Cao, Sam Ruiqing; Iansiti, Marco

Working Paper

Organizational Barriers to Transforming Large Finance Corporations: Cloud Adoption and the Importance of Technological Architecture

CESifo Working Paper, No. 10142

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Cao, Sam Ruiqing; Iansiti, Marco (2023) : Organizational Barriers to Transforming Large Finance Corporations: Cloud Adoption and the Importance of Technological Architecture, CESifo Working Paper, No. 10142, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/271786

CESifo WORKING PAPERS

10142
2022

Original Version: December 2022
This Version: February 2023

Organizational Barriers to Transforming Large Finance Corporations: Cloud Adoption and the Importance of Technological Architecture

Sam (Ruiqing) Cao, Marco Iansiti

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10142*

Organizational Barriers to Transforming Large Finance Corporations: Cloud Adoption and the Importance of Technological Architecture

Abstract

This paper studies the impact of technological architecture around data storage and processing on the performance of large financial corporations after being exposed to more stringent data privacy regulations. A modular approach to cloud adoption – which reflects in the lack of data interoperability and reliance on microservices architecture – significantly constrains corporations' ability to adapt after the GDPR became enforceable. We hypothesize that a modular approach to cloud adoption leads to uncontrolled scaling and data silos that hinder coordination and regulatory compliance. Using a difference-in-differences regression design, we find that establishment revenues lower by 30% among corporations substantially exposed to GDPR. Other corporations do not experience similar losses. We also find evidence consistent with theory using two alternative measures based on cloud vendor configurations.

Keywords: technological architecture, cloud computing, ecosystem, GDPR, organisation design.

*Sam (Ruiqing) Cao**
Stockholm School of Economics / Sweden
sam.cao@hhs.se

Marco Iansiti
Harvard Business School / USA
miansiti@hbs.edu

*corresponding author

February 2023
We thank Disha Monga and Jiayu Lih for excellent research assistance. We are grateful to Alexandria Sheng, Ellen Wu, Amanda Pratt, Ross Sullivan, and Keystone Strategy for implementing the survey and collecting the data. Satya Nadella, Kinney Zalesne, and Greg Shaw at Microsoft provided invaluable input and helped fund the data collection process. We are grateful to the participants in the Organization Design Community Idea Development Workshop and the MISQ Author Development Workshop for very useful feedback, as well as James Bessen, Tim DeStefano, Chiara Farronato, Chris Forman, Emilie Fröberg, Shane Greenstein, Karim Lakhani, Magnus Mähring, Kristina McElheran, Frank Nagle, Glen Weyl, Feng Zhu, and Letian Zhang for very helpful comments. One of us (Marco Iansiti) has worked with many companies in the IT sector and in traditional industries on the topic of digital transformation.

1. Introduction

The rapid diffusion of digital technologies requires organizational adaptation to be productive (Adner, Puranam, & Zhu, 2019; Bresnahan, 2019; Kretschmer & Khashabi, 2020; Agrawal, Gans, & Goldfarb, 2021). The proliferation of data leads to unprecedented scale advantages, especially among digital natives that increasingly disrupt traditional industries. On the other hand, large traditional corporations face various challenges in transforming their businesses to adopt digital technologies (Hanelt, Bohnsack, Marz, & Marante, 2021). A crucial goal for transformation is building systems that can process large amounts of data and develop capabilities in an agile but robust fashion. The key infrastructural technology enabling such capabilities is the cloud, where public cloud providers rent computing capacities to adopting firms, which can drastically improve local efficiencies in scaling up data storage and processing. However, while the cloud benefits small and young firms (Jin & McElheran, 2017; DeStefano, Kneller, & Timmis, 2020), there is little conclusive evidence that cloud adoption always adds value to large enterprises and under what organizational contexts the benefits may outweigh costs of adoption.

Filling the gaps in this body of scholarly work requires understanding the organizational aspects of cloud adoption. Large enterprises are not single-unit firms but can be viewed as complex business ecosystems (Ethiraj, 2007; Kapoor & Adner, 2012; Kapoor & Lee, 2013; Lee & Kapoor, 2017; Kapoor, 2018; Agarwal & Kapoor, 2022; Furr, Ozcan, & Eisenhardt, 2022). In the context of cloud adoption, different functional units' decisions to migrate applications to the cloud can be independent across establishments. They make heterogeneous decisions regarding which cloud providers to use and how to combine different technologies (or stacks) to build specific capabilities that meet various demands and pressures, such as regulatory compliance or market competition. Big data ecosystems form around these establishments and their technology stacks. Value creation in such ecosystems depends on complex interactions and complementarities among actors. The technological architecture of such an ecosystem is crucial to defining the system-level goals, members' roles, standards, and interfaces of an ecosystem (Teece, Pisano, & Shuen, 1997), guiding the interactions between actors within the system, and determining value

creation as different components in the system coordinate around tasks (Jacobides, Cennamo, & Gawer, 2018; Agarwal & Kapoor, 2022). Public cloud vendors and on-prem data centers are similar to complementors in a platform ecosystem, which interface with each other in complex ways and are complementary to the technological architecture. The technological architecture determines the patterns in which complementary technologies are incorporated with other components within the system, whether they connect through interfaces or become fully integrated into the system. The technological architecture dictates the interfaces that allow actors to connect and co-evolve with the organization (Boudreau, 2010; Kapoor & Agarwal, 2017).

This paper explores *enterprise data architecture*, which is a type of digital infrastructure that describes how data is collected, stored, transformed, distributed, and consumed within the organization (Dallemule & Davenport, 2017). We focus on two data architecture design features – data interoperability and microservices architecture. Data interoperability is the ability to drive data flows across different functional units and technical systems. It involves a technological architecture that breaks down data silos and systems that maintain "a single source of truth" across data sources. Microservices architecture is a technological architecture that features loosely coupled and independent sub-services, each performing a single function and maintaining its separate database. The microservices architecture underpins a modular system (Baldwin & Clark, 2000; Schilling, 2000; Agrawal, Gans, & Goldfarb, 2021), which can lead to uncontrolled local scaling without clear data ownership and create data silos across the organization that are hard to combine. On the other hand, data interoperability allows firms to control internal data sources and keep track of them to meet the requirements of data audits or data subject requests, which is particularly important in a heavily regulated industry.

The technological architecture can give rise to ecosystem-level interaction mechanisms between actors, which introduce bottlenecks and constrain value creation (Kapoor & Furr, 2015; Adner & Kapoor, 2016; Agarwal & Kapoor, 2022). These bottlenecks can become particularly salient when the external environment undergoes sudden changes that require organizations to adapt their technical systems. In the

realm of big data ecosystems, the most impactful regulatory shock in recent years is the European Union's introduction of the General Data Protection Regulation (GDPR), which governs how companies that deal with customers located in the EU/EEA can collect and process user data, as well as transfer them across geographic boundaries. The highly complex regulatory landscape poses barriers to adaptation. However, firms that fail to adapt are at risk of significant financial penalties if they fail to comply with the new data laws. Past research shows that ecosystem structure and technological architecture can significantly affect organizations' ability to adapt to sudden environmental changes (Levinthal, 1997; Siggelkow & Levinthal, 2003; Aggarwal & Wu, 2015; Burford, Shipilov, & Furr, 2022). We hypothesize that lack of data interoperability and reliance on the microservices architecture lower performance outcomes and negatively affect firms' ability to adapt and comply with data privacy regulations. More generally, a modular approach to cloud adoption and a modular technological architecture that underpins the data ecosystem across establishments are detrimental to organizations because they can lead to data silos and wasted spending on the cloud due to lack of coordination across establishments.

We find empirical results consistent with these hypotheses, using establishment-year panel data on 94,539 observations from 2016 to 2019 for 23 large financial services corporations with operational headquarters in the United States. We use the Aberdeen Computer Intelligence (CI) database to measure annual revenue, employment, and IT staff size at the establishment level. Data architecture variables – data interoperability and microservices architecture – are based on survey data collected by Keystone Strategy and Microsoft Corporation for independent purposes. We measure corporations' exposure to GDPR compliance risks using information about each corporation's geographic distribution of annual total sales, office space, and employment in their annual reports between 2016 and 2019. We use the difference-in-differences regression design to estimate the effects of GDPR enforcement on revenue performance at the establishment level. Our empirical findings show that when the technological architecture does not drive data interoperability or is built on microservices, establishment revenue lowers by about 30% after GDPR enforcement among corporations highly exposed to the EU/EEA (and hence compliance risks). Among

other establishments, we find null results. Using two alternative measures based on cloud vendor configurations associated with a modular approach to technological architecture – the average number of vendors per establishment and intrafirm vendor diversity – we again find evidence consistent with theoretical hypotheses. Our study contributes to the literature on the importance of technological architecture in managing business ecosystems characterized by interdependencies across components (Baldwin & Clark, 2000; Kapoor & Lee, 2013; Baldwin & Woodard, 2009; Kapoor & Agarwal, 2017; Jacobides, Cennamo, & Gawer, 2018; Cennamo, Ozalp, & Kretschmer, 2018), and organizational design for driving firm innovation and transformation (Adner, Puranam, & Zhu, 2019; Kretschmer & Khashabi, 2020; Eklund & Kapoor, 2022; Agarwal & Kapoor, 2022).

2. Theory Development

Data is a valuable strategic resource and dynamic capability (Penrose, 1959; Barney, 1986; Teece, Pisano, & Shuen, 1997). In data-intensive industries, the benefits of data accrue with its scope and scale, which leads to learning and network effects (Farboodi, Mihet, Philippon, & Veldkamp, 2019; Hagiu & Wright, 2020; Gregory, Henfridsson, Kaganer, & Kyriakou, 2021). Complementary to increasingly available data sources is cloud computing technology, an IT service model that delivers on-demand computing resources from a shared pool of distributed hardware and software across disparate environments and locations (Marston, Li, Bandyopadhyay, Zhang, & Ghalsasi, 2011; Benlian, Kettinger, Sunyaev, & Winkler, 2018), which provides the technological foundation for storing and processing such data at scale.

Many enterprises have recently moved their IT workload from on-prem data centers into the cloud. While the cloud can drastically increase the capabilities of large corporations, it can also lead to wasted spending on cloud computing costs without optimizing for efficiency gains. A recent survey shows that enterprises (employing more than 1,000 people) across the United States, Europe and Asia report significant cloud spending in 2021 – 36% of the respondents spend more than $12 million annually, and 83% spend more than $1.2 million annually (Flexera, 2022). These firms cite wasted cloud spending as an important concern as computing costs continue to rise. Hence, understanding what works for cloud adoption and

potential pitfalls leading to wasted resources is crucial to firms formulating cloud strategies to control costs while optimizing for the best use cases and capabilities.

There is no conclusive evidence of whether cloud computing must always benefit large enterprises despite its widespread use (Bloom & Pierri, 2018). Short-term and incremental approaches to cloud adoption create barriers to scaling (Arora, Bawcom, Lhuer, & Sohoni, 2022). Uncontrolled local scaling can result in global inefficiencies, such as latency problems with managing complex tasks that require coordination across disparate systems (Levinthal & Wu, 2010; Hecker & Kretschmer, 2010; Giustiziero, Kretschmer, Somaya, & Wu, 2022). Financial services firms identify the most crucial challenges in migrating workloads to the cloud to include the technical feasibility of migrating an application, understanding interdependencies that make up the application, and selecting suitable virtual machine instances in the cloud to suit workflow needs (Flexera, 2022).

2.1 Importance of Technological Architecture to Establishment Performance

The increasing availability of large volumes of data and the proliferation of cloud computing constitute contextual conditions that trigger new ways of organizing (Bresnahan & Greenstein, 2001; Yoo, Henfridsson, & Lyytinen, 2010; Bresnahan, 2019; Agrawal, Gans, & Goldfarb, 2021; Wessel, Baiyere, Ologeanu-Taddei, Cha, & Blegind-Jensen, 2021; Hanelt, Bohnsack, Marz, & Marante, 2021). For large incumbent corporations that were not “digital-native” firms, the accumulation of data comes with costs as organizational challenges begin to mount when firms need to handle complex IT systems that include decade-old applications that run alongside modern technologies in both on-prem and cloud environments.

Large corporations consist of many establishments, functional areas, product lines, and customer segments (Karim, Lee, & Hoehn-Weiss, 2022). Their technological systems around big data analytics are complex ecosystems that involve many different data sources, infrastructural environments, and software stacks. In platform-based ecosystems, the platform owner determines the architecture design and sets rules for complementors to participate in the ecosystem (Gawer & Cusumano, 2002; Kapoor, 2018; Cennamo, Ozalp, & Kretschmer, 2018). The technological architecture dictates the interfaces that allow actors to

connect and co-evolve with the platform (Boudreau, 2010; Kapoor & Agarwal, 2017). In this literature, complementors are crucial actors within platform ecosystems, and the platforms' choices regarding how it interacts with complementors are crucial to both the platform and the complementors (Gawer & Henderson, 2007; Kapoor & Agarwal, 2016, 2017; Wen & Zhu, 2017; Jacobides, Cennamo, & Gawer, 2018; Rietveld, Schilling, & Bellavitis, 2019). Creating an alignment structure that ensures joint value creation and mitigates conflicts over value capture is vital for strategic interactions between platform and complementors (Gawer & Henderson, 2007; Casadesus-Masanell & Yoffie, 2007; Kapoor, 2013; Kapoor & Lee, 2013; Adner, 2017).

A natural analogy emerges between platform ecosystems and large corporations: the technological architecture around data collection, processing, and usage shapes the interaction across disparate establishments and technical systems, including public cloud vendors and on-prem data centers that serve a similar role to complementors in the platform ecosystems literature. Similar to complementarities and interdependencies between ecosystem actors (Gawer & Henderson, 2007; Kapoor & Lee, 2013; Hannah & Eisenhardt, 2018), the corporations' big data ecosystems are shaped by the technological architecture, which influences value creation and interactions among actors within the ecosystem, by providing a structure to the interdependencies across these actors (Kapoor, 2018). Such technological architecture shapes how establishments and units choose particular cloud vendors and interface with other technologies.

More formally, the technological architecture describing how data is collected, stored, transformed, distributed, and consumed within the organization defines the enterprise data architecture, which is a type of digital infrastructure (Tilson, Lyytinen, & Sørensen, 2010; Henfridsson & Bygstad, 2013; Dallemule & Davenport, 2017) that encompass material systems (e.g., cloud storage and API interfaces) as well as non-material capabilities (e.g., tracking and encryption.) It coordinates the ecosystem of interconnected technologies and infrastructural tools within the corporation that jointly contribute to value creation and capture from dynamic and heterogeneous data sources. Patterns of interactions of ecosystem actors can introduce bottlenecks that constrain value creation (Kapoor & Furr, 2015; Adner & Kapoor, 2016).

Ecosystem structure and technological architecture can affect how firm performance adjusts after abrupt changes in the external environment (Levinthal, 1997; Aggarwal & Wu, 2015; Burford, Shipilov, & Furr, 2022). By guiding how actors within the ecosystem innovate and incorporate new technologies, the technological architecture can facilitate ecosystem-level mechanisms that impact firm performance outcomes by tackling these bottlenecks (either preventing or exacerbating them). Such bottlenecks become especially salient during times of sudden change in the external environment. Firms need to adapt to such changes while maintaining continuity in business activities. Such changes can strain operations enormously when firms' existing technical systems are inertial. Losses from failing to adapt can be much more pronounced compared to the slow performance deterioration in regular times.

2.2 Data Interoperability and Regulatory Compliance: Adapting to Privacy Regulation Shock

Privacy regulations, especially the European Union's General Data Protection Regulation (GDPR), constitute the most profound regulatory shocks with wide-ranging effects across data-intensive firms worldwide in recent years. GDPR established standard rules around data processing to protect consumers' privacy applicable not only to firms operating in the EU/EEA but also to firms outside the area that cater to customers in the EU/EEA or collect data about them (Peukert, Bechtold, Batikas, & Kretschmer, 2022). Ensuring compliance may require corporations to change various practices and technological implementations of their technology systems.

Many organizations have struggled tremendously with GDPR compliance, especially concerning cloud deployment. In response to the regulation, large cloud vendors respond by expending large amounts of resources on projects that help firms navigate GDPR compliance risks. Microsoft has put more than 1,600 engineers on GDPR-related projects, and Amazon Web Services introduced over 500 new features and services on security and compliance. When enterprises use cloud service providers for data storage and processing needs, both parties are responsible for compliance as "data controllers" and "data processors" under privacy regulations. When firms design a data system that incorporates on-prem resources and cloud providers, one of the most important aspects has become how the system may protect the privacy of data

subjects, which may mean that existing infrastructure that is not up to this task may have to be removed or replaced to ensure compliance with the regulation.

Compliance requires setting up robust systems and changing business processes, which can be costly and time-consuming. Third-party audits are necessary to ensure commitment to compliance, and spending resources to fulfill data-subject requests can further lower operational efficiency. Data can move quickly in the cloud across geographic locations and jurisdictions where different laws apply or simultaneously reside in multiple locations. This introduces complexities as cloud vendors may operate in various jurisdictions where data subjects (e.g., consumers) have different rights according to the regulation. Some regulations specifically govern data transfers between regions, e.g., Standard Contractual Clauses (SCC), also known as the EU Model Clauses, which are model contracts for transferring personal data from the EU/EEA. The complex landscape of data regulation makes it particularly difficult to measure, let alone ensure compliance. In addition, GDPR may only be one of many different privacy regulations applicable to companies in the United States at national and state levels, e.g., the California Consumer Privacy Act of 2018 (CCPA) and the Clarifying Lawful Overseas Use of Data Act (CLOUD Act). Simultaneously satisfying all the obligations under different regulatory frameworks may be difficult for organizations seeking to design and implement privacy compliance programs.

Ensuring regulatory compliance and controling the risks of violation require the organization to have the ability to track all its data sources and the movement of data within the organization. Data interoperability, which refers to the "functionality of information systems to exchange data and to enable sharing of information" (European Data Protection Supervisor), is meant to give data subjects enhanced control of their data and foster competition across platforms and digital services. Firms' internal data interoperability ensures that data can flow between departments and functional areas, allowing the firms to maintain "a single source of truth" regarding various metrics implied by the data. This allows organizations to have a holistic view of all their data sources and maintain the ability to track data systems robustly. This allows firms to meet demands from data subjects and third-party audits to provide truthful reports of the

organization's data assets, monitor information flows, and control compliance risks. Hence, data interoperability positively affects performance outcomes and lowers the costs associated with adaptation to meet compliance requirements for more stringent regulations. We propose:

H1: Among establishments of financial services corporations exposed to significant data regulation compliance risk, those that belong to corporations using standardized service interfaces to facilitate data interoperability experience significantly less severe revenue losses than other establishments.

2.3 Microservices Architecture and Adapting to Privacy Regulation Shock

A popular data architecture design is the microservices architecture, consisting of loosely coupled and self-contained sub-services. Each sub-service implements a single function, and all the components work jointly to form a complete application. In contrast to the traditional monolithic architecture with a single codebase and tightly coupled components, the microservices architecture maintains separate data stores and codebases for each independent sub-service. The entire architecture involves multiple microservices that form an ecosystem of building blocks to orchestrate the processing and movement of data.

Microservices is a lightweight architecture that is easier to scale and compatible with rapid scaling of cloud adoption due to its flexibility and suitability for interfacing with external technology providers (Bharadwaj, El Sawy, Pavlou, & Venkatraman, 2013; Giustiziero, Kretschmer, Somaya, & Wu, 2022). The microservices architecture constitutes a modular system consisting of loosely coupled components that can be taken apart and recombined into new configurations (Simon, 1962; Langlois & Robertson, 1992; Ulrich, 1995; Sanderson & Uzumeri, 1995; Baldwin & Clark, 2000; Schilling, 2000; Campagnolo & Camuffo, 2010; Agrawal, Gans and Goldfarb, 2021). Such modular systems have high degrees of openness and are conducive to strategies that involve value co-creation with external actors. It is also the technological foundation of the increasingly popular approach of "cloud-native" applications and data platforms built by stacking different components together, including open-source tools, each serving a single functional purpose such as frameworks (e.g., Presto), data pipeline orchestration (e.g., Airflow), building real-time streaming applications (e.g., Apache Kafka).

However, in many cases, microservices architecture may be harmful to organizations. In large corporations, the degree of modularity across establishments affects how they coordinate to work on joint tasks (Baldwin & Clark, 2000; Schilling, 2000; Englmaier, Foss, Knudsen, & Kretschmer, 2019). Under the microservices architecture, components do not share the execution contexts and lack a shared database. This can create various issues. It can introduce latency in applications that depend on multiple microservices because processing requests involving multiple services takes time. Communication between these services is required since they do not share information otherwise. If components within the system do not communicate, distributed data sources can become siloed. This leads to variations of truths that are not accountable by any data owners; the firm keeps multiple copies of the data object, which diverge after being modified by different processes. More generally, the technological architecture's global ramifications are important to consider, in addition to local efficiencies.

The microservices architecture can constrain adaptation to regulatory shocks, which involve stringent requirements for controlling internal data. Data siloes are associated with a lack of interoperability, making it impossible for the monitoring system to cover all data sources and enforce governance at the level of the whole organization. Lack of visibility makes it technically difficult to track data needs and cloud costs and ensure data integrity at each interface where data is transferred from one microservice to another. The lack of clearly defined ownership constitutes an organizational problem with decentralized data platforms and engineering teams. All these factors raise compliance risks and can be costly to firm performance.

Generally, interdependencies spanning component boundaries are more difficult to manage when firms face negative shocks and are constrained by such interdependencies to adapt effectively (Aggarwal & Wu, 2015). This is particularly problematic when modularity is associated with uncontrolled local scaling, which introduces complex interdependencies that create obstacles to making the necessary technical changes required to adapt to regulatory shocks. Interfaces between modular components become bottlenecks to adaptation when coordination hubs are congested by complex sourcing relationships (Zhou & Wan, 2017).

The microservices architecture can also introduce security risks, as it increases the number of entry points into the system.

In addition, the rise of container technology poses technical difficulties to migrating legacy applications. These legacy applications are often not implemented via microservices as they were written long ago (Flexera, 2022). Business users and IT professionals used to handling on-prem workloads may find it challenging to understand the microservices architecture, which requires familiarity with not only the overall application but also each sub-service that constitutes the entire application. Hence, it is technically more demanding to develop and manage microservices architecture than traditional monolithic architectures. For all these reasons, the microservices architecture can constrain the adaptation required to comply with more stringent data privacy regulations. Hence, we propose the following:

H2: Among establishments of financial services corporations exposed to significant data regulation compliance risk, those that belong to corporations with data fabric or platforms built on the microservices architecture experience significantly more severe revenue losses than other establishments.

2.4 Average Establishment Cloud Vendor Interoperability

A large majority of enterprise cloud adopters use multiple cloud providers. While top public cloud vendors – Amazon Web Services, Microsoft Azure, and Google Cloud Platform dominate the cloud market, none of them is strictly optimal for meeting all cloud computing needs of enterprise users. However, using multiple vendors for different needs may impose technical challenges. The cloud vendors do not naturally cooperate by default, as they all compete to have users choose their services over their competitors.

Cloud vendor interoperability is crucial to successfully scaling applications across multiple cloud vendors. This is especially relevant for large corporations, which may face even more daunting coordination challenges across multiple locations and functional units. As the scope of services required for the business increases, choosing the appropriate cloud vendor for different needs becomes more important, as prices and capabilities can differ substantially across cloud vendors. A multi-cloud strategy involves selecting different

cloud services across multiple vendors based on optimizing pricing and needs. By having the flexibility to shift workloads across multiple cloud providers, firms can avoid vendor lock-in and improve disaster recovery and operational efficiency in migrating data and applications.

The technical challenge that a successful multi-cloud strategy overcomes is to eliminate data silos and drive interoperability across establishments. This involves having an API access point for each cloud infrastructure, which creates an interface through which different data sources and cloud providers can achieve a single source of truth. Hence, when firms use multiple cloud vendors at the same establishment, an API platform that drives data interoperability is highly complementary for executing an effective multi-cloud strategy. Apart from eliminating data silos, it can also improve data governance and lowers compliance risks of data regulations. When regulatory changes restrict what firms are allowed to do with one vendor, they may shift the workload to a different vendor. By diversifying risks across multiple vendors, firms suffer less loss from having to adapt to ensure regulatory compliance.

Hence, we propose:

H3a: A lack of data interoperability across functional units and the presence of a data platform built on microservices architecture are associated with a lower likelihood of having a multi-cloud strategy and a lower average number of vendors per establishment.

H3b: Among establishments of financial services corporations exposed to significant data regulation compliance risk, those that belong to corporations where the average establishment uses a smaller number of cloud providers experience significantly more severe revenue losses than other establishments.

2.5 Intrafirm Diversity in Cloud Vendor Configurations and Uncoordinated Cloud Adoption Across Establishments

The modular approach to designing enterprise data systems, which often involves building data platforms on the microservices architecture, is associated with decentralized management and low-cost local scaling. However, modular architecture can lead to uncontrolled local scaling. Local decisions do not

account for coordination with other establishments, resulting in suboptimal patterns at the organizational level. Second, local changes to data may result in confusion in ownership when multiple users modify and keep copies of the original data. When two teams make uncoordinated decisions to edit the same source data, it can lead to variations of truths that will eventually hurt business performance.

We can spot signs of modular and decentralized architecture in the configuration of cloud vendors across establishments within a corporation. Different cloud vendor configurations across establishments indicate fragmented adoption that is not controlled or planned centrally. For example, when cloud adoption decisions are made locally and independently, the resulting choices of cloud providers can vary widely across different establishments. Different establishments end up using very different vendors. These technical decisions are associated with either a lack of intention to drive toward interoperability or practical difficulties in achieving interoperability.

When large corporations have a diverse base of external stakeholders, data interoperability can facilitate standardization efforts that allow data infrastructure to be managed appropriately in a controlled manner (Star & Ruhleder, 1996; Constantinides & Barrett, 2015). Regardless of the particular technical solutions applied to designing the technological architecture, driving data interoperability across functional units (and hence across cloud operators and establishments) often lead to standardization in technology provider choices across establishments. Data interoperability forces firms to develop simpler global structures of their various data sources and infrastructural tools. Robust internal integration allows firms to align data sources and capabilities through centralized platforms. This is particularly crucial to heavily regulated industries and organizations where internal control is important to organizational performance. Failing to coordinate cloud vendor configurations across locations can raise technical difficulties in driving toward interoperability and worsen performance outcomes, especially when firms must adapt to comply with more stringent data regulations.

Hence, we propose:

H4a: A lack of data interoperability across functional units and presence of data platform built on microservices architecture are associated with more fragmented cloud adoption and higher intrafirm diversity in cloud vendor configurations.

H4b: Among establishments of financial services corporations exposed to significant data regulation compliance risk, those that belong to corporations with higher intrafirm cloud vendor diversity experience significantly more severe revenue losses than other establishments.

3 Data and Methods

3.1. Measuring Cloud Adoption and Other Establishment-Level Variables

To conduct the empirical analyses, we use data on establishment-level variables for 23 large financial services corporations with operational headquarters in the United States. Among these corporations are some of the largest U.S. banks, asset management firms, and insurance companies, each with hundreds and thousands of branches across all U.S. states. Our primary data source is the Aberdeen Computer Intelligence (CI) database, which contains establishment-level information that allows us to measure variation in performance outcomes and technology adoption within the same corporation across establishments and locations.

3.1.1. Cloud Adoption and Vendor Choices

Large corporations in the financial services sector adopt cloud computing at very high rates and predominantly choose services from the top three public cloud vendors – Amazon, Microsoft, and Google. According to a survey by Flexera, 79% of large enterprises use Amazon Web Services, 76% use Microsoft Azure, and 49% use Google Cloud Platforms as their primary cloud provider (Flexera, 2022). Other market surveys and industry reports show similar rates of large enterprises adopting services from the top three public cloud vendors (e.g., Shah, 2018; Furber, 2020). In addition, firms often use more than one cloud but choose primary and subsidiary providers from the same small set of top vendors. Bank regulators, including the Federal Reserve and the FDIC in the United States, and the European Banking Authority have raised

concerns about financial institutions' reliance on a few large providers to meet technology needs and voiced concerns about managing associated security and compliance risks (Nguyen, 2022). These cloud vendors generate hundreds of billions of annual revenues by renting server capabilities to other firms.

We obtain data on cloud adoption among establishments of large financial services corporations from the Aberdeen Computer Intelligence (CI) database. The database contains detailed information about technology providers across U.S. firms, with particularly comprehensive coverage of establishments since 2017. To measure cloud adoption and vendor choices for each establishment, we focus on all cloud technologies provided by Amazon (AWS and related technologies), Microsoft (Azure and Microsoft System Center), and Google (Google Cloud Platform and Compute Engine). Figure 1 panel (a) shows that almost 90% of the establishments in the data adopted cloud services from at least one of the three largest providers. However, the figure may disproportionately reveal the cloud adoption rates of a few very large corporations, as each establishment counts equally toward the aggregate statistic, and some corporations have many more establishments than others. Specifically, the three largest institutions – Wells Fargo, Bank of America, and JPMorgan Chase – jointly account for 70% of overall establishments in the sample.

[Insert Figure 1 here]

Figure 1, panel (b) shows that when we exclude these three largest corporations from the sample, the average cloud adoption rate was below 70% in 2019. In addition, many establishments using public providers in 2018 no longer report the presence of any public cloud providers in 2019, suggesting the prevalence of reverse adoption following GDPR enforcement in mid-2018.

The data is much more suited for measuring vendor configurations (and hence multi-cloud strategies) and less suited for measuring the shares of public versus private clouds (and hence hybrid cloud strategies). The top three public cloud vendors also offer private cloud stacks. Though firms may use a mixture of private and public clouds, the data does not distinguish between them when the same vendor provides both. For example, an establishment using Microsoft System Center can involve both Azure public cloud and private cloud stacks. However, no further information is available about the fraction of workloads run in

the public versus the private cloud. On the other hand, there is much less uncertainty in measuring specific vendor choices, as it is apparent in the data when firms use services from Amazon or Microsoft. Hence, we focus on the information at the vendor level and the presence of each of the top three public cloud vendors.

3.1.2. Establishment Revenue and Control Variables

The primary performance outcome is the annual revenue observed in the Aberdeen Computer Intelligence (CI) database. All the revenues are measured in millions of U.S. dollars and rounded to the nearest integer amount. The precise outcome variable equals the logarithm of the dollar value of the revenue at the establishment level. The Aberdeen Computer Intelligence (CI) database updated its data collection methodology in 2017, which led to improved data quality and broader coverage of establishments compared to before 2017. For example, in 2016, more than 95% of all establishments had zero revenue, most of which were missing values instead of actual zeros. In contrast, less than 10% of establishments recorded zero revenue in 2017, similarly for each year after that. Hence, we drop establishment-year observations with zero revenues in 2016 and only include establishments with non-zero 2017 revenues. In the final sample, the autocorrelation in log revenues is in a reasonable range across all years (around 0.9).

We also measure two establishment-level control variables: total employment and IT staff size. Employment equals the number of employees, and the IT staff size is measured in ranges from the following: 1-4, 5-9, 10-24, 25-49, 50-99, 100-249, 250-499, 500-1000, and 1000 or more. We take the mid-points of each interval (except for "1000 or more", which we take as 1000) as the estimated number of IT employees for each establishment.

To link a corporation to establishment-level data, we first find the unique establishment identifier (the DUNS number[1]) of its corporate headquarter by manually looking up the corporation's name through Mergent Intellect (a business intelligence web aggregator of company profiles). We then identify all the

[1] Developed by Dun & Bradstreet (D&B), the DUNS (Data Universal Numbering System) number assigns a unique identifier to each single business location (i.e., an establishment of a private business) in the United States.

establishments in the CI database with the same enterprise ID as the corporate headquarter and include them in the sample.

3.2. Measuring Exposure to the EU/EEA Market and GDPR Compliance Risks

European Union's General Data Protection Regulation (GDPR) has been the most comprehensive privacy regulation introduced to harmonize privacy laws and enforcement throughout the European Union in recent years (Peukert, Bechtold, Batikas, & Kretschmer, 2022). The GDPR establishes standard rules around data processing to protect consumers' privacy and define violations based on user consent. It substantially expanded previous privacy regulations in the EU/EEA regarding geographic reach and violation penalties. The 88-page text describes very complex data protection laws in detail, which requires a deep understanding of the law for firms to comply with the regulation. Many companies, particularly those that use cloud providers to store and process user data, have struggled with compliance since the GDPR became widely applicable. As potentially hefty fines in the case of violations impose significant risks to firms' profits, GDPR enforcement can negatively affect firm performance as firms must incur adaptation costs which sometimes involve substantially transforming organizational practices to comply with the regulation or face substantial financial risks associated with privacy violations.

The GDPR became enforceable on May 25, 2018, and applicable to companies conducting business in Europe and those providing products and services to customers who live in the EU/EEA. Some corporations studied in our sample are multinational companies with substantial operations in European countries and cross-border transactions involving EU/EEA residents. These activities fall under the jurisdiction of the GDPR and are exposed to substantial compliance risks. There is substantial variation in the exposure to GDPR compliance risks across the sample, which we measure below. On the other hand, the geographic distribution of business activities typically results from a string of expansion decisions made long ago and remains relatively stable in recent years.

To measure a corporation's exposure to GDPR compliance risks, we manually collect data about business activities in the EU/EEA from each corporation's annual reports (10-Ks) between 2016 and 2019.

Some companies only report information about the EMEA region (which does not only include Europe but also includes the Middle East and Africa). Other companies may only report information about the Americas and Asia, from which we infer the upper bound of the share of activities in the EU/EEA. To ensure consistent measurement across the sample, we code each variable from the following list by looking up relevant information in the annual reports:

- mentioning of the GDPR (usually in the context of discussing compliance)
- Share of total sales in the EU/EEA (or EMEA) if available, otherwise share of total sales outside the Americas
- Share of office space or number of offices in the EU/EEA (or EMEA) if available, otherwise share of office space or number of offices outside the Americas
- Share of employees in the EU/EEA (or EMEA) if available, otherwise share of employees outside the Americas

We define ***high exposure to GDPR compliance risks*** as satisfying the following two criteria: (1) the annual report in at least one year from 2016 – 2019 mentions the GDPR, and (2) activity share in the EU/EEA or EMEA is 15% or higher (if available), otherwise activity share outside the Americas is 40% or higher (if data on EU/EEA or EMEA are not available). Some annual reports only contain partial information about activity shares, so we focus on what is available. We collapse each of the above variables (if available) across years and take the simple average across variables to define an overall activity share. Table 1 lists all corporations in the sample and information determining each corporation's exposure to GDPR compliance risks.

[Insert Table 1 here]

3.3. Survey-Based Measurement of Enterprise Data Architecture Characteristics

The enterprise data architecture is the technological backbone that aligns data-related capabilities with business needs (Saldanha, Lee, & Mithas, 2020; Correani, De Massis, Frattini, Petruzzelli, & Natalicchio,

2020). We obtain information about enterprise data architecture using survey data collected by Keystone Strategy LLC and Microsoft Corporation for independent purposes unrelated to this study. Data collection involved teams conducting in-person interviews with senior corporate leaders who oversee the digital initiatives of the entire organization. We map technical information from the survey to theoretical concepts relevant to eco-system modularity. We focus on two variables: *data interoperability* and *microservices architecture*, and describe each of them below.

Data Interoperability. Data interoperability refers to the capability of the technological architecture to drive coordination among different components within the ecosystem through service interfaces. The survey questions on data interoperability are as follows: *"Do you use services interfaces or APIs to publish departmental (e.g., not application) information between departments and/or functional areas?"*, and if the answer is affirmative, a second question is asked: *"Do you have a standardized convention for these APIs, such that data can flow between different departments (e.g., are you driving towards interoperability)?"* We code the $Interoperability$ variable as 1 if the answer to the second question is affirmative, and 0 if it is negative or missing (which happens when the response to the first question is negative).

Microservices Architecture. Microservices characterize a technological architecture where a collection of loosely coupled and self-contained sub-services form an entire system. Each sub-services operates independently to implement a single business function and maintains its own separate database. The microservices architecture is often the technical foundation for orchestrating cloud-native application platforms and container deployment. The survey question on microservices architecture is as follows: *"Is your data fabric or platform built on a microservices architecture?"* We code the $Microservices$ variable as 1 if the answer to the question is affirmative, and 0 otherwise.

3.4. Generalized Measurement of Data Architecture Modularity using Cloud Vendor Data

While the technological architecture can be measured by surveying people involved in designing the relevant systems, we can also infer these designs using data on the observed implementation or configurations of core technology components of these systems. Establishment-level data on the presence

of cloud products and vendors contain diagnostic information about the underlying technological architecture that led to specific configurations or patterns of adoption. We use information about exactly what cloud providers are present within each establishment to summarize various dimensions of cloud vendor configurations at the organization level. More specifically, we derive two corporation-level variables based on the cloud vendor choices *within an establishment* and *across establishments within the organization*, both of which are associated with the modularity and interoperability of the underlying design of the technological architecture.

Average Number of Cloud Vendors Per Establishment. The first variable measures the average within-establishment cloud adoption decision in relation to multi-cloud strategy and cloud vendor interoperability. The variable simply counts the number of different cloud vendors at each establishment (among the top-3 public cloud providers, Amazon, Microsoft, and Google), and calculates the average number of establishment-level vendors to derive an overall indicator of the prevalence of multi-cloud approach (and hence cloud vendor interoperability).

Intrafirm Cloud Vendor Diversity Across Establishments. The second variable measures the extent to which establishments make cloud adoption decisions independently, hence resulting in observed heterogeneity in vendor choices across establishments. More formally, we represent *cloud vendor configuration* at an establishment j of corporation i by the vector $\boldsymbol{\theta}_{ij}$ as the 3-dimensional vector as there are three largest public cloud providers (Amazon, Microsoft, and Google) that account for the lion's share of the enterprise cloud market. The k-th element (k=1, 2, or 3) of the vector is a binary indicator of whether an establishment adopted the k-th cloud vendor. We then define the average configuration across all establishments in a corporation as $\overline{\boldsymbol{\theta}}_i$. We denote the total number of establishments by N_i, and the cosine similarity between the average and the establishment-specific cloud vendor configuration as $\cos\left(\boldsymbol{\theta}_{ij}, \overline{\boldsymbol{\theta}}_i\right)$.

Based on these notations, we construct the $IntrafirmVendorDiversity$ variable, which measures the extent to which cloud service needs are met by the same set of vendors or span many different

configurations. To operationalize the measure, we apply the methodology of measuring information novelty on networks to our setting (Holtz et al., 2020; Aral & Dhillon, 2022) as follows.

$$IntrafirmVendorDiversity_i = \frac{1}{N_i}\sum_{j=1}^{N_i}\left[1 - \cos\left(\boldsymbol{\theta}_{ij}, \overline{\boldsymbol{\theta}_i}\right)\right]^2 \quad (1)$$

The measure ranges from 0 to 1, and larger values indicate higher intrafirm diversity in cloud vendor configurations across establishments. When the value of this variable is 0, all establishments adopting cloud have the exact same cloud vendor configuration; when establishments have very different cloud vendor configurations, the value of this variable becomes larger (and up to 1). This measure is particularly relevant for large corporations spanning hundreds and thousands of establishments. A high diversity in intrafirm cloud vendor configurations indicates fragmented cloud adoption decisions that do not require coordination across locations, which is diagnostic of a lack of interoperability across functional units, or a decentralized approach to cloud adoption decisions.

3.5. Estimation Strategy

We use a difference-in-differences regression design to estimate the differential impact of GDPR enforcement on revenue performance across an unbalanced panel of financial services establishments from 2016 to 2019. Equation 2 below describes the regression model. The outcome variable (Y_{ijt}) is the logarithm of the annual revenue measured in U.S. dollars at the firm-establishment (ij) and year (t) level. The treatment variable ($HighGDPR_i$) indicates high exposure to GDPR compliance risks which varies at the corporation (i) level, defined using business activity shares in the EU/EEA or EMEA market in Section 3.2. The event time variable $Post_t$ equals 1 when the year is 2018 or later, and 0 otherwise. The model controls for firm-establishment fixed effects ϕ_{ij} and year fixed effects η_t. Robust standard errors are clustered at the treatment level (by corporation). The interaction effect β_4 estimates the impact of GDPR enforcement on establishment revenues when firms are highly exposed to GDPR compliance risks. We can estimate this effect in different subsamples and compare the coefficient estimates across these subsamples.

$$Y_{ijt} = \beta_1 HighGDPR_i + \beta_3 Post_t + \beta_4 HighGDPR_i \times Post_t + \phi_{ij} + \eta_t + \epsilon_{ijt} \quad (2)$$

To show how the impact of GDPR on establishment performance may vary across features of the technological architecture around data storage and processing, we split the sample according to the values of corporation-level data architecture variables. For example, we estimate Model 2 separately on corporations with $Interoperable$ equal to 1 (IO=1) and on corporations with $Interoperable$ equal to 0 (IO=0). Then we compare the two coefficients β_4 obtained from estimating the model on the two subsamples and test whether they are statistically significantly different from each other and zero. We can also investigate how effects accrue over time by estimating an event-study variant of the model, replacing the $Post_t$ dummy with separate indicators for each year from 2016 to 2019 and omitting the year 2017 prior to GDPR enforcement as the baseline category for identification.

We also use an alternative model in Equation 3 to estimate the differential change in establishment revenues after GDPR enforcement by features of the enterprise data architecture (e.g., interoperability and microservices) or generalized measures of technological architecture modularity. Instead of defining the treatment variable as having high exposure to GDPR compliance risks, we estimate Model 3 on the subsample of corporations with high exposure (i.e., substantial activity shares in the EU/EEA market). We estimate the model using different treatment variables ($Architecture_i$), and the interaction effect γ_4 provides an alternative for calculating the differential impact of GDPR enforcement to comparing coefficients β_4 obtained from Model 2 on different subsamples defined by data architecture variables.

$$Y_{ijt} = \gamma_1 Architecture_i + \gamma_3 Post_t + \gamma_4 Architecture_i \times Post_t + \varphi_{ij} + \mu_t + \nu_{ijt} \quad (3)$$

4. Empirical Results

4.1. Descriptive Statistics

Table 2 shows descriptive statistics and pairwise correlations between the variables used in estimating regression models in Section 3.5 on the sample consisting of an unbalanced panel of 94,539 firm-establishment-year observations across 23 large financial services corporations (see the list in Table 1). Six

of these corporations are determined (by criteria in Section 3.2) to have high exposure to GDPR: American International Group, Chubb Limited, Citigroup, Mastercard Incorporated, MoneyGram International, and The Goldman Sachs Group. They account for about 14% of the firm-establishment-year observations in the full sample. Nine corporations use standardized service interfaces to facilitate data interoperability between functional areas. Ten corporations build data fabric or platforms on the microservices architecture.

[Insert Table 2 here]

Table 2, panel (a) shows that the median observation in the sample has an annual revenue of $3 million. About 56% of the observations belong to corporations that use standardized service interfaces to facilitate data interoperability between functional areas. About 58% of the observations belong to corporations that build data fabric or platforms on the microservices architecture. Table 2, panel (b) shows a substantial negative correlation of -0.382 between $Interoperable$ (IO) and $Microservices$ (MS). The survey-based data architecture variables are uncorrelated with establishment characteristics such as revenue, employment, and IT staff size.

4.2. Enterprise Data Architecture and Establishment Performance After GDPR Enforcement

In this section, we report the empirical results based on estimating the regression models in Section 3.5 and compute the p-value for each coefficient. Tables 3 and 4 focuses on comparing the effects of GDPR enforcement on post-shock performance across different subsample splits based on features of the enterprise data architecture. All columns control for firm-establishment fixed effects, year fixed effects, and time-varying employment and IT staff size.

4.2.1. Effects of Data Interoperability

In Table 3, results estimated from Model 2 in Section 3.5 are reported on samples split into corporations with data interoperability equal to *no* (columns 1–2) and *yes* (columns 3–4). The model coefficients are estimated separately for each sub-sample. Column 1 shows that the effect of GDPR enforcement on establishment revenue is an estimated 30% loss relative to establishments unaffected by GDPR (p-value<0.001) across corporations without data interoperability. In contrast, column 3 shows a

null effect of the regulatory shock across corporations with data interoperability. To ensure that the results are robust and apply widely to most firms in the sample, instead of being driven by the largest corporations, we report the results from estimating the same regression models after excluding the three largest corporations, each accounting for 20% or more of the overall establishments. After applying this restriction, the regressions yield similar regression results (columns 2 and 4) compared to those estimated on the full sample.

[Insert Table 3 here]

In column 5, we show regression estimates from Model 3 in Section 3.5, based on the subsample of corporations with high exposure to GDPR. Relative to establishments of corporations without data interoperability, the revenue change after GDPR enforcement is 31% higher among establishments of corporations with data interoperability. These results suggest that data interoperability has a large positive effect on the performance outcome of corporations exposed to compliance risks associated with GDPR enforcement.

Figure 2 illustrates the estimates graphically and reports the estimated log revenue in each year before and after GDPR enforcement, with 2017 excluded as the baseline year. The 95% confidence intervals are plotted as bars around each coefficient estimate. Panel (a) shows results estimated on the subsample consisting of corporations with data interoperability, and panel (b) shows results estimated on the subsample consisting of corporations without data interoperability. Before GDPR enforcement in 2018, establishment performance appeared to be on a similar trajectory between high GDPR exposure firms and other firms, which is evidence that support the parallel trends assumption. Starting in 2018, the establishments of high-exposure firms experienced significant revenue losses without data interoperability, as shown in panel (a). On the other hand, no statistically detectable difference is observed in panel (b), which compares the performance trajectories between high-exposure firms and other firms with data interoperability. These results are consistent with *Hypothesis H1* in Section 2.2.

[Insert Figure 2 here]

4.2.2. Effects of Microservices Architecture

In Table 4, results are reported on samples split into corporations with microservices architecture equal to *yes* (columns 1–2) and *no* (columns 3–4) and estimate coefficients separately for each sub-sample. Column 1 shows that the effect of GDPR enforcement on establishment revenue is an estimated 25% loss relative to establishments unaffected by GDPR (p-value<0.001) across corporations with microservices architecture. In contrast, column 3 shows a null effect of GDPR enforcement across corporations without microservices architecture. To ensure that the results are robust and apply widely to most firms, instead of being driven by the largest corporations, we estimate the regression model after excluding the three largest corporations, each accounting for 20% or more of the overall establishments. After applying this sample restriction, the regressions yield similar results (columns 2 and 4) to those estimated on the full sample.

[Insert Table 4 here]

In column 5, we show the regression results from estimating Model 3 in Section 3.5 on the subsample of corporations with high exposure to GDPR. Relative to other establishments, the revenue suffers a 23% more severe drop after GDPR enforcement among establishments of corporations with microservices architecture. These results suggest that microservices architecture has a large negative effect on the performance outcome of corporations exposed to compliance risks associated with GDPR enforcement.

Figure 3 illustrates the estimates graphically and reports the estimated log revenue in each year before and after GDPR enforcement, with 2017 excluded as the baseline year. The 95% confidence intervals are plotted as bars around each coefficient estimate. Panel (a) shows results estimated on the subsample consisting of corporations with microservices architecture, and panel (b) shows results estimated on the subsample consisting of corporations without microservices architecture. Before GDPR enforcement in 2018, establishment performance appeared to be on a similar trajectory across corporations with different levels of GDPR exposure, which is evidence that supports the parallel trends assumption. In 2018 and after, establishments of high-exposure firms with microservices architecture experienced significant revenue losses. On the other hand, no statistically detectable difference is observed in the performance between

high-exposure firms and other firms without microservices architecture since 2018. These results are consistent with *Hypothesis H2* in Section 2.3.

[Insert Figure 3 here]

4.3. Relating Cloud Strategies to Enterprise Data Architecture

This subsection explores the relationship between enterprise data architecture and within-organization cloud vendor configurations across establishments. We show that within-establishment cloud vendor interoperability and intrafirm vendor diversity (defined in Section 3.4) vary significantly across survey-based measures associated with a modular approach to enterprise data architecture design, indicated by both a lack of standardized interfaces for driving data toward interoperability and the presence of data platform built on a microservices architecture. On the other hand, a modular approach to cloud adoption leads to a lack of vendor interoperability within establishments and uncoordinated cloud adoption decisions across establishments. Our data are consistent with the arguments above.

Figure 4 uses violin plots to contrast the distributions of the cloud vendor variables across different sample splits. Each panel compares the distribution of an outcome variable between the six corporations without data interoperability and using microservices architecture (IO=0, MS=1) and the other 17 corporations. The outcome variable in panel (a) is the average number of cloud vendors per establishment. The outcome variable in panel (b) is the intrafirm diversity of cloud vendor configurations across establishments defined in Equation 1 of Section 3.4.

[Insert Figure 4 here]

These plots show that the corporations without data interoperability and using microservices architecture (IO=0, MS=1) have a substantially *lower* average number of cloud vendors per establishment and a substantially *higher* intrafirm diversity in cloud vendor configurations across establishments. These patterns are not driven by one or two outliers or skewed distributions. Instead, they reflect shifts across the entire distribution.

Table 5 shows results from two-sample t-tests across the same sample split underlying Figure 4. The difference in the means of the average number of cloud vendors per establishment between samples is -0.723 (p=0.013). The difference in the means of the intrafirm cloud vendor diversity between samples is 0.015 (p=0.006). These numbers are consistent with Figure 4 and show the substantial differences in means across the subsamples, which are statistically significant despite the small sample size of only 23 corporations. On the other hand, none of the other corporation-level characteristics, including firm size and revenue, show significant differences between the subsamples. In other words, these corporations appear balanced across other dimensions. These statistics support *Hypotheses H3a* in Section 2.4 and *Hypothesis H4a* in Section 2.5.

[Insert Table 5 here]

4.4. Generalized Measurement of Data Architecture Based on Cloud Vendor Configurations

In the previous subsection, we revealed correlational patterns that indicate a strong association between a modular approach to cloud adoption and enterprise data architecture designs that indicate a lack of data interoperability and reliance on microservices. Hence, we can use cloud vendor data to derive generalized measures of modularity of large corporations' technological architecture around data storage and processing by using cloud vendor data. This subsection shows substantial evidence that these generalized measures are indeed associated with heterogeneous impacts of GDPR enforcement on revenue performance.

Table 6 reports results from estimating Models 2 and 3 in Section 3.5. on subsamples of corporations with low (<1.5) versus high average per-establishment number of cloud vendors. Column 1 shows that across corporations with *low* cloud vendor interoperability, the effect of GDPR enforcement on establishment revenue is an estimated 29% loss relative to establishments unaffected by GDPR (p-value<0.001). In comparison, column 2 shows a null effect of the regulatory shock across corporations with high cloud vendor interoperability on average. Column 4 shows the results from estimating Model 3 on the subsample of corporations with high GDPR exposure. The sizes of the treatment group (low cloud vendor interoperability) and control group underlying the estimation are balanced: about 20% of overall

establishments are in the treatment group. Establishments of corporations with low cloud vendor interoperability experience 24% lower performance after GDPR enforcement than those with high cloud vendor interoperability.

[Insert Table 6 here]

In Table 7, we report results from estimating the models on subsamples of corporations with high (>0.01) versus low intrafirm cloud vendor diversity across establishments. Column 1 shows that across corporations with *high* intrafirm cloud vendor diversity, the effect of GDPR enforcement on establishment revenue is an estimated 25% loss relative to establishments unaffected by GDPR (p-value<0.001). In comparison, column 2 shows a null effect of the regulatory shock across corporations with low intrafirm cloud vendor diversity. Column 4 shows the results from estimating Model 3 on the subsample of corporations with high GDPR exposure. The sizes of the treatment group (high intrafirm vendor diversity) and control group underlying the estimation are balanced: about 21% of overall establishments are in the treatment group. Establishments of corporations with high intrafirm vendor diversity experience 30% lower performance after GDPR enforcement than those with low intrafirm vendor diversity.

[Insert Table 7 here]

To ensure that the results are robust and apply widely to most firms, instead of being merely driven by the largest corporations, we report the results from estimating the same regression models after excluding the three largest corporations, each accounting for 20% or more of the overall establishments. After applying this sample restriction, estimation yields similar results to those based on the full sample, as shown in column 3 of both Tables 6 and 7. All these corporations belong to high cloud vendor interoperability and low intrafirm vendor diversity groups.

Figure 5 illustrates the estimates graphically and reports the estimated log revenue in each year before and after GDPR enforcement (2017 is excluded as the baseline year), using an event-study variant of Model 2. The 95% confidence intervals are plotted as bars around each coefficient estimate. Panel (a) shows results

estimated on corporations with low (<1.5) cloud vendor interoperability and panel (b) shows results estimated on corporations with high (>0.01) intrafirm cloud vendor diversity. Before GDPR enforcement in 2018, the trajectories of establishment performance appeared similar between groups, consistent with the parallel trends assumption. Starting in 2018, the exposed establishments experienced significant revenue losses among firms with low cloud vendor interoperability or high intrafirm vendor diversity. These results support *Hypotheses H3b* in Section 2.4 and *Hypothesis H4b* in Section 2.5.

[Insert Figure 5 here]

5. Conclusion and Discussion

While ecosystems are a popular approach taken by studies of technological innovation in various industries, there has been little research applying the lens of ecosystems to analyzing digital transformation in large corporations. We fill this research gap by focusing on enterprise cloud computing, which provides the technological foundation fueling complex systems of infrastructural tools and software applications that together form an ecosystem of big-data technologies at the enterprise level. The technological architecture around data storage and processing guides the patterns of interaction among actors within a corporation's big-data ecosystem. It can shape the ecosystem's degree of modularity and the extent to which components are loosely coupled or tightly integrated. Our findings show that a modular approach to orchestrating corporation-level big-data ecosystems, characterized by a lack of data interoperability and reliance on microservices architecture, can severely constrain performance outcomes, especially in a changing regulatory landscape that requires organizational adaptation to ensure compliance. We focus on the natural experiment of a single regulatory shock – GDPR enforcement – which had a profound impact beyond the EU/EEA on firms worldwide holding and processing customer data, allowing us to cleanly identify the impact of enterprise data architecture on post-shock establishment revenue performance. A modular technological architecture facilitates local scaling that can lead to rapid expansion in cloud adoption within an establishment. However, it introduces barriers to interoperability across functional areas, limiting downstream capabilities and applications that require coordination. The lack of clear data ownership and

visibility into data sources can further lead to inconsistent copies of the same data source and exacerbate wasted cloud costs and privacy violation risks.

This paper contributes primarily to the literature on the impact of technological architecture on ecosystem performance (Baldwin & Clark, 2000; Baldwin & Woodard, 2009; Adner & Kapoor, 2010; Kapoor & Adner, 2012; Kapoor & Lee, 2013; Lee & Kapoor, 2017; Kapoor & Agarwal, 2016; Karim, Lee, & Hoehn-Weiss, 2022; Holgersson, Baldwin, Chesbrough, & Bogers, 2022). Our analyses recognize the connection between innovation ecosystems and digital infrastructure (Broadbent, Weill, & St. Clair, 1999; Tilson, Lyytinen, & Sørensen, 2010; Henfridsson & Bygstad, 2013), both of which constitute complex systems hosting interdependent components that interact in diverse ways and jointly create value. Our novel contribution lies in linking these two perspectives to analyze large corporations with many different functional areas and establishments which adopt cloud computing to form the technological foundation of powerful systems that can utilize large amounts of data and fuel downstream applications in big-data analytics and machine learning. We identify modular architecture as a bottleneck to adaptation, with an especially detrimental impact on performance after a shock affects the allocation of value within the ecosystem (Ethiraj & Levinthal, 2004; Baldwin, 2015; Karim, Lee, & Hoehn-Weiss, 2022; Eklund & Kapoor, 2022; Agarwal & Kapoor, 2022; Koçak, Levinthal, & Puranam, 2022).

This paper also contributes to the literature on organizational adaptation in the context of scaling digital resources (Adner, Puranam, & Zhu, 2019; Kretschmer & Khashabi, 2020; Giustiziero, Kretschmer, Somaya, & Wu, 2022). The diffusion of cloud computing constitutes contextual conditions that shape digital transformation and drive organizational change (Stieglitz, Knudsen, & Becker, 2016; Benlian, Kettinger, Sunyaev, & Winkler, 2018; Hanelt, Bohnsack, Marz, & Marante, 2021). While researchers have begun to examine how organizational features, especially decentralized structures and openness of platform ecosystems, affect innovation activities, our novel contribution is in applying the lens of ecosystem architecture to examine the phenomenon of cloud adoption and digital transformation, where a natural

mapping occurs from technology stacks and configurations in practice to theoretical concepts of modularity and decentralization.

The paper also contributes to the literature on the heterogeneous impact of regulation on firm performance (Elliehausen & Kurtz, 1988; Campbell, Goldfarb, & Tucker, 2015; Bessen, Impink, Reichensperger, & Seamans, 2020; Peukert, Bechtold, Batikas, & Kretschmer, 2022; Burford, Shipilov, & Furr, 2022). While previous research focuses on "scale economies" where the quantity and scope of firms' products and services explain differential effects of regulation, we build on these analyses and recognize that modular firms (which can be large corporations and not only small firms) may experience disproportionately worse performance. Our findings suggest a previously underexplored layer of heterogeneity, rooted in the complex interactions within large corporations supported by the technological architecture, which co-evolves with technology implementation, e.g., the usage of microservices architecture, cloud vendor choices and usage of API infrastructure to drive interoperability.

Our findings offer important insights to chief technology officers and system architects at large corporations. Digital transformation among large corporations depends on contextual conditions that increasingly involve the diffusion of cloud computing. Almost all large enterprises adopt public cloud services to some extent, and many spend millions of dollars on cloud computing costs every year. Many IT leaders are concerned about wasted spending on the cloud and the high costs of renting infrastructure from public vendors. The expansion of data privacy regulations in recent years, e.g., GDPR, further complicate the challenges of effectively orchestrating big-data ecosystems to drive value creation within the organization. Hence, setting up the technological backbones and effectively managing big-data ecosystems becomes increasingly crucial. Our results suggest viable pathways to avoid making technological architecture choices that can introduce bottlenecks to adaptation to comply with regulation and control costs of scaling data storage and processing.

References

Adner, R. (2017). Ecosystem As Structure: An Actionable Construct for Strategy. *Journal of Management*, *43*(1), 39–58.

Adner, R., & Kapoor, R. (2010). Value Creation in Innovation Ecosystems: How the Structure of Technological Interdependence Affects Firm Performance in New Technology Generations. *Strategic Management Journal*, *31*(3), 306–333.

Adner, R., & Kapoor, R. (2016). Innovation Ecosystems and the Pace of Substitution: Re-Examining Technology S-Curves. *Strategic Management Journal*, *37*(4), 625–648.

Adner, R., Puranam, P., & Zhu, F. (2019). What Is Different about Digital Strategy? From Quantitative to Qualitative Change. *Strategy Science*, *4*(4), 253–261.

Agarwal, S., & Kapoor, R. (2022). Value Creation Tradeoff in Business Ecosystems: Leveraging Complementarities While Managing Interdependencies. *Organization Science*.

Aggarwal, V., & Wu, B. (2015). Organizational Constraints to Adaptation: Intrafirm Asymmetry in the Locus of Coordination. *Organization Science*, *26*(1), 218–238.

Agrawal, A. K., Gans, J. S., & Goldfarb, A. (2021). AI Adoption and System-Wide Change. *National Bureau of Economic Research.*

Aral, S., & Dhillon, P. S. (2022). What (Exactly) Is Novelty in Networks? Unpacking the Vision Advantages of Brokers, Bridges, and Weak Ties. *Management Science.*

Arora, C., Bawcom, A., Lhuer, X., & Sohoni, V. (2022, August 3). Three big moves that can decide a financial institution's future in the cloud. *McKinsey Digital*. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-big-moves-that-can-decide-a-financial-institutions-future-in-the-cloud

Baldwin, C. Y. (2015). Bottlenecks, Modules and Dynamic Architectural Capabilities. *Harvard Business School Finance Working Paper*, *15–028*.

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

Baldwin, C. Y., & Woodard, C. J. (2009). The Architecture of Platforms: A Unified View. *Platforms, Markets and Innovation*, *32*, 19–44.

Barney, J. B. (1986). Types of Competition and the Theory of Strategy: Toward an Integrative Framework. *Academy of Management Review*, *11*(4), 791–800.

Benlian, A., Kettinger, W. J., Sunyaev, A., & Winkler, T. J. (2018). The Transformative Value of Cloud Computing: A Decoupling, Platformization, and Recombination Theoretical Framework. *Journal of Management Information Systems*, *35*(3), 719–739.

Bessen, J. E., Impink, S. M., Reichensperger, L., & Seamans, R. (2020). GDPR and the Importance of Data to AI Startups. *NYU Stern School of Business.*

Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. V. (2013). Digital Business Strategy: Toward a Next Generation of Insights. *MIS Quarterly*, 471–482.

Bloom, N., & Pierri, N. (2018). Cloud Computing Is Helping Smaller, Newer Firms Compete. *Harvard Business Review*, *94*(4).

Boudreau, K. (2010). Open Platform Strategies and Innovation: Granting Access vs. Devolving Control. *Management Science*, *56*(10), 1849–1872.

Bresnahan, T. (2019). *Artificial Intelligence Technologies and Aggregate Growth Prospects.*

Bresnahan, T. F., & Greenstein, S. (2001). The Economic Contribution of Information Technology: Towards Comparative and User Studies. *Journal of Evolutionary Economics, 11*, 95-118.

Broadbent, M., Weill, P., & St. Clair, D. (1999). The Implications of Information Technology Infrastructure for Business Process Redesign. *MIS Quarterly*, 159–182.

Burford, N., Shipilov, A. V., & Furr, N. R. (2022). How Ecosystem Structure Affects Firm Performance in Response to a Negative Shock to Interdependencies. *Strategic Management Journal*, *43*(1), 30–57.

Campagnolo, D., & Camuffo, A. (2010). The Concept of Modularity in Management Studies: A Literature Review. *International Journal of Management Reviews*, *12*(3), 259–283.

Campbell, J., Goldfarb, A., & Tucker, C. (2015). Privacy Regulation and Market Structure. *Journal of Economics & Management Strategy*, *24*(1), 47–73.

Casadesus-Masanell, R., & Yoffie, D. B. (2007). Wintel: Cooperation and Conflict. *Management Science, 53*(4), 584–598.

Cennamo, C., Ozalp, H., & Kretschmer, T. (2018). Platform Architecture and Quality Trade-Offs of Multihoming Complements. *Information Systems Research*, *29*(2), 461–478.

Constantinides, P., & Barrett, M. (2015). Information Infrastructure Development and Governance as Collective Action. *Information Systems Research*, *26*(1), 40–56.

Correani, A., De Massis, A., Frattini, F., Petruzzelli, A. M., & Natalicchio, A. (2020). Implementing a Digital Strategy: Learning from the Experience of Three Digital Transformation Projects. *California Management Review*, *62*(4), 37–56.

Dallemule, L., & Davenport, T. H. (2017). What's Your Data Strategy. *Harvard Business Review*, *95*(3), 112–121.

Daniel, C. (2022, November 7). Microsoft Azure Revenue and Growth Statistics. *The Signhouse Blog*. https://www.usesignhouse.com/blog/microsoft-azure-stats

DeStefano, T., Kneller, R., & Timmis, J. (2020). *Cloud Computing and Firm Growth.*

Eklund, J., & Kapoor, R. (2022). Mind the Gaps: How Organization Design Shapes the Sourcing of Inventions. *Organization Science*, *33*(4), 1319–1339.

Elliehausen, G. E., & Kurtz, R. D. (1988). Scale Economies in Compliance Costs for Federal Consumer Credit Regulations. *Journal of Financial Services Research*, *1*(2), 147–159.

Englmaier, F., Foss, N. J., Knudsen, T., & Kretschmer, T. (2018). Organization Design and Firm Heterogeneity: Towards an Integrated Research Agenda for Strategy. *Organization Design*.

Ethiraj, S. K. (2007). Allocation of Inventive Effort in Complex Product Systems. *Strategic Management Journal*, *28*(6), 563–584.

Ethiraj, S. K., & Levinthal, D. (2004). Modularity and Innovation in Complex Systems. *Management Science*, *50*(2), 159–173.

Farboodi, M., Mihet, R., Philippon, T., & Veldkamp, L. (2019). Big Data and Firm Dynamics. *AEA Papers and Proceedings*, *109*, 38–42.

Flexera. (2022). *State of the Cloud Report: The post-pandemic world comes into focus and FinOps practices gain momentum.* https://resources.flexera.com/web/pdf/Flexera-State-of-the-Cloud-Report-2022.pdf

Furber, S. (2020, August 18). As "big tech" dominates cloud use for banks, regulators may need to get tougher. *S&P Global Market Intelligence*. https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/as-big-tech-dominates-cloud-use-for-banks-regulators-may-need-to-get-tougher-59669007

Furr, N., Ozcan, P., & Eisenhardt, K. M. (2022). What Is Digital Transformation? Core Tensions Facing Established Companies on the Global Stage. *Global Strategy Journal*, *12*(4), 595–618.

Gawer, A., & Cusumano, M. A. (2002). *Platform Leadership: How Intel, Microsoft, and Cisco Drive Industry Innovation* (Vol. 5). Harvard Business School Press.

Gawer, A., & Henderson, R. (2007). Platform Owner Entry and Innovation in Complementary Markets: Evidence from Intel. *Journal of Economics & Management Strategy*, *16*(1), 1–34.

Giustiziero, G., Kretschmer, T., Somaya, D., & Wu, B. (2022). Hyperspecialization and Hyperscaling: A Resource-Based Theory of the Digital Firm. *Strategic Management Journal*.

Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2021). The Role of Artificial Intelligence and Data Network Effects for Creating User Value. *Academy of Management Review*, *46*(3), 534–551.

Hagiu, A., & Wright, J. (2020). *Data-Enabled Learning, Network Effects and Competitive Advantage.*

Hanelt, A., Bohnsack, R., Marz, D., & Antunes Marante, C. (2021). A Systematic Review of the Literature on Digital Transformation: Insights and Implications for Strategy and Organizational Change. *Journal of Management Studies*, *58*(5), 1159–1197.

Hannah, D. P., & Eisenhardt, K. M. (2018). How Firms Navigate Cooperation and Competition in Nascent Ecosystems. *Strategic Management Journal*, *39*(12), 3163–3192.

Hecker, A., & Kretschmer, T. (2010). Outsourcing Decisions: The Effect of Scale Economies and Market Structure. *Strategic Organization*, *8*(2), 155–175.

Henfridsson, O., & Bygstad, B. (2013). The Generative Mechanisms of Digital Infrastructure Evolution. *MIS Quarterly*, 907–931.

Holgersson, M., Baldwin, C. Y., Chesbrough, H., & M. Bogers, M. L. (2022). The Forces of Ecosystem Evolution. *California Management Review*, *64*(3), 5–23.

Holtz, D., Carterette, B., Chandar, P., Nazari, Z., Cramer, H., & Aral, S. (2020). The Engagement-Diversity Connection: Evidence from a Field Experiment on Spotify. *In Proceedings of the 21st ACM Conference on Economics and Computation*, 75–76.

Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a Theory of Ecosystems. *Strategic Management Journal*, *39*(8), 2255–2276.

Jin, W., & McElheran, K. (2017). Economies Before Scale: Survival and Performance of Young Plants in the Age of Cloud Computing. *Rotman School of Management Working Paper*.

Kapoor, R. (2013). Persistence of Integration in the Face of Specialization: How Firms Navigated the Winds of Disintegration and Shaped the Architecture of the Semiconductor Industry. *Organization Science, 24*(4), 1195–1213.

Kapoor, R. (2018). Ecosystems: Broadening the Locus of Value Creation. *Journal of Organization Design*, *7*(1), 1–16.

Kapoor, R., & Adner, R. (2012). What Firms Make vs. What They Know: How Firms' Production and Knowledge Boundaries Affect Competitive Advantage in the Face of Technological Change. *Organization Science*, *23*(5), 1227–1248.

Kapoor, R., & Agarwal, S. (2017). Sustaining Superior Performance in Business Ecosystems: Evidence from Application Software Developers in the iOS and Android Smartphone Ecosystems. *Organization Science*, *28*(3), 531–551.

Kapoor, R., & Furr, N. R. (2015). Complementarities and Competition: Unpacking the Drivers of Entrants' Technology Choices in the Solar Photovoltaic Industry. *Strategic Management Journal*, *36*(3), 416–436.

Kapoor, R., & Lee, J. M. (2013). Coordinating and Competing in Ecosystems: How Organizational Forms Shape New Technology Investments. *Strategic Management Journal*, *34*(3), 274–296.

Karim, S., Lee, C. H., & Hoehn-Weiss, M. (2022). Task and Resource Bottlenecks: A Holistic Examination of Task Systems Through an Organization Design Lens. *Strategic Management Journal*.

Koçak, Ö., Levinthal, D. A., & Puranam, P. (2022). The Dual Challenge of Search and Coordination for Organizational Adaptation: How Structures of Influence Matter. *Organization Science*.

Kretschmer, T., & Khashabi, P. (2020). Digital Transformation and Organization Design: An Integrated Approach. *California Management Review*, *62*(4), 86–104.

Langlois, R. N., & Robertson, P. L. (1992). Networks and Innovation in a Modular System: Lessons from the Microcomputer and Stereo Component Industries. *Research Policy*, *21*(4), 297–313.

Lee, J. M., & Kapoor, R. (2017). Complementarities and Coordination: Implications for Governance Mode and Performance of Multiproduct Firms. *Organization Science*, *28*(5), 931–946.

Levinthal, D. A. (1997). Adaptation on Rugged Landscapes. *Management Science, 43*(7), 934–950.

Levinthal, D. A., & Wu, B. (2010). Opportunity Costs and Non-Scale Free Capabilities: Profit Maximization, Corporate Scope, and Profit Margins. *Strategic Management Journal*, *31*(7), 780–801.

Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., & Ghalsasi, A. (2011). Cloud Computing—The Business Perspective. *Decision Support Systems*, *51*(1), 176–189.

Nguyen, L. (2022, January 3). Banks Tiptoe Toward Their Cloud-Based Future: Cloud computing is slowly changing how Wall Street banks handle their business, but concerns with security remain. *New York Times*. https://www.nytimes.com/2022/01/03/business/wall-street-cloud-computing.html

Park, Y., & Mithas, S. (2020). Organized Complexity of Digital Business Strategy: A Configurational Perspective. *MIS Quarterly*, *44*(1).

Penrose, E. T. (1959). The Theory of the Growth of the Firm. *New York: John Wiley & Sons Inc.*

Peukert, C., Bechtold, S., Batikas, M., & Kretschmer, T. (2022). Regulatory Spillovers and Data Governance: Evidence from the GDPR. *Marketing Science*, *41*(4), 746–768.

Rietveld, J., Schilling, M. A., & Bellavitis, C. (2019). Platform Strategy: Managing Ecosystem Value Through Selective Promotion of Complements. *Organization Science*, *30*(6), 1232–1251.

Sabherwal, R., Sabherwal, S., Havakhor, T., & Steelman, Z. (2019). How Does Strategic Alignment Affect Firm Performance? The Roles of Information Technology Investment and Environmental Uncertainty. *MIS Quarterly*, *43*(2), 453–474.

Saldanha, T. J., Lee, D., & Mithas, S. (2020). Aligning Information Technology and Business: The Differential Effects of Alignment During Investment Planning, Delivery, and Change. *Information Systems Research*, *31*(4), 1260–1281.

Sanderson, S., & Uzumeri, M. (1995). Managing Product Families: The Case of the Sony Walkman. *Research Policy*, *24*(5), 761–782.

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.

Shah, A. (2018, September 25). Microsoft Azure: The only consistent, comprehensive hybrid cloud. *Microsoft Azure Blog*. https://azure.microsoft.com/en-us/blog/microsoft-azure-the-only-consistent-comprehensive-hybrid-cloud/

Siggelkow, N., & Levinthal, D. A. (2003). Temporarily Divide to Conquer: Centralized, Decentralized, and Reintegrated Organizational Approaches to Exploration and Adaptation. *Organization Science*, *14*(6), 650–669.

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

Star, S. L., & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, *7*(1), 111–134.

Stieglitz, N., Knudsen, T., & Becker, M. C. (2016). Adaptation and Inertia in Dynamic Environments. *Strategic Management Journal*, *37*(9), 1854–1864.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic Capabilities and Strategic Management. *Strategic Management Journal*, *18*(7), 509–533.

Tilson, D., Lyytinen, K., & Sørensen, C. (2010). Research Commentary—Digital Infrastructures: The Missing IS Research Agenda. *Information Systems Research*, *21*(4), 748–759.

Ulrich, K. (1995). The Role of Product Architecture in the Manufacturing Firm. *Research Policy*, *24*(3), 419–440.

Wessel, L., Baiyere, A., Ologeanu-Taddei, R., Cha, J., & Blegind-Jensen, T. (2021). Unpacking the Difference Between Digital Transformation and IT-Enabled Organizational Transformation. *Journal of the Association for Information Systems*, *22*(1), 102–129.

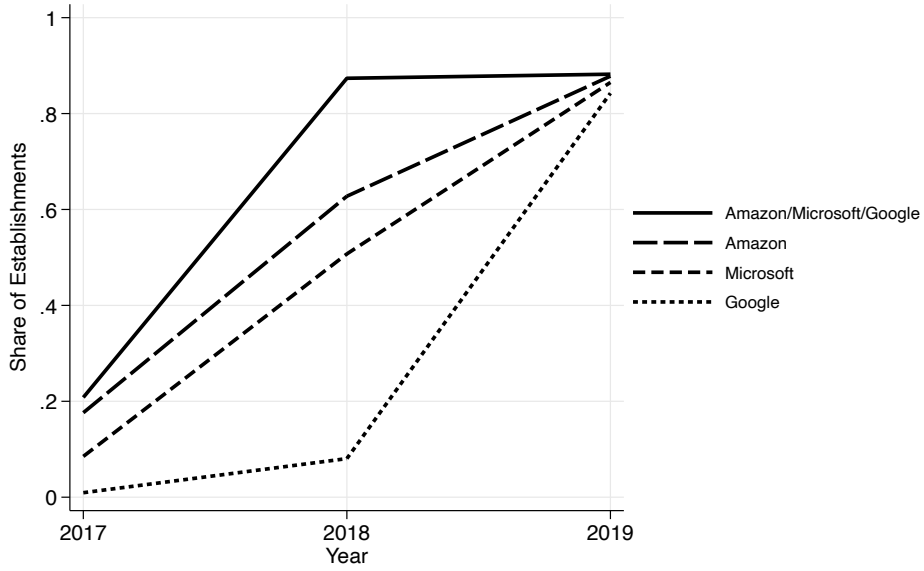
Wen, W., & Zhu, F. (2019). Threat of Platform-Owner Entry and Complementor Responses: Evidence from the Mobile App Market. *Strategic Management Journal, 40*(9), 1336–1367.

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). Research Commentary—The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. *Information Systems Research*, *21*(4), 724–735.

Zhou, Y. M., & Wan, X. (2017). Product Variety, Sourcing Complexity, and the Bottleneck of Coordination. *Strategic Management Journal*, *38*(8), 1569–1587.

Figure 1: Share of Establishments Adopting Cloud from 2017 to 2019, Overall and by Vendor. This figure shows the adoption rates of the leading cloud vendors (Amazon, Microsoft, and Google) at the establishment level from 2017 to 2019. Panel (a) focuses on the full sample, and panel (b) focuses on the sample that excludes the three largest establishments, each accounting for 20% or more of overall establishments.

(a) Full Sample



(b) Sample Excluding Largest Corporations

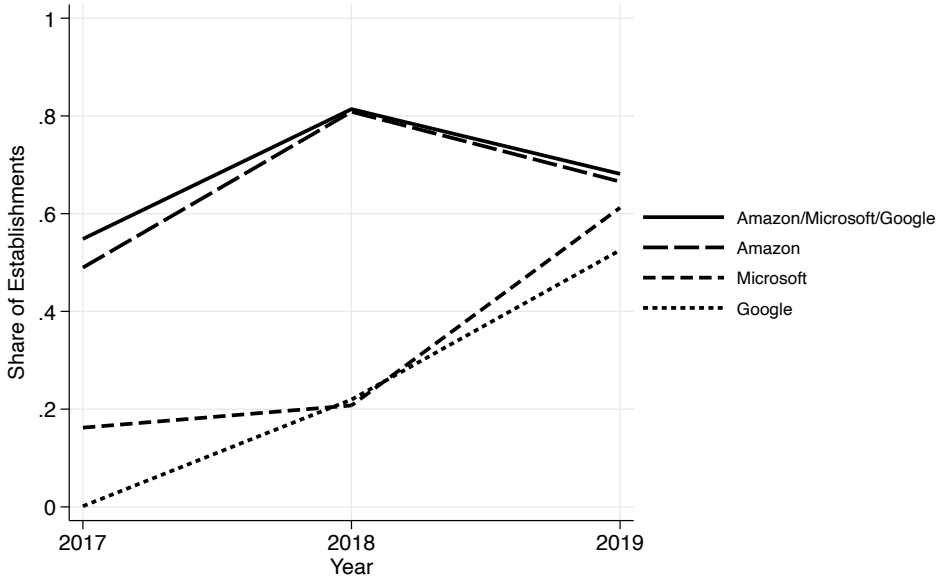
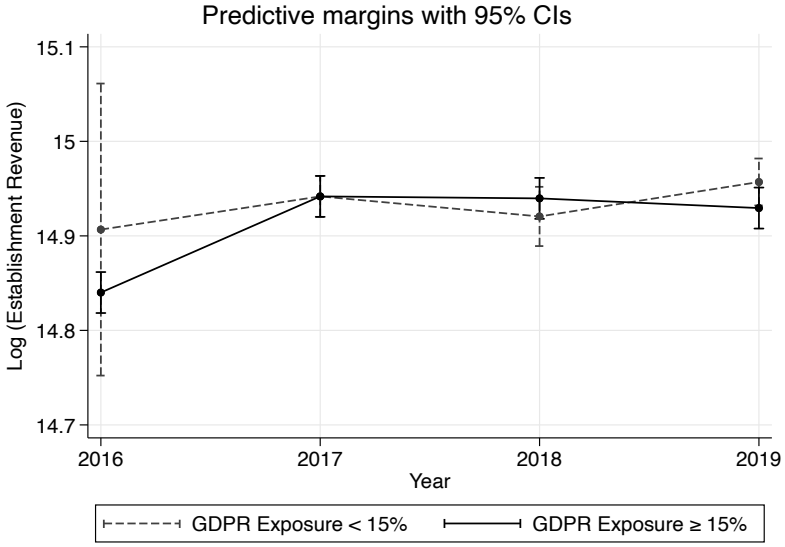


Figure 2: Revenue Response to GDPR Enforcement by Data Interoperability. The plotted estimates and the 95% confidence intervals are from an event-study version of estimating Model 2 from panel (a) of Section 3.5. The year before the event serves as the excluded baseline. Treatment is defined as having at least 15% of business activities in the EU/EEA. The model controls for establishment fixed effects and year fixed effects. Robust standard errors are clustered at the treatment level by corporation. The subfigures in panel (a) and panel (b) compare establishments of firms that drive toward data interoperability and other establishments.

(a) Interoperability = 1



(b) Interoperability = 0

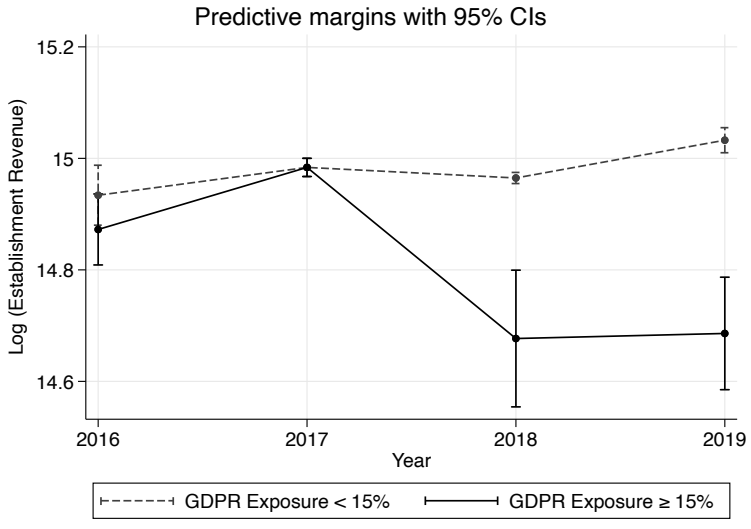
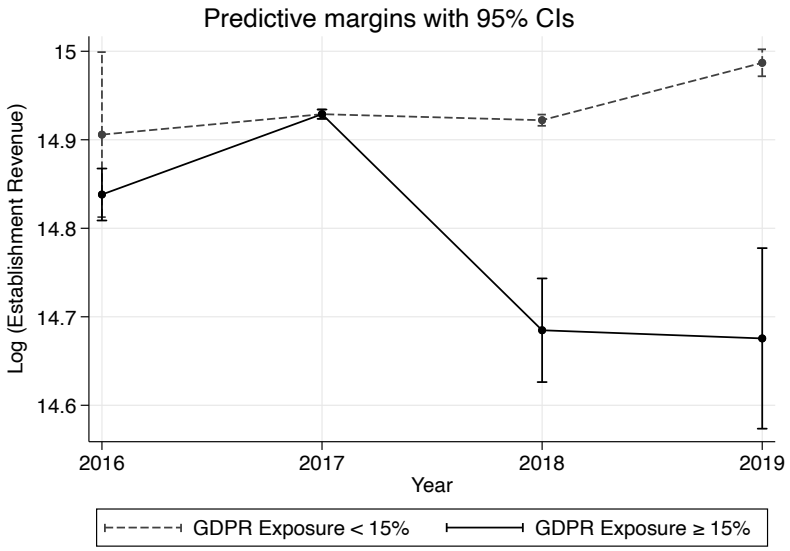


Figure 3: Revenue Response to GDPR Enforcement by Microservices Architecture. The plotted estimates and the 95% confidence intervals are from an event-study version of estimating Model 2 from Section 3.5. The model controls for establishment fixed effects and year fixed effects. The year before the event serves as the excluded baseline. Treatment is defined as having at least 15% of business activities in the EU/EEA. Robust standard errors are clustered at the corporation level (the same as GDPR exposure which defines the treatment). The subfigures in panel (a) and panel (b) compare establishments of firms that build their data platforms on the microservices architecture and other establishments.

(a) Microservices = 1



(b) Microservices = 0

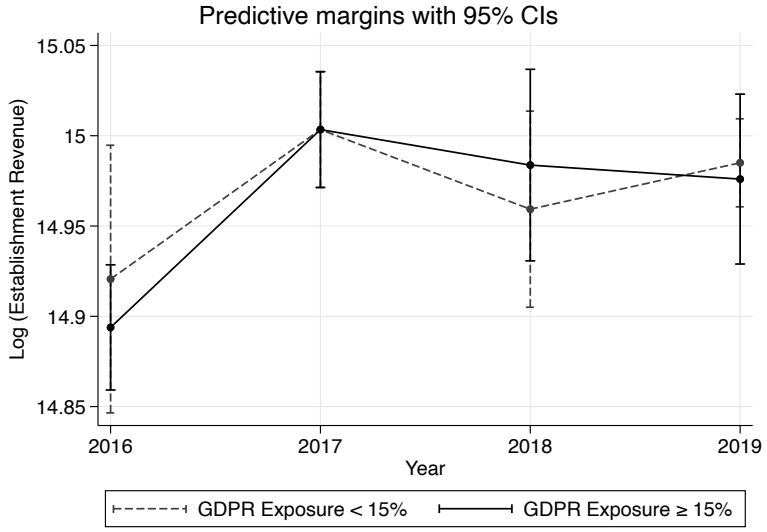
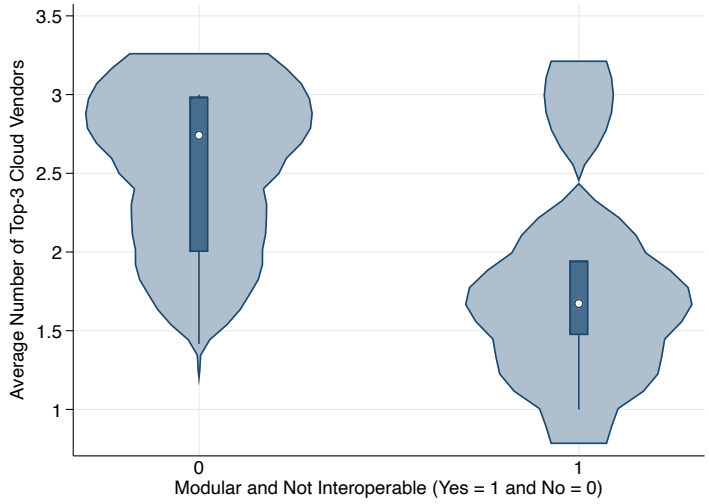


Figure 4: Enterprise Data Architecture and Cloud Vendor Configurations. This figure shows the violin plots to compare the distributions for two cloud vendor configuration variables at the corporation level across an indicator that equals 1 when interoperability = 0 and microservices = 1. The violin plots show the probability distribution, highlighting the median and interquartile range and 1.5x interquartile range of each variable across samples. Panel (a) shows the plots for the average number of cloud vendors (among Amazon, Microsoft, and Google) per establishment, and panel (b) shows the plots for the intrafirm cloud vendor diversity across establishments within the corporation (defined in Equation 1 of Section 3.4).

(a) Average Number of Cloud Vendors Per Establishment



(b) Intrafirm Cloud Vendor Diversity Across Establishments

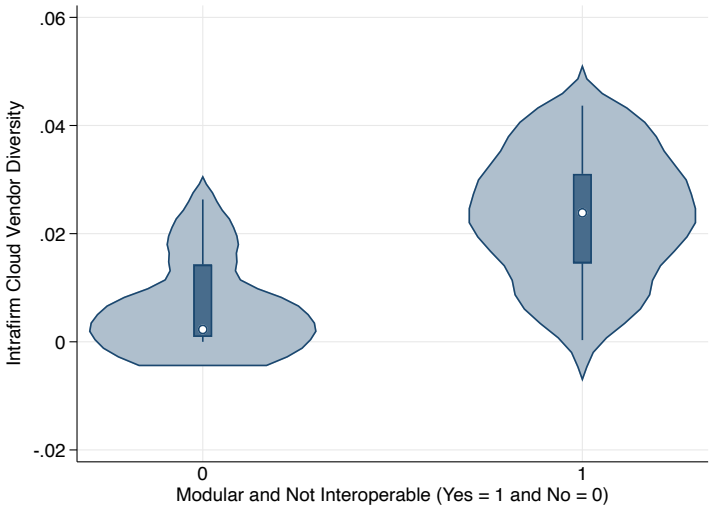
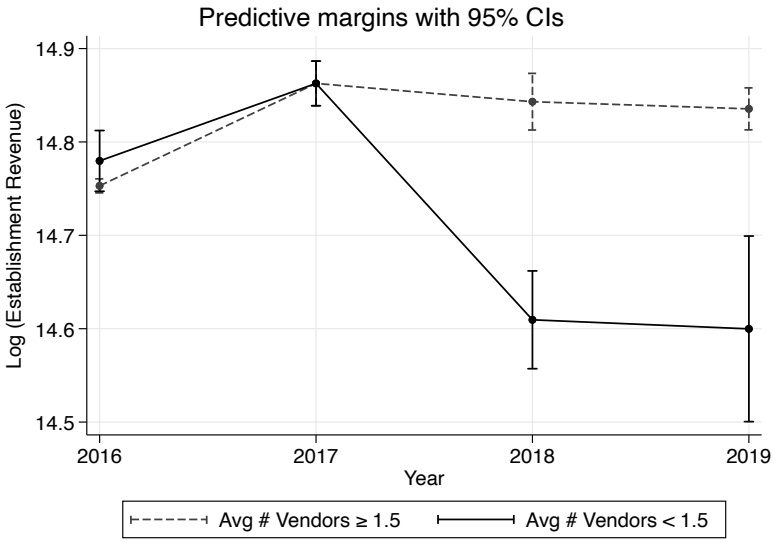


Figure 5: Revenue Response to GDPR Enforcement by Cloud Vendor Interoperability and Intrafirm Vendor Diversity. The plotted estimates and the 95% confidence intervals are from an event-study version of estimating Model 3 from Section 3.5. The year before the event serves as the excluded baseline. The model controls for establishment fixed effects and year fixed effects. Robust standard errors are clustered at the corporation level (the same as GDPR exposure which defines the treatment). All establishments in the sample are substantially exposed to GDPR compliance risks.

(a) Low vs. High Cloud Vendor Interoperability Within Establishment



(b) High vs. Low Fragmentation in Cloud Adoption Across Establishments

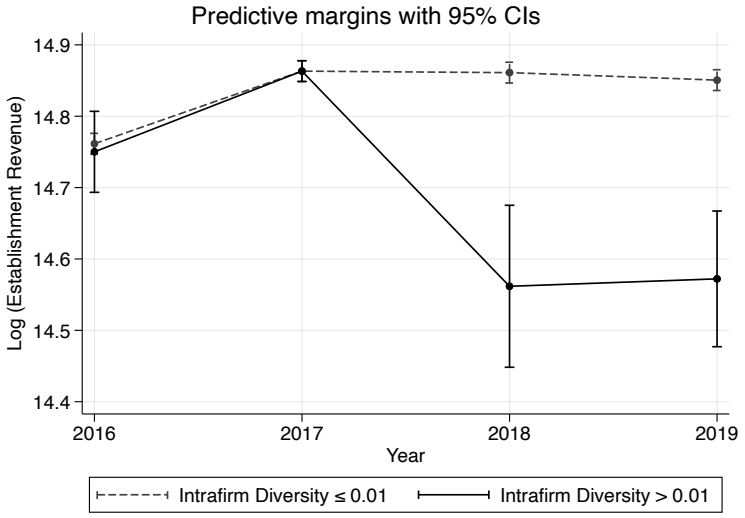


Table 1: GDPR Exposure of Large Finance Corporations in the Sample. This table lists all the large finance corporations in the sample and the variables for determining exposure to GDPR compliance risks. A corporation is defined (see Section 3.2) as having high exposure to GDPR if it meets the following criteria: (1) annual reports mention GDPR in at least one year between 2016 and 2019; (2) at least 15% of business activities (average across total sales, office space, and employment) are in the EU/EEA if data is available, or at least 40% of business activities outside the Americas if data on EU/EEA is not available.

Company	Mention GDPR	EU/EEA Share≥15%	Share Outside Americas ≥40%	High Exposure
American Express Company	1	n.a.	0	0
American International Group, Inc.	1	n.a.	1	1
Assurant, Inc.	1	0	0	0
Bank of America Corporation	1	0	0	0
Capital One Financial Corporation	1	0	0	0
Centene Corporation	0	n.a.	0	0
Chubb Limited	1	1	1	1
Citigroup Inc.	1	1	1	1
Eaton Vance Corp.	0	n.a.	0	0
Fifth Third Bancorp	0	n.a.	0	0
JPMorgan Chase & Co.	1	0	0	0
Lincoln National Corporation	0	n.a.	0	0
Mastercard Incorporated	1	n.a.	1	1
MetLife, Inc.	1	0	1	0
MoneyGram International, Inc.	1	n.a.	1	1
Prudential Financial, Inc.	1	0	0	0
The Bank of New York Mellon Corporation	0	0	1	0
The Charles Schwab Corporation	0	n.a.	n.a.	0
The Goldman Sachs Group, Inc.	1	1	1	1
The Hartford Financial Services Group, Inc.	1	0	0	0
The Travelers Companies, Inc.	0	n.a.	0	0
UnitedHealth Group Incorporated	1	n.a.	0	0
Wells Fargo & Company	0	n.a.	0	0

Table 2: Sample Descriptive Statistics and Correlations. Panel (a) describes the variables in the full regression sample – an unbalanced panel containing 94,539 establishment-year observations of 23 large financial services corporations between 2016 and 2019. Panel (b) shows the pairwise correlations between variables.

(a) Descriptive Statistics

	Mean	SD	P10	P50	P90	#Obs.
Log (Revenue)	14.957	1.273	13.816	14.914	16.213	94539
Log (Employment)	2.698	0.853	1.792	2.773	3.497	94539
Log (IT Staff Size)	1.049	0.690	0.000	1.099	1.099	94539
GDPR Exposure	0.347	0.476	0.000	0.000	1.000	94539
Microservices	0.580	0.494	0.000	1.000	1.000	94539
Interoperable	0.559	0.497	0.000	1.000	1.000	94539
Avg. # Cloud Vendors	2.874	0.349	2.741	2.996	2.999	94539
Intrafirm Vendor Diversity	0.003	0.007	0.000	0.000	0.010	94539

(b) Correlations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Log (Revenue)	1.000							
(2) Log (Employment)	0.822	1.000						
(3) Log (IT Staff Size)	0.653	0.821	1.000					
(4) GDPR Exposure	-0.037	-0.046	-0.068	1.000				
(5) Microservices	-0.022	-0.010	0.048	0.111	1.000			
(6) Interoperable	-0.013	0.019	-0.050	0.489	-0.382	1.000		
(7) Avg. # Cloud Vendors	-0.180	-0.138	-0.102	-0.229	-0.017	0.196	1.000	
(8) Intrafirm Vendor Diversity	0.152	0.092	0.068	0.323	-0.093	-0.087	-0.807	1.000

Table 3: Data Interoperability and Establishment Revenue after GDPR Enforcement. This table shows estimates from difference-in-differences models based on Model 2 (columns 1 – 4) and Model 3 (column 5) in Section 3.5. The outcome variable is the establishment-level log revenue measured annually. In each panel, columns 1 – 4 estimate the differential effects by interacting the post-event dummy with the binary indicator of high GDPR exposure. Column 5 estimates the differential effects by interacting the post-event dummy with the data interoperability indicator. The sample underlying each column is described in the table headers. "Excl.Big" indicates that the sample excludes the largest three corporations (see Section 3.1.1). All models are estimated in panel regressions with firm-establishment fixed effects and year fixed effects. Robust standard errors are clustered at the corporation level. $***p < 0.01$; $**p < 0.05$; $*p < 0.1$.

Dependent Variable	**Log (Revenue)**				
	(1)	(2)	(3)	(4)	(5)
Sample	IO=0	IO=0, Excl.Big	IO=1	IO=1, Excl.Big	High GDPR
(Year ≥ 2018) × High GDPR Exposure	-0.298*** (0.060)	-0.257*** (0.065)	0.011 (0.022)	0.126 (0.144)	
(Year ≥ 2018) × Interoperable					0.308*** (0.060)
Log (Employment)	0.765 (0.460)	0.623 (0.653)	0.678 (1.013)	1.435 (1.503)	0.754 (0.610)
Log (IT Staff Size)	0.112 (0.317)	-0.103 (0.327)	-0.095 (0.750)	-1.076 (1.291)	-0.353 (0.497)
FE: Firm-Establishment	Y	Y	Y	Y	Y
FE: Year	Y	Y	Y	Y	Y
Observations	41719	13388	52820	15339	13315
R^2	0.934	0.923	0.948	0.923	0.922

Table 4: Microservices Architecture and Establishment Revenue after GDPR Enforcement. This table shows estimates from difference-in-differences models based on Model 2 (columns 1 – 4) and Model 3 (column 5) in Section 3.5. The outcome variable is the establishment-level log revenue measured annually. In each panel, columns 1 – 4 estimate the differential effects by interacting the post-event dummy with the binary indicator of high GDPR exposure. Column 5 estimates the differential effects by interacting the post-event dummy with the microservices architecture indicator. The sample underlying each column is described in the table headers. "Excl.Big" indicates that the sample excludes the largest three corporations (see Section 3.1.1). All models are estimated in panel regressions with firm-establishment fixed effects and year fixed effects. Robust standard errors are clustered at the corporation level. $***p < 0.01$; $**p < 0.05$; $*p < 0.1$.

Dependent Variable	**Log (Revenue)**				
	(1)	(2)	(3)	(4)	(5)
Sample	MS=1	MS=1, Excl.Big	MS=0	MS=0, Excl.Big	High GDPR
(Year ≥ 2018) × High GDPR Exposure	-0.245***	-0.230***	0.017	0.084	
	(0.030)	(0.046)	(0.044)	(0.082)	
(Year ≥ 2018) × Microservices					-0.230***
					(0.043)
Log (Employment)	0.728	0.802	0.659	0.971	0.739
	(0.453)	(0.744)	(0.953)	(1.054)	(0.620)
Log (IT Staff Size)	-0.019	-0.536	0.011	-0.404	-0.343
	(0.316)	(0.397)	(0.675)	(0.742)	(0.510)
FE: Firm-Establishment	Y	Y	Y	Y	Y
FE: Year	Y	Y	Y	Y	Y
Observations	54855	7769	39684	20958	13315
R^2	0.948	0.924	0.931	0.922	0.921

Table 5: Corporation Level Two-Sample T-Tests. This table shows results from testing the balance between corporations by their data architecture designs. We split the sample across whether a corporation is non-interoperable (IO=0) and uses microservices architecture (MS=1). Out of the 23 corporations in the sample, six are in the "*yes*" subsample (non-interoperable and use microservices architecture), and 17 are in the "*no*" subsample (interoperable or does not use microservices architecture). We calculate the difference in the means across the two subsamples and the p-value associated with the difference for each variable. The corporations have similar means across all the listed characteristics except cloud configurations (top two variables) related to the degree of modularity measured by the configuration of cloud vendor choices.

Indicator (IO=0 and MS=1)	*Yes* (6) Mean	*No* (17) Mean	*Diff.* Yes-No	*p-value*
Average # Cloud Vendors	1.793	2.516	-0.723**	0.013
Intrafirm Vendor Diversity	0.023	0.007	0.015***	0.006
High GDPR Exposure	0.333	0.353	-0.020	0.935
Log (# Establishments)	5.534	4.933	0.601	0.538
Share Amazon/Microsoft/Google	0.791	0.791	0.001	0.995
Avg. Log (Revenue)	16.248	15.884	0.364	0.460
Avg. Log (Employment)	3.383	3.233	0.150	0.645
Avg. Log (IT Staff Size)	1.613	1.443	0.170	0.507

Table 6: Cloud Vendor Interoperability and Establishment Revenue after GDPR Enforcement. This table shows estimates from difference-in-differences models based on Model 2 (columns 1 – 3) and Model 3 (column 4) in Section 3.5. In each panel, columns 1 – 3 estimate the differential effects by interacting the post-event dummy with the binary indicator of high GDPR exposure. Column 4 estimates the differential effects by interacting the post-event dummy with the low cloud vendor interoperability indicator. The sample underlying each column is described in the table headers. All models are estimated in panel regressions with firm-establishment fixed effects and year fixed effects. Robust standard errors are clustered at the corporation level. $***p < 0.01; **p < 0.05; *p < 0.1$.

Dependent Variable	**Log (Revenue)**			
	(1)	(2)	(3)	(4)
Sample	Avg Vendors <1.5	Avg Vendors ≥1.5	Avg Vendors ≥1.5	High GDPR
Exclude Big	No	No	Yes	No
(Year ≥ 2018) × High GDPR Exposure	-0.292***	-0.008	0.062	
	(0.036)	(0.027)	(0.061)	
(Year ≥ 2018) × Low Avg # Vendors				-0.241***
				(0.039)
Log (Employment)	0.726	0.916	1.277	0.740
	(0.641)	(0.864)	(1.361)	(0.620)
Log (IT Staff Size)	-0.524	0.152	-0.392	-0.340
	(0.590)	(0.387)	(0.619)	(0.510)
FE: Establishment	Y	Y	Y	Y
FE: Year	Y	Y	Y	Y
Observations	3431	91108	25296	13315
R^2	0.885	0.947	0.933	0.921

Table 7: Intrafirm Cloud Vendor Diversity and Establishment Revenue after GDPR Enforcement. This table shows estimates from difference-in-differences models based on Model 2 (columns 1 – 3) and Model 3 (column 4) in Section 3.5. In each panel, columns 1 – 3 estimate the differential effects by interacting the post-event dummy with the binary indicator of high GDPR exposure. Column 4 estimates the differential effects by interacting the post-event dummy with the high intrafirm cloud vendor diversity indicator. The sample underlying each column is described in the table headers. All models are estimated in panel regressions with firm-establishment fixed effects and year fixed effects. Robust standard errors are clustered at the corporation level. $***p < 0.01$; $**p < 0.05$; $*p < 0.1$.

Dependent Variable	**Log (Revenue)**			
	(1)	(2)	(3)	(4)
Sample	Intrafirm Vendor Diversity ≥ 0.01	Intrafirm Vendor Diversity ≥ 0.01	Intrafirm Vendor Diversity < 0.01	High GDPR
Exclude Big	No	No	Yes	No
(Year ≥ 2018) × High GDPR Exposure	-0.247***	0.006	0.079	
	(0.064)	(0.014)	(0.068)	
(Year ≥ 2018) × High Vendor Diversity				-0.302***
				(0.058)
Log (Employment)	0.768	0.653	1.027	0.754
	(0.736)	(0.708)	(1.082)	(0.610)
Log (IT Staff Size)	-0.335	0.069	-0.534	-0.355
	(0.468)	(0.436)	(0.762)	(0.496)
FE: Establishment	Y	Y	Y	Y
FE: Year	Y	Y	Y	Y
Observations	6003	88536	22724	13315
R^2	0.908	0.947	0.928	0.922