Rahlmeier, Nils; Hopf, Konstantin

Bridging Fields of Practice : How Boundary Objects Enable Collaboration in Data Science Initiatives

Date of secondary publication: 05.12.2024

Accepted Manuscript (Postprint), Conferenceobject

Persistent identifier: urn:nbn:de:bvb:473-irb-982156

Primary publication

Rahlmeier, Nils; Hopf, Konstantin (2024): Bridging Fields of Practice : How Boundary Objects Enable Collaboration in Data Science Initiatives, in: Wirtschaftsinformatik 2024 Proceedings, New York: AIS, https://aisel.aisnet.org/wi2024/55/.

Bridging Fields of Practice: How Boundary Objects Enable Collaboration in Data Science Initiatives

Research Paper

Nils Rahlmeier[1,2], Konstantin Hopf[1]

[1] University of Bamberg, Information Systems and Energy Effieicnt Systems, Bamberg, Germany, konstantin.hopf@uni-bamberg.de
[2] e.venture consulting, Berlin, Germany, nils.rahlmeier@e-vc.org

Abstract. Data-intensive technologies draw high investments. Yet, data science projects are reported to suffer from poor collaboration, unrealistic expectations, and difficulties in realizing practical solutions between business and data science units. Moving beyond the currently prevalent approach to study data science practices, our study emphasizes the use of boundary objects between data science and collaborating fields. We interviewed collaborators from diverse fields in six organizational data science initiatives. Our inductive analysis of this rich data source uncovered six distinct mechanisms and six archetypes of boundary objects in data science projects. While archetypes that we label Alignment, Temporary, Collaboration, and Outcome are procedural and appear in selective stages of the data value creation process, the archetypes Infrastructure and Upskilling support projects along the value creation process. The archetypes and their mechanisms inform the management of data science initiatives, help to advance boundary object theory, and provide instruments to study data science initiatives.

Keywords: Management of Data Science Initiatives, Boundary Objects, Knowledge Sharing, Organizational Development

1. Introduction

Companies have been investing heavily in data-intensive technologies (e.g., big data, business analytics, artificial intelligence) in recent years, hoping to add value by making more effective, objective, and neutral decisions (Chen et al., 2012; Müller et al., 2018; Wu et al., 2019) or increasing their efficiency (Krakowski et al., 2022). Yet, many organizations struggle to carry out the data science initiatives that are necessary to adopt and leverage these technologies (Joshi et al., 2021; Someh et al., 2023). Many such projects are reported to be ad-hoc initiatives that never leave the pilot phase (Fountaine et al., 2019). Literature cites several reasons why data science initiatives fall short of delivering business value. For example, poor collaboration between business and data science units, unrealistic expectations about the technology, and the long way from "doing data science" to implementing operational solutions (Glushko, 2023; Joshi et al., 2021). Many challenges can be directed to a mismanagement of the interfaces between data science and other multidisciplinary collaborators (Hopf et al., 2023).

The field of data science has evolved over the past decade (Davenport and Patil, 2012, 2022). It is characterized by interdisciplinary knowledge work (Someh et al., 2023) and follows a non-linear explorative path involving cross-disciplinary expertise from multiple actors from different fields of practice (Avnoon, 2021; Parmiggiani et al., 2022; Pawlowski and Robey, 2004). Several studies demonstrate that data science projects are qualitatively different from other IT or non-IT projects, mainly because they use data as their material. Data is a constantly changing resource, which leads to constantly changing and unstable outputs of data science projects (Hopf et al., 2024; Parmiggiani et al., 2022). The tools used also differ, as many data science projects rely on learning algorithms that automatically extract patterns from large amounts of data (Faraj et al., 2018). This often leads to “black boxes” in which many outputs and data representations are difficult to understand and interpret (Anthony, 2021). In conventional software, by contrast, all instructions must be explicitly programmed.

At the center of this complex work environment is the *data scientist*, who appears to be the most important link between the different fields of practice (Shollo et al., 2022). Originally described as barely existent experts, the “hybrid of data hacker, analyst, communicator, and trusted adviser. The combination is extremely powerful—and rare" (Davenport and Patil, 2012: 73), these individuals are key players in creating value from data (Shollo et al., 2022) by integrating different parts of organizations.

Information systems (IS) research has made efforts to understand the practices that data scientists employ while they interact with technology (e.g., Grønsund and Aanestad, 2020), domain experts (e.g., Stice-Lusvardi et al., 2023; van den Broek et al., 2021), and managers (Hopf et al., 2023) separately. Yet, our knowledge of how data scientists work together with the variety of other actors in organizations is limited. This lack of clarity makes it difficult to establish appropriate working conditions for effective data science work and to manage collaboration in data science initiatives.

Our study addresses this knowledge gap by exploring the underlying dynamics of collaboration in data science projects. We selected *boundary objects* (Bowker and Star, 1999; Star and Griesemer, 1989) as the unit of analysis, which are objects “intersecting social worlds … *and* satisfy the informational requirements of each of them” (Star and Griesemer, 1989: 393). These objects are exchanged between data scientists and other actors from diverse fields of practice. Effective boundary objects in place decrease the need for coordination and allow shifting concerns between local and global focus (Huvila et al., 2017). They establish focal points that foster collaboration and bridge the gap between various knowledge domains. Boundary objects are an established theoretical frame in an organization and IS research when it comes to understanding and improving the collaboration of different stakeholders. Thus, this paper explores the question: *How do data scientists collaborate with other fields of practice (e.g., management, frontline employees, and IT) through the use of boundary objects?*

Our research uncovers six mechanisms and archetypes of boundary objects, and we explain their functions within data science projects and organizations. We position the identified boundary objects within the value creation process and emphasize the significance of approaching and developing a specific subset of them strategically. Furthermore, our research highlights the data scientist's pivotal role in bridging different fields of practice through the use of boundary objects while also serving as an educator within the organization. Before we give a detailed account of our study and its result, we introduce the phenomenon of data science and the theory of boundary objects.

2. Theoretical Background

2.1. Management of Data Science Initiatives

The scholarly discourse on the management and strategy related to data-intense technologies began a decade ago with the realization that “big data” exists in corporate IT systems (e.g., Chen et al., 2012). Literature acknowledges that a complex process is necessary to create value from data (Sharma et al., 2014; Thiess and Müller, 2018) and that advanced analytics (using statistics and machine learning) is necessary to lift any potential from data. So far, IS literature has examined issues related to technology design (e.g., Kane et al., 2021), the interaction of individuals with the technology (Grønsund and Aanestad, 2020; van den Broek et al., 2021), or how data-intensive technologies might change the work of individuals (Waardenburg et al., 2022).

Increasing interest is paid to the work of data scientists, who are the actors in organizations that develop data-intensive applications (Parmiggiani et al., 2022). Yet, empirical studies come to partly conflicting results: On the one hand, studies put data scientists into a strong position by demonstrating that they deliberately include and exclude domain experts or ML models to develop AI applications (van den Broek et al., 2021) or that are the key resource orchestrators in corporate AI initiatives (Shollo et al., 2022). On the other hand, Stice-Lusvardi et al. (2023) find that data scientists compromise their “good” practices in order to get their work accepted by others. Other studies show that the identity of data scientists is subject to constant change because of the emergence of their occupation (Avnoon, 2021; Vaast and Pinsonneault, 2021).

To create effective work environments and successfully manage data science initiatives, a more advanced understanding of the data science work boundaries with other fields of practice is necessary. To operationalize this investigation, we leverage boundary objects as the unit of analysis.

2.2. Boundary Objects

When heterogeneous actors from different social worlds collaborate, a central tension appears between divergent viewpoints and the need for generalizable findings. Star and Griesemer (1989) found that individuals use boundary objects to manage this tension to achieve a common goal or interest. These objects must be “plastic enough to adapt to local needs and the constraints of the several parties employing them, yet robust enough to maintain a common identity across sites” (Star and Griesemer, 1989: 393).

They identified four types of boundary objects in their study,[1] and over time, several additional types and classifications of boundary objects from different fields were contributed (Levina and Vaast, 2005; Star, 2010).

Boundary objects are artifacts that convey status information, discussions, and agreements (e.g., Doolin and McLeod, 2012) across the boundaries of organizations and stakeholder groups. Levina and Vaast (2005) identify the active use of an object by project participants from different fields of practice as a criterion that separates boundary objects from other types of objects. This is reflected in the concept of "designated boundary objects" versus "boundary objects-in-use" (Levina and Vaast, 2005). The latter class describes objects that emerge locally and create usefulness in a joint field. Marabelli et al. (2017) expanded the theory by examining temporal aspects of boundary objects. Although boundary objects have been analyzed in the context of information systems (Marabelli et al., 2017) and agile software development projects (e.g., Ben Chouikha and Dakhli, 2015; Zaitsev et al., 2016), their role in data science project environments has yet to be explored.

3. Method

We conducted an exploratory study of six cross-functional data science projects. At the core of our study are data scientists and their work relations to one or two collaborators from different fields of practice, which we refer to as a case. In total, we conducted 14 semi-structured interviews between Jan and Feb 2023 (average duration: 52 minutes, with a standard deviation of 8 minutes, in total 12 hours).

Study Design—A comprehensive understanding of boundary objects requires soliciting the perspectives of representatives from different social worlds. Consequently, we sought to represent *at least two perspectives on the same project from different fields of practice* to avoid bias that could arise by only interviewing data scientists. Hence, our study centers around one data scientist and at least one collaborator (see Table 1).

Sampling Strategy—We tried to achieve a broad coverage of boundary objects by using a maximum variation sampling (Patton, 2002: 243) and involved different organizational contexts and participants at various levels. Although all employers of the interview partners had locations in Germany, almost all of them have an international presence, which resulted in half of the interviews being conducted in German and the other half in English. We contacted 161 individuals via personal contact or searching for data scientists' profiles on LinkedIn, and received a response from 75 of them. Given that all potential interviewees had to find colleagues they have worked with in a cross-functional team, we could schedule 18 interviews. Four of them (where we had only a data scientist but no collaborator) served as pilot tests to refine our interview guide and identify potential issues. These four pilot interviews were excluded from the

[1] Without the intention of being exhausting, Star and Griesemer (1989) describe four boundary object types: *Repositories* (ordered pile of data or shared resource collections used by multiple stakeholders), *Ideal types* (conceptual models that capture essential features), *Coincident boundaries* (points where different perspectives overlap and require coordination), and *Standardized forms* (agreed-upon formats that facilitate communication and collaboration).

analysis sample. Prior to the interview, we had a phone call with the data scientist to determine if the person had been involved in a data science project (currently or in the past) and if other colleagues were available to be interviewed as part of the study. After conducting around ten interviews and covering four cases, recurring themes began to emerge, and the acquisition of new information during the latter interviews declined.

Structure and Contents of the Interviews—The interviews were structured around using and managing boundary objects in one data science project of the company. The interview guide remained relatively stable over time and had three sections: First, we asked contextual questions regarding the person, their organization, and the focal project. Second, we clarified project-related boundaries and objects. Third, we delved into the use and management of boundary objects to conduct an in-depth discussion of selected objects. To facilitate this discussion, the objects and project participants were listed on a shared screen, and we discussed the use, creation, and management of these objects in relation to the project participants. Third, we asked about the roles and boundary activities that the interview partners performed in the team.

Table 1. Case Overview and Interview Participants

#	**Participants** (Interview duration in minutes)	**Industry**	**Size**	**Project goal**	**Location of data science**
1	Product Owner (IP12, 51); Sen. Data Analyst (IP 13, 50); Data Science Lead (IP14, 56)	E-Commerce	>2500	Classify and recommend relevant influencer videos that feature products for users of an e-commerce app.	Internal
2	Product Owner (IP3, 51); Data Scientist (IP4, 52); Data Scientist (IP5, 51)	E-Commerce	500 - 2500	Predicting trustworthiness of customers to offer additional payment methods; Advanced forecasting for employee capacities and product quantities.	Internal
3	Lead Data Scientist (IP1, 63); Software Developer (IP2, 39)	Optical technologies	>2500	Automated forwarding of online submitted contact forms based on document understanding to internal recipients.	Internal
4	Project Manager (IP7, 47); Data Scientist (IP6, 44)	Analytics Consulting	500 – 2500	Digitization and document understanding of analog forms for a retail bank.	External
5	Manager (IP10, 43); Data Scientist (IP11, 61)	Analytics Consulting	>2500	Advanced time-series models for revenue forecasts with automated retraining.	External
6	Managing director (IP8, 65); Data Scientist/Software Engineer (IP9, 47)	Optical Engineering	<25	Predictive adjustment of production parameters on the basis of sensor data in order to increase quality of output.	External

Data Analysis—Both audio and screen sharing was recorded and transcribed verbatim, leading us to 205 pages of original text. We analyzed all data (including the notes and tables created and used within the interview) using the software MAXQDA. We used a multi-stage coding process, starting with initial open coding and establishing a focus for subsequent rounds by comparing it to the literature. We then did five (partly iterative) rounds of open, axial, and selective coding, addressing disagreements between researchers and constant comparison to ensure our theory was grounded in data:

Codig step 1: Initial screening and open coding—We started with open coding to give “extraordinary voice” to our informants while also aligning ourselves closely with their unique terminologies (Gioia et al., 2013). We came to a high number of 1192 open codes that covered diverse topics, e.g., agile working, collaborative tools, and culture.

Coding step 2: Focused and axial coding to identify boundary object instances—In our next iteration, we included our theoretical focus on boundary objects and first grouped all codes related to what indicates a boundary object. Our coding process encountered difficulties in classifying boundary objects, and we began with a large set of over 80 objects—a challenge that other studies also had to cope with (Star, 2010). We started with a subsequent axial coding using researcher-centric notions (Gioia et al., 2013) to develop categories representing boundary objects or the functions of boundary objects in data science. To enhance validity and reliability, we established clear criteria of inclusion—objects that are positioned and used over the boundaries of different social worlds (Star and Griesemer, 1989); neither used solely as pure input or output and objects that are in use (Levina and Vaast, 2005)—and exclusion—objects that are only created and used by one person, are used on a small scale (Star, 2010), objects not plastic enough (Star and Griesemer, 1989) and objects that are collaborative tools. With these criteria, we reduced the inventory to 54 instances of boundary objects.

Coding step 3: Finding mechanisms of boundary objects—In addition to the codes that helped us to find instances of boundary objects, we analyzed the remaining codes to understand the functions of boundary objects. That is, how boundary objects help individuals to coordinate and communicate across diverse fields of practice. We grouped the functions into mechanisms of boundary objects. Thereby, we tried to eliminate deviations in the formed dimensions and achieve theoretical saturation (Glaser and Strauss, 1967). In the end, we identified six mechanisms of how boundary objects facilitate collaboration between different fields of practice in data science.

Coding step 4: Conceptualization of boundary objects—In our next analysis step, we aimed to aggregate the boundary object instances into a low number of archetypes of boundary objects. To do so, we engaged in an iterative selective coding process to group together boundary objects that serve the same mechanism (as identified previously).

Coding step 5: Value creation mapping—In our last analysis step, we mapped the discovered boundary object (archetypes) onto the data value creation process (Sharma et al., 2014; Thiess and Müller, 2018). This approach allowed us to take an all-encompassing perspective beyond a mere operational view to encompass a strategic and organizational perspective. We assigned the archetypes by reviewing the literature on the various phases of value creation and the challenges at the transitions (see Figure 3).

4. Findings

The data science projects in our sample spanned across several fields of practice with different understandings of data science work. Yet, 13 out of 14 interviewees indicated a joint field could be established, also through the use of boundary objects. Before we describe archetypes of those, we describe mechanisms that boundary objects fulfilled.

4.1. Mechanism of Boundary Objects in Data Science

In our third analysis step, we inductively derived the boundary object mechanisms reported by informants from the data and coded them as aggregate dimensions:

1. *Understand & Define the Problem (UDP)*, identify and understand the client's business needs, and agree on and specify requirements.
2. *Coordination & Management (CM)*, reviewing and coordinating tasks and capacities, maintaining central project information for documentation and orientation, and summarizing and communicating project status.
3. *Create Common ground (CCG)*, actively train collaborators, and improve intra-team relationships.
4. *Solve problems (SP)*, establish a framework for collaborative work, regular reviews and challenges of work in progress, and discuss the current solution approach.
5. *Integrate Experience (IE)*, collaboratively integrating customer understanding, discussing and negotiating project-wide solution approach, and involving expert communities in the exchange.
6. *Share Results (SR)*, delivering actionable insights to the client, enriching their knowledge, and making results understandable for their decision-making.

4.2. Archetypes of Boundary Objects

Each archetype that we found describes a set of boundary object instances that effectively fulfill the same core function. The identified archetypes are shown in Table 2 and we describe each of them briefly in the next section. To better represent the complexity of the objects in each archetype, we visually represent a frequency distribution of their functions (according to the six mechanisms mentioned above). For these figures, we determine the relative frequency of codes regarding mechanisms in each archetype after the third coding step. That means we counted the number of times each mechanism was mentioned for each archetype and divided this by the total number of mechanism codes per archetype. We show these numbers in a radar chart in Table 2 using the abbreviated mechanisms.

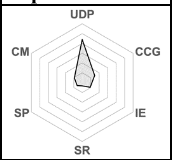
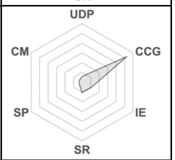
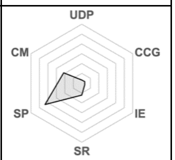
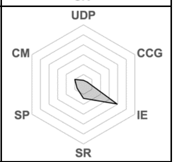
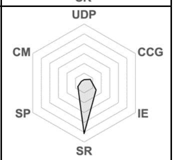
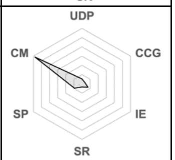
4.3. Boundary Objects Along the Data Value Creation Process

The archetypes stretch throughout the entire trajectory of data science projects. While the Alignment (I) and the Upskilling (II) archetypes particularly supported the pre-development phase of projects, the Temporary (III) and Infrastructural (VI) archetypes were mainly present in the development phase of projects, the Collaboration (IV) and Outcome (V) archetypes in the refinement and deployment phase of projects. We use the data value creation process (Sharma et al., 2014; Thiess and Müller, 2018) to locate the main use of this archetype along the process model, as we illustrate in Figure 3.

The *Alignment Archetype (I)* consists of boundary objects that help reach a consensus on defining a problem across diverse fields of practice (e.g., requirements document, problem statement). The archetype supports and engages the articulation of each field of practice's unique view, goals, and priorities, and makes their perspective visible to the whole team. The archetype helps to overcome the *"huge communication gap between the business and the data scientist. With incomplete understanding ... you start working ... and when you present results, you will get the feedback: 'But that's not what we were interested in.'" (Data scientist, IP3)* The boundary object instances of the

Alignment type are created by the client or collaboratively at the start of a project. After fast-paced development, these objects serve as orientation throughout the whole project. They are not actively terminated, rather no longer used as the project progresses, but may be transformed into other boundary objects.

Table 2. Boundary Object Archetypes

Type and Description	Examples	Functions (Categories)	Impact
I. Alignment: Support and represent the process of finding and formulating the consensus on the problem definition from different social worlds.	Requirement Catalogue, Service Agreement, Problem Statement	1) Identifying and understanding the business needs of the client 2) Agree on & specify requirements	
II. Upskilling: Boundary objects whose content is purposefully designed and used to convey one's own social world to another.	Upskilling Documents/Slides, Explanation/Interpretation aids	1) Actively train the client 2) Improve team-client relationship.	
III. Temporary: Temporary outcomes, used as a facilitator for multiple people to work together on complex problems.	Mock-Ups, Sketches, Working Documents	1) Framework for collaborative work 2) Review and challenge work in progress 3) Discuss current solution approach	
IV. Collaboration: Content encourages different social worlds to discuss their views with the aim to integrate and enrich knowledge from other social worlds.	Dashboards/Metrics Ideation Catalogue, User Stories, Project Goals, Review Presentation	1) Incorporating customer's business understanding 2) Project-wide discussion and negotiation of solution approach 3) Exchange and involvement of expert communities	
V. Outcome: Pool the output of other objects and grow over the course of the project from a preview to the finished project deliverable.	Application, final analysis presentation, Use Case Documentation, Data as a result	1) Presenting results for customer's decision-making. 2) Handover results and enrich knowledge 3) Making insights understandable for the client	
VI. Infrastructure: Content represents a comprehensive overview as orientation across all social worlds nevertheless, each social world can adapt the content to their own.	Roadmap, Backlog, Tickets, Project Charter	1) Review and coordination of tasks and capacities 2) Central project information for documenta orientation 3) Summarize and communicate project status	

The *Upskilling Archetype (II)* is used to communicate one's social world to others. These boundary objects come in the form of introductory slides, interpretation aids, etc. They actively train collaborators to build a common understanding and to improve team-client relationships to foster communication, collaboration, expectations, and trust between the data science team and the client. A senior data scientist and manager described the effect of actively upskilling the client: *"We need their [business side] help regarding internal drivers to create a feeling for what is important, to enable them to work with us, but also to make the collaboration more productive." (IP10)* Upskilling the client is about enabling the team to work more productively on the project while preparing them for long-term use of the project outcome. The boundary objects also increase the acceptance of the project and motivation of all participants by generating transparency. By actively training and involving the client, the boundary objects can foster shared ownership of the project and promote a sense of partnership. Upskilling boundary objects are created by experts independent of specific projects and are continuously developed based on years of professional experience.

The *Temporary Archetype (III)* facilitates collaboration on complex problems among multiple people. Examples are mock-ups and sketches that help to "*break down problem complexity. When you ... have to get a feel for [a problem]. ... Sometimes, it is also a matter of simply discussing what the best way to program something or to explain something quickly is, there is a visualization option." (Data scientist, IP11)* These boundary object instances are focused on improving the quality and effectiveness of teamwork within a complex problem-solving environment. By encouraging critical thinking, open communication, and constructive feedback using these boundary objects, teams can work together to develop more sophisticated and effective solutions. Temporary problem-solving objects are created spontaneously and unplanned during the project's progress, are updated continuously while in use. They are used only for a short time after their creation and, if deemed unhelpful, are terminated immediately. If they served their purpose, they may be transferred to other boundary objects.

The *Infrastructure Archetype (IV)* refers to critical components of effective project management (e.g., project roadmaps, product backlog). *"We use them to send to higher management, to newcomers, to people who are interested in the project, show them where we are, what we've done, what it looks like." (Product Owner, IP13)* They enable team-internal and project-wide evaluation of progress, coordinating tasks, and capacities. They provide orientation, communicate project status, documentation, and ensure access to necessary resources. They can also be used as an abstract high-level snapshot of the whole data science initiative to inform new team members or further uninvolved fields of practice. The Infrastructure boundary objects are either pre-existing templates in the departments adopted by the data science team or created anew by them. These objects become more specific and granular during the project, forming sub-types tailored to more specific use cases. They are continuously used and extended throughout the project and are not actively terminated but may no longer be used or transferred to another object when their purpose is fulfilled.

The *Collaboration Archetype (V)* facilitates discussion and the integration of knowledge from different social worlds (e.g., dashboards, user stories). *"We present something, and the [business side] will give their opinion ..., further suggestions ..., maybe 'you can also add this parameter', or 'from our experience, we have seen that this is something which might improve the detection rate'." (Data scientist, IP3)* These boundary objects facilitate project-wide discussion and negotiation of the solution approach by engaging team members to leverage their diverse expertise and identify potential issues. Collaboration archetypes are created within the project process by the data science team or collaboratively with the client for specific use cases. The content of these objects undergoes continuous development during the project through evaluation, reflection, consolidation, adaptation, and refinement while the format remains consistent. They are used continuously at specific points in the project and serve as a basis for initiating further steps. Once their purpose has been fulfilled and their content transferred to another object, they are dropped.

The *Outcome Archetype (VI)* merges and combines boundary objects that were created earlier and evolves from a preview to the final deliverable (e.g., application, final analysis). They are linked by their focus on effectively communicating knowledge to the client. Presenting results for customer decision-making involves clear and concise communication to enable informed decision-making. A data scientist explained *"At the end of the day it is about investments. For us, it was our final deliverable. And for the customer, it is just the next means for the subsequent communication meetings." (Data scientist, IP9)* Handing over results and enriching knowledge involves the transfer of information in an understandable way. Prioritizing clear communication and ongoing support enables customers to make informed decisions and fully utilize the insights gained, leading to better outcomes. The Outcome boundary objects are continuously developed across the project, refined, streamlined, and finalized by the data science team. Upon completion, the object is handed over and responsibility is transferred to the appropriate individuals or teams.

4.4. Procedural and Supporting Archetypes

We observed that certain aspects of data science extend beyond operational activities during a project's life cycle and are better viewed with a strategic perspective on value creation for the organization. Thus, we mapped the primary use of the six archetypes to phases in the data value creation process (Sharma et al., 2014; Thiess and Müller, 2018), as we show in Figure 3. As a result, we found four archetypes whose primary functions support the transition in the value creation (we describe those as *procedural archetypes*). The two remaining archetypes (Infrastructure and Upskilling) act *supportively* and throughout the entire process.

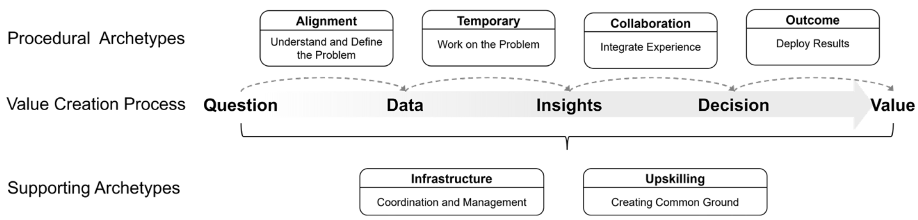


Figure 3. Overview Procedural and Supporting archetypes

Supporting archetypes are just as essential for the overall project outcome as the other archetypes, yet, they have an "indirect" impact on the transitions in the value creation in data science. These archetypes are employed to enhance the performance of other archetypes or to mitigate their weaknesses. The use of Infrastructure and Upskilling boundary objects often appeared in combination with boundary objects of the other archetypes. Surprisingly, we found that boundary objects with the main function of educating collaborators from other fields of practice come along with all other boundary object archetypes. With regard to the scoping phase of a project (where mainly Alignment boundary objects are used), a senior data scientist described the importance of using Upskilling boundary objects to create common ground: *"You cannot really get everyone on board without [upskilling]. That is essential for the project to succeed. … If people do not understand it, they will not use it."* (IP10) Furthermore, upskilling has positive side effects, as it enables the business side to imagine data science applications: "*Initially, the data science team had to approach the business department. Now the business department is actively approaching the data science team for assistance in obtaining and understanding data to gain insights."* (IP5)

5. Discussion

Our investigation is—to the best of our knowledge—the first study that unfolds the nature and use of boundary objects in the field of data science. It is therefore a contribution to the current discourse on the effective organization of data science work (Parmiggiani et al., 2022). We found six *mechanisms* and six *archetypes* of boundary objects that are different from earlier boundary object archetypes that were suggested for other fields (e.g., Ben Chouikha and Dakhli, 2015; Star and Griesemer, 1989).

Knowledge of the functions and use of domain-specific boundary objects helps researchers to further study the domain and practitioners to effectively design and use these boundary objects. While previous studies on data science initiatives focused on the relation and interactions between data scientists and experts (Stice-Lusvardi et al., 2023; van den Broek et al., 2021) or the interaction of data scientists with intelligent technologies (e.g., Grønsund and Aanestad, 2020), the boundary objects we identified allows to better manage and study of the interaction of data scientists with other fields of practice on a more granular level. Thus, *the identified boundary objects are instruments that future studies can use as sensitizing devices*.

In addition, we made two observations that inform the theory of boundary objects. First, one property of boundary objects is that they work as "a 'lowest common denominator' which satisfies the minimal demands of each world by capturing properties that fall within the minimum acceptable range of all concerned worlds" (Star and Griesemer, 1989: 404). Thereby, the objects can bridge boundaries and manage conflicting views. In our study, however, data scientists struggled to create objects that serve as the lowest common denominators when making insights understandable to collaborators with low data skills. Instead of "dumping down" the complexity to match the lowest denominator, data scientists started to train their collaborators, for example, using Upskilling boundary objects. Hence, in their effort to collaborate with other social worlds, data scientists have to identify and raise the lowest denominator and evaluate how they *can keep a certain required level of the complexity of their own work environment* across multiple social worlds. This is important for data science because many models are very complex or black boxes, making it essential for data scientists to explain limitations and potential risks associated with their use.

Second, our study found that some boundary objects fulfilled functions that were, at first sight, not reflected in their primary function. Temporary objects such as mock-ups and sketches are primarily utilized as frameworks for collaborative work on complex problems and for challenging current working results. However, beyond their primary functions, these objects also serve to provide orientation in project planning. They spark extensive project-wide reviews and coordination of tasks and capacities, which are integral aspects of the primary functions of infrastructural archetypes. These secondary functions and contributions of temporary boundary objects are sustained and transferred through their use in renewing and iterating infrastructural objects. Thus, it is crucial to perceive boundary object archetypes not only as discrete entities but also in relation to each other and how they interact within their social contexts.

6. Conclusion and Future Research

Our study demonstrates that a focus on boundary objects helps to better understand the current (sometimes unsatisfactory) outcomes of data science initiatives. This is because collaboration can be broken down and explored to the expectations as well as value propositions of individual detailed activities for interdisciplinary collaboration.

Implications for practice—The six mechanisms and six archetypes of boundary objects for data science projects are helpful in practice in three ways. First, the mechanisms and archetypes help to classify artifacts in data science projects as boundary objects. Practitioners can thereby better understand the functions of the different artifacts. Our findings can also be used to identify potential missing boundary object types that practitioners have not deployed. Second, data scientists can improve the design of boundary objects using the identified mechanisms to leverage them to enhance collaboration across different stages of data science initiatives, e.g., by formulating requirements for collaboration documents. Third, activities such as actively training non-data scientist staff and building working relationships with various other departments should be recognized by management and included in the scope of data scientist activities.

Limitations and future research—Future research should extend our study to enhance the empirical basis of our findings. Our data analysis indicated a point of theoretical saturation already in the 14 interviews with informants from six organizations. For a more thorough understanding of boundary work in data science, our research approach should be continued. We cannot exclude a bias in the selection of our cases and informants, as all companies were located in Germany. It was very hard to find cases where the data scientist was confident or willing to approach other colleagues to join the study. Consequently, we may have sampled participants where a healthy work environment (or a more open organizational culture) was already established, and the relationships may have been stronger than in the average organizational setting. Further analysis should also focus on the temporal dynamics of boundary objects, which will help to identify success factors to design and realize boundary objects for the effective management of data science projects.

References

Anthony C (2021) When Knowledge Work and Analytical Technologies Collide: The Practices and Consequences of Black Boxing Algorithmic Technologies. *Administrative Science Quarterly* 66(4). SAGE Publications Inc: 1173–1212.

Avnoon N (2021) Data Scientists' Identity Work: Omnivorous Symbolic Boundaries in Skills Acquisition. *Work, Employment and Society* 35(2): 332–349.

Ben Chouikha M and Dakhli SBD (2015) Knowledge Sharing in Computerization Projects: An Approach Based on Boundary Objects. *MCIS 2015 Proceedings*. Epub ahead of print 2015.

Bowker GC and Star SL (1999) *Sorting Things out: Classification and Its Consequences*. Inside technology. Cambridge, Mass: MIT Press.

Chen H, Chiang R and Storey V (2012) Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly* 36(4): 1165–1188.

Davenport TH and Patil DJ (2012) Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*. Epub ahead of print 1 October 2012.

Davenport TH and Patil DJ (2022) Is Data Scientist Still the Sexiest Job of the 21st Century? *Harvard Business Review*, 15 July. Available at: https://hbr.org/2022/07/is-data-scientist-still-the-sexiest-job-of-the-21st-century (accessed 17 May 2024).

Doolin B and McLeod L (2012) Sociomateriality and boundary objects in information systems development. *European Journal of Information Systems* 21(5): 570–586.

Faraj S, Pachidi S and Sayegh K (2018) Working and organizing in the age of the learning algorithm. *Information and Organization* 28(1): 62–70.

Fountaine T, McCarthy B and Saleh T (2019) Building the AI-Powered Organization. *Harvard Business Review*, 1 July. Available at: https://hbr.org/2019/07/building-the-ai-powered-organization (accessed 19 February 2020).

Gioia DA, Corley KG and Hamilton AL (2013) Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. *Organizational Research Methods* 16(1): 15–31.

Glaser B and Strauss A (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine Transaction.

Glushko RJ (2023) Seven ways to make a data science project fail. *Data and Information Management* 7(1): 100029.

Grønsund T and Aanestad M (2020) Augmenting the algorithm: Emerging human-in-the-loop work configurations. *The Journal of Strategic Information Systems* 29(2): 101614.

Hopf K, Müller O, Shollo A, et al. (2023) Organizational implementation of AI: Craft and mechanical work. *California Management Review* 66(1).

Hopf K, Joshi M, Stelmaszak M, et al. (2024) Crafting Ever-Changing Data Products: Towards a Human-Centered Process Model of Data Work. In: *ECIS 2024 Proceedings*, Paphos. Cyprus, 14 June 2024. AIS electronic library. Available at: https://aisel.aisnet.org/ecis2024/track01_peoplefirst/track01_peoplefirst/3.

Huvila I, Anderson TD, Jansen EH, et al. (2017) Boundary objects in information science. *Journal of the Association for Information Science and Technology* 68(8): 1807–1822.

Joshi MP, Su N, Austin RD, et al. (2021) Why So Many Data Science Projects Fail to Deliver. *MIT Sloan Management Review*, 2 March. Available at: https://sloanreview.mit.edu/article/why-so-many-data-science-projects-fail-to-deliver/ (accessed 28 July 2021).

Kane GC, Young AG, Majchrzak A, et al. (2021) Avoiding an oppressive future of machine learning: a design theory for emancipatory assistants. *MIS Quarterly* 45(1): 371–396.

Krakowski S, Luger J and Raisch S (2022) Artificial Intelligence and the Changing Sources of Competitive Advantage. *Strategic Management Journal*. Wiley. Epub ahead of print 6 February 2022. DOI: 10.1002/smj.3387.

Levina and Vaast (2005) The Emergence of Boundary Spanning Competence in Practice: Implications for Implementation and Use of Information Systems. *MIS Quarterly* 29(2): 335.
Marabelli M, Newell S and Vaast E (2017) Boundary objects survival over time: insights from the field of healthcare coordination. In: *Academy of Management Proceedings*, 2017.
Müller O, Fay M and Brocke J vom (2018) The Effect of Big Data and Analytics on Firm Performance: An Econometric Analysis Considering Industry Characteristics. *Journal of Management Information Systems* 35(2): 488–509.
Parmiggiani E, Østerlie T and Almklov PG (2022) In the Backrooms of Data Science. *Journal of the Association for Information Systems* 23(1): 139–164.
Patton MQ (2002) *Qualitative Research and Evaluation Methods*. 3 ed. Thousand Oaks, Calif: Sage Publications.
Pawlowski SD and Robey D (2004) Bridging User Organizations: Knowledge Brokering and the Work of Information Technology Professionals. *MIS Quarterly* 28(4): 645–672.
Sharma R, Mithas S and Kankanhalli A (2014) Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems* 23(4): 433–441.
Shollo A, Hopf K, Thiess T, et al. (2022) Shifting ML value creation mechanisms: A process model of ML value creation. *The Journal of Strategic Information Systems* 31(3): 101734.
Someh I, Wixom B, Davern M, et al. (2023) Configuring Relationships between Analytics and Business Domain Groups for Knowledge Integration. *Journal of the Association for Information Systems* 24(2): 592–618.
Star SL (2010) This is Not a Boundary Object: Reflections on the Origin of a Concept. *Science, Technology, & Human Values* 35(5): 601–617.
Star SL and Griesemer JR (1989) Institutional Ecology, `Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science* 19(3): 387–420.
Stice-Lusvardi R, Hinds PJ and Valentine M (2023) Legitimating Illegitimate Practices: How Data Analysts Compromised Their Standards to Promote Quantification. *Organization Science*. INFORMS. Epub ahead of print 30 January 2023. DOI: 10.1287/orsc.2022.1655.
Thiess T and Müller O (2018) Towards Design Principles for Data-Driven Decision Making – an Action Design Research Project in the Maritime Industry. In: *ECIS 2018 Proceedings*, Portsmouth, UK, 2018. AIS electronic library.
Vaast E and Pinsonneault A (2021) When Digital Technologies Enable and Threaten Occupational Identity: The Delicate Balancing Act of Data Scientists. *MIS Quarterly* 45(3): 1087–1112.
van den Broek E, Sergeeva A and Huysman M (2021) When the machine meets the expert: An ethnography of developing AI for hiring. *MIS Quarterly* 45(3): 1557–1580.
Waardenburg L, Huysman M and Sergeeva AV (2022) In the Land of the Blind, the One-Eyed Man Is King: Knowledge Brokerage in the Age of Learning Algorithms. *Organization Science* 33(1). INFORMS: 59–82.
Wu L, Hitt L and Lou B (2019) Data Analytics, Innovation, and Firm Productivity. *Management Science* 66(5): 2017–2039.
Zaitsev A, Tan B and Gal U (2016) Collaboration Amidst volatility: the Evolving Nature of boundary Objects in Agile Software Development. In: *ECIS 2016 Proceedings*, 2016.