6-18-2022

Unleashing the Potential of External Data: A DSR-based Approach to Data Sourcing

Pavel Krasikov
University of Lausanne, pavel.krasikov@unil.ch

Markus Eurich
University of Lausanne, markus.eurich@unil.ch

Christine Legner
HEC Lausanne, christine.legner@unil.ch

UNLEASHING THE POTENTIAL OF EXTERNAL DATA: A DSR-BASED APPROACH TO DATA SOURCING

Research Paper

Pavel Krasikov, Faculty of Business and Economics (HEC), University of Lausanne, Switzerland, pavel.krasikov@unil.ch

Markus Eurich, Faculty of Business and Economics (HEC), University of Lausanne, Switzerland, markus.eurich@unil.ch

Christine Legner, Faculty of Business and Economics (HEC), University of Lausanne, Switzerland, christine.legner@unil.ch

Abstract

External data has become an indispensable pillar in state-of-the-art decision-making and value creation in an enterprise context. Despite the increasing motivation to use external data, information systems (IS) research still lacks an adequate data sourcing perspective. This study aims to address this gap by investigating the practical challenges in this emerging field and developing a reference process for sourcing and managing external data. To this end, we adopt a design science research approach leveraging collaboration with practitioners from nine high-profile companies. Our findings contribute to the scarce body of knowledge on data sourcing in IS by proposing explicit prescriptions in the form of a reference process for sourcing and managing external data.

Keywords: External Data, Data Sourcing, Reference Process, Design Science.

1 Introduction

"With our own data we can only look internally. We need to see industry benchmarks, regional trends, what waves we can ride on; we derive competitive advantage by getting data from outside and enhancing our own data" (Chief Data Officer of Flagstar Bank, as cited in Belissent, 2019). As reflected in this quotation, external data has become essential to practitioners' decision-making and value-creation processes. A common way to define external data involves the notion of data coming from outside the company (Arndt and Gersten, 2001; Hopf, 2019; Strand and Syberfeldt, 2020). An increasing number of studies show that combining internal data with external data makes it possible to "compete on analytics" (Strand and Carlsson, 2008), enrich business processes, decrease internal data curation efforts, and create new services (Baud *et al.*, 2002; Baecke and Van den Poel, 2011; Schatsky *et al.*, 2019; Strand and Syberfeldt, 2020).

Despite the increasing motivation to use external data, most companies are sourcing data ad-hoc and have not yet established professional sourcing practices. Based on a survey of 100 medium-to-large American companies, the external data provider Explorium (2021) reports that 79% of organizations consider external data to be very valuable. At the same time, 77% of them lack an understanding of external data sourcing processes. Davenport *et al.* (2021) observe that external data is largely unmanaged within enterprises; they propose that sourcing high-quality data "builds on the process and supplier management techniques used by manufacturers of physical products." According to Jarvenpaa and Markus (2020), organizations source data for various purposes, but information systems (IS) research – and data management, in particular – lacks a focus on the data sourcing perspective. To address this gap in research, we ask the following two research questions:

What is the current status and challenges in sourcing external data in enterprises?

How should enterprises source and manage external data?

To account for the practical relevance of the topic (March and Smith, 1995), we adopt a design science research (DSR) approach to construct an artifact that "says how to do something," in line with Gregor (2006)'s type-V theory. Our study is embedded in a multiyear research program in the field of data management, which gives us privileged access to data experts from more than 20 multinational companies. Following the methodological steps suggested by Peffers *et al.* (2007), we develop a reference process for sourcing and managing external data. Our findings contribute to the scarce literature on data sourcing by proposing explicit prescriptions in the form of a reference process as a generic procedure for evidence-based IS research (Goeken, 2011). The designed reference process aims to solve the increasingly relevant organizational problems and generalize the process sequence in data sourcing and its elements, such as activities and milestones (Wilmsen *et al.*, 2020).

The remainder of this paper is structured as follows: Section 2 introduces the concept of sourcing in IS research and elaborates on data sourcing. Section 3 outlines our research design and process. Section 4 presents our findings and elaborates on the phases of the reference processes for sourcing and managing external data. In section 5, we summarize and discuss our findings.

2 Related Work

The sourcing of external data rarely appears in academic literature, and Jarvenpaa and Markus (2020) have called for research in this domain. In this section, we compare the different types of sourcing (data, IS, IT, and strategic sourcing) with the respective definitions, objects of sourcing, and underlying processes (see Table 1). We conclude that data sourcing approaches resemble existing strategic sourcing processes, but no links have been drawn between the two.

2.1 Sourcing in IS

"Sourcing is the act through which work is contracted or delegated to an external or internal entity that could be physically located anywhere" (Oshri *et al.*, 2015, p. 2). In an enterprise setting, accelerated technological change coupled with the growing importance of supply chain management has helped **strategic sourcing** to evolve from buying to a critical area of strategic management (Rafati and Poels, 2015). Strategic sourcing covers spend analysis, supplier selection and qualification, contract and relationship management, and analytics for the associated decision-making processes.

Sourcing decisions and their success have emerged as fundamental issues in IS sourcing (Watjatrakul, 2005). Kotlarski *et al.* (2018) define **IS sourcing** as "a broad umbrella term that refers to the contracting or delegating of IS- or IT-related work (e.g., an ongoing service or one-off project) to an internal or external entity (a supplier)." Concerning the mentioned information technology (IT) term, "**IT sourcing** research is a multi-disciplinary research endeavor that examines the organizational impacts of contracting-out IT functions to a third-party provider, from a technology and business perspective" (Sesay and Ramirez, 2016).

Building on the seminal review of the IT outsourcing literature (Lacity *et al.*, 2010), the same authors' subsequent study of business service sourcing (Lacity *et al.*, 2016) finds consistent evidence that transaction costs (and reduction of costs in general) were determinative for sourcing decisions. Sourcing is a transaction, but when it comes to terminology, "the words buying, purchasing, procuring, and sourcing are used as synonyms, referring to a transaction where a particular good is transferred between two organizations" (Schneider *et al.*, 2013). In their review of recent developments in outsourcing in the IT business service, Könning *et al.* (2019) use "sourcing" and "outsourcing" interchangeably. The authors point to the paradigm shift from cost reduction as a traditional motivator of the sourcing decision, to other stimuli, such as expertise, skill, quality improvement, and focus on core capabilities (Könning *et al.*, 2019). While IS outsourcing decisions have been widely discussed in the literature (Clark *et al.*, 1995; Lacity *et al.*, 2010; Oshri *et al.*, 2015; Könning *et al.*, 2019; Nevo and Kotlarsky, 2020), in the IS infrastructures context (Lyytinen *et al.*, 2017), e.g., hardware and software, the decisions on data sourcing are often seen as a routine taking place in the background (Jarvenpaa and Markus, 2020).

Jarvenpaa and Markus (2020) argue that, despite the rich body of knowledge on IT and business sourcing services, *data* in IS sourcing research remains unaddressed. Building on the definition by Kotlarski *et al.* (2018), the authors proclaim **data sourcing** as "procuring, licensing, and accessing data from an internal or external entity" (Jarvenpaa and Markus, 2020).

	IS sourcing	IT sourcing	Strategic sourcing	Data sourcing
Definition	"…a broad umbrella term that refers to the contracting or delegating of IS- or IT-related work (e.g., an ongoing service or one-off project) to an internal or external entity (a supplier)" (Kotlarsky *et al.*, 2018)	"…the delegation, through a contractual arrangement, of all or any part of the technical resources, human resources, and the management responsibilities associated with providing IT services to an external vendor" (Clark *et al.*, 1995)	"…a critical area of strategic management that is centered on decision-making regarding an organization's procurement activities such as spend analysis, capability sourcing, supplier selection and evaluation, contract management and relationship management" (Rafati and Poels, 2015)	"…procuring, licensing, and accessing data (e.g., an ongoing service or one-off project) from an internal or external entity (supplier)" (Jarvenpaa and Markus, 2020)
Sourcing processes	Make the sourcing decision, design contractual, structures, and manage the sourcing relationship (Kotlarsky *et al.*, 2018)	IT business services outsourcing processes (Könning *et al.*, 2019)	Identify needs, gather information about relevant factors, evaluate and select suppliers, evaluate best sourcing alternative, contracting (Nordin and Henrik, 2008; Ribas *et al.*, 2021)	Fragmented approaches: Data brokers: Acquire, integrate, assess, sell (Strand and Carlsson, 2008) Clarify data needs, data acquisition, and data application (Wang *et al.*, 2020) Find, assess, decide how to use, understand, obtain/purchase (Sun *et al.*, 2021) Open data: screen, assess, and prepare open data for use (Krasikov *et al.*, 2021)

Table 1. Prior research on data, IT, IS, and strategic sourcing

2.2 Data sourcing

Recent literature (Sun *et al.*, 2021) mentions three major options for data sourcing (see Table 2): conventional sourcing, crowdsourcing, and cloud-based approach.

Types	Definition	Sources
Conventional	Collection of data from a variety of sources, typically involving finding, obtaining/purchasing, assessing, integrating, and using the data.	Strand and Carlsson, 2008; Wang *et al*., 2020; Krasikov *et al*. 2021; Sun *et al*., 2021
Crowd-based	General public (i.e., the crowd) collectively contributes to the generation, aggregation, and processing of the data for its further use.	Deutch and Milo, 2012; Amsterdamer and Milo, 2015; Satish and Yusof, 2017; Lukyanenko *et al*., 2019; Sun *et al*., 2021
Cloud-based	Data sourcing from cloud platforms that provide access via dedicated interfaces.	Sun *et al*., 2021

Table 2. Types of data sourcing

Conventional data sourcing focuses on obtaining data from a variety of sources, which typically involves finding, obtaining/purchasing, assessing, integrating, and using the data. Despite its relevance, it has only played a minor role in the scarce literature on the topic. For instance, Strand and Carlsson (2008) study how data brokers (named syndicate data suppliers) source external data. On a high level, their activities involve acquiring data from various sources, integrating data into internal databases, refining, enriching, and then selling and delivering data to respective clients. This data acquisition perspective adopts a specific example of how external data can be sourced, addressing only the external data made available by specialized data providers. In a recent study, Wang *et al*. (2020) develop a reference model for knowledge-driven data provision processes in a data engineering environment. The model proposes three key phases: clarification of data needs based on the company's business activities, data acquisition based on the defined requirements and criteria, and data application that would satisfy the knowledge needs. Since the proposed model focuses on supporting data provision in data mining projects, it does not particularly address the acquisition of external datasets from specialized commercial providers. In the specific setting of agro-geoinformatics, Sun *et al*. (2021) claim that "conventional sourcing depends on human surveyors, is often labor-intensive, and has very tedious administrative processes." The authors highlight the importance of standardization in data sourcing, such as using standard formats and access methods to reach the openly available online resources. Since open data is often positioned among the sourcing candidates for companies (Zuiderwijk *et al.*, 2015; Hopf, 2019; Roeder *et al.*, 2020; Strand and Syberfeldt, 2020), a meta-analysis by Krasikov *et al.* (2021) reviews the existing studies on open data processes from the publisher and consumer perspective. They define the following core sourcing phases: screen, assess, and prepare open data for use. Overall, we note that although the data sourcing steps provide a general structure, they focus on specific scenarios of external data use (acquisition of paid sources, knowledge discovery, agro-geoinformatics, or open data) and do not fully address the enterprise perspective on data-sourcing activities.

The concept of crowdsourcing as an "emerging data procurement paradigm that engages Web users to collectively contribute and process information" (Amsterdamer and Milo, 2015) has received more interest in prior IS research. It implies outsourcing tasks to the network of people, such as freelancers, via digital labor marketplaces (Nevo and Kotlarsky, 2020). Thus, it is the responsibility of "the crowd to generate or source data" (Deutch and Milo, 2012). When it comes to companies collecting customer experience data (e.g., Amazon and Netflix), crowdsourcing makes it possible to improve the initial product offering (Satish and Yusof, 2017). Lukyanenko *et al*. (2019) underscore the value of crowdsourced user-generated content for which companies develop custom information systems.

Similarly, the concept of cloud-based data sourcing stems from the notion of cloud sourcing in IS, which refers to a form of outsourcing of IT resources to the cloud service providers (Muhic and Johansson, 2014; Lacity *et al.*, 2016; Schneider and Sunyaev, 2016). Cloud-based data sourcing implies that the data is hosted on the clouds, and access is provided to the user via dedicated interfaces and tools, e.g., Amazon Web Services and Microsoft Azure (Sun *et al.*, 2021).

Concerning external data sourcing, practitioners' insights should not be neglected. Forrester's approach to external data sourcing (Belissent, 2019) allocates roles along certain process steps; for instance, data hunters identify and evaluate potential data sources and data architects verify the "fit" of the external data, while procurement draws up the data sourcing contract. Aaser and McElhaney (2021) present McKinsey's view on how companies should harness the power of external data, where companies are initially advised to establish a dedicated data sourcing team that would take care of finding, accessing, procuring, integrating, reviewing, managing, and using external data.

2.3 Research gap

Compared with the rich body of knowledge on IT (out)sourcing, data sourcing in IS research has not been adequately explored. Some recent developments in the field take a step forward to properly define the concept (Jarvenpaa and Markus, 2020), but a more thorough understanding of the specific challenges associated with data sourcing and the ways to address them is still lacking. Despite the increasing motivation to use external data, data sourcing processes in enterprises and related practices have only occasionally been addressed in the academic literature. In the enterprise setting, only one academic study elaborates on a method addressing the sourcing processes concentrating on open data (Krasikov *et al.*, 2021). Today, only consulting reports elaborate beyond the nominal sequence of external data sourcing and managing by proposing dedicated roles. Nonetheless, these suggestions remain at the conceptual level and have not been integrated into the academic body of research.

3 Research Design and Process

To address this gap in research, we adhere to design science research by following the methodology formulated by Peffers *et al.* (2007) and seek to design actionable guidance for sourcing and managing external data by using a rigorous research process. McCarthy *et al.* (2020) argue that the successful identification of relevant real-world problems in DSR relies on the engagement of stakeholders (i.e., practitioners) in all phases, starting with the initial problem identification phase. Our multi-year research project debuted in February 2020, when we formed an expert group with practitioners from nine high-profile companies to investigate the challenges related to external data sourcing and management. These practitioners come from different sectors, including pharmaceutical, manufacturing, transportation, consumer goods, and insurance. They were experienced professionals who were already using external data and were involved in the related activities or initiatives in their companies. Figure 1 presents the summary of our research process.

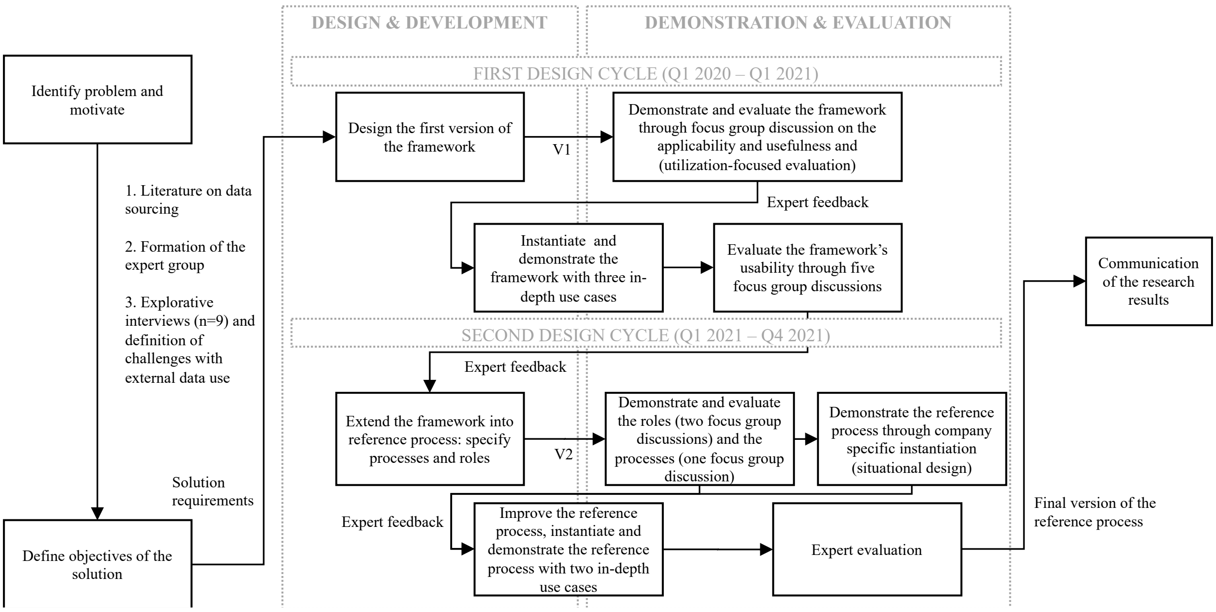

Figure 1. Research process

Following the six steps of the DSR processes model (Peffers *et al.*, 2007) and triggered by an industry need, we chose an objective-centered solution as a research entry point. We started by identifying the current status of the challenges that practitioners face when using external data. To this end, we conducted nine exploratory semi-structured interviews of 30 minutes with each of the companies mentioned above to understand their initial situation.

In the first design cycle, which lasted from February 2020 to January 2021, we focused on designing an artifact that would help companies address the challenges. This iteration resulted in a framework that depicted the five main phases in external data sourcing and managing. It was presented during a half-day focus group session, where the participants used it to document their own sourcing activities for different types of external data (open, paid, shared, and social media data). The framework was further instantiated for three different use cases, which are highly relevant to the focus group's participants. The first use case relied on using commercial data about companies' affiliations (Dun & Bradstreet, 2018) to map complex organizational relationships within a corporate structure for the company's business partners. The second use case aimed to optimize the logistics process by maintaining dangerous goods classifications and validate it by using openly available sources. The third use case targeted the improvement of corporate sustainability reporting activities with the help of external (shared) data among the different stakeholders (e.g., logistics, procurement, packaging, etc.). These external data use cases were discussed during five separate focus group meetings, which allowed us to consolidate the findings, agree on the applicability and usefulness of the artifact, and initiate a new design phase incorporating the collected feedback.

In the second design cycle (February–November 2021), we included the participants' feedback, refined the main components of the framework, and focused on roles and more granular processes for external data sourcing and management. This cycle marked the transformation from a more generic framework toward a refined artifact. We further demonstrated the applicability of the reference process with two additional use cases of common interest for practitioners. The fourth use case focused on using the public holiday data in sales and marketing analytics to account for the discrepancies between different markets and their geographical granularity. The fifth use case was dedicated to the Web scraping of product reviews from online marketplaces to enhance customer analytics. These scenarios were discussed in three separate focus group meetings. To specify the roles affected by sourcing and managing external data, we analyzed practitioner publications (Belissent, 2019; Schatsky *et al.*, 2019; Aaser and McElhaney, 2021) and conducted four interviews with more experienced companies from the composed expert group. After defining the emerging roles and their corresponding tasks, we held two focus group sessions to understand the additional roles involved in sourcing and managing activities within all the participating companies. Another session was held to discuss and validate the underlying activities and techniques for each of the main phases of the reference process. Based on these outcomes, one of the authors took part in a company-specific instantiation within one of the expert group companies as a form of naturalistic evaluation (Venable *et al.*, 2012). Based on the framework and the defined activities and roles, the company adapted its own process model for sourcing and managing external data, thereby making it possible to refine the reference process. Specifically, the sourcing phase was split into screening and assessment to further specify the underlying activities of each component. This ex-ante evaluation allowed us to conclude the applicability, utility, and flexibility of our artifact in the enterprise context.

Subsequently, the reference process was further consolidated, and its separate components (phases, activities, roles, and milestones) were discussed and evaluated in four individual sessions with two practitioners from two companies and two external data experts. These two experts have distinctive domain knowledge, while the practitioners have shown interest in applying the findings to the respective companies. The sessions concluded with a questionnaire to evaluate the abovementioned components of the reference process along the typical criteria (Prat *et al.*, 2015) by using a five-point Likert scale. Firstly, we asked the participants to evaluate the relevance of the identified challenges, where the respondents overall agreed with identified challenges. Secondly, the participants were asked to evaluate the relevance of the meta-requirements toward the design of our reference process. The respondents agreed and strongly agreed with the formulated meta-requirements, which are presented below in section 4.1. Thirdly, the participants strongly agreed (3/4) that the reference process addresses the real problem and helps manage external data from the sourcing request to the end-of-life, agreed (2/4) and strongly agreed (2/4) that the process ensures that the sourced external data is trusted, compliant, transparent, and of high quality, and agreed (4/4) that it helps clarify roles and responsibilities for sourcing and managing external data. In addition, the reference process was found to be understandable and useful (strongly agree 3/4 and agree 1/4), and applicable to the context of one's company (strongly agree 2/4, agree 1/4, and 1/4 neither agree nor disagree). Fourthly, the participants evaluated the contents of the reference process: They strongly agree (4/4) that the phases of the process are complete, agree (3/4) that roles are appropriate and help clarify the responsibilities for external data, strongly agreed (3/4) that the proposed milestones are appropriate, and shared the opinion that the proposed process variations make the process simpler and more flexible.

4 Reference Process for Sourcing and Managing External Data

4.1 Challenges and requirements

Given the benefits of external data use and the increasing interest in developing use cases, companies are making an effort to source external data. The insights from our exploratory interviews showed that practitioners primarily source commercial data sources from the data providers, followed by freely available open data. Nonetheless, our interviewees confirmed that data sourcing processes are, at present, largely unmanaged and subject to multiple challenges (see Table 3). The lack of knowledge about sourcing and managing external data appeared to be the most prominent challenge, coupled with the absence of standards and good practices. Practitioners also emphasized the issue of purchasing external data multiple times within the company, which results in additional efforts and fees. Furthermore, the respondents view the unknown quality of external datasets (e.g., in terms of correctness, completeness, format, consistency) and the lack of trust in external data sources as challenges. In addition, missing knowledge about usage rights, licenses, compliance (e.g., GDPR), and the related legal aspects were troubling. Finally, the lack of transparency about the contents of external data sources raised concerns among the interviewed practitioners. Based on these insights, we formulate the meta-requirements (MR), which are abstract enough to address the class of artifacts but are tied directly to the solution objective:

- MR1 – The artifact should help manage external data from sourcing requests to the end-of-life.
- MR2 – The artifact should clarify roles and responsibilities for external data.
- MR3 – The artifact should ensure that the sourced external data is trusted, compliant, transparent, and high-quality.

Table 3 provides an overview of the challenges with the corresponding meta-requirements. The analysis of the current state and the challenges allowed the expert group to reach a consensus that the outcome of this research should produce a reference process for sourcing and managing external data.

External data sourcing and managing challenges	Meta-requirement
Lack of knowledge about external data sourcing and managing	MR1, MR2
No standards for sourcing and managing external data	MR1
No overview of which external datasets have been sourced and are used by whom	MR1, MR2
Unknown quality of external datasets (e.g., correctness, completeness, format, consistency)	MR3
Lack of trust in external data	MR3
Missing knowledge about usage rights, licenses, compliance (e.g., GDPR), and legal aspects	MR3
Lack of transparency about the contents of external data sources	MR3

Table 3. Meta-requirements for external data sourcing and managing reference process

4.2 Reference process

Reference models are typically developed in design and evaluation cycles and are considered important artifacts in IS research (Winter and Schelp, 2006). They were found to be specifically relevant for knowledge accumulation in data management (Legner *et al.*, 2020). In an enterprise setting, a reference process aims to generalize the usual process sequence and its elements, such as activities and milestones (Becker *et al.*, 2007; Wilmsen *et al.*, 2020). To address our specific research objectives, we designed a reference process that provides actionable guidelines supporting companies with external data sourcing and managing.

We iteratively developed and evaluated our reference process for sourcing and managing external data through two major design cycles. Based on the literature regarding data sourcing (see section 2.2) and the empirical evidence from the design cycles, we identified six core phases of the process: start, screen, assess, integrate, manage and use, and retire. Each process step contains a clear input, a set of underlying activities, and related roles and techniques. It ends with a defined milestone, allowing a progress review along the reference process. The sequence of the phases is nominal, allowing for the simultaneous execution of activities if the necessary conditions are met. Table 4 provides an overview of the reference process.

	Input	Activities	Techniques	Roles	Output/Milestone
Start	External data request	- Define relevant datasets or data needs - Specify the business context - Specify the target system *Variants:* - If use case already exists – skip business requirements - If datasets are known – skip to assessment phase	- Specification of the context requiring new data - Definition of relevant business concepts	- Requestor (data/business) - Data/business analyst - IT specialist	M1: Documented external data use case with a documentation template
Screen	External data use case	- Search for suitable datasets - Identify relevant sources - Locate dataset candidates - Align with data requirements - Check if the data was not already sourced within the organization	- Rely on the metadata provided by the publisher to identify trustable sources - Leverage from dedicated data portals, search engines, and expert knowledge	- Data hunter/data steward	M2: Candidate datasets identified and documented in a list with names of datasets, publishers, and data sources
Assess	Identified datasets	- Define assessment criteria - Verify the budget for the desired datasets - Execute assessment along the defined criteria *Variants:* - If no or not enough datasets fulfilled the criteria – revisit criteria - If no datasets were selected – revisit screening phase	- Three-level assessment: metadata, schema, and content	- Data steward - Procurement - Compliance officer	M3: Selected datasets, purchase order/contact with data provider
Integrate	Selected datasets	- Access external datasets - Identify relevant business concepts - Define the owner of newly identified data	- Rely on knowledge graph mapping approaches	- IT specialist - Data steward	M4: Integrated external datasets
Manage and use	Integrated external datasets	- Establish governance for external data - Use external data - Manage updates - Monitor the use *Variants:* - Continuous use of external data, unless no termination of use case is planned	- Ensure clear guidelines for the use of external data	- Requestor (data / business) - External data expert - Compliance officer	M5: Decision to terminate the use case
Retire	Termination decision	- Decide how external data is treated at the end-of-life *Variants:* - If new data is needed for the use case – back to screening phase - If use case is retired but not the data – back to start phase	- Adhere to a secure data archiving approach within the company	- Data steward - Procurement	If the dataset is no longer used, it should be retired

Table 4. Reference process for sourcing and managing external data

4.2.1 Start

The first phase is triggered by a request for external data and aims to define and document the motivation for sourcing the new data. As seen in section 2.2, clarification of data needs (Wang *et al.*, 2020) makes it possible to initiate the sourcing process and lays the groundwork to define the concrete use case–related requirements and identification of the sourcing goal. The process originates with the receipt of the external data demand for a potential use case. Since this is an initiation phase, there are several possibilities for how the activities can develop: Variant 1a – the use case does not exist, and the external datasets are unknown; Variant 1b – the use case already exists, but the external datasets are unknown; Variant 1c – the use case already exists, and the datasets are known.

In the first variant, since the use case does not exist, the business requirements for the use of external data should be defined. This includes the trigger and purpose of the use case, as well as responsible functions and teams within the organization. Since the potential external datasets are still unknown, the data requirements must be defined. This comprises potentially relevant sources, their geographical and temporal coverage, as well as their respective levels of granularity. In addition, it is necessary to specify relevant data objects (internal and external business concepts/attributes) that serve as primary data requirements. Furthermore, the system requirements need to be specified, namely the target system for onboarding external data and its required format. These activities enable our reference process and directly address the MR1. Variant 1b becomes relevant when companies have already formulated precise use cases involving external data and can skip the business requirements formulation, starting directly with data requirements. Variant 1c implies that the initiation phase is no longer needed, since the company is assumed to have a clear picture of their business, data, and system requirements already, which allows them to skip directly to the assessment phase (see section 4.2.3). The milestone (M1) reached in this phase represents a documented use case for external data, encompassing the requirements mentioned above.

4.2.2 Screen

The screening phase primarily targets the identification of relevant sources and underlying datasets. Finding relevant sources appears to be a challenge beyond our expert group – for instance, 74% of respondents in a survey conducted by Explorium (2021) confirm they are "not sure what to look for." Searching proves to be even more challenging when it comes to the resources freely available online (Krasikov *et al.*, 2021), as opposed to a more context-specific offering provided by data brokers. Therefore, the input of this phase builds on the previously defined external data use cases, particularly the primary data requirements. As one of the common data sourcing steps (Wang *et al.*, 2020; Sun *et al.*, 2021), screening intends to locate relevant external data sources. These possible sources could include open data portals, traditional search engines or dedicated dataset search engines, data providers and brokers, shared platforms, social media data, and expert knowledge (Strand and Carlsson, 2008; Roeder *et al.*, 2020; Strand and Syberfeldt, 2020; Krasikov *et al.*, 2021). Once candidate sources have been identified, relevant datasets candidates are located, in line with the primary data requirements of the previous phase. To get a better understanding of the data, dataset samples can be viewed directly at the source level, if available, or requested from the data provider. Moreover, referring to the internal data catalog is crucial to verify whether the data has not already been sourced within the organization. The milestone (M2) in this phase represents a list with names of the candidate datasets, along with publisher details, to keep record of the source. Furthermore, contracts and service-level agreements (SLA) negotiations must be conducted in line with procurement guidelines.

4.2.3 Assess

This phase aims to assess the different elements of the data based on the predefined selection criteria. The distinction in the assessment criteria relies on whether external data is paid or non-paid. Namely, our insights from the focus group and the evidence from Belissent's (2019) guide show that paid data requires the procurement team's involvement to select the datasets that align with the allocated budget and fulfill the use case requirements. Based on the feedback from the focus group discussion, we adopt a multifaceted assessment approach (Krasikov *et al.*, 2021), since it addresses the core data-related challenges (see section 4.1) and proposes three levels of assessment. The first level (metadata) conveys a multitude of information, such as access conditions, licensing information (permissions and prohibitions), publishing details (publisher, publishing date, update cycle), and general content-related information (language, geographical and temporal coverage, number of records and attributes). The second (schema level) is important to verify whether the needed attributes (see section 4.2.1) are present in the preselected external datasets. The third level (content assessment) investigates the dataset contents in terms of typical data quality dimensions and their metrics, such as completeness, uniqueness, and validity. We identify three possible variants of how the assessment criteria can be met: Variant 3a – assessment criteria are met, and the necessary number of datasets is selected; Variant 3b – no (or not enough) datasets passed assessment criteria, and the assessment criteria should be revisited; Variant 3c – no datasets passed the assessment criteria, and the screening phase should be revisited.

This phase concludes with the achievement of M3 with a list of selected datasets that have passed the defined assessment criteria. The process flexibility brought about by these variations is important since the evidence shows the quality of external data may vary. Thus, the selection criteria are subject to revisions on the company's side (Variant 3b) to make the use case feasible. If no datasets pass the assessment criteria, the screening phase must be revisited to identify further candidate datasets.

4.2.4 Integrate

Selected datasets serve as input for the integration phase to onboard the external data into the company's systems and to document and prepare them for further use. This phase's activities begin by accessing the external data via a proposed interface. Along with an overview of the target integration system, the use case requirements help when choosing the appropriate access interface (e.g., download or API). Integration efforts are a challenging endeavor because of the external datasets' heterogeneity. The onboarding activities include thorough documentation of selected external datasets, which provides complete metadata information about the sources and their contents – specifically, the attributes. Our approach relies on the use of knowledge graphs to ensure the business concepts in external data correspond to the internal ones and can proceed by mapping both in a common data model. Knowledge graphs are considered an appropriate way of integrating heterogeneous datasets (Bizer *et al.*, 2009; Paulheim, 2016) and have proved to be effective in the context of open data (Krasikov *et al.*, 2021). In case there are multiple external datasets with semantically corresponding attributes, these sources should be combined by using a common data model. Newly onboarded external data requires ownership attributed to a dedicated data owner. The phase ends with M4 once external datasets have been integrated.

4.2.5 Manage and use

When the external datasets have been integrated, it is crucial to ensure their successful use. Our findings show that, following the integration of external data into the internal systems, its management should not diverge from the process established for internal data. This implies that external data is considered to be internal once it has been fully onboarded (see section 4.2.4). Nonetheless, provided the similarities between internal and external data, upon the integration of the latter, attention should be paid to the aspects that deviate. External data may need to receive updates from the original source and, thus, the update cycles should be managed by establishing datasets' versions and onboarding new data accordingly (see section 4.2.4). During its use, user feedback is collected to maintain the quality of the external data and identify improvement opportunities. Two scenarios were identified regarding the use of external data: Variant 5a – end date for the use of external data is anticipated; Variant 5b – continuous use of external data with no retirement planned. In case a decision to terminate the use of external data is taken and M5 is reached, the next phase will be embraced. The use of external data will continue until a retirement decision is taken.

4.2.6 Retire

The final phase aims to formalize the end-of-life of external data. The following variants summarize our findings of the possible scenarios in this phase: Variant 6a – the use case and the dataset are retired; Variant 6b – the use case remains, but new datasets are needed; Variant 6c – the dataset remains, but a new use case is needed.

Variant 6a assumes the complete termination of the process, where the use of external data is completely discontinued. Archiving the downloaded external data is a possibility when there are no contracts associated with its use. Conversely, it is important to monitor the termination terms of contracts/subscriptions in order not to bear the costs for the non-used data. Variant 6b implies that the process goes back to the screening phase (see section 4.2.2) and the search for new datasets begins. By contrast, if the use case is retired (6c) but the datasets remain in the internal systems, the process restarts with the definition of a new use case in the initial phase, respecting the possible Variant 1b.

4.3 Roles

Our findings, as well as the insights from the practitioners' reports, show that assigning the dedicated organizational roles along the process is essential for sourcing and managing external data (Belissent, 2019; Aaser and McElhaney, 2021; Explorium, 2021). Based on our learnings from the evaluation rounds and the insights from the company-specific instantiation of the process, we have identified that most of the tasks can be assumed by already existing roles within the company. We learned that this encompasses not only the data governance roles but also the associated functions such as procurement and compliance. For instance, the contracting process and negotiations with external data providers would rely on the procuring officer's efforts, while assessment would be the main task of a data steward, who ensures the data meets the desired standards of quality and is accessible for use within the company (van Donge *et al.*, 2020). However, given the new activities related to the searching and screening of the relevant external data sources, we have identified an emerging role of a **data hunter**, who also acts as a domain expert. A data hunter's main responsibility is to find and review external data to ensure its fit within the defined use case (Belissent, 2019; Aaser and McElhaney, 2021). Therefore, they accumulate the expertise to manage external datasets upon its integration and can also serve as a company-internal single point of contact. Table 5 summarizes the roles for external data sourcing and management based on the (academic and practitioner) literature, expert input, and insights from the company-specific instantiation of the process.

Role	Responsibilities for core activities in the process
Requestor (data/business)	- Submits the request for external data - Uses the external data for selected use cases
Data analyst & business analyst	- Defines business and data requirements for future use cases - Identifies new use cases for external data, develops proofs-of-concept, and conducts analytics
Data architect/engineer	- Defines system requirements for external data integration
Data hunter/external data expert	- Partners with business to find, review, and manage external data (Aaser and McElhaney, 2021) - Monitors external data sources for relevant data to enable use cases - Acts as company-internal single point of contact for sourcing and managing external data - Provides first-level support internally, acts as an internal consultant for external data-related topics
Data steward	- Assesses the quality and fit of external datasets
Procurement	- Negotiates contracts and SLAs, analyzes the pricing conditions and contract termination
Compliance officer	- Ensures that external data is collected with appropriate permissions (contract and license agreements) and used in accordance with the applicable data laws and internal policies

Table 5. Roles involved in external data sourcing and managing

Our findings help to distinguish two configurations for the roles in the reference process. In the first and most common role model, existing roles undertake new activities proposed by the process. By contrast, the second model assumes that new tasks are taken over by the emerging role of data hunter or external data expert. Therefore, we conclude that new tasks relating to external data sourcing and managing can be either delegated to a new role or taken on by existing roles.

5 Discussion

While there are high expectations on external data's potential to fill gaps for reasonable decision-making and value creation, up until now there is no systematic approach to external data sourcing.

Previous studies on conventional, crowd-based, and cloud-based data sourcing provide insights on generic steps of data sourcing, e.g., data demand, data acquisition, and data application (Wang *et al.*, 2020). Studies dedicated to external data address relevant data sourcing activities, like identifying and evaluating potential data sources (Strand and Carlsson, 2008; Belissent, 2019). However, they do not provide reference processes but rather address external data activities from a data role perspective, e.g., the range of tasks for a data hunter (Belissent, 2019). The major advantages of the reference process that we have developed for sourcing and managing external data are (1) its ability to guide enterprises in their sourcing activities in a systematic way with clear milestones, and (2) its foundation on essential design and evaluations principles:

(1) *Guidance*: The reference model addresses the challenges associated with external data, such as the uncertainty of external datasets quality. Unexplored external datasets require a more thorough assessment than with traditional quality metrics (Zhang *et al.*, 2019). Since external data creators and publishers are detached from their users (i.e., enterprises), the latter have limited knowledge about the data's characteristics and underlying quality. For the case of repurposed data, it is essential to adopt an approach that provides multiple perspectives on the sourced data (Krasikov *et al.*, 2021).

(2) *Design principles*: An important design principle is to ensure that the obtained contents of the sourced external data are clear, and similar (or even the same) data has not already been integrated within the company. This principle builds on the transparency aspect of the external data (e.g., in terms of its provenance and trustworthiness) as one of the fuels for the digital economy. Having clear documentation of candidates and existing external data sources helps adhere to this principle. Since data cataloging activities are often seen as volunteer efforts (Jarvenpaa and Markus, 2020), this principle helps enforce their importance, particularly when it comes to the sourcing of new, unknown data within the internal systems. Another design principle is that at the handover, "reliable and relevant information is clearly communicated" (Maher *et al.*, 2013). We addressed this design principle with the proposed milestones which present well-defined decision points that allow the involved parties to effectively communicate based on standardized performance criteria. Furthermore, to overcome the shortcomings of linear phase models, the reference process offers flexibility and process variability by formulating 11 possible variations of how the activities in the phases can be executed. They are also positioned as entry points, considering the current situation of the sourcing activities within the company. "Flexibility is an important characteristic of a reference process because it describes the ease with which a reference process accommodates and adapts to changes of the process requirements" (Wilmsen *et al.*, 2020).

6 Conclusion

Although the use of external data is not new and has been mentioned in the enterprise context since the late 1990s (Čas and Meier, 1999), its sourcing is often associated with simply "getting the data" without any further specifications on how exactly this can be accomplished. In scientific literature, however, a systematic approach to sourcing and managing external data was lacking until now.

Accordingly, we propose a reference process for sourcing and managing external data that guides enterprises through unknown strides and that is methodologically well founded on design science research. In our industry–research collaboration, we leverage design science research to build a reference process that supports the external data sourcing and managing activities in the enterprise setting. Our reference process comprises six core phases which are described by the means inputs, activities and process variations, techniques, roles, and milestones.

To the best of our knowledge, this is the first study that addresses external data sourcing from a design science research perspective. This means that we do not only focus on the necessity of performing certain activities, but develop and refine a reference process in iterative design cycles. We managed to win four companies that are experienced and advanced in the field of data management to evaluate the reference process. Their feedback has been favorable with regards to the process's relevance, understandability, and usefulness.

From an academic perspective, our findings synthesize and expand the scarce body of knowledge on data sourcing by proposing a reference process as a generic procedure (Goeken, 2011). We notice that the proposed reference process shows commonalities with strategic sourcing processes, but also uncovers data sourcing specificities. The latter include, for instance, the importance of semantics and concept mapping to integrate external data.

From the practitioners' perspective, the designed reference process aims to solve the increasingly relevant organizational challenges and contributes to the professionalization of external data sourcing. This enables a shift from ad-hoc sourcing practices to a well-defined approach for sourcing and managing external data.

Since we are among the first to explore this area, certainly there are limitations, and further research is needed to gain a deeper understanding of the domain. Although our findings were well-perceived throughout the instantiations and the focus group discussions, no large-scale demonstrations or evaluations have been conducted yet. Thus, we foresee future research activities to apply the reference process in diverse use cases and enterprise contexts, which would help generalize our findings and identify situational configurations. While our study focused on a reference process, it provides some first insights on the emerging roles in the context of external data sourcing and management. Studying both aspects would allow to gain a broader perspective on external data governance mechanisms. Another promising avenue for future research is an opportunity to explore and discuss the peculiar nature of digital data as semantic resources, drawing upon emerging literature on the topic (Aaltonen and Penttinen, 2021; Aaltonen *et al.*, 2021).

The reference process for sourcing and managing external data can support data hunters and decision makers in organizing their activities, but it is not a foregone conclusion, thus specific characteristics of enterprises' data foundation must always be included.

References

Aaltonen, A., Alaimo, C. and Kallinikos, J. (2021) "The Making of Data Commodities: Data Analytics as an Embedded Process", *Journal of Management Information Systems*. Routledge, 38(2), pp. 401–429.

Aaltonen, A. and Penttinen, E. (2021) "What Makes Data Possible? A Sociotechnical View on Structured Data Innovations", in *Proceedings of the 54th Hawaii International Conference on System Sciences*.

Aaser, M. and McElhaney, D. (2021) *Harnessing the power of external data*. McKinsey, pp. 1–7. Available at: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/harnessing-the-power-of-external-data (Accessed: November 11, 2021).

Amsterdamer, Y. and Milo, T. (2015) "Foundations of Crowd Data Sourcing", *ACM SIGMOD Record*, 43(4), pp. 5–14.

Arndt, D. and Gersten, W. (2001) "External Data Selection for Data Mining in Direct Marketing.", in *Proceedings of the Sixth International Conference on Information Quality*, pp. 44–61.

Baecke, P. and Van den Poel, D. (2011) "Data augmentation by predicting spending pleasure using commercially available external data", *Journal of Intelligent Information Systems*, 36(3), pp. 367–383.

Baud, N., Frachot, A. and Roncalli, T. (2002) "Internal Data, External Data and Consortium Data - How to Mix Them for Measuring Operational Risk", *SSRN Electronic Journal*, pp. 1–18.

Becker, J., Delfmann, P. and Knackstedt, R. (2007) "Adaptive Reference Modeling: Integrating Configurative and Generic Adaptation Techniques for Information Models", in Becker, J. and Delfmann, P. (eds.) *Reference Modeling*. Heidelberg: Physica-Verlag HD, pp. 27–58.

Belissent, J. (2019) *The Insights Professional's Guide To External Data Sourcing*. Forrester Research, pp. 1–17. Available at: https://www.forrester.com/report/The-Insights-Professionals-Guide-To-External-Data-Sourcing/RES139331 (Accessed: May 14, 2019).

Bizer, C., Heath, T. and Berners-Lee, T. (2009) "Linked Data - The Story So Far", *International Journal on Semantic Web and Information Systems*, 5(3), pp. 1–22.

Čas, K. and Meier, M. (1999) "Integration of Internal and External Data for Marketing Management", in Zupančič, J., Wojtkowski, W., Wojtkowski, W. G., and Wrycza, S. (eds.) *Evolution and Challenges in System Development*. Boston, MA: Springer US, pp. 489–503. Available at: http://link.springer.com/10.1007/978-1-4615-4851-5_45 (Accessed: October 31, 2020).

Clark, T. D., Zmud, R. W. and Mccray, G. E. (1995) "The Outsourcing of Information Services: Transforming the Nature of Business in the Information Industry", *Journal of Information Technology*. SAGE Publications Ltd, 10(4), pp. 221–237.

Davenport, T., Evgeniou, T. and Redman, T. C. (2021) "Your Data Supply Chains Are Probably a Mess. Here's How to Fix Them.", *Harvard Business Review*, 24 June. Available at: https://hbr.org/2021/06/data-management-is-a-supply-chain-problem (Accessed: October 20, 2021).

Deutch, D. and Milo, T. (2012) "Mob data sourcing", in *Proceedings of the 2012 International Conference on Management of Data*. Scottsdale, Arizona, USA: ACM Press, p. 581.

van Donge, W., Bharosa, N. and Janssen, M. F. W. H. A. (2020) "Future government data strategies: data-driven enterprise or data steward?: Exploring definitions and challenges for the government as data enterprise", in *Proceedings of the 21st Annual International Conference on Digital Government Research*. Seoul Republic of Korea: ACM, pp. 196–204.

Dun & Bradstreet (2018) *Global Family Tree Report*. Available at: https://docs.dnb.com/onboard/en-GB/Business/Reports/global-family-tree-report (Accessed: November 17, 2021).

Explorium (2021) *State of External Data Acquisition*. Available at: https://www.explorium.ai/resource/explorium-2021state-of-external-data-acquisition/ (Accessed: October 22, 2021).

Goeken, M. (2011) "Towards an Evidence-based Research Approach in Information Systems", in *Proceedings of the 32nd International Conference on Information Systems*. Available at: https://aisel.aisnet.org/icis2011/proceedings/researchmethods/10.

Gregor, S. (2006) "The Nature of Theory in Information Systems", *MIS Quarterly*, 30(3), pp. 611–642.

Hopf, K. (2019) *Predictive Analytics for Energy Efficiency and Energy Retailing*. PhD thesis, University of Bamberg. Available at: https://fis.uni-bamberg.de/handle/uniba/45617 (Accessed: April 25, 2021).

Jarvenpaa, S. L. and Markus, M. L. (2020) "Data Sourcing and Data Partnerships: Opportunities for IS Sourcing Research", in *Information Systems Outsourcing*. Springer (Progress in IS), pp. 61–79. Available at: http://link.springer.com/10.1007/978-3-030-45819-5_4 (Accessed: April 21, 2021).

Könning, M., Westner, M. and Strahringer, S. (2019) "A Systematic Review of Recent Developments in IT Outsourcing Research", *Information Systems Management*, 36(1), pp. 78–96.

Kotlarsky, J., Oshri, I., Dibbern, J. and Mani, D. (2018) *MIS Quarterly Research Curation on IS Sourcing*. Available at: https://www.misqresearchcurations.org/blog/2018/6/26/is-sourcing (Accessed: October 14, 2021).

Krasikov, P., Legner, C. and Eurich, M. (2021) "Sourcing the Right Open Data: A Design Science Research Approach for the Enterprise Context", in *The Next Wave of Sociotechnical Design*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 313–327.

Lacity, M. C., Khan, S. A. and Yan, A. (2016) "Review of the empirical business services sourcing literature: an update and future directions", *Journal of Information Technology*, 31(3), pp. 269–328.

Lacity, M. C., Khan, S., Yan, A. and Willcocks, L. P. (2010) "A Review of the it Outsourcing Empirical Literature and Future Research Directions", *Journal of Information Technology*. SAGE Publications Ltd, 25(4), pp. 395–433.

Legner, C., Pentek, T. and Otto, B. (2020) "Accumulating Design Knowledge with Reference Models: Insights from 12 Years' Research into Data Management", *Journal of the Association for Information Systems*, 21, pp. 735–770.

Lukyanenko, R., Parsons, J., Wiersma, Y. and Maddah, M. (2019) "Expecting the Unexpected: Effects of Data Collection Design Choices on the Quality of Crowdsourced User-Generated Content", *MIS Quarterly*, 43(2), pp. 634–647.

Lyytinen, K., Sørensen, C. and Tilson, D. (2017) "Generativity in digital infrastructures: A research note", in *The Routledge Companion to Management Information Systems*. Routledge.

Maher, B., Drachsler, H., Kalz, M., Hoare, C., Sorensen, H., Lezcano, L., Henn, P. and Specht, M. (2013) "Use of Mobile Applications for Hospital Discharge Letters – Improving Handover at Point of Practice", *International Journal of Mobile and Blended Learning*, 4, pp. 19–42.

March, S. T. and Smith, G. F. (1995) "Design and Natural Science Research on Information Technology", *Decision Support Systems*, 15(4), pp. 251–266.

McCarthy, S., Rowan, W., University College Cork, Lynch, L., University College Cork, Fitzgerald, C., and University College Cork (2020) "Blended Stakeholder Participation for Responsible Information Systems Research", *Communications of the Association for Information Systems*, 47(1), pp. 716–742.

Muhic, M. and Johansson, B. (2014) "Cloud Sourcing – Next Generation Outsourcing?", *Procedia Technology*, 16, pp. 553–561.

Nevo, D. and Kotlarsky, J. (2020) "Crowdsourcing as a strategic IS sourcing phenomenon: Critical review and insights for future research", *The Journal of Strategic Information Systems*, 29(4), p. 101593.

Nordin, F. and Henrik, A. (2008) "Business service sourcing: A literature review and agenda for future research", *International Journal of Integrated Supply Management*, 4(3/4), pp. 378–405.

Oshri, I., Kotlarsky, J. and Willcocks, L. P. (2015) *The Handbook of Global Outsourcing and Offshoring*. Edited by I. Oshri, J. Kotlarsky, and L. P. Willcocks. London: Palgrave Macmillan UK.

Paulheim, H. (2016) "Knowledge Graph Refinement", *Semantic Web*, 8(3), pp. 489–508.

Peffers, K., Tuunanen, T., Rothenberger, M. A. and Chatterjee, S. (2007) "A Design Science Research Methodology for Information Systems Research", *Journal of Management Information Systems*, 24(3), pp. 45–77.

Prat, N., Comyn-Wattiau, I. and Akoka, J. (2015) "A Taxonomy of Evaluation Methods for Information Systems Artifacts", *Journal of Management Information Systems*, 32(3), pp. 229–267.

Rafati, L. and Poels, G. (2015) "Towards Model-Based Strategic Sourcing", in Oshri, I., Kotlarsky, J., and Willcocks, L. P. (eds.) *Achieving Success and Innovation in Global Sourcing: Perspectives and Practices*. Cham: Springer International Publishing (Lecture Notes in Business Information Processing), pp. 29–51.

Ribas, I., Lusa, A. and Corominas, A. (2021) "Multi-step process for selecting strategic sourcing options when designing supply chains", *Journal of Industrial Engineering and Management*, 14(3), pp. 477–495.

Roeder, J., Muntermann, J. and Kneib, T. (2020) "Towards a Taxonomy of Data Heterogeneity", in *Proceedings of the 15th International Conference on Wirtschaftsinformatik*. Potsdam, Germany, pp. 293–308.

Satish, L. and Yusof, N. (2017) "A Review: Big Data Analytics for enhanced Customer Experiences with Crowd Sourcing", *Procedia Computer Science*, 116, pp. 274–283.

Schatsky, D., Camhi, J. and Muraskin, C. (2019) *Data ecosystems: How third-party information can enhance data analytics*. Deloitte. Available at: https://www2.deloitte.com/content/dam/insights/us/articles/4603_Data-ecosystems/DI_Data-ecosystems.pdf (Accessed: February 19, 2020).

Schneider, C. O., Bremen, P., Schönsleben, P. and Alard, R. (2013) "Transaction cost economics in global sourcing: Assessing regional differences and implications for performance", *International Journal of Production Economics*, 141(1), pp. 243–254.

Schneider, S. and Sunyaev, A. (2016) "Determinant factors of cloud-sourcing decisions: reflecting on the IT outsourcing literature in the era of cloud computing", *Journal of Information Technology*, 31(1), pp. 1–31.

Sesay, A. and Ramirez, R. (2016) "Theorizing the IT Governance Role in IT Sourcing Research", in *Proceedings of the 22nd Americas Conference on Information Systems*, p. 10.

Strand, M. and Carlsson, S. A. (2008) "Provision of External Data for DSS, BI, and DW by Syndicate Data Suppliers", in *Proceedings of the 2008 conference on Collaborative Decision Making: Perspectives and Challenges*. IOS Press, pp. 245–256.

Strand, M. and Syberfeldt, A. (2020) "Using external data in a BI solution to optimise waste management", *Journal of Decision Systems*, 29(1), pp. 53–68.

Sun, Z., Di, L., Fang, H., Guo, L., Tan, X., Jiang, L. and Chen, Z. (2021) "Agro-geoinformatics Data Sources and Sourcing", in Di, L. and Üstündağ, B. (eds.) *Agro-geoinformatics: Theory and Practice*. Cham: Springer International Publishing, pp. 41–66.

Venable, J., Pries-Heje, J. and Baskerville, R. (2012) "A comprehensive framework for evaluation in design science research", in *Design Science Research in Information Systems. Advances in Theory and Practice*. Springer, pp. 423–438. Available at: http://link.springer.com/chapter/10.1007/978-3-642-29863-9_31 (Accessed: May 11, 2016).

Wang, W. M., Preidel, M., Fachbach, B. and Stark, R. (2020) "Towards a Reference Model for Knowledge Driven Data Provision Processes", in Luis M. Camarinha-Matos, Hamideh Afsarmanesh, and Angel Ortiz (eds.) *Boosting Collaborative Networks 4.0*. Cham: Springer International Publishing (IFIP Advances in Information and Communication Technology), pp. 123–132.

Watjatrakul, B. (2005) "Determinants of IS sourcing decisions: A comparative study of transaction cost theory versus the resource-based view", *The Journal of Strategic Information Systems*, 14(4), pp. 389–415.

Wilmsen, M., Gericke, K., Jäckle, M. and Albers, A. (2020) "Method for the Identification of Requirements for Designing Reference Processes", *Proceedings of the Design Society: DESIGN Conference*. Cambridge University Press, 1, pp. 1175–1184.

Winter, R. and Schelp, J. (2006) "Reference Modeling and Method Construction: A Design Science Perspective", in *Proceedings of the 2006 ACM Symposium on Applied Computing*. New York, USA: ACM (SAC '06), pp. 1561–1562.

Zhang, R., Indulska, M. and Sadiq, S. (2019) "Discovering Data Quality Problems: The Case of Repurposed Data", *Business & Information Systems Engineering*, 61(5), pp. 575–593.

Zuiderwijk, A., Janssen, M., Poulis, K. and van de Kaa, G. (2015) "Open Data for Competitive Advantage", in *Proceedings of the 16th Annual International Conference on Digital Government Research*, pp. 79–88.