ICIS 2022 Proceedings

Digital Innovation, Entrepreneurship, and New Business Models

Dec 12th, 12:00 AM

Data Liquidity: Conceptualization, Measurement and Determinants

Gabriele Piccoli
Lousiana State University, gpiccoli@cct.lsu.edu

Joaquin Rodriguez
Grenoble Ecole de Management, joaquin.rodriguez@grenoble-em.com

Ida Asadi Someh
University of Queensland, i.asadi@business.uq.edu.au

Barbara Wixom
MIT Sloan School of Management, bwixom@mit.edu

Data Liquidity: Conceptualization, Measurement and Determinants

Short Paper

Gabriele Piccoli
Louisiana State University,
Baton Rouge, Louisiana, US
University of Pavia, Pavia, Italy
gpiccoli@lsu.edu

Joaquin Rodriguez
Grenoble Ecole de Management,
12, rue Pierre Sémard,
38000 Grenoble, France
joaquin.rodriguez@grenoble-em.com

Ida Asadi Someh
University of Queensland, Brisbane,
St Lucia 4072
Queensland, Australia
i.asadi@business.uq.edu.au

Barbara Wixom
MIT Sloan Center for Information
Systems Research
100 Main St, Cambridge, MA 02141 US
bwixom@mit.edu

Abstract

Despite the rhetoric that "data is the new oil" organizations continue to face challenges in data monetization, and we don't have a reliable way to measure how easily data assets can be reused and recombined in value creation and appropriation efforts. Data asset liquidity is a critical, yet underexamined, prerequisite for data monetization initiatives. We contribute to the theorizing process by advancing a definition, conceptualization, and measurement of data liquidity as an asset level construct. Based on interviews with 95 Chief Data and Analytics Officers from 67 distinct large global organizations, we identify three determinants of data liquidity: inherent asset characteristics, structural asset characteristics, and asset environment characteristics. We theorize the existence of equifinal configurations that yield liquid data assets, configurations that should prove helpful to academics and practitioners seeking to understand data liquidity and its impact on firms' data monetization efforts as well as society at large.

Keywords: Data Liquidity, Data Monetization, Data Reuse, Data Recombination

Introduction

The business press has embraced the rhetoric that "data is the new oil" for more than a decade (Palmer, 2006; Parkins, 2017). Yet, successfully monetizing data remains an elusive goal for most organizations (Wixom & Ross, 2017) who find data to be context specific and trapped in organizational siloes (Wixom et al., 2021). We define data monetization as the *direct or indirect process of converting data into quantifiable business value* (Gartner, 2022; Wixom & Farrell, 2019). One of the key challenges in data monetization is the ability to bridge the gap between discovery, the identification of data assets of potential value, and the use of such data assets in value creating initiatives (Wixom & Piccoli, 2020). When the gap is shortened significant opportunities for data monetization emerge, leading to organizational and societal benefits. For example, by requiring payers to create Application Program Interfaces (APIs) that expose health patient records, the US Interoperability and Patient Access rule promotes more equitable and efficient health care systems. Similarly, Open Banking fostered a wave of innovation in the European banking industry by requiring large banks to share with third-party providers customer account information via open, secure, and standardized interfaces.

In the Information Systems (IS) literature, data has long been recognized as a strategic resource (Piccoli & Ives, 2005). Traditionally, companies have created value from data by supporting strategic business processes (Chen et al., 2012), improving their decisions when responding to the competitive environment (Rainer & Watson, 1995), or increasing the efficiency of their operations (Melville et al., 2004). More recently, technological advancements such as cloud computing, multicluster technology, and shared data architectures facilitated the emergence of novel data monetization strategies (Wixom & Ross, 2017). Yet, such monetization strategies are contingent on the degree to which data assets are easily actionable within and across organizations (Wixom & Piccoli, 2020).

Previous research on IT infrastructure flexibility (Byrd & Turner, 2000) and data warehousing (Wixom & Watson, 2001) recognizes the importance of improving retrieval and flow of data within and across organizations. This stream however, focused on data used in combination with specific software applications, in a technological environment where it was "not easy to separate one from the other" (Byrd & Turner, 2000, p. 194). More recently, IT infrastructure research accounts for the decoupling of software, hardware, and data components (Mikalef et al., 2021; Tiwana et al., 2010). Proponents of this view suggest that organizational applications and data can be architected as a set of distinct atomic components that can be "recombined and restructured to quickly construct new solutions" (Mikalef et al., 2021, p. 518). Thus, they contribute a view of data no longer tied to specific applications, but rather independently designed for multiple planned and unplanned future uses. Taking this view focuses the analysis on data assets, conceptualized as *cohesive sets of related data* (Piccoli et al., Forthcoming; Wixom et al., 2021). Data asset-level interest has grown along with the growing investigation of analytics (Park et al., 2017) and with the emergence of contemporary data asset phenomena, such as data exchanges (Moyano et al., 2020).

In line with recent research developments, our work conceptualizes data as decoupled from specific software applications or technologies. However, rather than studying data as one of the many elements of an organizations' IT infrastructure, we focus on data assets as the unit of interest. This approach has several advantages, chief amongst them the ability to fill an important gap in the literature. The IS literature has mostly treated IT as an organizational-level construct that included hardware, network, software, and data. Doing so, limits our ability to uncover the idiosyncrasies related to value creation through data. Previous work does demonstrate the importance of information and analytics capabilities in firm performance (Park et al., 2017) and details data monetization capabilities and practices (Wixom & Ross, 2017). However, there is no rigorous investigation explaining *how* organizations can increase data liquidity, the "ease of data asset reuse and recombination" (Wixom et al., 2021). Applying governance rules or technical specifications across all organizational data is cost prohibitive and arguably unnecessary for maximizing value creation opportunities. Adopting a finer-grained view helps uncovering how organizations can increase reuse and recombination of *specific* data assets (e.g., strategic data assets) with high value creation potential.

While it is too early to advance a theory of data liquidity, we intend this paper to contribute a useful early step in the theorizing process (Weick, 1995) by advancing a definition, conceptualization, and measurement of data liquidity as an asset level construct. Such definition is grounded in multiple rounds of interviews with 95 Chief Data and Analytics Officers from 67 distinct large global organizations. Secondly, we identify three determinants of data liquidity: (1) inherent asset characteristics, (2) structural asset characteristics, and (3) asset environment characteristics. Rather than relying on process- or variance-based approaches to support our theorizing, we find a configurational approach (Park et al., 2020) more conducive to identifying combinations of characteristics that need to co-occur for data assets to achieve high liquidity. The advantage of the configurational method is in its capacity to model and capture complex causality in the interconnected set of causal conditions that together produce the outcome. This is important because we theorize the existence of equifinal configurations that yield liquid data assets. The configurations should prove helpful to academic research, by illuminating causal pathways that foster high data liquidity. They should also be useful to practicing managers in need of templates to adopt for evaluating competing courses of actions to unlock the value potential of data assets.

The paper is organized as follows. In the next section we conceptualize data liquidity and lay the theoretical framework for creating data liquidity measures. Next, we present preliminary measures, and we illustrate the research design we use to uncover equifinal configurations of data asset characteristics and firm resources. The paper concludes with expected findings.

Related Literature

Data Liquidity

In finance, asset liquidity stems from "the ease with which the firm's assets can be sold on a secondary market" (Morellec, 2001, p. 174). Liquidity measurement and monitoring is critical for organizations to ensure the firm remains solvent and can continue to meet its short-term obligations. An "asset is liquid if it can be converted into cash quickly and at low cost" (Gopalan et al., 2012). Cash is the most liquid financial asset because it can be immediately deployed in any financial transaction. Liquid assets are valuable because finding a potential user for the asset is easier, indicating that the asset owner can easily raise capital (Pham et al., 2018). Thus, financial liquidity does not imply or require any value creation. Instead, it measures the ease with which a firm can convert an asset from a form of capital that cannot be immediately used in economic transactions to a form of capital that is recognized as a legal tender.

Following the logic of asset liquidity, recent research defines *data liquidity* as the *ease of data asset reuse and recombination* (Wixom et al., 2021). Thus, data liquidity measures the readiness with which potential users of a data asset can deploy it in value creating (or value destructing) initiatives. As with financial asset liquidity, data liquidity is a continuum. Liquid data is readily available to potential users who can easily deploy it to execute initiatives while illiquid data impose obstacles on those seeking to utilize them. As with financial liquidity, data liquidity does not imply the creation of value. Instead, data liquidity makes a data asset more attractive for potential users by lowering the barriers for its reuse and recombination into data monetization initiatives. In other words, data liquidity is an enabler of data monetization, not a determinant or a prerequisite. There are many examples of highly liquid data asset that have dubious value creation potential, such as a REST API that returns random quotes by Keyne West (i.e., kanye.rest).

Reuse is the action of *using* something again, more than once. Economic literature defines data as a non-rival good (Jones & Tonetti, 2020), meaning that it is not consumed by use. Thus, data assets are highly reusable, at least theoretically. However, the marginal cost of enabling prospective users to obtain a data asset or enabling data asset reuse in multiple use cases can vary significantly depending on the asset's inherent characteristics (e.g., regulated personally identifiable information) or the way in which the data asset is structured and governed. Fostering high reuse has long been a recognized enabler of data monetization by firms that curate high quality, valuable data assets. These organizations have strong incentives to maximize reuse across different customers and use cases. For example, consider the case of the data asset underlying Google Maps. The data asset (i.e., the maps) is exposed and made available to the largest possible number of users; Alphabet serves map information more than two billion times a day to thousands of customers (e.g., Uber). However, reusability can be restricted by the scope of users or use cases the data asset was designed for or by burdensome and lengthy processes required for accessing and obtaining the data. For example, new users or new use cases might require ad-hoc coordination about specific needs and the maintenance of additional file formats or delivery schedules. Moreover, due to privacy, security, financial and other legal restrictions, firms must enforce strict vetting procedures before granting access to data assets containing personal identifiable or otherwise sensitive information.

Recombination is the process of combining things again or differently. Data monetization initiatives typically require the joining and connection of multiple data assets. For example, PepsiCo, the multinational food, snacks and beverage corporation, had to aggregate data across divisions and external sources when launching PEPWORX℠, a suite of tools that allowed retailers to optimize store-level assortment at a granular location level (Wixom, 2019). Recombination is not always feasible and, even when possible, it often engenders high costs because of interdependencies between data assets. Data assets encode representations of the world, therefore their size or state changes when new records are added or deleted, or when existing records are updated. When data assets depend on other applications or business processes, recombination is complex and can require costly movement, replication, and transformations of data before integration becomes feasible. For example, data warehouses are specially prepared repositories that result from the combination of multiple source systems and transformation processes (Wixom & Watson, 2001). Moreover, heterogeneous standards across data assets might hinder the possibility of recombination (Byrd & Turner, 2000). A data asset stored in an Oracle SQL schema requires specialized technology and knowledge to combine with a collection stored in MongoDB.

The ease of reuse and recombination of a data asset represent related yet distinct dimensions of data liquidity. An increase in one of the two dimensions increases data liquidity, but the drivers of increased reuse likely differ, at least in part, from those of increased recombination. For example, it is possible that a data asset can theoretically scale to large number of users and a wide variety of use cases, but proprietary technology or minimal semantic standards may make it very difficult to integrate it with other data assets. To lower coordination costs and increase local autonomy, many organizations do not enforce enterprise or industry standards when capturing data across departments, unit, or regions. As a result, while the ease of reuse of the local data can be high, data recombination across different systems requires the implementation of expensive re-engineering and standardization processes. At the same time, a highly standardized data asset that can easily be recombined with other data might require lengthy permissioning and vetting processes before new use cases can be carried out. For example, the data owner can require the submission of a formal request that result in a negotiation process where the use case must be approved, and service level agreements defined. In short, ease of reuse and recombination are formative dimensions of data liquidity, and the characteristics of the data can influence both dimensions differently.

Despite its importance to practitioners and researchers, the literature lacks a reliable way to measure data liquidity. Consequently, little research has investigated how firms can maximize the liquidity of their data assets. We contribute to fill this measurement and theoretical gap by grounding the work in the theory of modularity (Baldwin & Clark, 2000; H. Simon, 1962) and digital resources (Piccoli et al., Forthcoming).

Modularity

An object is an enduring structured collection of elements that persist in time (Faulkner & Runde, 2013). All social, biological, or technological objects can be conceptualized as a system of components, or modules, that are both interdependent and independent (H. Simon, 1962). Modularity is a design principle that intentionally increases the independence of a design's elements (Sanchez & Mahoney, 1996) and represents "the degree to which a system's components can be separated and recombined" (Schilling, 2000, p. 312). In this context, "a module is a unit whose structural elements are powerfully connected among themselves and relatively weakly connected to elements in other units" (Baldwin & Clark, 2000, p. 63). Modules are designed to deliver functionalities that enable an entity, such as a product or system, to operate (Pil & Cohen, 2006). Modularity is a degree, rather than a binary condition. Greater modularity in a system's architecture or product design enables separation of the system's components and facilitates their recombination into new configurations (Schilling, 2000). Conversely, an integrated or monolithic architecture is more challenging to decompose because the interdependencies between the parts of the system are complex and overlapping (Baldwin & Clark, 2000; Schilling, 2000).

Previous research shows how products (Parnas, 1972; Pil & Cohen, 2006), organizations (Greeven et al., 2021), or IT infrastructures and platform (Henfridsson & Bygstad, 2013; Tiwana et al., 2010) can increase *reuse* and *recombination* of their components by enforcing a modular architecture. By doing so, they decompose the system into elements that perform as modules within a larger system architecture (Schilling, 2000). Modular components can be easily reused across product families and/or product generations when they are used "within the range of variations allowed by the modular product architecture" (Sanchez & Mahoney, 1996, p. 66).

With respect to digital technology, reuse is a core principle in software development. It eliminates redundancy by allowing for the utilization of existing software code in different programs and applications (Fichman & Kemerer, 1993). Software modules are routinely reused and recombined. The extent of recombination of useful modules in modern applications and software systems becomes evident when problems occur, like the outages ensuing from a server-side update of Facebook's iOS SDK which exposed the vulnerability of entire services (e.g., Spotify, Doordash). More recently the discovery of a vulnerability in the Apache Log4J Java framework exposed thousands of applications and programs that used it as a module in their design, forcing their owners to scramble to patch their solutions. To the extent that modules are "compatible with the overall system architecture, they may be recombined easily with one another" (Schilling, 2000, p. 2000) to create novel configurations. For example, Express, the web framework for node.js, can be recombined with other node packages that perform different functions to create a multitude of software solutions.

The independence of modular software components is enabled by information hiding (Parnas, 1972), through abstraction (Baldwin & Clark, 2000) and encapsulation (Schilling, 2000). Information hiding ensures that modular components interact with one another strictly through their interface (Schilling, 2000), which serves as the "preestablished way to resolve potential conflicts between interacting parts of the design" (Baldwin & Clark, 2000, p. 73). When abstracted and encapsulated by a comprehensive interface, reuse and recombination of modular components does not require ad-hoc integration or coordination activities on the part of module designers (Sanchez & Mahoney, 1996). This is because designers of separate modular components need no specific knowledge or information about the inner working of modules they need to reuse or integrate with, as long as they maintain adherence to the interface specifications (Baldwin & Clark, 2000).

Programmatic Bitstring Interfaces

As a design principle, modularity can be applied to both material and digital objects (Kallinikos et al., 2013; H. A. Simon, 1996). Material objects have spatial attributes, such as shape or volume (e.g., Ikea furniture). Digital objects are those nonmaterial or hybrid objects "whose component parts include one or more bitstrings" (e.g., Express node.js package) (Faulkner & Runde, 2019, p. 1285). Bitstrings, separated in program files and data files, are "the sequences of 1's and 0's used in computing to represent information in binary form" (Faulkner & Runde, 2019, p. 804).[1] Data assets are nonmaterial digital objects. Importantly however, while data assets stored by way of digital information technology are by definition digital objects, not all data assets are designed as modular components. Most importantly, not all data assets have a programmatic bitstring interface. For example, many legacy point of sale (POS) systems in restaurants have no software interface for data extraction, forcing operators to access granular purchase data by perusing pdf representations of the guest check, making the data difficult to reuse and recombined.

Recent research introduced the notion of digital resources, as those "digital objects that a) are *modular*, b) *encapsulate objects of value*, namely assets and/or capabilities, c) and are accessible by way of a *programmatic bitstring interface*" (Piccoli et al., Forthcoming, p. 5). A subset of digital resources, digital data assets, are modular components that encapsulates a data asset and makes it accessible for reuse and recombination through a programmatic bitstring interface. Several implications follow from this definition. First, configuring a digital data asset requires the conscious design decision to abstract and encapsulate the data so as to enforce information hiding. Second, the data asset configured as a modular component must be made available for reuse and recombination through a programmatic bitstring interface that regulates both the technical and governance aspects of data asset access and provision. Modern cloud-based restaurant POS systems, such as Toast and Odoo, do expose the granular menu item purchase data via an API that makes it easy for operators to reuse the data for analytics, not just operational needs. Moreover, some modern POS systems expose the data to ecosystem partners who reuse it and recombine it with other data to offer new functionalities (e.g., customer relationship management, menu engineering and analysis). As these examples illustrate, when the interface to a digital data asset is fully inscribed in software (Kallinikos et al., 2013; Piccoli et al., Forthcoming), the parameters, inputs, and outputs of digital data assets are expressed as software and interactions with the module are programmatic, as opposed to manual (Piccoli et al., Forthcoming).

A data asset is only configured as a digital resource when its programmatic bitstring interface comprise the technical specifications (e.g., RESTful architecture), the contract structure (e.g., per-use or bulk billing), and the coordination mechanisms (e.g., notification services) that govern access to the data (Baldwin & Clark, 2000). Doing so, the digital interface fully regulates and enforces the technical and governance specifications for data reuse and recombination. For example, a data provider sharing a table on a data platform (e.g., Snowflake) can programmatically enforce access control by setting interface privileges that allow authorized users to obtain only specific data elements and partitions of the data asset. More importantly, the data provider can also monitor and enforce contractual agreements through the digital interface, such as allowing free queries, charging per query once a certain number is exceeded, or imposing rate limits that contractually cap total cost for different users.

In summary, we posit that the ease of reuse and recombination of a data asset will depend on the characteristics of the data as well as the ability of the organization to encapsulate the data asset as a modular

[1] As defined by Faulkner and Runde (2019), the term bitstring parallels the colloquial use of the term software, encompassing data and program files. Thus, while generally maintaining Faulkner and Runde's (2019) terminology, we sometimes substitute the term software.

component in a programmatic bitstring interface. In the remainder of this research in progress paper we describe the methodology, detail preliminary data liquidity measures and the design of the empirical study we use to uncover equifinal configurations of data asset characteristics that yield liquidity.

Methodology

To capture the inherent complexity associated with any digital phenomena (Park et al., 2020), we leverage a configurational research approach (Meyer et al., 1993; Mithas et al., 2022). In this methodology a configuration is developed as a "multidimensional constellation of conceptually distinct characteristics that commonly occur together" (Meyer et al., 1993, p. 1175). Generally, a configurational research approach helps researchers produce theories that capture and link multiple sets of interconnected elements (also referred to as characteristics or causal conditions) that must be simultaneously present for a certain outcome to occur (Ragin, 2008). The approach differs from the traditional process and variance models in conceptualizing causality. In the configurational approach, complex causality is captured by building casual relations that are conjunctural, equifinal, and asymmetric (Misangyi et al., 2017). In the specific context of our work, conjunctural logic shifts the attention away from individual data, data asset or organizational characteristics. Rather, it evaluates combinations of characteristics that together yield an empirically observed level of data asset liquidity. The equifinality explicitly allowed by the configurational approach allows for different configurations or causal pathways to explain high (or low) levels of data asset liquidity. This methodology aligns with our research objectives, because the asymmetry of the possible causal relations enables us to uncover characteristics that are causally related in one configuration but unrelated or inversely related in others. In other words, the presence of a characteristic for high liquidity data assets does not imply that its absence will leads to low liquidity – as any practicing manager will readily volunteer, there are different ways to achieve high data asset liquidity.

We designed a three-phased study for building configurations that explain the drivers of data liquidity: (1) executive discussion, (2) within-case analysis and (3) cross-case analysis.

Phase 1– Executive Discussion: As the research on data liquidity is only just emerging, we convened an online discussion with the Data Advisory Board members of the Center for Information Systems Research of the MIT Center for IS Research. The executive discussion was designed to help the research team in two important ways. First, the process helped to establish a conceptual domain for data liquidity, and second, the process proved the practical relevance of the concept as an emergent phenomenon.

The Board consists of 95 data executives representing 67 large companies headquartered around the globe. Most organizations were multi-national and for-profit, and the executives held Chief Data Officer, Chief Analytics Officer, or equivalent roles. Each executive was asked to answer two questions in an online discussion board: 1) If your CEO demanded that you report on your organization's data liquidity, what metric(s) would you use? 2) Now, focus on a single data asset that you believe can be repeatedly monetized. a) How could you measure how liquid it is? b) What can you do to make it more liquid? 44 executives from 40 distinct organizations answered. Two researchers analyzed the answers using NVIVO software to code distinct and recurring concepts and themes and to create an initial list of characteristics and outcomes of data liquidity. The researchers then presented a definition of data liquidity and an initial categorization of measures to the executives in two one-hour webinars (to accommodate different global time zones) during which the executives provided reactions and input. The exercise confirmed that: a) executives found the concept of data liquidity compelling but viewed it as difficult to achieve and measure; b) executives are actively investing in practices that they hope will increase data liquidity; c) the concept of data liquidity is distinct from data asset discovery and value creation/appropriation activities.

Phase 2– Within-case analysis: This phase buildt the factorial logic of our configurational approach (Park et al., 2020). We developed characteristics and outcomes that describe liquidity of data assets. We conducted 15 semi-structured interviews and collected data on 21 data assets (16 of which were described as having high liquidity and 5 of which were described as having low liquidity). As prescribed (Gioia et al., 2013) we initially developed key categories inductively to organize the data. We contrasted and compared various data categories and created abstract second-order themes, which were further grouped and synthesized into aggregate dimensions (see Table 1 for the resulting preliminary measures).

Phase 3– Cross-case analysis: We performed 20 additional data collection interviews (May-July 2022) on high and low liquidity data assets. This phase helped us build the combinatorial logic of configurational approach by transforming interview data into fuzzy sets aided by a calibration framework to identify measures/indicators for the attributes, define relevant variation in fuzzy-set membership and assign anchor points (Basurto & Speer, 2012). Fuzzy sets allow varying degrees of membership in a construct (e.g., fully in, fully out, and the points of ambiguity where there are neither fully in nor out). The anchor points for each variable determine the degree of membership (Ragin, 2008). After conducting the interviews and coding them against the characteristics and outcomes defined during phase 2, we ran the truth table algorithm to produce configurations.

Data Asset (DA) Characteristics	Preliminary Measure	Sample Question
Inherent Asset Characteristics	Sensitivity	Does the DA contain personal identifiable information?
	Value potential	Is the DA used to generate revenue streams?
Structural Asset Characteristics	Sensitivity remediation	Is the DA sensitivity addressed through asset manufacturing?
	Degree of productization	Does the DA expose data at the lowest level of granularity?
	Degree of standardization	Does the DA adhere to firm or industry standards?
Asset Environment Characteristics	Permissioning automation	To what extent is the provisioning process automated?
	Self-provisioning	Can users self-provision once permission is granted?

Table 1. Data Asset Characteristics and Preliminary Measures

Preliminary Outcomes and Conclusion

Despite the decade old rhetoric that "data is the new oil," organizations continue to face challenges in their pursue of data monetization initiatives. Understanding data asset liquidity is a critical, yet underexamined, prerequisite for data monetization because data that cannot be "actioned" is difficult to monetize (Wixom & Piccoli, 2020). We lack clear constructs and measures that capture how easily data assets can be reused and recombined. Conceptualizing and reliably measuring data liquidity are prerequisites steps to informing research and practice about the characteristics of liquid data assets and the actions organizations should take to increase the liquidity of strategic data assets (Wixom et al., 2021).

By departing from a focus of correlations and net effects, our preliminary findings provide useful boundary conditions about the characteristics of data assets that influence their liquidity. While the research is still in progress, our extensive fieldwork and discussions with executives suggests some preliminary observations and possible findings:

- **Equifinal configurations**. We theorize that causal conditions are likely to interact in complex ways leading to different configurations accomplishing the same level of data asset liquidity. Early results confirm our expectations as the data liquidity depends on inherent characteristics of the data, the structure of the data asset and organizational characteristics. In particular, preliminary findings suggest that firms might overcome data sharing problems related to the inherent sensitivity of the data asset through three different practices. First, they can manufacture the data asset to make the sensitive information illegible even when accessed. Common approaches are the tokenization or elimination of personally identifiable information (e.g., social security number). Second, firms can aggregate information to make it impossible to identify personal information from the data. Third, organizations can enforce complex and taxing permissioning processes that ensure that the data is only access by legitimate individuals.

- **Idiosyncratic configurations**. We also predict that the maximization of the two dimensions of liquidity, namely reuse and recombination, will result from distinct configurations. Early results suggest that some constraints are irrelevant for reuse, but can drastically hamper recombination. For example, one informant from a global commercial bank cited the example of cross border financial regulation as a major obstacle to recombination of customer data originating in different jurisdictions. Similarly, while the level of granularity of the data does not appear to impact the ease of reuse, preliminary analysis indicates that it is an important condition for achieving recombination. For example, highly refined "data products" likely foster reuse by being easily deployed in data wrapping or data sale initiatives (Wixom & Ross, 2017). Conversely, granular data assets are likely to lower the barriers to extensive recombination, particularly when coupled with extensive metadata and standardization. This result suggests that data aggregation can increase ease of reuse while hampering opportunities for recombination with other data assets.

Our preliminary results indicate that while data liquidity is a priority, many firms are still navigating the design challenges associated with data liquidity. Organizations we studied have made more progress in increasing the ease of reuse of their data assets, while recombination appears more difficult to attain. Only a small number of companies (i.e., three) have been able to design data assets that are both easy to reuse and recombine. While it is still early to identify specific configurations leading to different designs, early results corroborate our expectations about the design tension firms face. Yet, while perhaps more complex and costly, it is possible to achieve high reuse and recombination simultaneously.

References

Baldwin, C. Y., & Clark, K. B. (2000). *Design Rules: The Power of Modularity* (Vol. 1). MIT press.

Basurto, X., & Speer, J. (2012). Structuring the calibration of qualitative data as sets for qualitative comparative analysis (QCA). *Field Methods*, *24*(2), 155–174.

Byrd, T. A., & Turner, D. E. (2000). Measuring the flexibility of information technology infrastructure: Exploratory analysis of a construct. *Journal of Management Information Systems*, *17*(1), 167–208.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 1165–1188.

Faulkner, P., & Runde, J. (2013). Technological Objects, Social Positions, and the Transformational Model of Social Activity. *MIS Quarterly*, *37*(3), 803–818. https://doi.org/10.25300/MISQ/2013/37.3.06

Faulkner, P., & Runde, J. (2019). Theorizing the Digital Object. *MIS Quarterly*, *43*(4), 1279–1302. https://doi.org/10.25300/MISQ/2019/13136

Fichman, R. G., & Kemerer, C. F. (1993). Adoption of software engineering process innovations: The case of object orientation. *MIT Sloan Management Review*, *34*(2), 7.

Gartner. (2022). *Definition of Data Monetization—Gartner Information Technology Glossary*. Gartner. https://www.gartner.com/en/information-technology/glossary/data-monetization

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, *16*(1), 15–31.

Gopalan, R., Kadan, O., & Pevzner, M. (2012). Asset liquidity and stock liquidity. *Journal of Financial and Quantitative Analysis*, *47*(2), 333–364.

Greeven, M. J., Yu, H., & Shan, J. (2021). Why Companies Must Embrace Microservices and Modular Thinking. *MIT Sloan Management Review*, *62*(4), 1–6.

Henfridsson, O., & Bygstad, B. (2013). The Generative Mechanisms of Digital Infrastructure Evolution. *MIS Quarterly*, *37*(3), 907–931. https://doi.org/10.25300/MISQ/2013/37.3.11

Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the Economics of Data. *American Economic Review*, *110*(9), 2819–2858.

Kallinikos, J., Aaltonen, A., & Marton, A. (2013). The Ambivalent Ontology of Digital Artifacts. *MIS Quarterly*, *37*(2), 357–370. https://doi.org/10.25300/MISQ/2013/37.2.02

Melville, N., Kraemer, K., & Gurbaxani, V. (2004). Information Technology and Organizational Performance: An Integrative Model of IT Business Value. *MIS Quarterly*, *28*(2), 283–322. https://doi.org/10.2307/25148636

Meyer, A. D., Tsui, A. S., & Hinings, C. R. (1993). Configurational approaches to organizational analysis. *Academy of Management Journal*, *36*(6), 1175–1195.

Mikalef, P., Pateli, A., & van de Wetering, R. (2021). IT architecture flexibility and IT governance decentralisation as drivers of IT-enabled dynamic capabilities and competitive performance: The moderating effect of the external environment. *European Journal of Information Systems*, *30*(5), 512–540.

Misangyi, V. F., Greckhamer, T., Furnari, S., Fiss, P. C., Crilly, D., & Aguilera, R. (2017). Embracing causal complexity: The emergence of a neo-configurational perspective. *Journal of Management*, *43*(1), 255–282.

Mithas, S., Xue, L., Huang, N., & Burton-Jones, A. (2022). Editor's Comments: Causality Meets Diversity in Information Systems Research. *Management Information Systems Quarterly*, *46*(3), i–xvi.

Morellec, E. (2001). Asset liquidity, capital structure, and secured debt. *Journal of Financial Economics*, *61*(2), 173–206.

Moyano, J. P., Schmedders, K., & others. (2020). What Managers Need to Know About Data Exchanges: The Era of Big-data Silos is Fading. Shared Data is the Future. *MIT Sloan Management Review*, *61*(4), 39–44.

Palmer, M. (2006). *Data is the New Oil*. ANA Marketing Maestros. https://ana.blogs.com/maestros/2006/11/data_is_the_new.html

Park, Y., El Sawy, O. A., & Fiss, P. (2017). The role of business intelligence and communication technologies in organizational agility: A configurational approach. *Journal of the Association for Information Systems*, *18*(9), 1.

Park, Y., Fiss, P. C., & El Sawy, O. A. (2020). Theorizing the Multiplicity of Digital Phenomena: The Ecology of Configurations, Causal Recipes, and Guidelines for Applying QCA. *MIS Quarterly*, *44*(4).

Parkins, D. (2017). The world's most valuable resource is no longer oil, but data. *The Economist*. https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

Parnas, D. L. (1972). On the Criteria to Be Used in Decomposing Systems into Modules. *Communications of the ACM*, *15*(12), 1053–1058. https://doi.org/10.1145/361598.361623

Pham, L. T. M., Van Vo, L., Le, H. T. T., & Le, D. V. (2018). Asset liquidity and firm innovation. *International Review of Financial Analysis*, *58*, 225–234.

Piccoli, G., & Ives, B. (2005). IT-Dependent Strategic Initiatives and Sustained Competitive Advantage: A Review and Synthesis of the Literature. *MIS Quarterly*, *29*(4), 747–776. https://doi.org/10.2307/25148708

Piccoli, G., Rodriguez, J., & Grover, V. (Forthcoming). Digital Strategic Initiatives and Digital Resources: Construct Definition and Future Research Directions. *MIS Quarterly*, *Forthcoming*. https://doi.org/10.25300/MISQ/2022/17061

Pil, F. K., & Cohen, S. K. (2006). Modularity: Implications for Imitation, Innovation, and Sustained Advantage. *Academy of Management Review*, *31*(4), 995–1011. https://doi.org/10.5465/AMR.2006.22528166

Ragin, C. C. (2008). Measurement versus calibration: A set-theoretic approach. In *The Oxford handbook of political methodology*.

Rainer, R. K., & Watson, H. J. (1995). The keys to executive information system success. *Journal of Management Information Systems*, *12*(2), 83–98.

Sanchez, R., & Mahoney, J. T. (1996). Modularity, Flexibility, and Knowledge Management in Product and Organization Design. *Strategic Management Journal*, *17*(S2), 63–76. https://doi.org/10.1002/smj.4250171107

Schilling, M. A. (2000). Toward a General Modular Systems Theory and Its Application to Interfirm Product Modularity. *Academy of Management Review*, *25*(2), 312–334. https://doi.org/10.5465/AMR.2000.3312918

Simon, H. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, *106*(6), 467–482.

Simon, H. A. (1996). *The Sciences of the Artificial*. MIT press.

Tiwana, A., Konsynski, B., & Bush, A. A. (2010). Platform Evolution: Coevolution of Platform Architecture, Governance, and Environmental Dynamics. *Information Systems Research*, *21*(4), 675–687. https://doi.org/10.1287/isre.1100.0323

Weick, K. E. (1995). What theory is not, theorizing is. *Administrative Science Quarterly*, *40*(3), 385–390.

Wixom, B. H. (2019). *PepsiCo Unlocks Granular Growth Using a Data-Driven Understanding of Shoppers*. MIT CISR. https://cisr.mit.edu/publication/MIT_CISRwp439_PepsiCoDX_Wixom

Wixom, B. H., & Farrell, K. (2019). *Building Data Monetization Capabilities that Pay Off*. Building Data Monetization Capabilities That Pay Off. https://cisr.mit.edu/publication/2019_1101_DataMonCapsPersist_WixomFarrell

Wixom, B. H., & Piccoli, G. (2020). Actioned Analytics Pave the Way to New Customer Value. *MIT Sloan Management Review*. https://sloanreview.mit.edu/article/actioned-analytics-pave-the-way-to-new-customer-value/

Wixom, B. H., Piccoli, G., & Rodriguez, J. (2021). Fast-Track Data Monetization With Strategic Data Assets. *MIT Sloan Management Review*, *62*(4), 1–4.

Wixom, B. H., & Ross, J. W. (2017). How to monetize your data. *MIT Sloan Management Review*, *58*(3).

Wixom, B. H., & Watson, H. J. (2001). An empirical investigation of the factors affecting data warehousing success. *MIS Quarterly*, *25*(1), 17–41.