Data-push innovation beyond serendipity: The case of a digital platform making Earth Observation data fit into multiple use contexts

Raphaëlle Barbier, Skander Ben Yahia, Sylvain Lenfle, Benoit Weil

Technovation 132 (2024) 102992

Contents lists available at ScienceDirect

Technovation

journal homepage: www.elsevier.com/locate/technovation

Data-push innovation beyond serendipity: The case of a digital platform making Earth Observation data fit into multiple use contexts

Raphaëlle Barbier [a,*], Skander Ben Yahia [a], Sylvain Lenfle [b], Benoit Weil [a]

[a] Mines Paris – PSL University, France
[b] CNAM & Ecole Polytechnique, France

A B S T R A C T

The potential of data to circulate across organizations and sectors and stimulate innovation in multiple contexts has been largely acknowledged by practitioners and researchers. This has given rise to a specific form of innovation strategy, "data-push innovation", which consists of stimulating the use of existing data by third-party actors. However, how to manage such a strategy remains challenging. The paper explores this question by examining the longitudinal case study of an actor that has successfully stimulated the use of Earth observation data by multiple actors over the last 40 years. The paper offers several contributions to research in information systems and innovation management. First, the paper shows that data-push innovation can be fostered through the intentional design of a so-called "fit system" allowing data to be used in multiple contexts. Such a fit system can be built as a generic system, following similar "generification" strategies as those supporting platform or software development but with original patterns to adapt to the specificities of data-push innovation. Second, the paper characterizes the types of "boundary resources" needed to support this process. These boundary resources especially have a two-way resourcing function: they help third-party actors contribute to the fit system development, but they also allow the fit system owner to identify the knowledge boundaries preventing data from gaining meaning in new contexts. Third, the paper reveals an intriguing form of localized and nondominant platform leadership, focusing on gaining generative power rather than controlling power over the platform ecosystem.

1. Introduction

In recent years, the development of the IoT, of increasingly low-cost sensors and of computational capacities has dramatically increased the flow of data in almost every business, industry and research area. This "big data" phenomenon has been largely described in both innovation management literature (e.g. Cappa et al., 2021; Appio et al., 2021; Bharadwaj and Noble, 2017; Blackburn et al., 2017; George et al., 2014) and information systems (IS) literature (e.g. Günther et al., 2017; Abbasi et al., 2016; Gandomi and Haider, 2015; Chen et al., 2012).

Scholars have especially highlighted different forms of data-based innovation strategies to take advantage of the new possibilities offered by data. Among these different possibilities, scholars have especially unveiled certain forms of "data-push innovation" (Han and Geum, 2022; Trabucchi and Buganza, 2020) starting with existing data and focusing on stimulating the use of these data by third-party actors outside the organization's boundaries, which is also similarly referred to as "outbound data-based innovation" (Trabucchi et al., 2018). For example, Uber created a service called Uber Movement, providing anonymized data collected through their service of matching riders and drivers to help urban planning around the world – e.g., by offering insights to avoid traffic congestion caused by specific events or holiday traffic (Trabucchi and Buganza, 2020). Data-push innovation strategies seem all the more promising because data are acknowledged to be highly 'portable', thus having the broader potential of being used and reused across a large variety of contexts across organizations and industries (e.g. Günther et al., 2017; Lycett, 2013). However, how to manage such strategies remains insufficiently understood. Scholars point to the "challenging nature of the effort for identifying the potential [users of data]" (Trabucchi and Buganza, 2020, p. 9). As such, these strategies are described as merely "a matter of serendipity" (Trabucchi et al., 2018, p. 52), relying on the ability of potential customers to identify a "different hidden value within existing databases" (Trabucchi et al., 2018, p. 51). This paper aims to bridge the gap in how to manage data-push innovation strategies beyond serendipity. In particular, the paper proposes to leverage two main streams of literature that already provide important insights but have been insufficiently articulated in the specific context of data-push innovation.

On the one hand, information systems (IS) research has a long tradition in investigating digital innovation strategies by closely analyzing the underpinning system of interrelated components enabling information to be created, processed, distributed, and used by

* Corresponding author.
E-mail address: raphaelle.barbier@minesparis.psl.eu (R. Barbier).

https://doi.org/10.1016/j.technovation.2024.102992
Received 15 February 2022; Received in revised form 20 February 2024; Accepted 4 March 2024
Available online 10 March 2024

organizations to support multiple forms of activities. In particular, a recent and growing body of IS research elucidates the specificities of data compared to other forms of digital objects, such as software modules, computer programs, digitalized books or videos, and the resulting specificities of the innovation processes involved in creating value from data (Aaltonen et al., 2021; Mikalsen and Monteiro, 2021; Alaimo et al., 2020; Alaimo and Kallinikos, 2020; Østerlie and Monteiro, 2020; Monteiro and Parmiggiani, 2019; Aaltonen and Tempini, 2014). These studies suggest that data-push innovation, as a specific form of data-based innovation, especially needs a logic of meaning construction (e.g., Alaimo et al., 2020). Indeed, data are originally associated with a meaning that is linked to the purpose for which they are produced (e.g., anonymized data collected through Uber to match riders and drivers). However, beyond their initial context of production, these data are likely to be meaningless for others. In this respect, IS scholars highlight that the meaning of data needs to be reconstructed through a sophisticated process to make data valuable in a new context (e.g. Aaltonen et al., 2021; Alaimo et al., 2020; Aaltonen and Tempini, 2014). In line with IS research (e.g. Li and Nielsen, 2019; Dittrich, 2014), a "context of use", or "use context", refers to the specific context in which data eventually appear to be valuable for an actor having certain needs and practices (e.g., an advertising company that is willing to assess its advertising audiences and for which data from the telecommunication operator eventually prove to be valuable). As expressed by Li and Nielsen (2019) in the case of software development, "designers attempt to understand the users' current needs and practices, and predict and anticipate how the artifact to be designed can fit into this context" (p. 5).

Drawing upon the recent studies on data specificities (e.g. Aaltonen et al., 2021; Alaimo et al., 2020; Aaltonen and Tempini, 2014), it appears that (re)constructing the meaning of data involves building *technical components*, such as algorithms, to create meaningful metrics for a certain group of users (e.g., Aaltonen and Tempini, 2014); *socio--organizational components*, such as contracts with relevant stakeholders or collaborations, to build consensus across the industry on the legitimacy of the chosen metrics (e.g., Aaltonen et al., 2021); and *cognitive components*, such as knowledge on the sensors producing data and the potential contexts of use (e.g., Monteiro and Parmiggiani, 2019). The paper designates this set of components as the *"fit system", which is aimed at enabling data to gain meaning in a certain range of heterogeneous contexts, or in other words, enabling data to fit into a certain range of heterogeneous use contexts*. From this perspective, *fostering data-push innovation can be considered to involve designing a fit system between data and multiple use contexts*. However, this literature stream mainly considers a limited timeframe in a limited number of occurrences, merely focusing on fitting the data produced in a given context into one new use context or within one single sector, e.g., making data that are initially collected by a telecommunication operator meaningful for advertising companies (e.g. Aaltonen et al., 2021; Alaimo et al., 2020). Therefore, this literature offers a limited understanding of how data can be used and reused by a large variety of organizations across heterogeneous sectors.

In this respect, a second stream of literature offers complementary insights. Indeed, a long tradition of research in IS and innovation management has investigated how information systems could span a large variety of use contexts, for example, digital platforms such as Apple's iOS operating system (e.g. Constantinides et al., 2018; de Reuver et al., 2018; Tiwana et al., 2010), enterprise software planning (ERP) systems, or open data platforms (Bonina et al., 2021; Bonina and Eaton, 2020; Karhu et al., 2018; Ruijer et al., 2017). In this regard, scholars have elucidated so-called "generification" mechanisms that support the design of systems that are generic enough to be compatible with diverse use contexts and flexible enough to be further customized to provide a better fit for a specific use context (Gizaw et al., 2017; Hanseth and Bygstad, 2015; Silsand and Ellingsen, 2014; Monteiro et al., 2013; Pollock et al., 2007). However, for the specific case of data-push innovation, the relevance and forms of these mechanisms need to be better understood.

Drawing upon the respective contributions and limitations of these two streams of literature, this paper addresses the following research question: *how can generification mechanisms* support *the development of a fit system allowing data to be used in multiple contexts?* Following recommendations for further longitudinal empirical research capturing how data-based value creation processes unfold over time (Günther et al., 2017), this paper is based on a longitudinal case study of an actor that has successfully managed such generification mechanisms in the specific case of Earth Observation (EO) data, i.e., data that are used to capture the different phenomena related to Earth's physical, chemical and biological systems and are considered to be a promising open data resource (Borzacchiello and Craglia, 2012).

Our research primarily contributes to IS and innovation management research, advancing our understanding of data-push innovation. Aligning with a common approach in IS research, the strategy of an actor is primarily examined through the analysis of the information system (here, the fit system) the actor intends to develop. In this regard, the fit system can be understood as the means for the fit system owner to set a certain strategy and take on a certain role within its ecosystem. This analysis allows us to characterize one possible strategy to support data-push innovation beyond serendipity, coined "data genericity building", which consists of designing the fit system as a generic system. The paper shows the originality of the generification mechanisms when applied to the case of data-push innovation. These original patterns are intrinsically linked to the specificities of the meaning construction process underpinning data-push innovation, in contrast with other forms of digital innovation. The case study analysis also unveils an original form of strategic positioning that goes hand-in-hand with the development of the fit system, allowing a fit system owner who has limited power in a given ecosystem to contribute to the dynamics of this ecosystem in a localized and nondominant way.

These results have important implications for researchers and practitioners with regard to the specificities of data-push innovation within the broader field of digital innovation. From a theoretical perspective, the concept of a "fit system" offers a way of bridging the two IS streams of literature mentioned previously, which are highly complementary but remain thus far insufficiently articulated: on the one hand, there is long-standing research on building generic information systems that can serve a large variety of users; on the other hand, there is a more recent but growing body of research on characterizing data as specific digital artifacts, rather than more well-known forms such as software modules. Our case study highlights that the management of data-push innovation can be inspired by well-known strategies for platform or software development (aiming at developing generic systems) but requires additional adaptations to take into account the fact that data cannot 'travel' across use contexts as mere software modules to be customized and recombined. On a more practical basis, the paper depicts the efforts of generification through a threefold perspective: which generification mechanisms unfold, how they can be supported, and why they are important with regard to the strategy of the actor aimed at fostering data-push innovation.

2. Theoretical background

This section, which is organized into three parts, presents the conceptual lens that will be used in the case study analysis and discussion sections to analyze the generification mechanisms involved in designing a fit system between data and multiple use contexts. The first part elaborates on the composition of the fit system. The second part further characterizes its architecture and the associated nature of generification mechanisms. The third section synthesizes the main insights derived from the two previous parts.

2.1. Composition of the fit system

As introduced above, the underlying rationale of the fit system

consists of enabling data to gain a certain meaning in use contexts that are different from the ones for which the data were initially produced. As indicated by the recent body of studies on data as specific digital artifacts (Aaltonen et al., 2021; Mikalsen and Monteiro, 2021; Alaimo et al., 2020; Alaimo and Kallinikos, 2020; Østerlie and Monteiro, 2020; Monteiro and Parmiggiani, 2019; Aaltonen and Tempini, 2014), a *fit system is not automatically generated when data are produced but needs to be built through dedicated design efforts*. As recalled by Mikalsen and Monteiro (2021), there is sometimes a "misconception that data speak for themselves", which downplays the "considerable, ongoing work to craft data into data". In this respect, the case of network data produced by a telecommunications operator is particularly telling, as, for example, developed by Aaltonen et al. (2021). Data are initially produced for basic management tasks of the network infrastructure, as every click, call and message of the network subscribers are recorded. These data encode the identity of the subscriber, the time and type of the network operation and other details in the form of alphanumeric characters, such as 097369D2D7372762D31080000000000000001; 1;33668741168; 3322208; 6;20081101004923; 20081101004923; 20081101004923. As such, it clearly appears that those data cannot easily—or at least directly—fit into other use contexts beyond the boundaries of the telecommunication operator. These records of alphanumeric characters appear meaningless until they are transformed into metrics such as records of audience members and advertising audiences that can then fit into advertising companies' practices.

As mentioned in the introduction, three broad categories of components are especially critical for supporting the meaning construction process that is needed to make data fit into various use contexts. These categories are hereafter examined in more detail. First, the fit system includes *technical components*. These components consist of the technical infrastructure required for the collection, storage, processing and distribution of data (e.g. Abbasi et al., 2016). In particular, the algorithms and analytical methods used to process data into meaningful metrics play a critical role (e.g. Günther et al., 2017; Chen et al., 2012).

Beyond these technical components, the fit system also involves designing some forms of *socio-organizational components*, referring to the various forms of relationships to be built with relevant actors. As highlighted by Günther et al. (2017), creating value from data requires specific organizational models. This might involve internal relationships within a given organization, for example, taking the form of temporal and local arrangements between data analysts to discuss the meaning of data (e.g. Mikalsen and Monteiro, 2021). This might also involve external relationships with data producers and users or other relevant actors of the ecosystem, e.g., "to negotiate the adoption of metrics and their meanings, eliminate ambiguities, and build consensus across the industry or ecosystem in which they operate" (Aaltonen et al., 2021, p. 419).

Finally, *cognitive components* are also an essential part of the fit system. They specifically refer to knowledge on the contexts of data production and uses. Indeed, concerning the data production context, data can never be considered completely "raw" and are always formatted by the particular conditions under which they are produced (e.g. Gitelman, 2013). The representation of reality conveyed by data is guided by an initial purpose (Aaltonen et al., 2021; Mikalsen and Monteiro, 2021) that ultimately "sets the boundaries" of what can be later derived from data (Aaltonen and Tempini, 2014), hence "[circumscribing] the opportunities and limitations of data as resources" (Aaltonen et al., 2021, p. 418). Moreover, regarding the understanding of new potential use contexts, other authors have especially emphasized the crucial importance of "industry-specific expertise" (e.g. Schymanietz et al., 2022; Urbinati et al., 2019), which consists of knowledge on sectors and organizations that might benefit from data. This especially includes understanding the political, environmental, economic, and social aspects that might influence the relevance of data for a given organization or sector.

Notably, these categories of components should not be considered separately, as they are closely intertwined. Indeed, both technical and organizational components require adapted knowledge to be constructed. Moreover, technical and socio-organizational components are also closely linked (e.g., standards requiring relationships with the industry to be defined and implemented). All three dimensions thus need to evolve synchronously, as noted by Aaltonen et al. (2021), who called for "keeping several data sources, analytical tools, and organizational practices in sync" (p. 416).

2.2. Architecture of the fit system and generification mechanisms

To further understand how the fit system can enable data to be used in multiple contexts, it is important to consider the architecture of the fit system. The literature has already extensively highlighted modularity as a promising way of designing complex systems by developing and recombining modular components (Baldwin and Clark, 2000), which is especially prevalent in platform architectures (e.g. Gawer, 2014; Baldwin and Woodard, 2009). A platform architecture basically comprises three main elements: a generic core made of low-variability components, a periphery of complementary modules addressing the variety and variability of specific needs and practices associated with given use contexts, and interfaces setting the rules of interactions among components (Baldwin and Woodard, 2009). Depending on the nature of these interfaces, the peripheral components can be either built by the platform owner in cases of internal platforms that are limited to the boundaries of a given organization or by third-party innovators in cases of industry platforms that span a wider ecosystem of actors (Gawer, 2014). The term "ecosystem" has been widely used in management research (e.g. Jacobides et al., 2018). The present paper follows the "platform ecosystem" view mentioned by the latter authors, considering the "ecosystem" as the broad set of actors organizing around the platform and contributing to its functioning. Moreover, it is important to note that the fit system is a form of *innovation platform*, defined as "a technological foundation upon which a large number of [actors] can build further complementary innovations", and differing from transactional platforms that "create value by facilitating the buying and selling of existing goods and services", such as social networks or online marketplaces (Gawer, 2020).

Articulating a generic core and use-specific peripheral components especially allows the platform owner to manage the tension between standardization, which is necessary for a system to serve a multitude of organizations, and flexibility, which is necessary for a system to fit with the specificities of a given use context and to be open to future changes (e.g. Monteiro et al., 2013; Henfridsson and Bygstad, 2013; Hanseth et al., 1996). Scholars have further investigated how the generic core and peripheral components of a system could be designed over time, in particular through so-called *"generification" mechanisms* (Gizaw et al., 2017; Silsand and Ellingsen, 2014; Monteiro et al., 2013; Pollock et al., 2007). In particular, Gizaw et al. (2017) describe these mechanisms in an open and distributed context of innovation for the development of an open-source health information system, describing "open generification" as "establishing the necessary resources […] that enable locally situated developers to perform local innovations" (p. 635). Taking a dynamic view, the authors specifically show that generification involves two intertwined processes: *disembedding*, which is defined as "the process of lifting out local software requirements out of their contexts and abstracting them to serve diverse user needs across space and time" (p. 622), and *embedding*, which is defined as "pinning down the disembedded system back to situated realities" (p. 622).

This case of "open generification" seems particularly relevant for our investigation. Indeed, data-push innovation corresponds to a case of distributed innovation in which third-party actors (i.e., potential users of data) can be involved, as especially highlighted in recent cases of open data platforms (e.g. Bonina and Eaton, 2020; Ruijer et al., 2017). As such, it seems consistent to consider similar forms of generification mechanisms applied to the case of the fit system. In this regard, the definition of both disembedding and embedding processes can be

adapted as follows: *disembedding* refers to the process of making certain elements of the fit system serve diverse use contexts; *embedding* refers to the process of adding use-specific peripheral components to the fit system to better fit data into a given use context.

Regarding the interfaces between the generic core and peripheral components, the IS literature has specifically developed the concept of *"boundary resources"* (Gawer, 2020), e.g., in the case of Apple's iOS operating system (Eaton et al., 2015; Ghazawneh and Henfridsson, 2013). In the case of Apple described by the latter authors, these boundary resources are characterized by both a "resourcing" function and a "securing" function. The resourcing function designates the "process by which the scope and diversity of a platform is enhanced" (Ghazawneh and Henfridsson, 2013, p. 177). This involves fostering third-party actors' contributions to the expansion of the platform through the development of complementary modules (e.g., through Application Programming Interfaces (APIs) and Software Development Kits (SDKs)). In addition to a resourcing function, the boundary resources also have a "securing" function, enabling a form of control over the strategic components of the platform (e.g., through licenses defining what is allowed and what is not). Recent studies have unveiled a broader variety of boundary resources—such as information portals, documentation, helpdesks, and workshops—in cases where it might not be trivial for third-party actors to develop complementary components (Foerderer et al., 2019). Such forms of boundary resources appear to be critical for overcoming "knowledge boundaries" between the platform owner and third-party actors by "furnishing third parties with development knowledge in order to facilitate participation in development and innovation" (Foerderer et al., 2019, p. 120). This echoes similar findings on open data platforms, where boundary resources include multiple devices such as web portals and hackathons fostering the use of datasets by various actors or dataset templates to help data producers share their data in a standardized format (Bonina and Eaton, 2020). The main notions used to describe the architecture of the fit system are synthesized in Fig. 1.

2.3. Summary of theoretical background

Our research question has been formulated as follows: *how can generification mechanisms* support *the development of a fit system allowing data to be used in multiple contexts?* Several factors suggest that the forms of embedding-disembedding processes and boundary resources supporting the design of the fit system are likely to differ from reported cases of generification. Indeed, additional issues can be noted in the context of data-push innovation, calling for more intensive and probably new forms of embedding-disembedding processes.

First, one can note that reported cases of generification rely on already well-established design capacities of local developers (Gizaw et al., 2017), although these competencies might need to be enhanced (Silsand and Ellingsen, 2014). In contrast, in the context of data-push innovation, potential users might be highly unfamiliar with data (e.g. Magalhaes and Roseira, 2020; Jetzek et al., 2019; Janssen et al., 2012) and thus do not necessarily constitute an active and strongly-tied

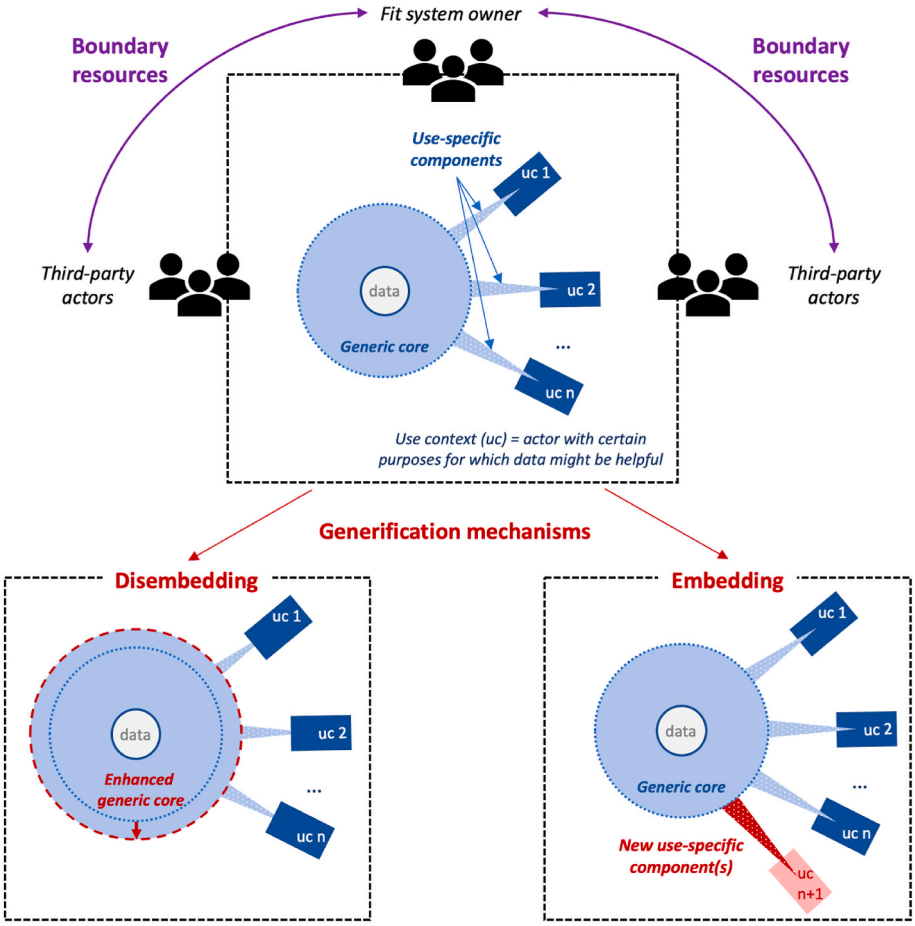

Fig. 1. Synthesis of the main concepts of our theoretical background. Representation of the **fit system** in pale blue, entailing a generic core (i.e., components that are common to several use contexts) and a periphery made of use-specific components (dotted blue spikes). The dark blue boxes represent the use contexts into which data are eventually used thanks to the design of the fit system. **Boundary resources** are represented in purple and support the interfaces between the fit system owner and the third-party actors of the ecosystem. **Generification mechanisms** are illustrated in the two lower boxes: **embedding** consists of adding new use-specific components to the fit system, and **disembedding** consists of enhancing the generic core of the fit system.

network of third-party innovators (Bonina and Eaton, 2020). In such contexts, it can thus be expected that the fit system owner needs to undertake more extensive embedding efforts compared to previous cases of generefication to balance the initial limited capacities of third-party actors. Second, beyond embedding processes, data-push innovation is also likely to require more extensive disembedding processes. These processes were already critical for other systems, such as ERPs, to ensure that these systems could fit multiple users. However, given the high 'portability' or 'liquidity' of data underlined by scholars and practitioners (e.g. Günther et al., 2017; Lycett, 2013), data have the potential to be used by an increasing number and variety of users, thus calling for amplifying disembedding processes.

These considerations suggest that a specific tension needs to be overcome in a context of data-push innovation: on the one hand, third-party actors are likely to require more embedding efforts from the fit system owner; on the other hand, and at the same time, the fit system owner is driven toward intensified disembedding efforts to address the large range of potential data users. It especially raises the question of how this double intensification can be managed by the fit system owner. Initially, the fit system owner does not seem particularly well positioned to do so due to its limited capacities to reach and mobilize third-party actors. Indeed, the variety and heterogeneity of sectors in which data could be used increases the chance for the fit system owner of primarily appearing as a new entrant with limited leadership in these sectors (e.g., a telecommunications operator primarily appears as a small player in the advertising sector).

These elements lead us to assume that original forms of generification mechanisms are likely to unfold to cope with the expected issues related to data-push innovation mentioned above. These mechanisms will be specifically examined in the discussion section by taking three complementary perspectives.

- The *'what'* perspective, describing the observed embedding and disembedding processes and their sequence over time;
- The *'how'* perspective, describing the actions and devices supporting these mechanisms, with a specific focus on boundary resources as encouraged by recent developments in IS research described above;
- The *'why'* perspective, describing the drivers pushing the fit system owner to undertake dedicated embedding-disembedding processes.

By taking these three complementary perspectives, analyzing the generification mechanisms of the fit system will also allow us to better understand the role and strategy of the fit system within its ecosystem.

3. Methodology

This research uses a qualitative methodology, relying on a longitudinal case study (Yin 2009). We investigate the case of an organization that has steered the development of a platform fostering the use of EO data across various types of actors (industries, public bodies, research communities) for more than 40 years. In particular, the case study is conducted following principles of collaborative management research (Shani et al., 2008). Collaborative management research is particularly adapted to investigate an ongoing phenomenon that organizations face in practice but that remains poorly understood. It especially involves "an organization […] undergoing development by adopting new structures and processes, [and] researchers [attempting] to provide knowledge, which is not readily accessible in the organization" (Pasmore et al., 2008, p. 13). In this respect, collaborative management research offers an interesting way to investigate the structures and processes put in place by an organization involved in data-push innovation. The paper specifically investigates the case of an organization that has successfully developed a fit system allowing data to be used in multiple contexts over time.

3.1. Overview of the Earth observation data ecosystem

Earth observation (EO) refers to the gathering of data about Earth's physical, chemical and biological systems collected through in-situ instruments and satellites and computed from large models such as those used in meteorology. EO data were initially produced mainly for scientific goals to monitor the planet and its environment, e.g., to measure and monitor the ocean, solar radiation reaching the ground, the composition of the atmosphere, and the status of vegetation. Different kinds of instruments are used for this purpose and include in-situ sensors (for example, floating buoys to monitor ocean currents, temperature and salinity or land stations that record air quality and rainwater trends), airborne sensors, or satellites. In recent years, the development of remote-sensing satellites and increasingly high-tech in-situ instruments has generated an increasing amount of data. Moreover, the European Union has significantly invested in making these scientific data a common good that is freely accessible to all potential users through dedicated "open-data" policies. Socioeconomic applications of these data are diverse and promising. They could benefit not only the EO scientific community but also public authorities, private companies, industry, universities, and citizens. In particular, EO data can aid these different actors in facing current socioenvironmental grand challenges, e.g., by providing the means of monitoring and responding to natural disasters (fires, floods, earthquakes and tsunamis), assessing the potential of renewable energy sources, managing freshwater supplies and agriculture, and addressing emerging diseases and other health risks. However, the actors of the EO community currently face difficulties in further stimulating the use of EO data in such a large variety of evolving use contexts and acknowledge the need to go beyond mere open-data policies through different types of capacity-building activities and new projects targeting the development of services based on EO data that could be more easily integrated into users' practices (Goor et al., 2021; Ranchin et al., 2021).

3.2. Empirical materials

This paper focuses on the longitudinal case study of the research center *O.I.E. (Observation, Impacts, Energy)* at MINES Paris – PSL (France). With a current team of 24 researchers, O.I.E. has developed several research activities since 1976 based on the exploitation of EO data in the fields of renewable resource evaluation (solar, wind, ocean, etc.), meteorology and climatology for energy, methodological development to assess the environmental impacts of renewable energy uses, and interoperability and dissemination of information through databases and Web services. The activities related to solar radiation illustrate particularly well the successful efforts of O.I.E. in continuously stimulating the use of EO data in a large variety of use contexts that are different from those of their original purposes. Indeed, the work of O.I.E. has basically consisted of taking satellite data initially developed for climate and meteorology purposes (Meteosat series of satellites) and transforming them into solar radiation data that can be further used in multiple other contexts through so-called "SoDa" (SOlar radiation DAta) Web services. These services are maintained and commercialized by a commercial company called "Transvalor", with whom O.I.E. has built a specific partnership since 2009. These services are currently used by thousands of users (76 000 unique visitors in 2020) and approximately 100 clients of paid-for services, spanning various sectors (e.g., solar energy industry agriculture, construction industry, health industry). The SoDa service is mostly used by companies that leverage SoDa for their own purposes or to build new services.

From this perspective, it can be considered that O.I.E. is building a fit system between EO data (in particular solar radiation data) and multiple use contexts. As such, it will be designated as a "fit system owner" in subsequent sections of the paper. Moreover, it is important to note that the case of O.I.E. offers a particularly interesting situation for precisely investigating the phenomena related to the design of the fit system.

Indeed, it corresponds to a form of a "pure" case where the rationale almost exclusively lies in creating the fit between existing data and users' existing capacities, without considering how new data could be produced or how users' practices should be dramatically transformed. Indeed, O.I.E. is not in a position to trigger the launch of new satellites in response to certain demands; this role is being taken by other industrial players that develop scientific instruments producing EO data, such as Airbus or Thales Alenia Space. Although O.I.E. is not responsible for the production of new data, it is important to note that the nature of the data used by O.I.E. does change over time. Indeed, datasets are continuously enriched with new measurements when instruments are running (e.g., every 15 min for Meteosat data), and new instruments are regularly created. Moreover, considering the side of data users, O.I.E. is primarily external to the sectors that might benefit from EO data. Therefore, O.I.E. cannot trigger deep reorganizations of these different sectors. As a result, O.I.E. has to account for the practices of the already well-established actors, both on the side of data production and data usage.

3.3. Data collection and analysis

Our empirical materials were collected from September 2018 to September 2023. Different forms of empirical evidence were exploited to ensure triangulation between sources (Yin, 2009), which are listed in Table 1. The interviews allowed us to build an in-depth understanding of O.I.E.'s activities over time. Secondary sources of data were used to enrich some aspects discussed during the interviews, especially to validate the temporality and exact content of O.I.E.'s activities. Regular informal interactions were also used as a way of enriching our interpretation of O.I.E.'s activities and clarifying some points of analysis.

These data were analyzed in a collaborative research setting involving both researchers and practitioners, which were, in this case, members of O.I.E. and other relevant partners, such as Transvalor (Shani et al., 2008). This setting especially aims to "reduce the likelihood of drawing false conclusions from the data collected, with the intent of both proving performance of the system [of action] and adding to the broader body of knowledge in the field of management" (Pasmore et al., 2008, p. 20). Following guidelines for collaborative research (Cirella et al., 2012; Pasmore et al., 2008), the analysis procedure consisted of *progressively building a shared interpretation of empirical data and findings between researchers and practitioners*. The analysis procedure thus involved two intertwined forms of actions: (1) continuously deepening the understanding of O.I.E.'s activities on the basis of interviews, notes taken after informal interactions, and secondary sources of data and (2) validating this understanding with practitioners through dedicated collaborative meetings, resulting in a shared interpretation of empirical data and findings, as detailed in Table 2. These meetings were organized as follows: the first phase consisted of a presentation made by our research team on our current understanding of O.I.E.'s activities and

Table 1
Sources of empirical data.

Interviews	Interviewed members of O.I.E. - Emeritus researcher and previous director of O.I.E. (involved in solar radiation research at O.I.E. since the 80s): *6 h in 2018 to go through O.I.E.'s overall history and origins* - Senior researcher and director of O.I.E. (PhD at O.I.E. 1991–1993, researcher at O.I.E since 1995): *2 h in Nov 2018, 2 h in Nov 2021, regular informal interactions since 2018* - Senior researcher in remote sensing & energy sector (PhD at O.I.E. 1996–2000, research engineer in the industry 2000–2007, O.I.E. researcher since 2007): *1 h in Nov 2018, 1 h in Nov 2019, 2 h in Nov 2021, regular informal interactions since 2018* - Research scientist with specific expertise in data infrastructures and Web services (involved in the team since the 80s): *2 h in July 2019, 2 h in Nov. 2021, regular informal interactions since 2018* - Researcher with specific expertise in ICTs (involved in the team since the 2000s): *2 h in July 2019* Transvalor's team involved in SoDa (in 2021) - Manager of the team since 2009: *2 h in Nov 2018* - Technical support and sales engineer: *2 h in Nov 2018* – 2 researchers in remote sensing for longer-term research projects: *2 h in Nov 2018 (only one of them)* – 2 R&D engineers for maintenance and operations: *2 h in July 2019 (only one of them)*
Secondary sources	**List of past and present research projects**: 56 projects extracted from O.I.E.'s website (related to solar resource assessment, interoperability and dissemination of data, instruments for Earth observation) **Scientific publications (conferences and journal papers)** accessible through the HAL open-access platform: - Related to Heliosat methods, especially (Cano et al., 1986; Rigollier et al., 2004; Blanc et al., 2011b; Lefèvre et al., 2013; Blanc and Wald, 2015; Tournadre et al., 2022) - Related to SoDa project and Web services, especially (Rigollier et al., 2000; Gschwind et al., 2006; Thomas et al., 2013) - Related to solar cadasters developed for different use contexts, especially (Blanc et al., 2011a; Ménard et al., 2013; Callegari et al., 2017; Blanc and Ménard, 2021)
Informal interactions	1st author visiting periods in O.I.E. research team: 3 weeks from July to August 2019 & 4 weeks in November 2021 (the notes taken during these periods were used to enrich an overall understanding of O.I.E.'s activities)

Table 2
Synthesis of the validation meetings implemented for data analysis.

	Collaborative validation meetings	Formalization of a validated and shared interpretation of data and findings
April 2019	**Objective:** validating with O.I.E. a first interpretation of empirical data and findings (built on secondary sources of data and interviews) **Participants:** – 3 members of O.I.E.: lab director, senior researcher in remote sensing & solar energy, research scientist expert in data infrastructures and Web services – 3 management researchers (including 1st & 4th authors)	**Visual mapping strategy:** 1st template representing the fit system as developed by O.I.E to enable the use of solar radiation data by ISWT **Narrative strategy:** document written by our research team, revised and validated by the participants, and additionally validated by two external reviewers (one expert in the EO field and one innovation management researcher)
June 2019	**Objective**: validating with Transvalor the interpretation of empirical data and findings previously discussed with O.I.E. **Participants:** – 3 members of O.I.E.: lab director, senior researcher in remote sensing & solar energy, research scientist expert in data infrastructures and Web services – 3 members of Transvalor: team manager, 2 researchers – 3 management researchers (including 1st & 4th authors)	**Visual mapping strategy:** 1st template representing the fit system as developed by O.I.E to enable the use of solar radiation data by ISWT **Narrative strategy:** document written by our research team, revised and validated by the participants, and additionally validated by two external reviewers (one expert in the EO field and one innovation management researcher)
January 2022	**Objective:** validating with O.I.E. an updated interpretation of empirical data and findings (enriched through interviews and regular informal interactions) **Participants:** – 3 members of O.I.E.: lab director, senior researcher in remote sensing & solar energy, research scientist expert in data infrastructures and Web services – 1 management researcher (1st author)	**Visual mapping strategy:** 2nd template representing the evolving composition of the fit system over time **Narrative strategy:** preliminary version of the present paper, revised and validated by O.I.E.

remaining questions and blind spots, and the second phase consisted of a semistructured discussion with the participating practitioners starting with their reactions to what was presented in the first phase. To support this process, two analytical strategies were specifically used. These included a visual mapping strategy, allowing the "simultaneous representation of a large number of dimensions" (Langley, 1999, p. 700), and a narrative strategy, involving the "construction of a detailed story from the raw data" (Langley, 1999, p. 695), as recommended for research with the goal of "understanding how things evolve over time and why they evolve in this way" (Langley, 1999, p. 692). Finally, the intermediary versions of the present paper written in 2022–2023 were systematically reviewed and validated by O.I.E.'s team.

More specifically, to characterize the nature of the fit system and identify the embedding and disembedding processes over time, the following process was applied.

First, we conducted an in-depth analysis of a recent case in which O.I.E. had successfully enabled data to fit into a new use context (a start-up called "In Sun We Trust" (ISWT) using solar radiation data to foster the installation of photovoltaic panels in urban areas). The different components of the fit system developed by O.I.E. were identified based on the collected empirical materials presented above. In particular, the interviews with O.I.E.'s members could be easily triangulated with existing online documentation (academic publications, communications on O.I.E.'s and ISWT's websites). The components of the fit system were categorized as "use-specific" if they had been specifically developed to adapt to ISWT's context or "generic" if they were used to fit data into multiple use contexts beyond ISWT. A first type of visual mapping template was especially used to support this process, allowing us to represent our understanding in a synthetic way and obtain O.I.E.'s feedback and validation (see Fig. 2). This first visual mapping template provides a *static view* of the fit system, i.e. the composition of the fit system captured at one moment in time.

Subsequently, the different components of the fit system were traced back into O.I.E.'s long-term trajectory starting in 1976. Specific attention was given to identifying when a given generic component was first built and determining whether it was built as a use-specific component or as a generic component at the outset. Disembedding processes could be identified when O.I.E.'s development efforts consisted in widening the range of use contexts for which a given component could be used. The embedding processes were associated with O.I.E.'s development efforts to further fit data into a given use context. A second type of visual mapping template was used to represent our analysis in a synthetic way and obtain O.I.E.'s feedback and validation (see Fig. 2). As opposed to the first visual mapping template focusing on the fit system at one moment in time, the second visual mapping template provides a *dynamic view* of the fit system, by showing the evolution of use-specific and generic components of the fit system over time. For simplification purposes, in this dynamic-view template, the components of the fit system are named based on the technical dimension of the fit system, but the other dimensions (socio-organizational, cognitive) are closely intertwined. For example, "SoDa Web services" refers to the technical artifacts but also embodies O.I.E.'s relationship with Transvalor and O.I.E.'s knowledge on solar radiation and energy sectors.

4. Case study analysis

The case study analysis of O.I.E. begins with a recent situation in which O.I.E. succeeded in fitting EO data into a given use context (ISWT). This allows us to precisely describe the components of the fit system in this context. Our analysis then goes back in time to investigate

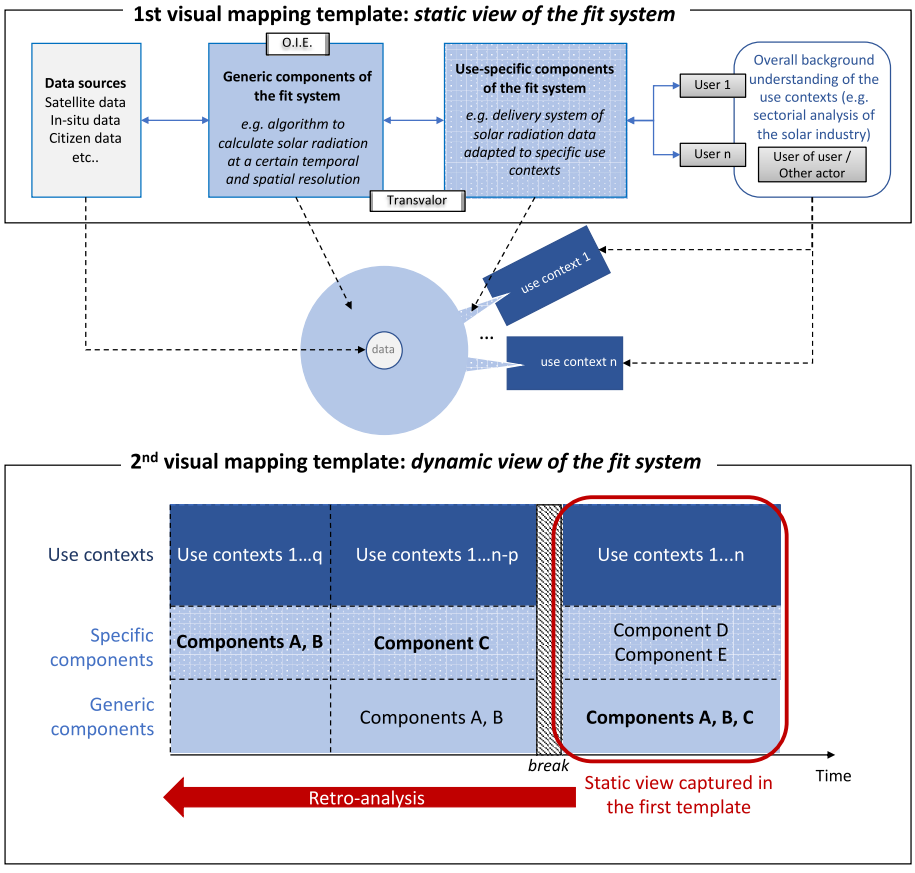

Fig. 2. Visual mapping templates used as supportive tools of the collaborative research process. The completed versions of these templates are later shown in the "Case study analysis" section (Figs. 4 and 6).

O.I.E.'s activities since 1976 and how they have contributed to designing the fit system involved in the case of ISWT. This retro-analysis allows us to further elucidate the embedding-disembedding processes, the boundary resources underpinning the design of the fit system over time and the reasons for such developments.

4.1. Characterizing the fit system underpinning a successful case of data-push innovation

Since 2014, O.I.E. has developed so-called "solar cadasters" at the urban scale for the start-up ISWT. Created in 2015 following the first interactions of the founders with O.I.E., this start-up aims to provide private individuals with a free support service facilitating the installation of photovoltaic (PV) systems on their roofs. This includes a simulation tool of economic profits provided by the installation of PV systems allowing electricity self-consumption or selling to the electric grid at the feed-in tariff (see Fig. 3) and the connection of interested clients with local PV installers, who are selected by ISWT for their validated trustworthiness. ISWT makes profits by selling the cadasters to municipalities or local authorities, as well as taking margins on the transactions between the clients and PV installers when a contract is eventually signed thanks to ISWT's service. At this stage, ISWT provides all the administrative and legal services required by the local public services and the French electricity provider EDF (declaration of the building work for the PV installation, connection to the grid, etc.). Since 2015, ISWT has progressively deployed its services all over French territory, triggering more than 3500 installations and becoming exclusive partners of more than 80 local authorities in five years. In 2019, ISWT merged with Otovo, a leading PV energy company in Scandinavia.

Our analysis focuses on the first years of ISWT development (from 2015 to 2019), in which O.I.E. and Transvalor played a critical role by providing ISWT with the solar cadasters used as a basis of the ISWT simulation tool (Callegari et al., 2017). These solar cadasters basically involve computing the long-term average solar radiation received by a given 1 m^2 surface in an urban area (e.g., the city of Nantes). The story of how these interactions started was recalled by one of the O.I.E.'s main researchers involved, as follows:

> "**Before working with ISWT, I had been working on solar cadasters for the Provence-Alpes-Côte d'Azur (PACA) region** in France at 200 m of resolution. And I gave a talk at a conference where I said: if we want to go from 200 m to 1 m, we need 3D models. […] Nicolas [one of the founders of ISWT] wanted to create a start-up to support the development of photovoltaics at the urban scale. He called me and […] and I told him that **I had already talked about the possibility of assessing solar radiation at an urban scale at this conference and that I had identified 3D data as the main issue.** So I told him it would be interesting to talk with IGN [French national mapping agency] and that **I could put him in contact with the research director of IGN that I knew**. […] ISWT was then hosted by IGN, providing ISWT with their data for the digital surface model, which IGN initially did not know what to do about." (Philippe)

This story provides a few important hints about the fit system built by O.I.E. in the ISWT case. First, it mirrors well the multifaceted composition of the fit system, involving the design of technical components (e.g., 3D data associated with a digital surface model), socio-organizational components (e.g., building a partnership between ISWT and IGN), and cognitive components (e.g., knowledge on the missing scientific gaps to move from a 200 m resolution to a 1 m resolution). These different components are more exhaustively presented in Table 3 and synthetically represented in Fig. 4 based on the static-view visual mapping template.

This analysis of the fit system in the case of ISWT suggests the existence of generic components. The story told by O.I.E. clearly indicates the use of preexisting components that had been previously built by O.I.E. to fit solar radiation data in other use contexts beyond ISWT, for example, for the PACA region. It is further confirmed by the in-depth analysis of the fit system unveiling components that had been built by O.I.E. over a long period of time, such as long-term solar resource datasets or O.I.E.'s relationship with the company Transvalor. However, considering the case of ISWT alone is not sufficient for understanding how these different components have been successfully built in the long run. To examine these aspects, we need to go back to O.I.E.'s history.

4.2. A dynamic view of the generification mechanisms supporting the design of the fit system between solar radiation data and multiple use contexts

The longitudinal analysis of O.I.E. highlights how the fit system has been designed over time through dedicated generification mechanisms contributing to progressively enriching the generic core of the fit system. To do so, O.I.E. conducted a sequence of (re)embedding and disembedding processes ('what' perspective), enabled by the progressive creation of various boundary resources ('how' perspective), and supported by several forms of drivers ('why' perspective). These different elements are hereafter described in more detail.

4.2.1. 'What' perspective

The sequence of embedding-disembedding processes conducted by O.I.E. is represented in Fig. 6. Embedding occurs when new use-specific

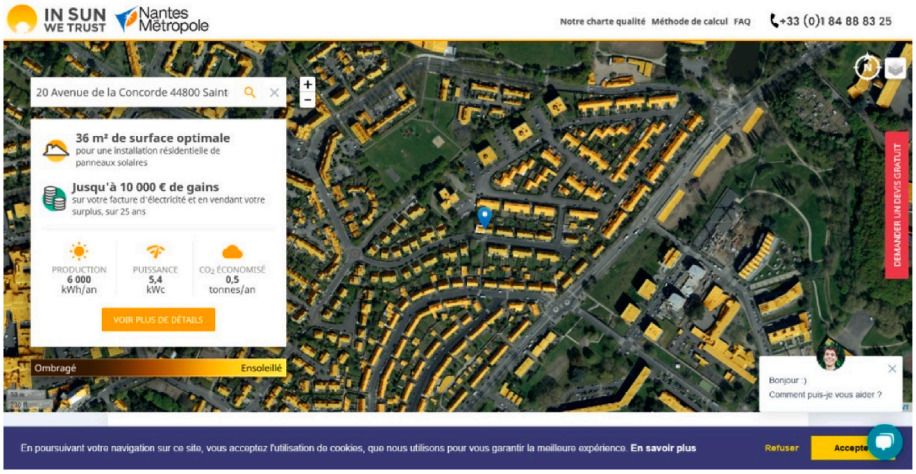

Fig. 3. ISWT simulation tool based on solar cadasters at the urban scale (Source: Blanc and Ménard 2021).

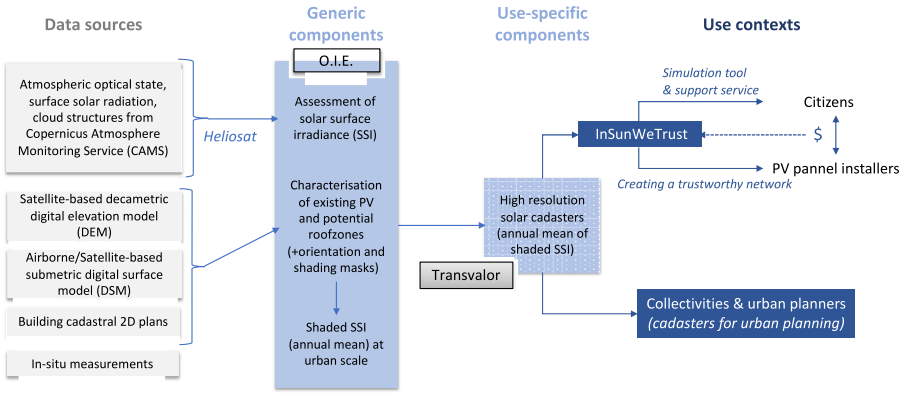

Fig. 4. Static view of the fit system as appearing in the case of ISWT (data sources used as inputs, their transformations into generic and use-specific components of the fit system, and the associated use contexts).

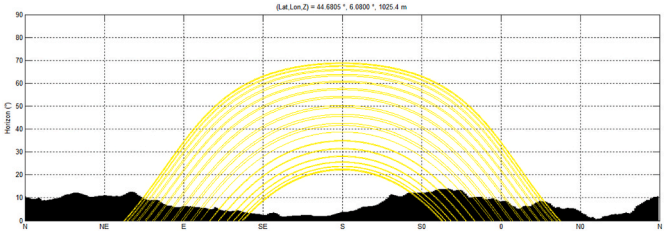

Fig. 5. Example of a horizon (black area) computed from the digital elevation model 'SRTM' at location (44.6805°N, 6.08°E). The daily trajectories of the sun are represented in yellow (three trajectories per month throughout the year). The x-axis is the azimuth orientation, beginning from the north. The y-axis is the elevation angle in degrees (Source: Blanc et al., 2011).

components are created. Disembedding occurs when the generic core is enhanced with new generic components. The following paragraphs go through the depicted transformations of the fit system, starting in 1976.

O.I.E.'s research work on solar radiation started within a project supported by the Solar Energy R&D Programme of the European Commission with the goal of building the first European Solar Radiation Atlas (Grüter et al., 1986). The project aimed to assess solar radiation reaching the ground more precisely and reliably, especially by taking advantage of new data from the Meteosat satellites. O.I.E. started to develop methods to transform Meteosat top-of-atmosphere radiance images into surface solar radiation data, giving birth to the first version of the so-called "Heliosat" methods (Cano et al., 1986), which have been continuously revised to take advantage of technical and scientific advances. The first developments made within the Solar Energy R&D Programme of the European Commission can be referred to as an initial *embedding* process. Indeed, this process contributes to building a fit system between Meteosat-based solar radiation data and a given use context (the European Commission aiming to build a European Solar Radiation Atlas). These developments were later complemented by *disembedding* processes aimed at making these first components useable in other contexts. This is, for example, expressed by O.I.E.'s previous director, who was involved in the early days of Heliosat:

> "The Heliosat principle can be summarized as follows: it consists of detecting an anomaly (e.g., a cloud) with regard to an evolving model of reference (the surroundings). So with the same conceptual framework underlying the Heliosat method [for solar radiation assessment], we made a system for the detection of forest fires. […] And I also had a discussion with motorway-operating companies about traffic monitoring applications […] **It's the Heliosat method that led me to think of these kinds of applications and what we can build on it.**" (Lucien)

This attempt at making solar radiation data useable in multiple contexts took a new dimension with the so-called SoDa project funded by the European Commission from 2000 to 2003. The goal of SoDa was to broaden the use of solar radiation data to new user communities by building "one-stop-shop" easy access based on web-service technologies and specific interfaces (Rigollier et al., 2000). This approach successfully aroused the interest of actors from heterogeneous sectors, such as solar energy, astronomy, air quality, building engineering, climatology, education, health, materials, meteorology, oceanography, agriculture, and agroforestry (Gschwind et al., 2006). These promising results led O.I.E. to dedicate significant efforts to sustaining these services after the

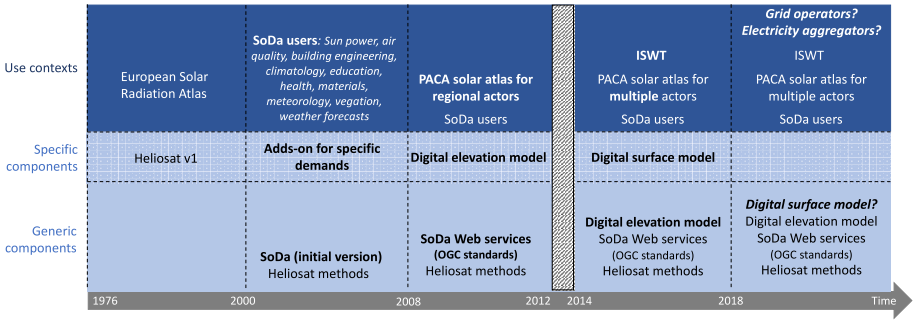

Fig. 6. Summary of the generification mechanisms observed in the case of O.I.E. over time ('what' perspective). The components of the fit system are named based on the technical dimension of the fit system, but the other dimensions (socio-organizational, cognitive) are also closely intertwined (e.g., "SoDa Web services" refers to the technical artifacts but also embodies O.I.E.'s relationship with Transvalor and O.I.E.'s knowledge on solar radiation and energy sectors). The elements written in italic with a question mark correspond to current O.I.E.'s development efforts (thus still ongoing and uncertain). The elements written in bold are the ones that have changed compared to the previous period of time.

Table 3
Composition of the fit system enabling solar radiation data to fit into the use context of ISWT.

	Components of the fit system built by O.I.E. underpinning the case of ISWT
Technical	***Long-term solar resource datasets***, based on the Helioclim-3 database or the CAMS radiation database of solar irradiation values (i.e., the amount of energy received per unit area during a given duration – measured in Wh/m2), with a 4–5 km spatial resolution estimated from data provided by the 'Meteosat Second Generation' satellite. This database is built and updated based on the Heliosat-2 algorithms developed by O.I.E. (Rigollier et al., 2004). The satellite-based dataset can also be calibrated with local in-situ measurements from meteorological stations. ***Downscaling algorithms*** to enhance spatial resolution from 5 km up to a few meters (required when considering roofs at urban scale). This especially requires taking into account the spatial and temporal variability of irradiation values due to the altitude, the local orientation, and the shadowing effect of the horizon caused by the surrounding orography (e.g. mountains but also buildings, vegetation and roof superstructure in an urban area). To calculate these so-called "shadow masks" (see Fig. 5), two models were used: a decametric resolution **digital elevation model (DEM)** describing the natural orography (mountains, hills, valley etc.), and a high-accuracy decimetric **digital surface model (DSM)** giving 3D description of buildings, vegetation and superstructures (provided by the IGN French national mapping agency). ***Roof location algorithms***, based on another dataset provided by IGN (high-accuracy map of building footprints to provide location and contours of corresponding roofs).
Organizational	***Relationship with Transvalor***: O.I.E. is in charge of designing the algorithms. Transvalor provides the Helioclim-3 database on an operational basis for a large range of actors and is in charge of operating the processing chain providing a city-scale solar cadaster as a final output based on the technical components, on an operational and repeated basis following the deployment of ISWT spreading to the whole French territory. ***Relationship with IGN and ISWT***: IGN supports ISWT and provides them with data. O.I.E. provides their technical expertise on how to process these data.
Cognitive	Knowledge of O.I.E. on the ***solar energy sector*** and on the ***data sources and processing techniques***, based on its long experience in the field.

official end of the project, thus officially creating the SoDa Service in 2003. O.I.E. ensured the maintenance of SoDa Service until it was later transferred to a commercial company in 2009. The creation of SoDa Service involved conducting parallel embedding and disembedding processes: *embedding* by making modifications to better fit with the contexts of the project's participants, as well as *disembedding* efforts. Indeed, as highlighted again by the previous director of O.I.E., O.I.E. intentionally avoided designing a system that would exclusively focus on a single sector:

> "It was easy to go from one application to another, because SoDa was not designed for the solar energy sector, but for solar radiation assessment." (Lucien)

A new sequence of embedding-disembedding processes was later encouraged by a new turn in O.I.E.'s history, when O.I.E. decided to entrust the SoDa Service to the company Transvalor in April 2009. Indeed, the growing number of user demands led O.I.E. to find external support for the operation, maintenance and commercialization of the services. Thanks to the partnership with Transvalor, SoDa could be provided to users reliably 24/7. The sustainability of these activities was ensured through a freemium business model involving the provision of free basic solar radiation data at a limited spatial and temporal resolution and more sophisticated paid-for services. Through the operationalization and commercialization of SoDa, the fit system has been enriched to further enable solar radiation data to address diverse use contexts. The fit system has sometimes been adapted further to specific use contexts through slight adaptations of the service to the requests of users (*embedding*). However, the fit system has also been made increasingly generic by using these requests to identify the possible transformations of SoDa that seem to be the most promising for several use contexts (*disembedding*). This disembedding process has also involved the complete redesign of SoDa to make it compliant with the standards recommended by the Open Geospatial Consortium (Thomas et al., 2013). These efforts have allowed the fit system to become interoperable with other products or systems and compatible with a wider range of use contexts.

In parallel with these advances related to SoDa operationalization, another sequence of embedding and disembedding processes can be identified with regard to the developments of solar cadasters. Indeed, downscaling algorithms leveraging a digital elevation model were developed by O.I.E. before partnering with ISWT in the context of building a solar cadaster for the PACA region (Blanc et al., 2011a; Ménard et al., 2013). These developments were associated with an *embedding* process, as they were initially triggered and funded by local and regional organizations and councils from 2008 to 2012. However, beyond this initial demand, O.I.E. later invested significant efforts in *disembedding* these developments by making these solar cadasters useable in a larger range of contexts, e.g., by enriching them with additional data (e.g., the distance to the nearest electric grid line) and integrating them into standard-compliant webservices, especially in the framework of an FP7 European project called ENDORSE. With the case of ISWT, a new phase of *embedding* unfolded, especially through the additional integration of a high-accuracy digital surface model to enhance the resolution of solar cadasters and make them relevant at the urban scale. The case of ISWT is not the end of the story, and new attempts at further *disembedding* ISWT-related developments can already be noted. Indeed, O.I.E. is currently exploring how these urban solar cadasters could potentially fit into a larger pool of use contexts, e.g., grid operators managing the increasing integration of PV in the grid.

4.2.2. How perspective

The analysis of O.I.E.'s history also shows that the generification mechanisms have been especially encouraged through the development of specific forms of *boundary resources*. These boundary sources are listed and further described in Table 4 (the "resourcing" and "securing" functions are further commented on in the discussion section). One can note that some of the boundary resources developed by O.I.E. are not only targeted at the potential complementors of the system (here, users of data). Indeed, some of them are aimed at supporting "peer developers", which are taking on similar endeavors of building fit systems between EO data and various use contexts in the solar energy sector or in different areas. Moreover, only one kind of boundary resources concerns the producers of data (here, operators of in-situ instruments). This can be explained by O.I.E.'s downstream position compared to these actors and, as a result, their limited ability to trigger transformations on this side.

4.2.3. Why perspective

Throughout its history, O.I.E.'s generification efforts have been fostered by a few noticeable drivers. A first driver lies in the *logic of science* underpinning O.I.E.'s activities as a research center since its creation, as explicitly stated by several interviewed members of O.I.E.:

> "As researchers, our role is to **crystallize basic building blocks at the highest level** and not to develop components that are too much oriented toward a specific domain.**"** (Lionel)

> "At the creation of our lab, **we benefited from computation means that were quite unique in France and even Europe**. Many scientists from various backgrounds used to come visit us to process their images based on the tools that had been installed here. […] So, **right from the beginning, we were working in a multidomain atmosphere:** geology, forestry, agriculture, meteorology, oceanography, and data processing. […] So **all the developments we made**

Table 4
Summary of boundary resources developed by O.I.E. over time ('how' perspective).

'How' perspective: *focus on boundary resources as means of supporting generification mechanisms*					
Starting date	Boundary resources	Targeted third-party actors	Resourcing function		Securing function
			For third-party actors	For the fit system owner	
2004	**Licenses granting access to Heliosat and derived methods** Licenses allowing users to freely use, share and adapt these products and tools, provided that they give credit to the licenses (Creative Commons or equivalent in the EO community, e.g. GEOSS Data Core Sharing principles)	Users&Peer developers	Allows users and peer developers to innovate on top of Heliosat methods.	-	Allows the fit system owner to increase its renown and legitimacy
2006	**Lead of a community of practice on renewable energies in the intergovernmental body Group on Earth Observations (GEO)** Aiming at sharing best practices among players aiming to bridge EO and energy sectors (later transformed in the GEO-VENER initiative)	Peer developers	Allows peer developers to benefit from the fit system owner's experience (sharing difficulties and best practices).	Allows the fit system owner to keep abreast of current practices, new developments and future trends.	Allows the fit system owner to foster the evolution of the ecosystem into certain directions.
2006	**Contribution to the International Energy Agency's implementing agreements related to solar energy** Supporting the emergence of a common vision of what defines the quality of solar resource assessment and forecasting methods or developing standardized and integrating procedures for data bankability. Currently participating to the PVPS program (Task 16) that aims to lower barriers and costs of grid integration of PV and lowering planning and investment costs for PV by enhancing the quality of the forecasts and the resources assessments.	Users & Peer developers	Allows users peer developers to benefit from the fit system owner 's experience (sharing difficulties and best practices).	Allows the fit system owner to keep abreast of current practices, new developments and future trends.	Allows the fit system owner to foster the evolution of the ecosystem into certain directions.
2008	**Webservice-energy SDI (Spatial Data Infrastructure)** Created by O.I.E. following recommendations from the GEO Standard and Interoperability Forum (SIF). The SDI aims at gathering, promoting and spreading EO data for the development of renewable energies based on open standards. The SDI ensures the access to these resources to potential third-party users based on open standards. However, it also plays a facilitating role for other peer developers of the EO community, as it gathers the services and products developed by several tens of organizations.	Users & Peer developers	Allows users and peer developers to innovate on top of existing EO data.	-	Allows the fit system owner to increase its renown and legitimacy
2009	**Helpdesk for SoDa service** Annual training session on the basics of solar radiation that has been organized for 30–40 international participants every year since 2013.	Users	Allows potential users to learn about solar radiation data and further leverage them.	Allows the fit system owner to better understand use contexts and identify the future development paths that seem the most promising.	
2013	**Solar Training** A training session on the basics of solar radiation organized for 30–40 international participants every year.	Users	Allows potential users to learn about solar radiation data and further leverage them.	Allows the fit system owner to test some of its assumptions related to the future development paths that seem the most promising.	
2019	**Jupyter Notebooks** Offers the possibility to play with parts of codes and visualize them through images or graphs. Initially, implemented by O.I.E. in the EU-funded project 'NextGEOSS', these notebooks have been further developed by O.I.E. as a way of supporting their interactions with potential users, serving as demonstration tools to explore new use cases based on the temporal variability of solar radiation (Blanc and Ménard, 2021).	Users	Allows potential users to get a better understanding of possible ways of using solar radiation data.	Allows the fit system owner to learn from users, by analyzing how the latter use the Jupyter Notebook and make it evolve.	
2022	**Data Management Plan self-assessment tool** Tool developed by O.I.E. within the EU-funded 'e-shape' project (Ranchin et al., 2021) to help data-based product or service developers to assess their current status and trajectory toward compliance with data sharing principles supporting standards.	Peer developers	Allows peer developer to progress in their implementation of data sharing principles.	-	Allows the fit system owner to foster the evolution of the ecosystem into certain directions.
2022	**Web application to visualize the homogeneity of data provided by the network of in-situ measurements**	Peer developers	Provides peer developers with an easier access to in-situ measurements encoded in	-	Allows the fit system owner to foster the evolution of the

(*continued on next page*)

Table 4 (*continued*)

'How' perspective: *focus on boundary resources as means of supporting generification mechanisms*					
Starting date	Boundary resources	Targeted third-party actors	Resourcing function		Securing function
			For third-party actors	For the fit system owner	
		& Data producers	standard forms. Allows in-situ data producers to understand the benefits of standards for in-situ data.		ecosystem into certain directions.
	Developed in order to facilitate the access and the use of in-situ measurements by other data-based product or service developers, in compliance with standards but also compatible with other nonstandard formats still used in the field (Blanc et al., 2022). Additionally, intended to be used by O.I.E. as a demonstration of what can be done with in-situ data encoded in standards, and potentially further convince the providers of in-situ data to better comply with standards.				
2023	**"libinsitu" an open source python library (under BSD licence imposing minimal restrictions on the use and distribution)** The library allows the transformation of heterogeneous time series of in-situ measurements into standard and interoperable well-formed binary file.	Peer developers	Provides peer developers with an easier way of encoding in-situ measurements in standard forms.	-	Allows the fit system owner to foster the evolution of the ecosystem into certain directions.

tended to be quite generic; we needed to conceptualize a lot of things." (Lucien)

This science-driven logic also appears in O.I.E.'s efforts in supporting peer developers, which can even be O.I.E.'s competitors in some cases. Indeed, as emphasized by the previous O.I.E. director:

> "A lot of methods currently built by our competitors are based on Heliosat or derivative versions. […] And **I initially educated many of those people who created competing databases**." (Lucien)

Generification has also been enabled by a few *organizational drivers*, especially O.I.E.'s capacity to identify the high potential of genericity of solar radiation data. In particular, this capacity involves ensuring a continuous watch on potential data uses and available technical means such as data production instruments, algorithms and IT capacities:

> "I constantly have a double view. A methodological view that is: '**given these data, how to take advantage of them?**' And another view that is: **'what are the available data sources that I could use to further fill in the gaps** I currently have in my methods and that could improve the service?'" (Philippe)

Moreover, *institutional factors* have also played an important role, especially the funders of the projects through which O.I.E. has progressively developed its fit system (mainly the European Commission), as well as international organizations such as the Group on Earth Observations (GEO). The strategy followed by O.I.E. has been strongly supported by these institutional actors, pushing for the development of a wide range of applications for EO data in various sectors, as well as the adoption of standards, as expressed by interviewees:

> "**When we joined the Group on Earth Observations (GEO), we became aware of the powerful role of standards**. Even if you merely comply with them and do not necessarily develop them, they have a multiplier effect on your capacity for action in the community." (Lucien)

This last verbatim also points to a *strategic driver*. Indeed, O.I.E. clearly acknowledges that the choice of complying with standards was partly driven by gaining a form of strategic advantage. This choice contributes to the generification of the fit system in two ways. From a technical perspective, the fit system becomes compatible with a wider range of use contexts. Moreover, from an organizational perspective, complying with standards increases O.I.E.'s legitimacy in the field and its ability to forge valuable partnerships, thus opening up new opportunities to further develop the fit system, as also highlighted in the following verbatim:

> "Standards have enabled small players like us to be part of big initiatives. They have played a crucial enabling role in creating new partnerships." (Lionel)

5. Discussion

The longitudinal analysis of O.I.E. have allowed us to identify the generification mechanisms that have underpinned the design of the fit system over time. This section discusses the similitudes and peculiarities of the patterns observed in the case of O.I.E. compared to the ones described by the literature for other forms of generic systems (e.g., digital platforms, software systems). These patterns are commented on considering the 'what', 'how', and 'why' perspectives on generification mechanisms subsequently. They are synthesized in Fig. 7.

5.1. *What are the observed forms of generification mechanisms? Disembedding as more closely intertwined with embedding*

The analysis of O.I.E. confirms the relevance of generification mechanisms in the case of data-push innovation. For other forms of generic systems studied in the literature, generification mechanisms involve both embedding and disembedding processes. The underlying rationale lies in the enrichment of the generic core, avoiding building components that are too oriented toward a specific domain. This directly resonates with previous studies on open ERP systems, in which the developers of the global generic system aim to "stay away from directly designing for particular local needs" (Gizaw et al., 2017, p. 635). The same authors also highlight that "although important, the ultimate goal of open generification is not to make a software work in a particular context; it is rather to take the working solution further to multiple other contexts" (p. 637).

However, a few peculiarities can be noticed in the case of O.I.E. compared to the literature on generification. As expected from our literature analysis, O.I.E. has no choice but regularly undertake significant embedding processes, e.g., solar cadasters for public authorities of a

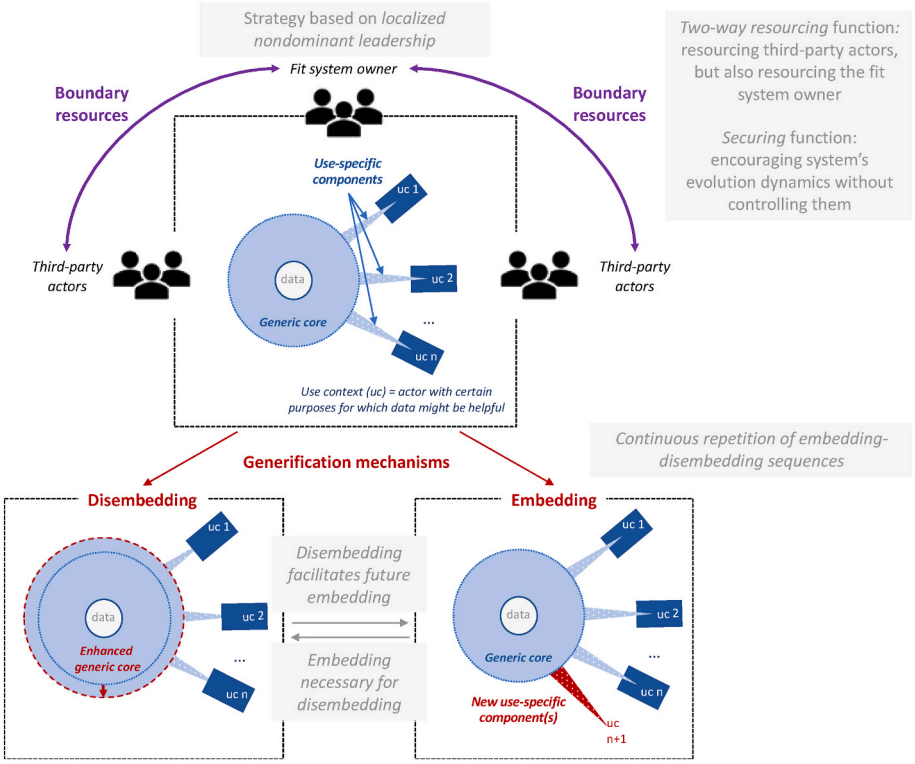

Fig. 7. Summary of the main outcomes (in gray boxes) related to what forms of generification mechanisms are involved in building the fit system, how these mechanisms can be supported by specific forms of boundary resources, and why the fit system owner is led to support such forms of generification mechanisms.

given region, then for ISWT at a finer resolution. This stems from the need to take into account the initial limited capacities of potential users to leverage data on their own. Therefore, although the ultimate goal lies in enriching the generic core of the fit system, O.I.E. also needs to design the necessary use-specific components until the point data can fit into a given use context or, in other words, until the point data fit with actors' capacities to further leverage them. However, in the case of O.I.E., embedding is not only driven by the mere objective of better serving the specificities of local use contexts to avoid a risk of 'misfit', as often suggested in the literature (e.g. Li and Nielsen, 2019). O.I.E.'s history reveals that embedding processes are always followed by disembedding processes. More precisely, use-specific components are designed by O.I.E. if they offer promising ways to further enrich the generic core. In other words, O.I.E. focuses on designing use-specific components that are deemed to have generic potential. For example, the case of ISWT gives the opportunity to develop new components (e.g., digital surface model) that are thought of by O.I.E. as having the potential to serve multiple other use contexts (e.g., grid operators). In the same vein, the concept of Heliosat algorithms was initially developed for building the European Solar Radiation Atlas but was also used by O.I.E. to explore a range of other use contexts, such as forest fire detection. This aspect seems quite original compared to usual considerations in generic software development, in which the adaptations made to fit the system into a given context at an implementation level do not directly contribute to building up the genericity of the system (e.g. Li and Nielsen, 2019). Taking Li and Nielsen's (2019) classification, the case of O.I.E. invites us to consider that, in a context of data-push innovation, the implementation-level design (corresponding to embedding processes) not only follows a 'design for use' approach, aimed at making the system fit into a given context but also contributes to a 'design for design' objective, aimed at building the infrastructure allowing further design at a later stage. In other words, *embedding processes appear to lay the foundation for future disembedding processes*.

This observation could seem paradoxical at first sight. Indeed, one could wonder why O.I.E. needs to go through heavy and regular embedding processes while its primary objective is driven toward genericity and would thus require a focus on disembedding. This peculiarity can be linked to the specificities of data-push innovation compared to other forms of digital innovation, stemming from specific properties of data compared to other digital artifacts. Indeed, the meaning and value of data for a specific use context are not given beforehand (e.g. Alaimo et al., 2020). It cannot be given solely by the fit system owner or by the potential users alone. Returning to the case analyzed by Alaimo et al. (2020), the telecommunications operator cannot imagine how its data can be meaningful and valuable for the advertising company without having strong interactions. Nor can the advertising company imagine by itself that there might be data owned by the telecommunication operation that support their business. Therefore, from the perspective of the fit system owner, *the generic components that will enable data to gain meaning in multiple use contexts are unknown from the outset.* For example, in the case of O.I.E., it is difficult to predict beforehand whether the digital surface model underpinning the development of urban solar cadasters should be made as a generic component. O.I.E. can assume the potential of genericity of such a component, but this assumption can only be later validated if the digital surface model also proves useful for other use contexts. The generic components are designed as such based on the fit system owner's understanding of the current range of potential use contexts. This explains why embedding and disembedding processes appear so closely intertwined in the case of O.I.E., especially in the following way: *disembedding cannot occur independently from embedding.* Alternatively, embedding can occur without disembedding as a necessary precondition. However, over time, the case of O.I.E. also shows that embedding is largely facilitated by previous efforts of disembedding. Indeed, disembedding results in building generic components that can be later reused (e.g., Heliosat algorithms, digital surface model, partnership with Transvalor), thus making embedding processes less cumbersome for a new use context.

5.2. How can generification mechanisms be supported with specific forms of boundary resources? A noticeable two-way resourcing function

To a certain extent, our results align with recent studies showing the need of a sophisticated range of boundary resources targeted at different types of actors of the platform ecosystem (e.g. Bonina and Eaton, 2020). In line with Foerderer et al. (2019), boundary resources can be seen as a way of bridging knowledge boundaries between the fit system owner and third-party actors. The case of O.I.E. confirms the latter authors' conclusions, showing that bridging knowledge boundaries does not only involve "providing complementors with access to a joint technological infrastructure, but [it] rather requires investments in cooperative relationships with complementors" (p. 140).

However, a few original patterns can also be noted. As exposed above, building a generic fit system involves continuously exploring multiple use contexts and undertaking numerous sequences of embedding-disembedding processes over time. This aspect has direct consequences on the functions of the boundary resources allowing the fit system owner to support such generification efforts. Indeed, there is a noticeable impact on the *resourcing* function of the boundary resources, which usually designates "the process by which the scope and diversity of a platform is enhanced" (Ghazawneh and Henfridsson, 2013, p. 177), especially by attracting third-party actors to join the development of the system and encouraging their innovation (Eaton et al., 2015). Noticeably, in the case of O.I.E., the resourcing function of some boundary resources unfolds in a double direction: *resourcing the third-party actors of the ecosystem*, as is usually the case (e.g. Foerderer et al., 2019) but also *resourcing the fit system owner itself*. Taking the example of O.I.E., this two-way resourcing function appears quite clearly for boundary resources that are specifically aimed at creating closer interactions with the potential users of data. It includes a helpdesk for the SoDa Service created in 2009, an annual training session on the basics of solar radiation organized every year since 2013, as well as Jupyter Notebooks used as demonstration tools since 2019. On the one hand, these boundary resources allow potential users to learn about solar radiation data and how to further innovate on top of the fit system. However, on the other hand, these boundary resources also allow O.I.E. to learn about the sectors in which these potential users evolve, their practices and possible requests. Based on this understanding, O.I.E can assess the most promising development paths, especially by identifying which use-specific and generic components are worth building.

This original form of boundary resources can be again explained by the unknown character of the meaning of data in a given new use context at the outset (e.g. Aaltonen et al., 2021; Mikalsen and Monteiro, 2021; Alaimo et al., 2020). Indeed, in the process of building the meaning of data for a given use context, both the fit system owner and the third-party actors need knowledge from the other side: the fit system owner needs to understand elements related to the use context (e.g., potential users' practices, rules and regulations in the sector), while the potential users need to obtain an understanding of what they could do with these new data. However, neither the fit system owner nor the third-party actors are able to exactly pinpoint the types of knowledge missing on both sides before they interact. In this respect, not only is the meaning of data initially unknown, but the knowledge boundaries between the fit system owner and the potential users are also largely unknown from the outset. This aspect offers an extension to the results of Foerderer et al. (2019): not only do the boundary resources contribute to bridging knowledge boundaries, but they also play an important role in *identifying these knowledge boundaries*.

Other interesting patterns can be noted regarding the *securing* function of the boundary resources. Usually, the securing function is associated with "the process by which the control of a platform and its related services is increased" (Ghazawneh and Henfridsson, 2013, p. 177). In the case of open data platforms, Bonina and Eaton (2020) depict securing as a way of regulating how the different members of the ecosystem carry out their respective tasks contributing to platform evolution. By contrast, the position of O.I.E. does not allow it to exert such a form of control. Indeed, O.I.E. has limited control over the players involved in producing data (e.g., big players from the space industry launching satellites) or using data (e.g., large industrial energy companies or public agencies involved in energy transportation and distribution). In such conditions, the securing function of boundary resources cannot consist of purely controlling the evolution dynamics of the system but rather consists of fostering certain forms of evolution dynamics. In the case of O.I.E., it entails at least two complementary aspects. The first consists of *increasing the fit system owner's legitimacy* in the ecosystem. For example, O.I.E. has allowed the reuse of Heliosat algorithms through Creative Commons licenses or equivalent and has also created a Spatial Data Infrastructure (Webservice-energy) to gather, promote and spread EO data for the development of renewable energies based on open standards. By making the outcomes of O.I.E.'s development efforts widely available, these boundary resources contribute to making O.I.E.'s recognition as an active player in the ecosystem. The second securing aspect consists of *encouraging* some activities of the actors of the ecosystem without directly controlling them, e.g., encouraging peer developers toward further compliance with standards. This aspect is illustrated by boundary resources such as O.I.E.'s participation in communities of practices, as well as the "libinsitu" open source python library, allowing peer developers to transform heterogeneous time series of in-situ measurements into standard and interoperable data files.

5.3. Why support such forms of generification mechanisms? A strategy relying on a form of localized and nondominant leadership

These patterns related to the generification mechanisms and boundary resources, as well as the drivers of generification mentioned in the findings, contribute to depicting the strategic position of the fit system owner within its ecosystem. It especially appears that O.I.E. has developed an intriguing leadership position that differs from historical cases of platform leadership described for technological platforms (Gawer and Cusumano, 2002) or well-known success stories of digital platforms such as Apple's iOS operating system (Eaton et al., 2015; Ghazawneh and Henfridsson, 2013).

Considering the ecosystems of actors in which solar radiation data are introduced, O.I.E. cannot be considered a hub firm with the ability to orchestrate overall ecosystem dynamics (e.g. Dhanaraj and Parkhe, 2006). However, O.I.E. can still be considered as having a certain form of *localized leadership* when leadership is considered to be "the exertion of influence in order to 'make things happen' […] despite a lack of formal authority" (Müller-Seitz, 2012, p. 429). Indeed, the analysis reveals that O.I.E. has been able to punctually implement "localized" actions of stimulation, consisting of stimulating some small-scale transformations with pinpointed actors of the ecosystem. For example, from the perspective of estimating solar radiation at an increasingly fine-grained resolution, O.I.E. was the one building connections between the founders of ISWT and the French national mapping agency that had key datasets. This localized leadership is directly rooted in the fit system built by O.I.E. over the years. It is only made possible because O.I.E. had already accumulated considerable experience in solar radiation estimation and could thus leverage existing components of the fit system.

Moreover, the leadership form developed by O.I.E. could be characterized as a form of *nondominant leadership* that focuses more on gaining generative power rather than controlling power. Both generativity and control have been extensively described as fundamental characteristics of digital platforms (Yoo et al., 2010). For a platform owner, gaining controlling power basically consists of strengthening its capacity to orchestrate the overall ecosystem around the platform, whereas gaining generative power involves strengthening its capacity to open up new innovation spaces for others. In the case of O.I.E., in addition to the focused actions of stimulation described above, this

nondominant logic especially appears in their use of standards and the distribution of captured value. Indeed, regarding standards, the IS literature has already shown how firms could establish an exclusive and powerful position in the ecosystem by influencing or imposing some forms of standards (Lyytinen and King, 2006; Yoo et al., 2005). In the case of O.I.E., the use of standards unfolds following a different logic. Indeed, O.I.E. does not aim to gain dominating control over the ecosystem based on these standards. O.I.E. has chosen to comply with existing standards recommended by authoritative agencies (e.g., OGC) in order to increase the genericity of the fit system. But, echoing other studies showing how generification can support standardization (Hanseth and Bygstad, 2015), O.I.E. also contributes to fostering the development and implementation of these standards among the different actors of the ecosystem to further enhance generativity. Moreover, O.I.E. does not aim to appropriate all the value generated by the development of the fit system. Instead, O.I.E. shares this value with a network of partners responsible for maintaining some building blocks of the fit system (e.g., Transvalor). This echoes the observation made by other scholars, who have argued that "by proactively investing in seeding an ecosystem through developing and cultivating new business partners while not trying to appropriate all the value allows a firm to have sustainable access to a greater variety of capabilities and resources" (Zeng and Glaister, 2018, p. 131).

6. Conclusion

The question of building generic systems serving a large variety of use contexts is not new. It has already been investigated by a long tradition of research on digital platforms, infrastructures or software systems (e.g. Foerderer et al., 2019; de Reuver et al., 2018; Gizaw et al., 2017; Eaton et al., 2015; Silsand and Ellingsen, 2014; Monteiro et al., 2013; Tilson et al., 2010; Pollock et al., 2007). In this context, Gizaw et al. (2017) raised the following issue as particularly critical: "To what extent can software 'travel' to organizations and countries for which it was not designed for?" (Gizaw et al., 2017). Similar questions hold today for data, particularly in current developments around data-push innovation: to what extent can data 'travel' to contexts across organizations and sectors for which they were not designed? Concrete examples of successful cases driven by companies such as Uber confirm the possibility of data to travel but also point to the difficulty of systematizing such approaches beyond serendipity (e.g. Trabucchi and Buganza, 2020; Trabucchi et al., 2018). As highlighted by a growing body of IS research on the specificities of data as digital artifacts (e.g. Aaltonen et al., 2021; Alaimo et al., 2020; Monteiro and Parmiggiani, 2019), specific issues stem from the fact that data cannot travel as mere software modules. These authors show that, beyond the recombinant logic that usually underpins the development of digital platforms or generic systems, data can travel across contexts provided that their meanings are built and rebuilt for the new contexts in which they are used.

Our paper aims to further illuminate what it means for data to travel, not only from an initial context of production to a new use context but also across multiple use contexts over time. This research is specifically based on the case study analysis of an actor (called O.I.E.) that has succeeded in stimulating the use of Earth observation data in multiple use contexts for more than 40 years. The paper particularly contributes to IS and innovation management research by providing a better understanding of how data-push innovation strategies can be managed beyond serendipity. The paper does not claim to have exhaustively unveiled all possible strategies but has elucidated a specific strategy that could be characterized as "data genericity building", which consists of intentionally designing a generic "fit system", comprising a set of technical, socio-organizational and cognitive components enabling data to be used in multiple contexts. This is consistent with other well-developed studies in innovation management on generic or general-purpose technologies, which are characterized by their "potential for pervasive use in a wide range of sectors and by their technological dynamism" (Bresnahan and Trajtenberg, 1995). Scholars have indeed shown that designing these generic technologies involves "intentionally designing common features that bridge the gap between a priori heterogeneous applications and technologies", rather than only relying on a trial-and-error approach where common features are randomly discovered (Hooge et al., 2016). This does not imply that serendipity cannot occur; however, a genericity-building strategy is all the more successful, as it is managed through a dedicated and thought-through strategy (Hooge et al., 2016; Kokshagina et al., 2016).

These results call for more precisely conceptualizing the "genericity" of data. Genericity can be fundamentally understood as the ability of data to fit into multiple use contexts. This paper invites us to consider it not as an intrinsic property of data but rather as a property resulting from the design of the fit system. From this perspective, data are considered all the more generic, as the residual design effort to fit data into a new use context is low, or, in other words, as the fit system comprises a large number of generic components that are common to different use contexts. One might think that it is simply a matter of finding the optimal components that are worth being made generic once and for all. However, the paper indicates that genericity should not be conflated with acontextuality. In other words, generic components are not designed independently from contexts. Generic components are progressively designed by understanding the multiplicity of contexts in which data could be used. As such, genericity is not a fixed property but a result of generification mechanisms that need to be continuously supported and adapted to take into account new arising contexts. The case of O.I.E. shows that these mechanisms can take original forms, compared to other well-known cases of information systems development, which is intrinsically related to the meaning construction process underpinning data-push innovation: the fit system is what supports the process of continuously building and (re)building the meaning of data in order for them to travel from one context to another. Continuous generification efforts allow data to progressively become more generic by being used in an increasing number of contexts. These results have important implications for scholars: to better understand how the high genericity potential of data can be unleashed, it is critical to pay close attention to the different forms of processes, competencies, and conditions required to sustain generification efforts over time.

These academic contributions also offer new insights for practitioners willing to implement data-push innovation strategies. Our results especially suggest that data-push innovation can be supported by specific forms of strategies to build generic fit systems. The paper gives concrete examples of some managerial practices involved in supporting such a process, especially elucidating a rich variety of boundary resources with original functions compared to the ones initially described for digital platforms. This paper also describes an original form of localized and nondominant leadership, which could inspire practitioners that have limited power in a given ecosystem but are willing to contribute to the dynamics of this ecosystem.

Some limitations are worth noting, leading to several research perspectives. Our paper has delved into the single case study case of O.I.E., which unveils an interesting strategy to support data-push innovation. Several factors especially played an important role in fostering generification in the case of O.I.E., such as the science-based nature of its activities and the favorable institutional context. One could wonder whether similar strategies could be supported in different conditions. Other types of strategies could also exist, and they deserve further investigation. For example, the literature on open data indicates the existence of a range of business models and actors involved in bridging the gap between data and use contexts (Magalhaes and Roseira, 2020; Janssen and Zuiderwijk, 2014). In this regard, recent projects in Europe to stimulate the use of EO data in multiple sectors suggest that diverse forms of actors might be concerned with similar strategies (Ranchin et al., 2021); these include private companies, public authorities, and meteorological institutes that already have a long history in developing EO-based weather forecasts and their applications (Lenfle and

Söderlund, 2022). It will thus be interesting to compare the different strategies, mechanisms and organizational characteristics that might exist depending on the profiles of these actors, as well as the conditions of their emergence and viability. Another research perspective lies in the temporal dimension of generification. The paper has mainly highlighted the dynamics of the fit system related to the transformations of use contexts. However, drastic changes of available data (e.g., following the launch of new satellites) are also likely to impact the design of the fit system. Our study could be extended to further analyze whether and how generification could also be relevant to cope with significant and repeated transformations of data over time.

CRediT authorship contribution statement

Raphaëlle Barbier: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Skander Ben Yahia:** Investigation, Data curation. **Sylvain Lenfle:** Supervision, Conceptualization. **Benoit Weil:** Validation, Supervision, Investigation, Conceptualization.

Data availability

Data will be made available on request.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 820852. The authors also warmly thank the whole team of O.I.E. and Transvalor for their involvement and our fruitful interactions, in particular P. Blanc, L. Ménard, T. Ranchin, and L. Wald, who have provided stimulating insights that have strongly contributed to this work.

References

Aaltonen, A., Alaimo, C., Kallinikos, J., 2021. The making of data Commodities: data analytics as an embedded process. J. Manag. Inf. Syst. 38, 401–429. https://doi.org/10.1080/07421222.2021.1912928.

Aaltonen, A., Tempini, N., 2014. Everything counts in large amounts: a critical realist case study on data-based production. J. Inf. Technol. 29, 97–110. https://doi.org/10.1057/jit.2013.29.

Abbasi, A., Sarker, S., Chiang, R.H.L., 2016. Big data research in information systems: toward an inclusive research agenda. J. AIS 17, 3. https://doi.org/10.17705/1jais.00423.

Alaimo, C., Kallinikos, J., 2020. Managing by data: algorithmic categories and organizing. Organ. Stud., 017084062093406 https://doi.org/10.1177/0170840620934062.

Alaimo, C., Kallinikos, J., Aaltonen, A., 2020. Data and value. In: Handbook of Digital Innovation. Edward Elgar Publishing, Cheltenham, UK.

Appio, F.P., Frattini, F., Petruzzelli, A.M., Neirotti, P., 2021. Digital transformation and innovation management: a Synthesis of existing research and an agenda for future studies. J. Prod. Innovat. Manag. 38, 4–20. https://doi.org/10.1111/jpim.12562.

Baldwin, C.Y., Clark, K.B., 2000. Design Rules: the Power of Modularity. MIT Press.

Baldwin, C.Y., Woodard, C.J., 2009. The architecture of platforms: a unified view. In: Platforms, Markets and Innovation. Edward Elgar, Cheltenham, pp. 19–44.

Bharadwaj, N., Noble, C., 2017. Finding innovation in data rich environments. J. Prod. Innovat. Manag. 34, 560–564. https://doi.org/10.1111/jpim.12407.

Blackburn, M., Alexander, J., Legan, J.D., Klabjan, D., 2017. Big data and the future of R&D management. Res. Technol. Manag. 60, 43–51. https://doi.org/10.1080/08956308.2017.1348135.

Blanc, P., Espinar, B., Gschwind, B., Ménard, L., Thomas, C., Wald, L., 2011a. High spatial resolution solar atlas in Provence-Alpes-Cote d'Azur. In: ISES Solar World Congress 2011. Kassel, Germany, 34552.

Blanc, P., Gschwind, B., Lefèvre, M., Wald, L., 2011b. The HelioClim project: surface solar irradiance data for climate applications. Rem. Sens. 3, 343–361. https://doi.org/10.3390/rs3020343.

Blanc, P., Ménard, L., 2021. Au-delà des cadastres solaires pour le développement du photovoltaïque urbain. www.construction21.org.

Blanc, P., Wald, L., 2015. L'estimation du rayonnement solaire au sol par la nouvelle méthode Heliosat-4. Météorologie 8, 53. https://doi.org/10.4267/2042/56839.

Bonina, C., Eaton, B., 2020. Cultivating open government data platform ecosystems through governance: lessons from Buenos Aires, Mexico City and Montevideo. Govern. Inf. Q. 37, 101479 https://doi.org/10.1016/j.giq.2020.101479.

Bonina, C., Koskinen, K., Eaton, B., Gawer, A., 2021. Digital platforms for development: foundations and research agenda. Inf. Syst. J. 31, 869–902. https://doi.org/10.1111/isj.12326.

Borzacchiello, M.T., Craglia, M., 2012. The impact on innovation of open access to spatial environmental information: a research strategy. Int. J. Technol. Manag. 60, 114. https://doi.org/10.1504/IJTM.2012.049109.

Bresnahan, T.F., Trajtenberg, M., 1995. General purpose technologies 'Engines of growth? J. Econom. 65, 83–108.

Callegari, D., Bodereau, N., Bourgeon, M., Ebel, A., Dion, E., Force, M., Séguin, E., Wey, E., Saboret, L., Gschwind, B., Blanc, P., 2017. Solar cadaster of Nantes metropole based on high resolution solar mapping at urban scale from 10 cm digital surface model for rooftop PV development. In: ICEM 2017. Bari, Italy.

Cano, D., Monget, J.-M., Albuisson, M., Guillard, H., Regas, N., Wald, L., 1986. A method for the determination of the global solar radiation from meteorological satellites data. Sol. Energy 37, 31.

Cappa, F., Oriani, R., Peruffo, E., McCarthy, I., 2021. Big data for creating and capturing value in the digitalized environment: unpacking the effects of volume, variety, and veracity on firm performance. J. Prod. Innovat. Manag. 38, 49–67.

Chen, H., Chiang, R.H.L., Storey, V.C., 2012. Business intelligence and analytics: from big data to big impact. MIS Q. 36, 1165–1188. https://doi.org/10.2307/41703503.

Cirella, S., Guerci, M., Shani, A.B., 2012. A process model of collaborative management research: the study of collective creativity in the luxury industry. Syst. Pract. Action Res. 25, 281–300. https://doi.org/10.1007/s11213-011-9220-x.

Constantinides, P., Henfridsson, O., Parker, G.G., 2018. Introduction—platforms and infrastructures in the digital age. Inf. Syst. Res. 29, 381–400. https://doi.org/10.1287/isre.2018.0794.

de Reuver, M., Sørensen, C., Basole, R.C., 2018. The digital platform: a research agenda. J. Inf. Technol. 33, 124–135. https://doi.org/10.1057/s41265-016-0033-3.

Dhanaraj, C., Parkhe, A., 2006. Orchestrating innovation networks. AMR (Adv. Magn. Reson.) 31, 659–669. https://doi.org/10.5465/amr.2006.21318923.

Dittrich, Y., 2014. Software engineering beyond the project – sustaining software ecosystems. Information and Software Technology, Special issue on Software Ecosystems 56, 1436–1456. https://doi.org/10.1016/j.infsof.2014.02.012.

Eaton, B., Elaluf-Calderwood, S., Sørensen, C., Yoo, Y., 2015. Distributed tuning of boundary resources: the case of Apple's iOS service system. MIS Q. 39, 217–243. https://doi.org/10.25300/MISQ/2015/39.1.10.

Foerderer, J., Kude, T., Schuetz, S.W., Heinzl, A., 2019. Knowledge boundaries in enterprise software platform development: antecedents and consequences for platform governance. Inf. Syst. J. 29, 119–144. https://doi.org/10.1111/isj.12186.

Gandomi, A., Haider, M., 2015. Beyond the hype: big data concepts, methods, and analytics. Int. J. Inf. Manag. 35, 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007.

Gawer, A., 2020. Digital platforms' boundaries: the interplay of firm scope, platform sides, and digital interfaces. Long. Range Plan., 102045 https://doi.org/10.1016/j.lrp.2020.102045.

Gawer, A., 2014. Bridging differing perspectives on technological platforms: toward an integrative framework. Res. Pol. 43, 1239–1249. https://doi.org/10.1016/j.respol.2014.03.006.

Gawer, A., Cusumano, M.A., 2002. Platform Leadership: How Intel, Microsoft, and Cisco Drive Industry Innovation. Harvard Business School Press, Boston.

George, G., Haas, M.R., Pentland, A., 2014. Big data and management. Acad. Manag. J. 57, 321–326. https://doi.org/10.5465/amj.2014.4002.

Ghazawneh, A., Henfridsson, O., 2013. Balancing platform control and external contribution in third-party development: the boundary resources model. Inf. Syst. J. 23, 173–192. https://doi.org/10.1111/j.1365-2575.2012.00406.x.

Gitelman, L., 2013. "Raw Data" Is an Oxymoron, Infrastructures Series. The MIT Press, Cambridge, Massachusetts ; London, England.

Gizaw, A.A., Bygstad, B., Nielsen, P., 2017. Open generification. Inf. Syst. J. 27, 619–642. https://doi.org/10.1111/isj.12112.

Goor, E., Freytag, I., Le Bouler, G., Dusart, J., Van Meerloo, M., Ollier, G., Ramhoer, J., 2021. Eurogeo - the European component of GEO. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. Presented at the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 290–292. https://doi.org/10.1109/IGARSS47720.2021.9554463.

Grüter, W., Guillard, H., Möser, W., Monget, J.M., Palz, W., Raschke, E., Reinhardt, R.E., Schwarzmann, P., Wald, L., 1986. Solar radiation data from satellite images: determination of solar radiation at ground level from images of the Earth transmitted by meteorological satellites. In: An Assessment Study, Solar Energy R&D in the Ec Series F, vol. 4. Springer Netherlands, p. 100.

Gschwind, B., Ménard, L., Albuisson, M., Wald, L., 2006. Converting a successful research project into a sustainable service: the case of the SoDa web service. Environ. Model. Software 21, 1555–1561. https://doi.org/10.1016/j.envsoft.2006.05.002.

Günther, W.A., Rezazade Mehrizi, M.H., Huysman, M., Feldberg, F., 2017. Debating big data: a literature review on realizing value from big data. J. Strat. Inf. Syst. 26, 191–209. https://doi.org/10.1016/j.jsis.2017.07.003.

Han, M., Geum, Y., 2022. Roadmapping for data: concept and typology of data-integrated smart-service roadmaps. IEEE Trans. Eng. Manag. 69, 142–154. https://doi.org/10.1109/TEM.2020.3013295.

Hanseth, O., Bygstad, B., 2015. Flexible generification: ICT standardization strategies and service innovation in health care. Eur. J. Inf. Syst. 24, 645–663. https://doi.org/10.1057/ejis.2015.1.

Hanseth, O., Monteiro, E., Hatling, M., 1996. Developing information infrastructure: the tension between standardization and flexibility. Sci. Technol. Hum. Val. 21, 407–426. https://doi.org/10.1177/016224399602100402.

Henfridsson, O., Bygstad, B., 2013. The generative mechanisms of digital infrastructure evolution. MIS Q. 37, 907–931. https://doi.org/10.25300/MISQ/2013/37.3.11.

Hooge, S., Kokshagina, O., Le Masson, P., Levillain, K., Weil, B., Fabreguettes, V., Popiolek, N., 2016. Gambling versus designing: organizing for the design of the probability space in the energy sector: gambling versus designing. Creativ. Innovat. Manag. 25, 464–483. https://doi.org/10.1111/caim.12178.

Jacobides, M.G., Cennamo, C., Gawer, A., 2018. Towards a theory of ecosystems. Strat. Manag. J. 39, 2255–2276. https://doi.org/10.1002/smj.2904.

Janssen, M., Charalabidis, Y., Zuiderwijk, A., 2012. Benefits, adoption barriers and myths of open data and open government. Inf. Syst. Manag. 29, 258–268. https://doi.org/10.1080/10580530.2012.716740.

Janssen, M., Zuiderwijk, A., 2014. Infomediary business models for connecting open data providers and users. Soc. Sci. Comput. Rev. 32, 694–711. https://doi.org/10.1177/0894439314525902.

Jetzek, T., Avital, M., Bjorn-Andersen, N., 2019. The sustainable value of open government data. JAIS 702–734. https://doi.org/10.17705/1jais.00549.

Karhu, K., Gustafsson, R., Lyytinen, K., 2018. Exploiting and defending open digital platforms with boundary resources: Android's five platform forks. Inf. Syst. Res. 29, 479–497. https://doi.org/10.1287/isre.2018.0786.

Kokshagina, O., Masson, P.L., Weil, B., Cogez, P., 2016. Portfolio management in double unknown situations: technological platforms and the role of cross-application managers. Creativ. Innovat. Manag. 25, 270–291. https://doi.org/10.1111/caim.12121.

Langley, A., 1999. Strategies for theorizing from process data. AMR (Adv. Magn. Reson.) 24, 691–710. https://doi.org/10.5465/amr.1999.2553248.

Lefèvre, M., Oumbe, A., Blanc, P., Espinar, B., Gschwind, B., Qu, Z., Wald, L., Schroedter-Homscheidt, M., Hoyer-Klick, C., Arola, A., Benedetti, A., Kaiser, J.W., Morcrette, J.-J., 2013. McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. Atmos. Meas. Tech. 6, 2403–2418. https://doi.org/10.5194/amt-6-2403-2013.

Lenfle, S., Söderlund, J., 2022. Project-oriented agency and regeneration in socio-technical transition: insights from the case of numerical weather prediction (1978–2015). Res. Pol. 51, 104455 https://doi.org/10.1016/j.respol.2021.104455.

Li, M., Nielsen, P., 2019. Making Useable Generic Software. A Matter of Global or Local Design? 10th Scandinavian Conference on Information Systems.

Lycett, M., 2013. 'Datafication': making sense of (big) data in a complex world. null 22, 381–386. https://doi.org/10.1057/ejis.2013.10.

Lyytinen, K., King, J.L., 2006. Standard making: a critical research frontier for information systems research. MIS Q. 30, 405–411. https://doi.org/10.2307/25148766.

Magalhaes, G., Roseira, C., 2020. Open government data and the private sector: an empirical view on business models and value creation. Govern. Inf. Q. 37, 101248 https://doi.org/10.1016/j.giq.2017.08.004.

Ménard, L., Wald, L., Blanc, P., Gschwind, B., 2013. Development and integration of a local solar atlas into a GEOSS compliant global spatial data infrastructure (GSDI). In: Presented at the 27th International Conference on Informatics for Environmental Protection, p. 132.

Mikalsen, M., Monteiro, E., 2021. Acting with inherently uncertain data: practices of data-centric knowing. J. Assoc. Inf. Syst. Online 22, 1715–1735.

Monteiro, E., Parmiggiani, E., 2019. Synthetic knowing: the politics of the internet of things. MIS Q. 43, 167–184. https://doi.org/10.25300/MISQ/2019/13799.

Monteiro, E., Pollock, N., Hanseth, O., Williams, R., 2013. From artefacts to infrastructures. Comput Supported Coop Work 22, 575–607. https://doi.org/10.1007/s10606-012-9167-1.

Müller-Seitz, G., 2012. Leadership in interorganizational networks: a literature review and suggestions for future research. Int. J. Manag. Rev. 14, 428–443. https://doi.org/10.1111/j.1468-2370.2011.00324.x.

Østerlie, T., Monteiro, E., 2020. Digital sand: the becoming of digital representations. Inf. Organ. 30, 100275 https://doi.org/10.1016/j.infoandorg.2019.100275.

Pasmore, W.A., Stymne, B., Shani, A.B., Rami, Mohrman, S.A., Adler, N., 2008. The promise of collaborative management research. In: Handbook of Collaborative Management Research. SAGE Publications, Inc., 1 Oliver's Yard, 55 City Road London EC1Y 1SP, pp. 6–32. https://doi.org/10.4135/9781412976671.n1.

Pollock, N., Williams, R., D'Adderio, L., 2007. Global software and its provenance: generification work in the production of organizational software packages. Soc. Stud. Sci. 37, 254–280. https://doi.org/10.1177/0306312706066022.

Ranchin, T., Ménard, L., Fichaux, N., Reboul, M., 2021. Partners, all the e-shape, 2021. E-Shape - EuroGEO showcases: application powered by Europe contribution to eurogeo and to the development of the EO industry. In: IEEE International Geoscience and Remote Sensing Symposium IGARSS. Presented at the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 293–295. https://doi.org/10.1109/IGARSS47720.2021.9553426.

Rigollier, C., Albuisson, M., Delamare, C., Dumortier, D., Fontoynont, M., Gaboardi, E., Gallino, S., Heinnemann, D., Kleih, M., Kunz, S., Levermore, M., Major, G., Martinoli, M., Page, J.H., Ratto, C., Reise, C., Remund, J., Rimoczi-Paal, A., Wald, L., Webb, A., 2000. Exploitation of distributed solar radiation databases through a smart network: the project SoDa. In: EuroSun 2000. Copenhagen, Denmark.

Rigollier, C., Lefèvre, M., Wald, L., 2004. The method Heliosat-2 for deriving shortwave solar radiation from satellite images. Sol. Energy 77, 159–169.

Ruijer, E., Grimmelikhuijsen, S., Hogan, M., Enzerink, S., Ojo, A., Meijer, A., 2017. Connecting societal issues, users and data. Scenario-based design of open data platforms. Govern. Inf. Q. 34, 470–480. https://doi.org/10.1016/j.giq.2017.06.003.

Schymanietz, M., Jonas, J.M., Möslein, K.M., 2022. Exploring data-driven service innovation—aligning perspectives in research and practice. J. Bus. Econ. 1–39. https://doi.org/10.1007/s11573-022-01095-8.

Shani, A.B., Rami, Mohrman, S.A., Pasmore, W.A., Stymne, B., Adler, N., 2008. Handbook of Collaborative Management Research. SAGE Publications, Inc. https://doi.org/10.4135/9781412976671, 1 Oliver's Yard, 55 City Road London EC1Y 1SP.

Silsand, L., Ellingsen, G., 2014. Generification by translation: designing generic systems in context of the local. J. Assoc. Inf. Syst. Online 15. https://doi.org/10.17705/1jais.00358.

Thomas, C., Saboret, L., Wey, E., Gschwind, B., Ménard, L., Wald, L., 2013. Benefits and limits of OGC-web services to the new SoDa service on solar energy. In: Page, B., Fleischer, A.G., Göbel, J., Wohlgemuth, V. (Eds.), 27th International Conference on Informatics for Environmental Protection. Hambourg, Germany, p. 149.

Tilson, D., Lyytinen, K., Sørensen, C., 2010. Digital infrastructures: the missing IS research agenda. Inf. Syst. Res. 21, 748–759. https://doi.org/10.1287/isre.1100.0318.

Tiwana, A., Konsynski, B., Bush, A.A., 2010. Platform evolution: coevolution of platform architecture, governance, and environmental dynamics. Inf. Syst. Res. 21, 675–687. https://doi.org/10.1287/isre.1100.0323.

Tournadre, B., Gschwind, B., Saint-Drenan, Y.-M., Chen, X., Amaro E Silva, R., Blanc, P., 2022. An alternative cloud index for estimating downwelling surface solar irradiance from various satellite imagers in the framework of a Heliosat-V method. Atmos. Meas. Tech. 15, 3683–3704. https://doi.org/10.5194/amt-15-3683-2022.

Trabucchi, D., Buganza, T., 2020. Fostering digital platform innovation: from two to multi-sided platforms. Creativ. Innovat. Manag. 29, 345–358. https://doi.org/10.1111/caim.12320.

Trabucchi, D., Buganza, T., Dell'Era, C., Pellizzoni, E., 2018. Exploring the inbound and outbound strategies enabled by user generated big data: evidence from leading smartphone applications. Creativ. Innovat. Manag. 27, 42–55. https://doi.org/10.1111/caim.12241.

Urbinati, A., Bogers, M., Chiesa, V., Frattini, F., 2019. Creating and capturing value from Big Data: a multiple-case study analysis of provider companies. Technovation 84 (85), 21–36. https://doi.org/10.1016/j.technovation.2018.07.004.

Yin, R.K., 2009. Case Study Research: Design and Methods. SAGE.

Yoo, Y., Henfridsson, O., Lyytinen, K., 2010. The new organizing logic of digital innovation: an agenda for information systems research. Inf. Syst. Res. 21, 724–735. https://doi.org/10.1287/isre.1100.0322.

Yoo, Y., Lyytinen, K., Yang, H., 2005. The role of standards in innovation and diffusion of broadband mobile services: the case of South Korea. The Journal of Strategic Information Systems, The Future is UNWIRED: Organizational and Strategic Perspectives 14, 323–353. https://doi.org/10.1016/j.jsis.2005.07.007.

Zeng, J., Glaister, K.W., 2018. Value creation from big data: looking inside the black box. Strat. Organ. 16, 105–140.