagmulder, "Interorganizational cost management and relational context", Accounting, Organizations & Society, ScienceDirect, 29(1), 2004, pp. 1-26.

[41] Hevner, A. R., "A three cycle view of design science research", Scandinavian Journal of Information Systems, AIS, 19(2), 2007, pp. 87-92.

[42] Flick, U., An Introduction to Qualitative Research, Sage Publications, Inc.: New York, 2018.

[43] Myers, M.D. and M. Newman, The qualitative interview in IS research: Examining the craft. Information and Organization, Elsevier, 17, 2007, pp. 2–26

[44] Zerche, J., I. Schmale, and J. Blome-Drees, Einführung in die Genossenschaftslehre, Oldenbourg, 2016.

[45] Mayring, P., Qualitative Content Analysis: Theoretical Foundation, Basic Procedures and Software Solution, SSOAR, 2014.

[46] Miles, M. B., A.M. Huberman and J. Saldana, Qualitative Data Analysis - A Methods Sourcebook, Sage Publications, 2013.