Capgemini
RESEARCH INSTITUTE

DATA SHARING MASTERS

How smart organizations use data ecosystems to gain an unbeatable competitive edge

EXECUTIVE SUMMARY

Data sharing ecosystems arise when organizations agree to share data and insights under locally applicable regulations to create new value for all participants. These ecosystems can take different forms: data brokerages providing aggregate data to its clients, reciprocal data sharing among supply chain partners, sharing of insights across sector boundaries, among others.

The potential financial benefits of data sharing ecosystems can reach 9% of annual revenue of an organization in total in the next five years (equivalent to over $940 million for a typical organization with annual revenues of $10 billion)

- By engaging in data ecosystems, organizations have, on average, improved customer satisfaction by 15%; improved productivity/efficiency by 14%; and reduced costs by 11% annually in the last 2-3 years

- We estimate that data sharing ecosystems have the potential to deliver significant financial benefits – including new revenue streams, cost savings, and productivity improvements – over the next five years based on a typical organization with $10 billion annual revenues:
  – Conservative scenario: $235 million (2.4% of annual revenue)
  – Optimistic scenario: $940 million (9.4% of annual revenue)

- Organizations extensively using external data (i.e., making use of more than 7 external data sources) also exhibit superior financial performance:
  – Up to 14x higher fixed asset turnover and 2x higher market capitalization.

Driven by business benefits, organizations are accelerating data ecosystem engagements

- 48% of organizations plan to launch new data ecosystem initiatives in the near future; 84% of them will do so within the next three years
  – Telecom (81%), banking (73%), and consumer goods (60%) are the sectors with the highest proportions of organizations that plan to launch new data ecosystem initiatives.

- One in four organizations (25%) will invest upwards of $50m in data ecosystems in the next 2-3 years.
  - These firms will come principally from the telecom (55%), banking (43%), and insurance (28%) sectors

  - At a geographic level, the US (47%), UK (36%), and Germany (27%) will have the largest share of firms making major investment commitments (those investing >$50m).

Data ecosystems cut across sectors, data domains, and value chains

- Data ecosystems are found and used across different sectors, data domains, and value chains

- The most common use cases among the 37 unique ecosystem use cases that we analyzed for this research are:
  - Tracking advertising efficiency and targeting
  - Sharing data and insights on new research and development
  - Sharing product and service consumption and usage data
  - Sharing aggregate mobility data for services such as smart cities, parking, mapping etc.
  - Sharing online purchase and transaction data.

Despite the commitment to data ecosystems, their full potential is still untapped

- We found that a majority of organizations (61%) primarily engage in data ecosystems involving simple data sharing and low levels of collaboration. On the other hand, only 14% of organizations are engaged in more collaborative models with extensive data sharing.

- Organizations involved in more collaborative data ecosystems have the potential to significantly outperform organizations in less complex and collaborative ecosystem models:
  - These collaborative leaders drive an additional 10 percentage points of financial advantage (including new revenue, higher productivity, and lower costs) in the next three years.

Following a clear data sharing roadmap can deliver a sustained advantage

We believe that organizations following a clear roadmap for data ecosystem engagements will build a sustained advantage for themselves and their partners. This roadmap involves:

- Formulating a data ecosystem strategy: clearly articulating the purpose of engaging in data ecosystems, identifying data-sharing use cases that create value, and establishing leadership and governance for data-sharing initiatives.

- Making key design decisions: deciding what data can be shared (sourced or supplied), who the trusted ecosystem partners will be, and which collaboration model and business model will work best.

- Kicking off implementation by running small-scale pilots, aligning required capabilities, analyzing processes that need to be transformed, and defining new internal ways of working.

- Sustaining the advantage by scaling up the use cases to their full potential; measuring and monitoring success; and moving up the data and insights value chain.

- Proactively addressing privacy, ethics, trust, and regulatory requirements.

INTRODUCTION

It is common knowledge that organizations are awash with data. However, the real value of data is derived not from its accumulation but from its selection and effective use. Our recent research on the data-powered enterprise has confirmed this: we found that only a minority of organizations (39%) are able to turn data-driven insights into a sustained competitive advantage.[1]

That research, however, also identified a high-performing cohort of organizations that are able to turn volumes of data into value. These "data masters" lead the way on a number of fronts. They possess the tools and technologies needed to analyze and apply data (data foundations or enablers). They also have the required people, skills, processes and organizational culture (data behaviors). And, these data masters significantly outperform other organizations on a range of financial and other performance parameters.

A common characteristic of data masters is their ability to use external data to enhance data-driven insights and decision making. The research revealed that a large majority of data masters are looking beyond traditional sources of external data and making use of non-traditional sources such as data aggregators and data disruptors (e.g. hyperscalers such as Google, Amazon, and Facebook). Yet, getting access to external data is challenging; relevant, good-quality external data is especially hard to find. This is where data ecosystems can help.

Data sharing ecosystems arise when organizations agree to share data and insights under locally applicable regulations to create new value for all participants.
Data ecosystems are helping organization to, for instance:

• Run analytics on logistics partner data to better optimize supply chain

• Build a truly end-to-end view of the sustainability impact of the new products before launching them

• Learn from drug discovery data from many Pharma companies to bring safer drugs to market faster

Data ecosystems have existed in some form for over a decade. In one of the earliest forms of for-profit data ecosystem, banks, credit-card companies, and lending institutes have been partnering with credit-reporting agencies and bureaus and sharing consumer-credit information. Today's data ecosystems have come a long way and possess several characteristics that are distinct from those in the past:

• Complexity (multi-party, heterogeneous, including data from individuals as well as from businesses)

• Transparency: enable trust and participants' sovereignty over their data

• Openness and dynamism (easier to join, participate and leave)

• Following modern standards and technology for interoperability

• Ability to build on trusted software infrastructure.

Examples include:

1. **data.europa.eu** provides access to open data made available across Europe to organizations and the general public. There are over 1,300,000 datasets from 36 countries, available across all major sectors, across 82 catalogues (e.g. portals of open data from various EU institutions, Member States, and government departments and agencies), ready for use via various open-data licenses.[2]

2. **Mayo Clinic** – a US-based provider of integrated healthcare services, education, and research – launched a data marketplace called the Clinical Data Analytics platform in 2020. The marketplace provides access to anonymized patient data, including disease patterns, diagnosis, treatments provided, and care plans, to healthcare organizations, providers, and life-sciences companies.[3] This assists organizations in discovering new drugs and more effective treatment plans, solving complex medical problems and, ultimately, saving patients' lives.

3. **Melloddy** is a partnership between 10 pharmaceutical companies – including Amgen, AstraZeneca, Novartis and Bayer – to share access to an extensive chemical library consisting of more than 10 million small molecules. The partners share data and use AI enabled by Nvidia to promote drug discovery and research & development.[4]

The last few years have seen rising interest among organizations in coming together to create these ecosystems. A number of factors have led to this:

- **From a business standpoint:**
  We found that 54% of organizations say that a renewed push to monetize there is major reason they will engage in data ecosystems. Since the dawn of the Big Data era, organizations have been amassing swathes of user data. Estimates suggest that the volume of data created and captured worldwide was about 59 zettabytes in 2020[5], which is equivalent to the storage space of 8 billion 128 GB iPhones. The combined value of the data economies of the EU27 countries alone is estimated to be €325bn in 2019[6]. This is expected to grow to over €550bn by 2025, representing 4% of the overall EU GDP. Yet, our research found that fewer than half (43%) of organizations are able to monetize their data and insights through products and services, and only 22% are able to quantify the value of data.

- **From a technology standpoint:**
  Estimates suggest that by 2023, there will be 3.6 networked devices and connections per person on Earth, up sharply by 1.5x since 2018[7]. These devices will give rise to vast amounts of data that will be invaluable to organizations. GSMA estimates that the global Internet of Things market will be worth $900 billion by 2025 – 3x its size in 2019, and a majority of this revenue will be generated by "applications, platforms and services that exploit the insights generated by data"[8]. With recent technological developments, it has now become possible for large organizations to share large amounts of data under a secure and real-time cloud-based data exchange e.g., with Snowflake, AWS, and Dawex. Solution providers – such as LeapYear and Harbr – allow collaborative insights from shared data without the need to share granular data.[9] Didier Navez, senior vice president of strategy and alliances at Dawex – a data exchange platform technology provider – told us: *"Across sectors, organizations are engaging more with data ecosystems. A confluence of factors is leading to this heightened activity. First, there's a growing realization that data can be more than an asset – it can become a product in itself. Second, technology is increasingly making it easy to collect, manage, and access new types of data. And third, there is a sharply growing need for exchanging data in trusted environments in compliance with relevant regulations."*

- **From a regulatory and sovereignty standpoint:**
  Rising citizen and consumer expectations around privacy and security of data have led to the adoption of a range of comprehensive regulations, such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), which define the rights of data subjects and how freely data can be shared across organizations. More regulations like these are on the horizon around the world. Alongside regulation, there are calls for data sovereignty (the capability of an individual or an organization to have control over their personal and business data)[10]. Can these developments spur innovation while ensuring data privacy and security? Most notably, in 2020 the EU published its European Strategy for Data[11], which aims to support Europe's competitiveness in the data economy, while also promoting data sovereignty[12] (read more on the EU data strategy in the boxout on page 24: "How are organizations preparing for the new EU data strategy?").

- **From a sustainability standpoint:**
  Organizations around the world are scrambling to make progress on sustainable development goals and climate targets. The adverse effects of climate change are increasing in frequency and impact. Three in five organizations (60%) cite willingness to make progress on sustainable development goals or climate change as a top driver to take part in data ecosystems. Among organizations that want to create societal and sustainability impact with data ecosystems, a majority of organizations focus on engaging in data ecosystems for sustainable development from an environmental, social, and governance (ESG) perspective (73%) and on enabling social inclusion for marginalized sections of society (65%).

So, how can data ecosystems help organizations derive competitive advantage? And can they help create new value in terms of innovative, data-based products and services, while benefitting society? To help answer these questions, we launched an extensive survey of 750 industry executives in 12 countries. In addition, we spoke with over 30 industry experts and academics (more details on the methodology are in the Appendix). Our findings examine the following key data ecosystem dimensions:

- How organizations are accelerating data ecosystem engagements in pursuit of significant business value

- How data ecosystems drive value across sectors

- How the full potential of data ecosystems remains untapped

- How organizations can design, implement and sustain a winning data ecosystem strategy.

What do we mean by data ecosystems?

For this research, we define data ecosystem as a partnership between multiple institutions to share and manage data to create new value that would not have been possible in the previous, siloed system. "Data ecosystems" is not a universal term and there are many different terms used to refer to data sharing, e.g. data estates, data marketplaces, data spaces, among others. This sharing happens in compliance with all local regulations and guidelines, and protects confidentiality and privacy, especially for end-consumer or citizen data. For instance, in a traditional data ecosystem:
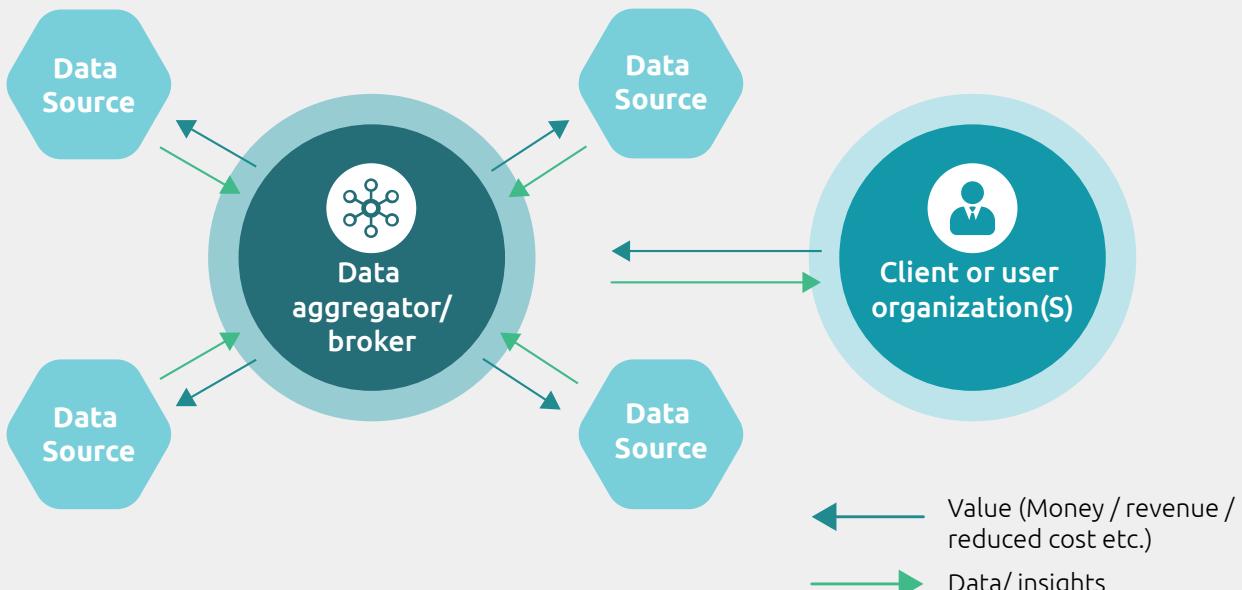
- Multiple organizations (data-source organizations, for example) pool their data on proprietary end-consumer relationships
- There could be one or more aggregator/broker organizations that facilitate this data sharing; data ecosystems may or may not have mediators, or one of the source organizations can serve as an aggregator
- The client organizations use data aggregated by the broker and, in exchange, provide some value in the form of monetary (revenue, income) or non-monetary incentives (e.g. reduced costs, reduced risks, improved customer experiences).

Newer technologies, such as homomorphic encryption and differential privacy, are enabling more detailed collaboration while meeting regulatory requirements around privacy and security. Differential privacy allows parties to give access to granular data for collaborative analysis without sharing the granular data themselves. Data-sharing is based on reciprocal value exchange, making shared data beneficial to all participants. Using a variety of sources also enhances diversity and inclusivity of information.

In this research, we analyze four broad types of data ecosystems (see Figures 1-4 below):

- Data brokerage and aggregation ecosystem
- Reciprocal data-sharing ecosystem
- Federated analytics ecosystem
- Collaborative data-supply-chain ecosystem.

Figure 1. Data brokerage and aggregation

Value (Money / revenue / reduced cost etc.)

Data/ insights
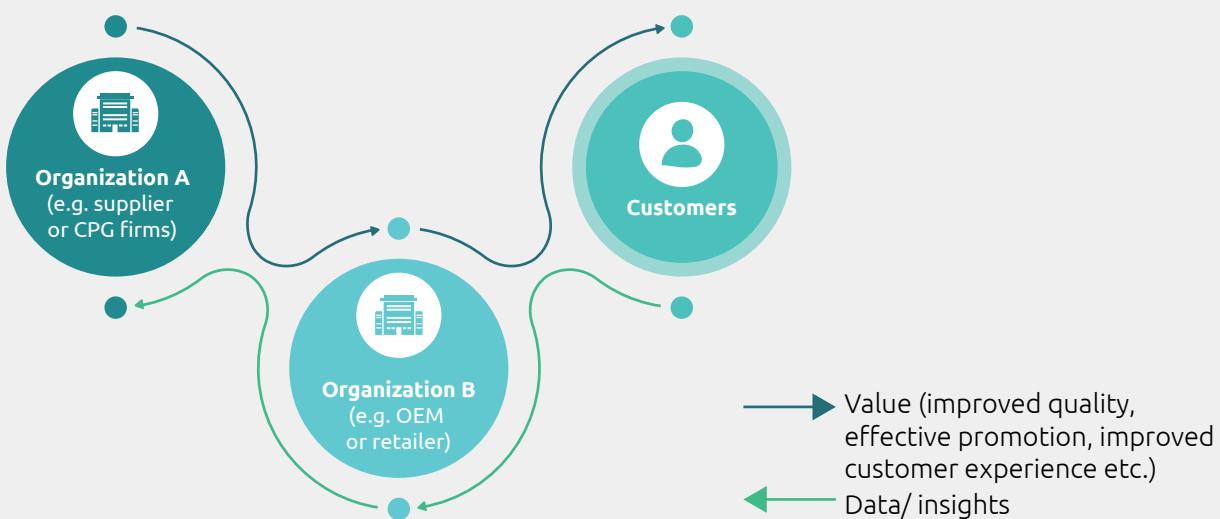
Source: Capgemini Research Institute analysis

I. Data brokerage and aggregation ecosystems (primary model for 55% of organizations in our research)

In this type of ecosystem, an organization has, or aggregates, data that holds direct value for its clients. Data, or the insights derived from it, are often commoditized. It can be collected from consumers or end-users directly through use of apps or online forms, or can be sourced via connected devices such as health trackers etc.

Example: Thinknum, a US-based data-aggregation platform, collects data from government sources, social media, and company websites. It helps companies such as retailers to select store locations based on their potential profitability.[13]

Figure 2. Reciprocal data-sharing configuration

- **Organization A** (e.g. supplier or CPG firms)
- **Organization B** (e.g. OEM or retailer)
- **Customers**

→ Value (improved quality, effective promotion, improved customer experience etc.)
← Data/ insights
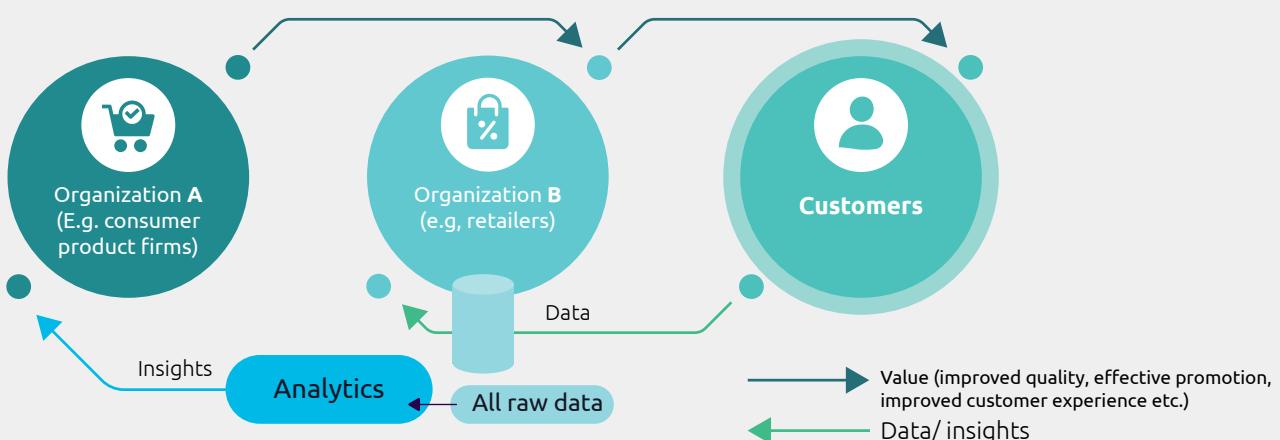
II. Reciprocal data sharing (Primary model for 20% of organizations in our research)

In this model, multiple companies work together to drive efficiency. The sharing of data results in the value of collaboration increasing. Usually, one of the partners is the "dominant" player but smaller players can also obtain value from sharing data. Often, these start with the sharing of information around terms agreed beforehand.

Example: A parts manufacturer keeps the original equipment manufacturer (OEM) informed on logistics. In return, the OEM informs them of the results of quality testing.

Figure 3. Federated analytics ecosystem

- Organization **A** (E.g. consumer product firms)
- Organization **B** (e,g, retailers)
- **Customers**

Insights
Data
Analytics
All raw data

→ Value (improved quality, effective promotion, improved customer experience etc.)
← Data/ insights
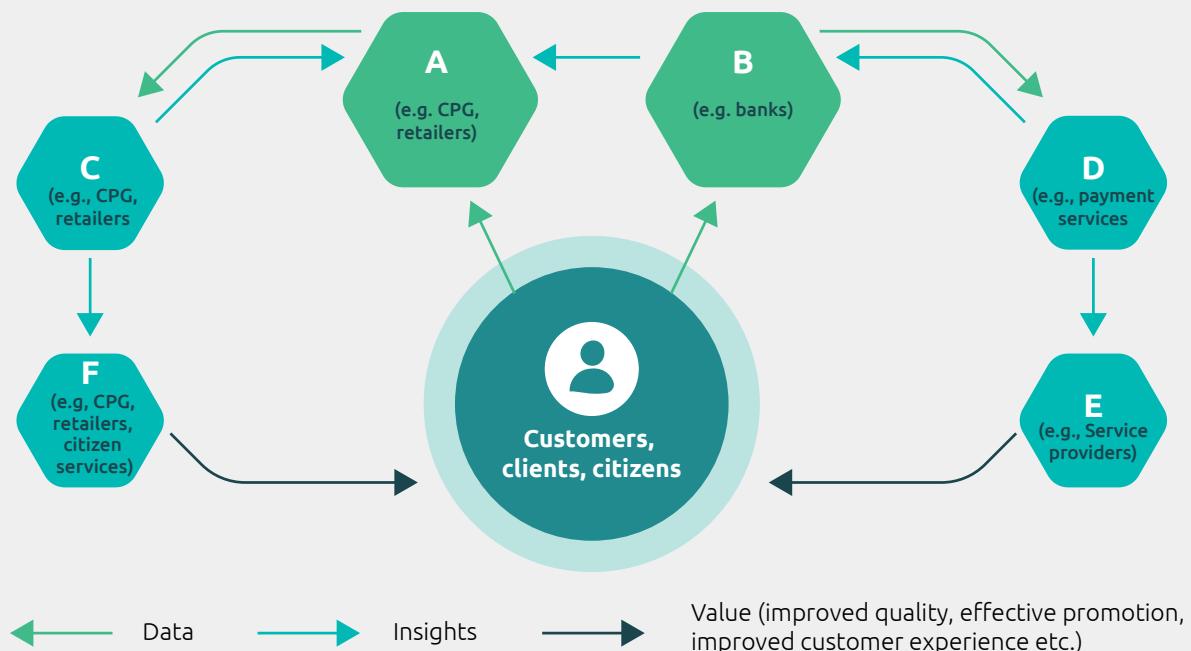
III. Federated analytics ecosystem (Primary model for 13% of organizations in our research)

The key objective of this model is to run analytics on and derive insights from another organization's data, which can't be directly shared or accessed owing to regulatory concerns, data volumes, or other reasons. The business model of this ecosystem is to sell access to data for running analytics and charging is based not only on data supplied but on the complexity and form of the analytics.

Example: NVIDIA's Clara is a federated learning framework that allows different organizations and hospitals to securely collaborate, train, and contribute to a global model. This allows hospitals to securely combine their local knowledge with external entities and reach more accurate diagnosis.[14]

Figure 4. Collaborative data-supply chain

CPG = Consumer product goods

Source: Capgemini Research Institute Analysis.
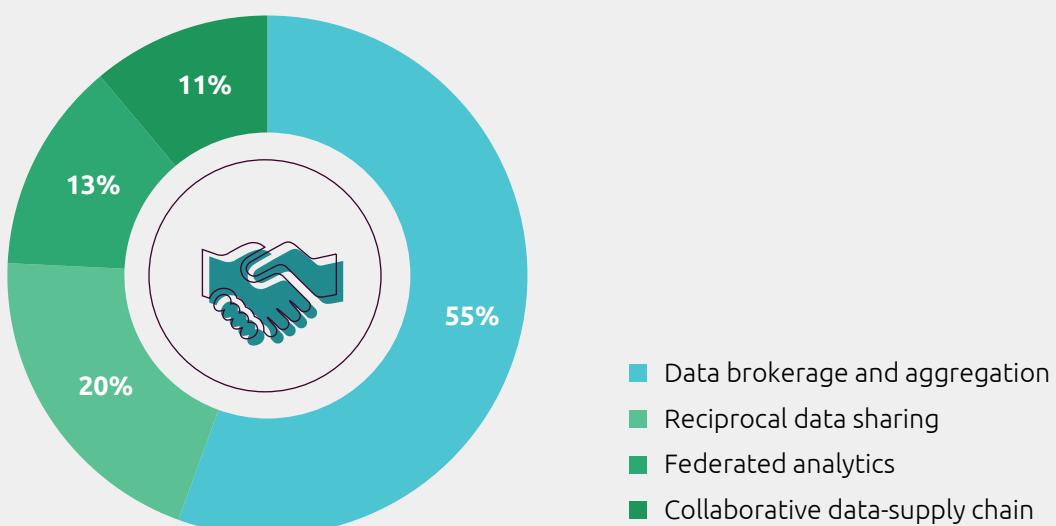
IV. Collaborative data-supply chain (Primary model for 11% of organizations)

This is the most collaborative model of data ecosystem partnership, in which multiple organizations serve a single customer or market. In this model, the collaboration between organizations itself becomes a new product or service, driving an outcome for the customer.

Example: The Open Carbon Footprint brings together different partners in a value chain to enable participants to accurately determine carbon footprint of end products.[15]

| Figure 5. | Distribution of primary data ecosystem models |

Which of the following is your organization's primary model to engage in data ecosystems?

- Data brokerage and aggregation
- Reciprocal data sharing
- Federated analytics
- Collaborative data-supply chain

Source: Capgemini Data Ecosystem Survey; N=750 organizations, April-May 2021.

ORGANIZATIONS ARE ACCELERATING DATA ECOSYSTEM ENGAGEMENTS IN PURSUIT OF SIGNIFICANT BUSINESS VALUE
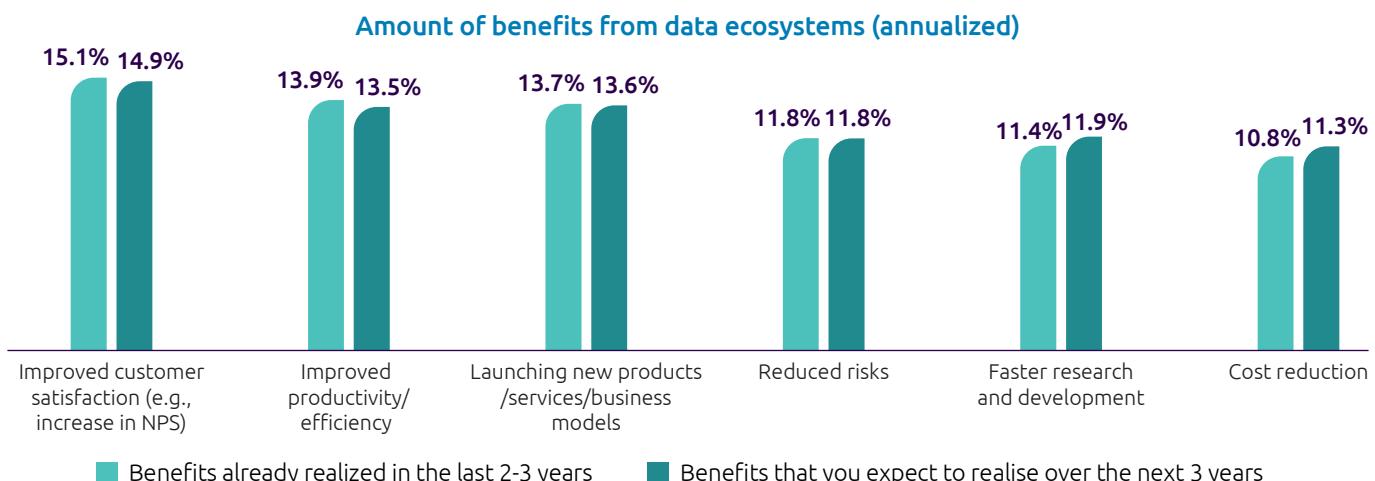
Data ecosystems have made a significant impact on multiple fronts, such as improving customer satisfaction, increasing operational productivity/efficiency, and reducing costs (see Figure 6). Examples include:

- HiLo, a joint initiative to improve risk modelling in the maritime industry, has been successful in reducing lifeboat accidents by 72% across a fleet of 900 ships; engine-room fires were reduced by 65% on a fleet of 1800 ships; and bunker spills reduced by 25%.[17] This is made possible by pooling data from member shipping companies on near-misses, incidents, and accidents. HiLo is able to use this pooled data to create predictive models that can generate

specific recommendations to improve onboard safety for each shipping company.

- Greiner Packaging International – an Austria-headquartered, large-scale packaging provider for the food, chemical, and healthcare industries, among others – was able to identify a potential reduction in fixed-asset investment of 35% by working in partnership with Spain-based logistics insight company, OBUU.[16] OBUU used data from Greiner's five manufacturing plants and analyzed key performance indicators (KPIs) to establish the efficiency of the latter's supply chain. This resulted in significant cost savings for Greiner, which can be used to drive operational efficiency in future.

Figure 6. Improved customer satisfaction and improved productivity/efficiency emerge as the top benefits of engaging in data ecosystems

Amount of benefits from data ecosystems (annualized)

	Improved customer satisfaction (e.g., increase in NPS)	Improved productivity/ efficiency	Launching new products /services/business models	Reduced risks	Faster research and development	Cost reduction
Benefits already realized in the last 2-3 years	15.1%	13.9%	13.7%	11.8%	11.4%	10.8%
Benefits that you expect to realise over the next 3 years	14.9%	13.5%	13.6%	11.8%	11.9%	11.3%

■ Benefits already realized in the last 2-3 years   ■ Benefits that you expect to realise over the next 3 years

Percentages indicate the amount of annualized benefit realized over the level of three years earlier or expected to be realized over today's levels.

Source: Capgemini Data Ecosystem Survey; N=750 organizations, April-May 2021. NPS stands for net promotor score – a key customer satisfaction metric.

Data ecosystems have led to significant gains in customer satisfaction, productivity, and innovation

Our survey shows that, in the last three years, organizations have achieved benefits of more than 10% (annualized) on parameters such as cost and risk reduction (see Figure 6). On parameters such as customer satisfaction, the statistic goes up to more than 15%, demonstrating the capability of data ecosystems to solve customer pain points. Organizations are optimistic about data ecosystems and expect these benefits to continue in the next three years.

Organizations participating in data ecosystems have the potential to gain financial benefits up to $940 million or over 9% of annual revenue in the next five years

To analyze the overall impact of data ecosystems on financials of organizations, we take the case of a typical organization in our study which has an average revenue of $10 billion. We estimated the impact of advantages arising from data ecosystem participation on three main financial areas over the next five years:

• New revenue opportunities from innovative products, services and business models

• Productivity improvements

• Cost savings.

While estimating the impact of data sharing in ecosystems on new revenues of organizations, we consider a two limiting factors:

1. The EU data market[18] study which estimates that the data economy will constitute 4% of the EU's GDP by 2025 in a baseline scenario. From this, we estimate that data sharing in ecosystems can potentially add 4% of additional revenue to an average organization in an optimistic scenario. We estimate this to result equally – 2% each – from new revenue potential and productivity improvement potential, and potential cost savings to impact 4% of the overall cost of operation.

2. From our discussions with industry experts and clients, we estimate that in a conservative scenario, the top line impact will be limited to 1% of organizational revenue. We estimate this 1% to equally come from new revenue opportunities (0.5%) and productivity improvements (0.5%), and potential cost savings to impact 1% of the overall cost of operation.

Our analysis shows that an organization in our survey with an average revenue of $10 billion participating in data ecosystems can potentially gain new revenue of up to $95 million in the conservative scenario and USD $378 million in the optimistic scenario, globally and cumulatively in the next five years (table 1). The details of this analysis are given in the appendix.

Table 1. Financial benefits of data ecosystems for an average organization with annual revenues of $9 billion

	Conservative estimate	Optimistic estimate
Average revenue of organizations in our survey, global (USD billion)	10.0	10.0
Timeline (years)	5	5
Impact of data ecosystems on total global revenue (based on economic estimates, details are in appendix)	0.5%	2%
Future benefits from launching new products/services/business models from data ecosystems (annualized, based on survey results)	13.6%	13.6%
Additional revenue from data ecosystem per company, cumulative in five years (USD M) (computed from above)	95	378

Continuing with similar calculations for productivity gain and cost savings, we calculate that organizations stand to gain nearly $235 million of financial benefits (2.4% of annual revenue) in a conservative scenario and up to $940 million (9.4% of annual revenue) in an optimistic scenario (see table below).

Financial head	Conservative scenario	Optimistic Scenario
Additional revenue (USD M)	95	378
Additional cost savings (USD M)	47	186
Additional productivity gain (USD M)	94	377
Financial advantage per organization, cumulative over the next five years (USD M)	235 (2.4% of revenue)	940 (9.4% of revenue)

Source: Capgemini Data Ecosystem Survey; N=750 organizations, April-May 2021. Capgemini analysis.

Organizations are planning to accelerate their data ecosystem engagements

We expect a significant acceleration in data ecosystem initiatives in the near future. Nearly half of the surveyed organizations (48%) are looking to enter into new ecosystems or initiatives and more than one in three organizations (36%) are working on strengthening their existing ecosystem initiatives (see Figure 7). Eighty-four percent of the organizations that intend to launch new data ecosystems initiatives plan to do so within the next three years.

Christina Poirson – Group chief data officer at Société Générale – outlines how regulatory initiatives are helping accelerate ecosystem initiatives: *"Data ecosystems are advancing quite rapidly. Clearly, in the EU we are witnessing a strong push from regulation to establish smoother data sharing in the European data space for financial services. It is very much aligned with our own perspective of maximizing the value of data while respecting the protection of sensitive data."*

The rise of the open banking ecosystem in the UK – comprising 300 members, including leading UK banks and FinTechs, and over 2.5m UK customers – is a perfect example of this growing interest in data ecosystems. The legislation that enabled open banking came into effect in January 2018, with the Payment Services Directive 2 (PSD2) coming into force across Europe at this time. Open banking allows customers to share access to their financial data securely with providers (banks and FinTechs, among others) to get new products and services, manage finances, access credit, and make payments. Banks and other financial institutions use application programming interfaces (APIs) to share consumer-consented information securely to generate new insights and services. In 2018, open-banking customers made 320,000 payments and generated 67m API calls. By 2020, this had risen to over 4m payments (13x) and 5.8bn API calls (87x)[19]. Today, the ecosystem has over and 2.5 million UK consumers and businesses using open-banking-related products and services.

At AXA Singapore, Priyanka Jha, manager, value proposition – told us: *"I definitely see an acceleration in data ecosystem initiatives and I believe there are three major drivers behind it. First, organizations are trying to advance their sustainability and social actions and commitments. Second, there is greater willingness to share data outside one's organization and a realization of the value contained in it. And third, there is availability of technologies and platforms to securely and privately share data in accordance with regulations that are getting stricter."*

The COVID-19 pandemic served as a catalyst for data sharing. Brands in China, for instance, shared their data on Alibaba's Freshippo platform and benefited from the services of their delivery partner, Cainiao, to move online quickly and continue operations.[20] The pandemic also led to an unprecedented rise in delivery volumes. Data sharing between Freshippo and enterprises such as restaurants helped onboard more than 5,000 temporary employees for delivery and picking tasks when their permanent workplaces faced government-mandated shutdowns.[21]
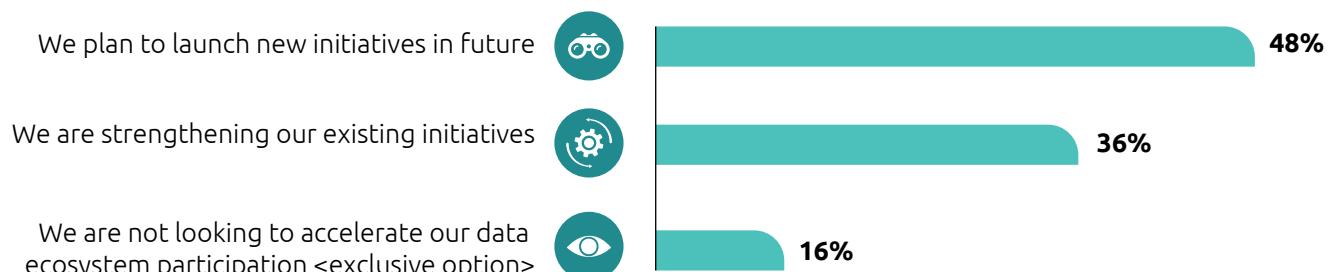
"At first, you need a trusted environment where organizations can come together to collaborate with faith in the onboarding and vetting of participants and their respective roles, and second, data exchange has to happen in a trusted manner, in compliance with regulations under mutually agreed terms and conditions, including transparency in data sharing and pricing."

- Didier Navez,
Senior vice president of strategy and alliances at Dawex

Figure 7. Nearly half of organizations are planning to increase their data ecosystem engagement in the near future

Are you planning to increase your organization's engagement with data ecosystems? (multiple selections)

We plan to launch new initiatives in future	**48%**
We are strengthening our existing initiatives	**36%**
We are not looking to accelerate our data ecosystem participation <exclusive option>	**16%**

Percentages indicate the amount of annualized benefit realized over the level of three years earlier or expected to be realized over today's levels

Source: Capgemini Data Ecosystem Survey; N=750 organizations, April-May 2021.

Consumer-facing sectors such as telecom, banking, consumer goods, and retail, are usually in the vanguard when launching new data ecosystem initiatives. These sectors have been front-runners in using data for decision making. Our research on data-powered enterprises has also found that telecom, insurance, and banking lead all other major sectors in data-driven decision making.
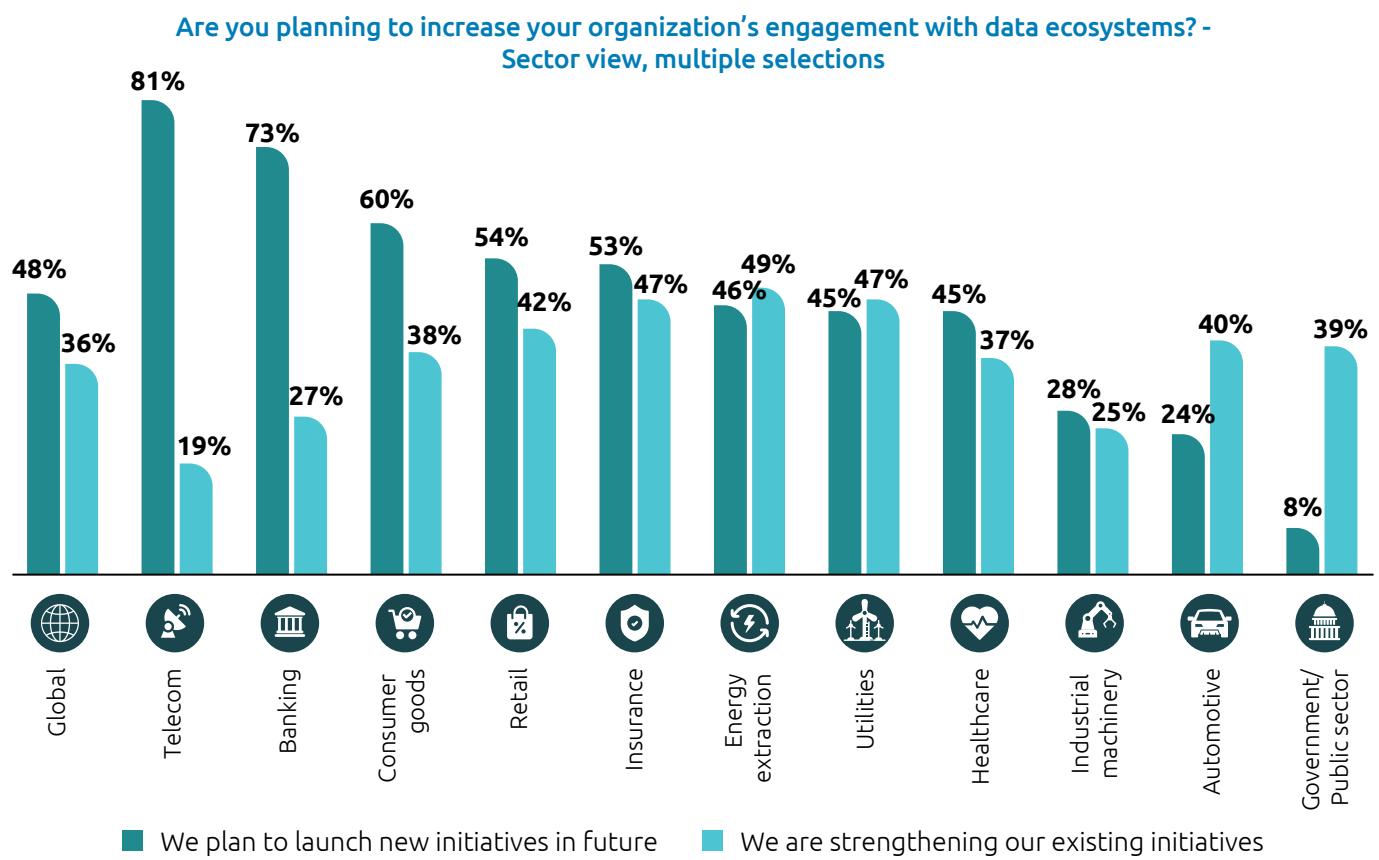
One in four organizations to invest more than $50m in data ecosystem initiatives in the next three years

A majority of organizations will invest between $10m and $50m in data ecosystem initiatives in the next 23 years. This includes all investments made towards acquisition of technology infrastructure, tools, talent and skills, and process re-engineering, among others.

Interestingly, a large share of organizations, one in four, will invest upward of $50m in the next three years. This cohort of organizations is likely to reap greater rewards over the next five years or so, as their new investments in data ecosystems start to bear fruit. On average, there will be an investment of $40 million per organization, though the level of investment varies greatly among sectors and countries. Among sectors, organizations planning to launch new initiatives (as seen earlier) are looking to invest large amounts as well. Fifty-five percent of telcos will be investing more than $50m, while 43% of banking-sector firms will do so (see Figure 9). Among countries, the US and UK will be the biggest spenders, with more than one in three organizations spending over $50m over the next three years.

Are you planning to increase your organization's engagement with data ecosystems? - Sector view, multiple selections

Sector	We plan to launch new initiatives in future	We are strengthening our existing initiatives
Global	48%	36%
Telecom	81%	19%
Banking	73%	27%
Consumer goods	60%	38%
Retail	54%	42%
Insurance	53%	47%
Energy extraction	46%	49%
Utilities	45%	47%
Healthcare	45%	37%
Industrial machinery	28%	25%
Automotive	24%	40%
Government/Public sector	8%	39%

■ We plan to launch new initiatives in future   ■ We are strengthening our existing initiatives

Source: Capgemini Data Ecosystem Survey, N=750 organizations, April-May 2021.

Figure 9. One in four organizations will spend over $50m on data ecosystem initiatives

What is the total investments made/planned to be made in your data ecosystem initiatives (in the next 2-3 years)?

Category	Percentage
Less than USD 10 Million	17%
USD 10 - 50 Million	51%
USD 50-100 Million	15%
More than USD 100 million	10%
Don't know	7%

25%
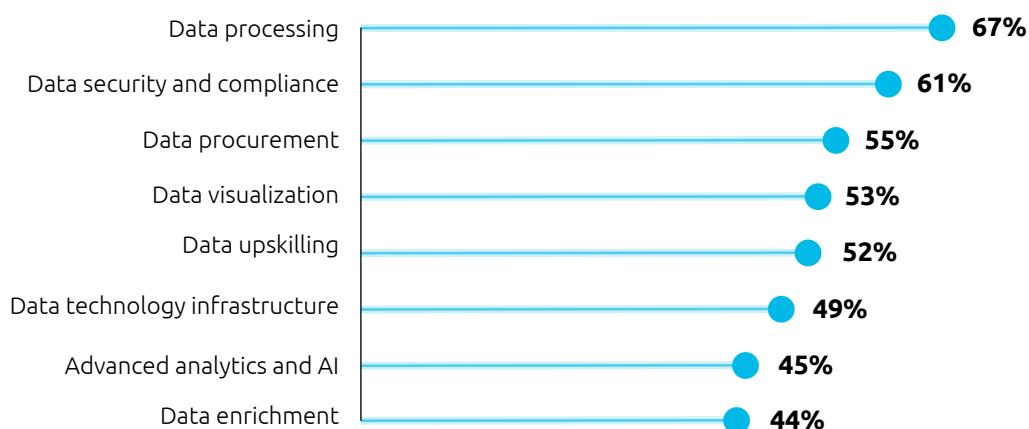
Annual investment made/planning to make in data ecosystem initiatives by sector

Sector	Don't know	Less than USD 10 Million	USD 10 - 50 Million	USD 50-100 Million	More than USD 100 million
Telecom		7%	39%	36%	19%
Banking		8%	48%	20%	23%
Insurance		4%	68%	17%	11%
Retail	3%	13%	56%	14%	13%
Global	7%	17%	51%	15%	10%
Energy extraction (including oil and gas, mining)	5%	17%	51%	14%	8%
Utilities (electricity, water, gas)		29%	50%	13%	8%
Consumer goods	1%	18%	62%	8%	12%
Automotive	21%	17%	43%	13%	5%
Healthcare	3%	26%	53%	13%	5%
Industrial machinery	17%	23%	47%	9%	4%
Government/ Public sector	21%	28%	44%	7%	

Legend:
- Don't know
- Less than USD 10 Million
- USD 10 - 50 Million
- USD 50-100 Million
- More than USD 100 million

*Arranged in descending order of total of investment between $50m-$100m and more than $100m

With regards to data sharing with other organizations via data ecosystems, which of these value-added services do you procure from third-party service providers?

Service	Percentage
Data processing	67%
Data security and compliance	61%
Data procurement	55%
Data visualization	53%
Data upskilling	52%
Data technology infrastructure	49%
Advanced analytics and AI	45%
Data enrichment	44%

Data processing: e.g., making data structure standardized, searchable or the ability to run crosstabs or user-defined functions instead of direct access to raw/sensitive data. **Data security and compliance:** Services to ensure compliance with local regulations e.g. for auditing, privacy (such as anonymization), privacy, and security related purposes. **Data procurement:** Acquiring relevant third-party data on a continuous basis. **Data visualization:** creating visuals of raw data, derived data or insights. **Data upskilling:** Training people on data technologies and hiring people skilled in data. **Data technology infrastructure:** platforms, data lakes, data warehouses, data centers etc. **Advanced analytics and AI:** generating insights from raw data using advanced analytics/Machine Learning/AI. **Data enrichment:** Including data cleansing, deduplication, validation etc. to improve data quality and veracity.

One in four organizations will spend over $50m on data ecosystem initiatives

Annual investment made/planning to make in data ecosystem initiatives by country

Country	Can't say/don't know	Less than USD 10 Million	USD 10 - 50 Million	USD 50-100 Million	More than USD 100 million
US	2%	6%	52%	24%	15%
UK	1%	4%	56%	19%	20%
France	1%	14%	49%	20%	15%
Germany	1%	8%	60%	12%	19%
Global	7%	17%	51%	15%	10%
EU	7%	22%	50%	13%	8%
Netherlands	7%	22%	51%	16%	4%
Norway	5%	23%	53%	11%	8%
Sweden	9%	28%	50%	11%	3%
Spain	11%	34%	42%	11%	3%
Australia	11%	19%	57%	8%	5%
China	27%	25%	38%	3%	8%
India	16%	19%	57%	8%	0%
Italy	21%	40%	37%	3%	0%

Legend:
- Can't say/don't know
- Less than USD 10 Million
- USD 10 - 50 Million
- USD 50-100 Million
- More than USD 100 million

*Arranged in descending order of total of investment between $50m-$100m and more than $100m.

Source: Capgemini Data Ecosystem Survey; N=750 organizations, April-May 2021

Organizations that make extensive use of external data enjoy a financial performance premium

Our research on "*The data-powered enterprise*" was a deep-dive into data practices and related organizational culture. This research found a small cohort (~16%) of "data masters" – companies whose key characteristic is extensively leveraging external data in decision making.

This research also found that, on average, data masters earn 70% more revenue per employee; have 245% higher fixed-asset turnover; and have 22% higher profitability. We re-analyzed the data from this study in the context of external data sharing in ecosystems. In our data-powered enterprise study, we had found 14 unique external data sources are utilized by organizations (the full list of these data sources is in the appendix). For this new analysis we segregated organizations into two buckets – one where organizations leverage up to half of the sources available to them (i.e. up to seven) and the other where organizations leverage more than half of the data sources (more than seven). Figure 10 shows the key results.

Our analysis found that organizations that use more than seven data sources have nearly 14x the fixed-asset turnover and 2x the market capitalization compared to organizations who do not use any external data for decision making. They also have 37% more revenue per employee.

Figure 10. Organizations leveraging a greater number of external data sources exhibit superior financial performance

Number of external data sources used for decision making	Market capitalization	Fixed-asset turnover (FY 2019/20)	Revenue per employee (FY 2019/20)
Out of 14 possible data sources available to them*	Percentages indicate the degree of financial out performance over organizations that do not use external data for their decision making		
Organizations using more than 7 external data sources	119%	1,256%	37%
Organizations using 1-7 data sources	56%	461%	2%
Organizations not using external data sources (base)	0%	0%	0%

How data ecosystems drive value across sectors

Data ecosystems offer benefits on multiple fronts. In this section, we provide three illustrations of how data ecosystems can benefit various partners within a range of parameters. As illustrated below (see Figure 11, 12, and 13), data sharing can lead to benefits for all participants in the ecosystems. While sharing data, organizations must also ensure that the strictest rules are adopted to ensure interoperability, data privacy, confidentiality, and security (see boxout on "Preserving data privacy in ecosystems", which outlines privacy-preserving practices, techniques, and principles).

Data ecosystems span across citizen and consumer domains

Data can be used in a variety of domains to achieve value, from improving the customer experience to operating efficiency.

Take the intersection of automotive with various other domains, for example. As Figure 11 shows, data aggregated from vehicles can not only help automotive firms unlock an additional revenue stream, but also have a positive societal impact and, at the same time, offer a better driving experience. Automotive and mobility-services firms can collect data on commercial-vehicle fleets and consumer vehicles. This data includes aggregated information about destinations, times of travel, frequency of travel, average speeds, etc. For instance:

Data for societal good

The aggregated data is then shared with smart cities for a variety of uses, such as:

- A detailed flow of traffic data can help cities plan traffic flow
- Data on congestion and traffic movement can help identify areas that require parking spaces
- Identifying traffic hotspots and popular destinations can help cities plan citizen services

Smart cities, with the help of the Internet of Things (IoT) and emerging technologies, can share up-to-date, accurate data on current parking availability, traffic conditions, and road blockages with vehicle owners. Automotive firms that use this technology can offer an enhanced customer experience, including a reduction in time spent in traffic. This has the added benefit of lower levels of pollution.
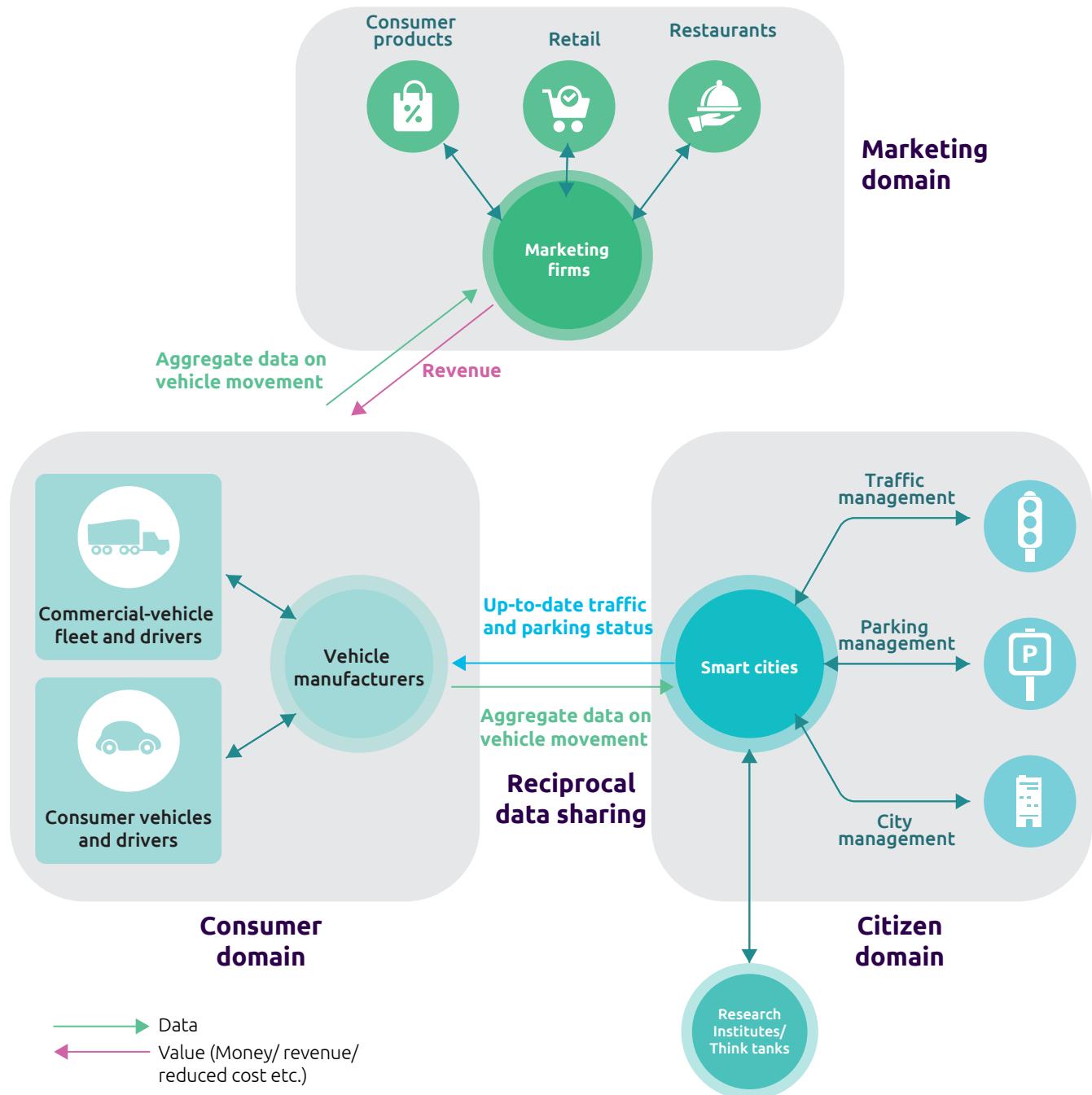
Data as a revenue stream

The aggregated data can be commoditized and offered to third parties. For instance, consumer product firms and advertising agencies can use traffic-flow data to ascertain which locations within the city would provide the highest number of viewers for advertisements. These firms benefit in the form of more effective marketing campaigns and automotive firms unlock an additional revenue stream.

25%

Share of organizations that will invest more than $50 million in data ecosystems in the next 2-3 years

Figure 11. Data ecosystems cut across various models

Marketing domain

Consumer products

Retail

Restaurants

Marketing firms

Aggregate data on vehicle movement

Revenue

Consumer domain

Commercial-vehicle fleet and drivers

Consumer vehicles and drivers

Vehicle manufacturers

Reciprocal data sharing

Up-to-date traffic and parking status

Aggregate data on vehicle movement

Smart cities

Citizen domain

Traffic management

Parking management

City management

Research Institutes/ Think tanks

Data

Value (Money/ revenue/ reduced cost etc.)

Source: Capgemini Research Institute analysis

Data ecosystems cut across various sectors

Data ecosystems allow organizations to collaborate and share data across sectors. Telecom firms, for instance, have troves of data that can unlock significant value for related sectors For example, Figure 12 shows how data from telecom firms can help multiple other sectors.

Retail. Aggregated telecom data offers street-level insights into customer movement. This can help retailers trace customer paths and draw more elaborate customer-journey maps by shedding light on which stores/places customers visited before and after visiting the retailer's store. This insight, cross-referenced with demographic data, can help retailers plan store layouts and make effective marketing decisions.

Sharing retail data with CPG companies offers them a better predictability for their product to optimise the level of stocks, avoid stock shortage thanks to a better anticipation of sales. For fresh food, a better anticipation of sales and a better prediction of required assortment reduce massively the volume of waste.

Governments agencies. Telecom data can be a vital to public sector and government agencies. Aggregate and anonymized data can help governments track citizen
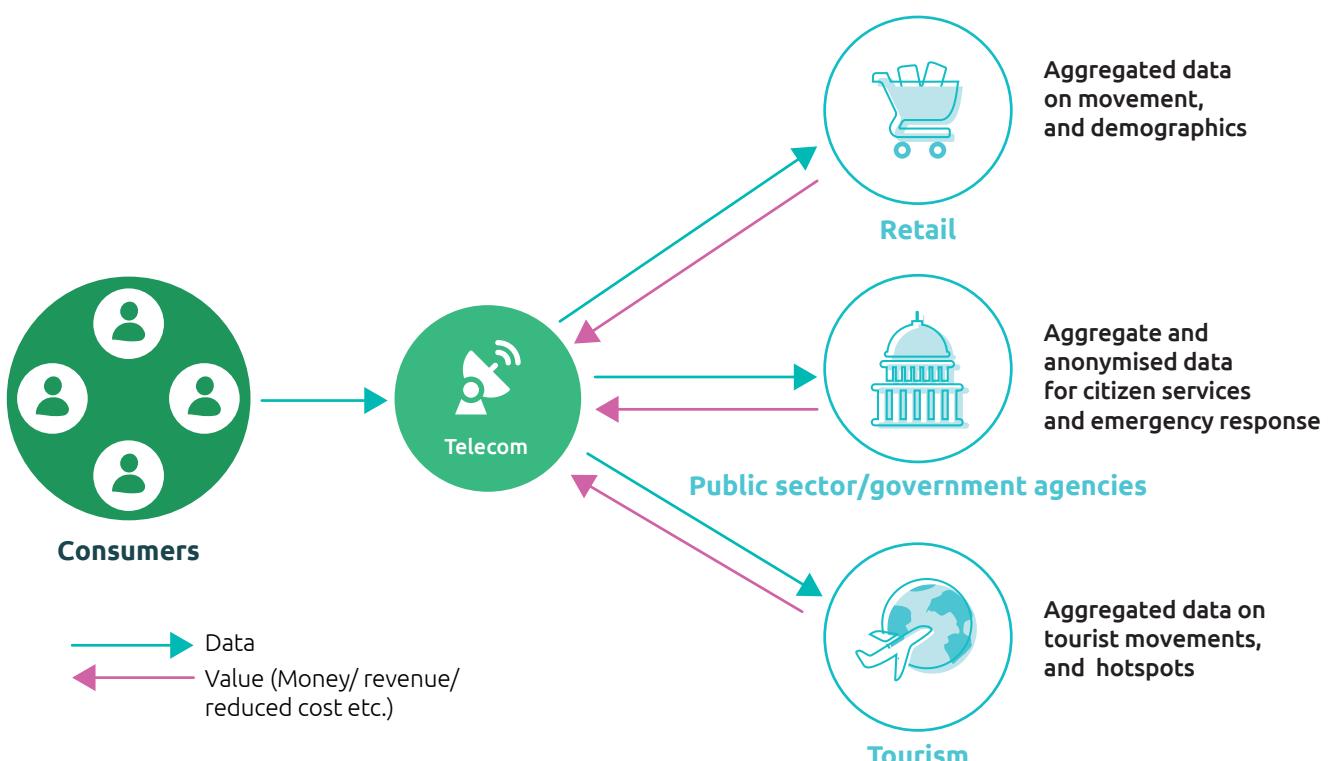
movement, migration patterns, response to emergencies, etc. Telecom data has been a vital component of government response to the COVID-19 pandemic too. Vodafone Analytics Platform provides Lombardy's regional government with heat maps to show changes in population movement before and after containment. Separately, they are also working with authorities to better predict the impact of public policy decisions on the future spread of the virus.[22]

Tourism. Telco data can shed light on the origins of tourists and their most visited eateries and retail sites. This can be vital for local restaurants, since they can tailor their offerings based on tourist preferences. Hotels can also plan inventory and manage resources better. Tourist sites can use this data to get a more accurate picture of average tourist flows, duration of stay, etc.

Data ecosystems enable more efficient dissemination of public services

Seamless data sharing between public entities can help cities make the transition to "smart" status and offer citizens the opportunity to participate in the development of their cities. Private enterprises can also take part in these ecosystems as a way of augmenting their services to citizens.

Figure 12. Data ecosystems offer benefits across various sectors
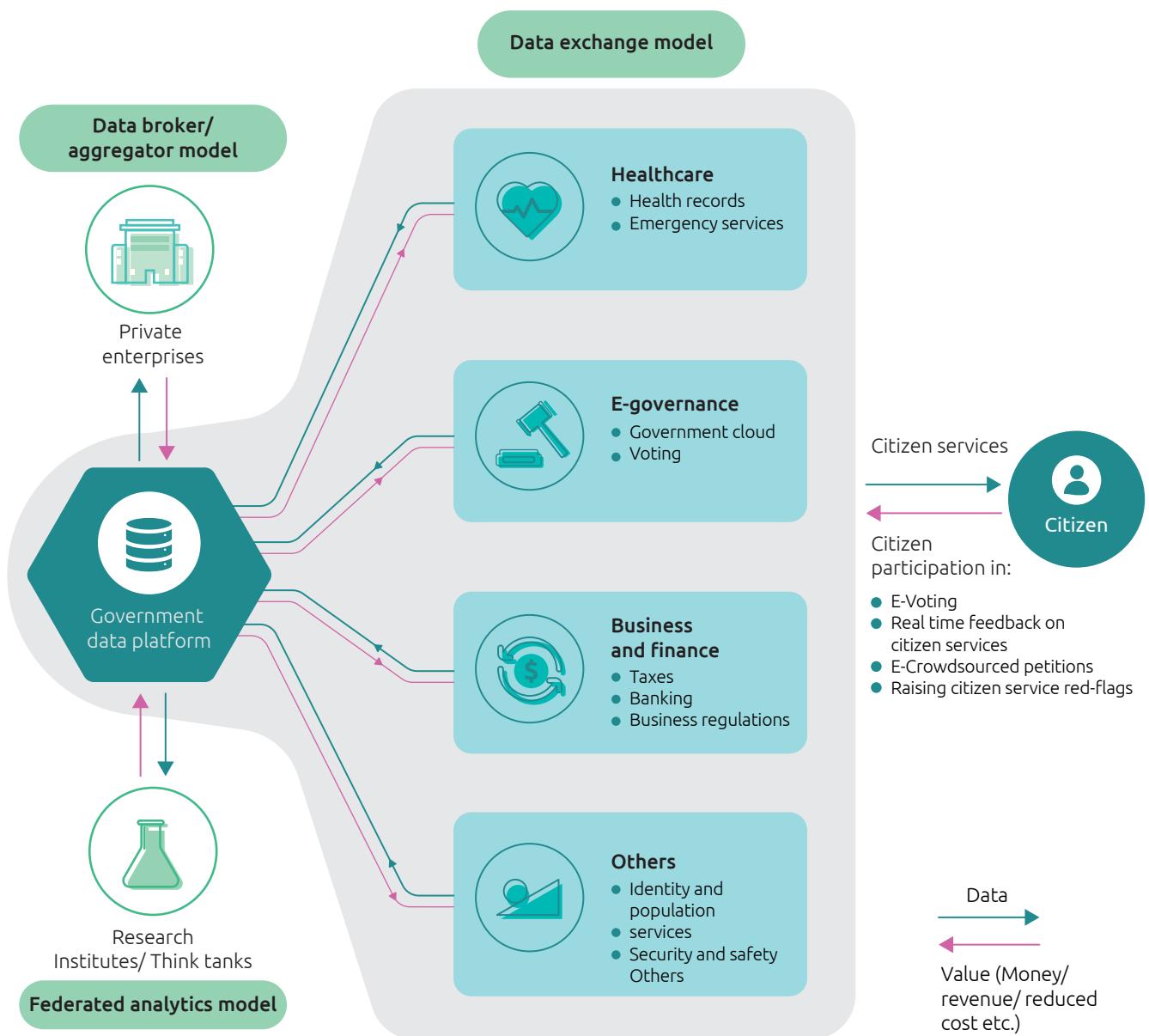
Source: Capgemini Research Institute analysis

Citizen services related to governance, safety, business, and finance can be onboarded to a data-exchange platform to enable data sharing and seamless delivery of e-services to citizens. This also enables citizens to voice their opinions, questions, and concerns at any time, from anywhere. For instance, citizens can immediately raise red flags about non-operating citizen services; crowdsource concerns in a more transparent manner; and highlight pain points that affect a large proportion of citizens.

The mobility is also accelerated by data partnership, federating all the kind of transportation, including the green ones (scooters or bikes) to provide the best mobility services.

Estonia has implemented e-Estonia and offers 99% of public services online, with 24/7 availability. The building blocks of this ecosystem include e-identity, healthcare, security and safety, and e-governance. Implementing the data exchange has led to large gains. Estonia has already saved 844 years of working time, with more than 99% of banking transaction and tax declarations carried out online. Apart from this, 44% of Estonians used i-Voting to cast votes for the Estonian parliament in 2019.[23]

Figure 13. Data ecosystems enable more efficient delivery of public services via various models
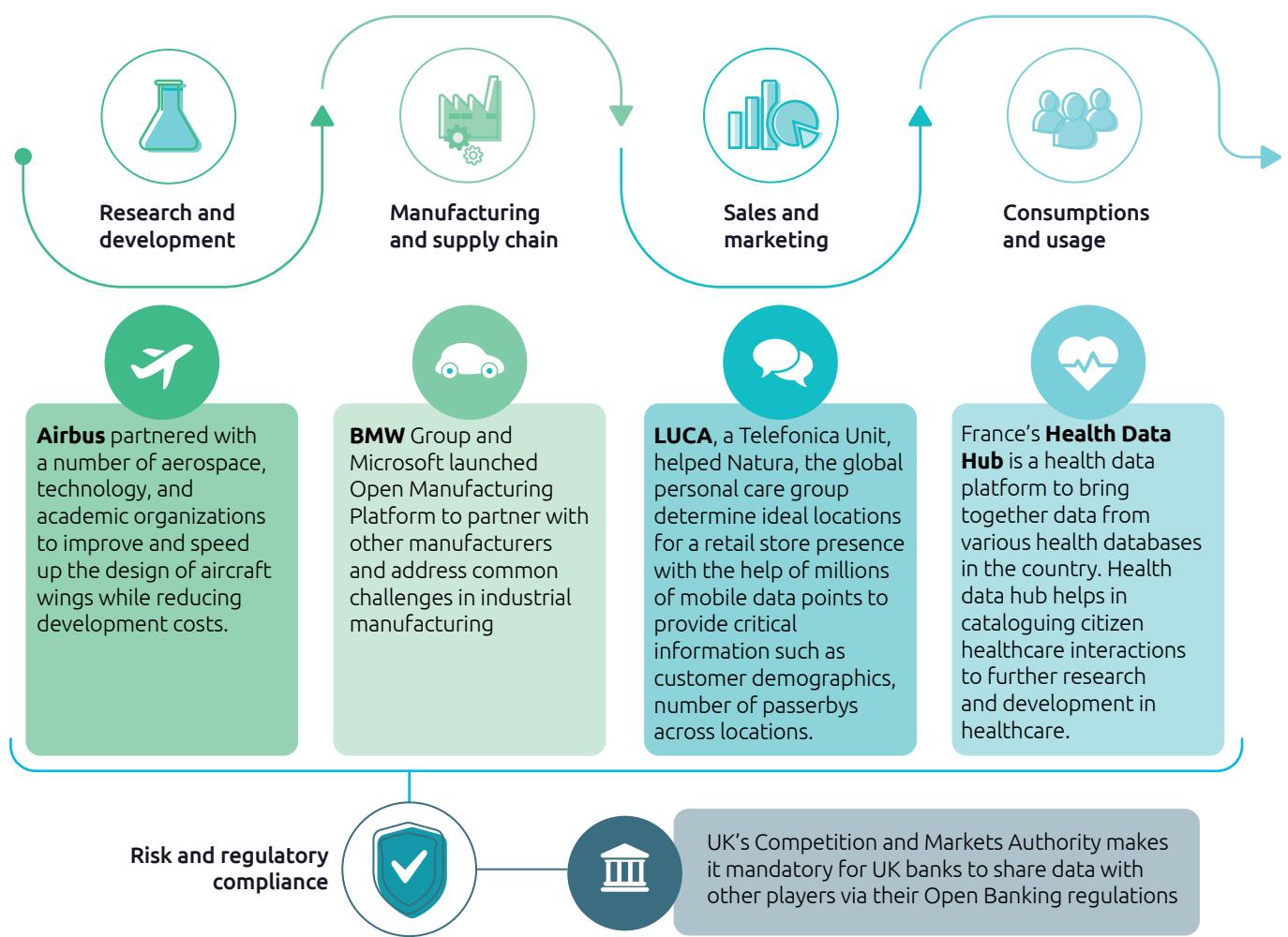
Source: Capgemini Research Institute analysis

The most common use cases across sectors span from research and development to consumption of products and services

While the potential of data ecosystem use cases remain limitless across domains in an organization, we collected a repository of 37 use cases across sectors for which organizations usually engage in data ecosystems. Among them, our survey revealed use cases that are most commonly adopted by organizations.(see Figure 14).

Figure 14. Most common data-sharing use cases by value-chain section

Research and development

Manufacturing and supply chain

Sales and marketing

Consumptions and usage

Airbus partnered with a number of aerospace, technology, and academic organizations to improve and speed up the design of aircraft wings while reducing development costs.

BMW Group and Microsoft launched Open Manufacturing Platform to partner with other manufacturers and address common challenges in industrial manufacturing

LUCA, a Telefonica Unit, helped Natura, the global personal care group determine ideal locations for a retail store presence with the help of millions of mobile data points to provide critical information such as customer demographics, number of passerbys across locations.

France's **Health Data Hub** is a health data platform to bring together data from various health databases in the country. Health data hub helps in cataloguing citizen healthcare interactions to further research and development in healthcare.

Risk and regulatory compliance

UK's Competition and Markets Authority makes it mandatory for UK banks to share data with other players via their Open Banking regulations

Source: The Open Data Institute, "The value of sharing data in supply chain optimization: A case study in the aerospace sector," March 2020. Microsoft, "Microsoft and the BMW Group launch the Open Manufacturing Platform", April 2019. Think Big, "Planning the location of a retail site with the data," February 2021. Health Data Hub, FAQ, Accessed June 2021. Wired, "What is Open Banking and PSD2? WIRED explains", April 2018.

Tracking advertising efficiency and targeting

This is the most common use case for several consumer-facing sectors, including consumer goods(48% of organizations are implementing it at full scale), retail (51%) and telecom (87%). Collaborating across sectors presents a unique opportunity to gain a clearer picture of customers' likes and dislikes and a day in the life of an organization's typical customer. Insights yield concrete benefits by making possible more precise changes in organizations' marketing decisions, achieving higher efficiency. At Starbucks, Megan Brown – director, knowledge management & data literacy – told us: *"Our core use cases for external data sharing revolve around three broad areas: analyzing customer purchase behavior; customer segmentation for enhanced personalization; and media mix modeling to better align our media spend with their returns. In addition, there are several uses for external data coming from our ecosystem partnerships that help us in niche areas of business."*

LUCA, a Telefonica Unit, helped Natura, the global personal care group determine ideal locations for a retail store presence with the help of millions of mobile data points to provide critical information such as customer demographics, number of passerbys across locations. These insights can also help retailers take into account days and hours of operation.[24]

Sharing data and insights into new research and development

The healthcare and manufacturing sectors depend on effective research and development (R&D). Our survey found that 41% of organizations across sectors are running pilots of this use case while 20% have moved beyond pilot to scale stem. In both these sectors, data ecosystems enable R&D to emerge as the most common use case.

Airbus's Skywise is a leading open-data platform for the aviation industry. Its ecosystem of partners includes over 9000 connected aircraft from over 100 airlines around the world and over 18,000 unique users. Using data and insights from the Skywise program, a US flagship carrier was able to save $13m in fuel costs per year. Another airline was able to reduce reliability reporting time from three weeks to one day.[25]

Europe Bioinformatics Institute(EBI) launched the COVID-19 data platform in collaboration with the European Commission, European Open Science Cloud, ELIXIR Europe, Universities, and healthcare centers enable rapid access to datasets and results related to the COVID-19 pandemic to accelerate research and support of diagnostics, therapeutics, and vaccines.[26]

Sharing real-time energy/water-supply and consumption data

The energy and utilities sector requires highly efficient forecasting methods to reduce energy waste and increase profits. With the help of emerging technologies and the ability to collect granular data, energy and utility firms can obtain a clearer picture of energy-consumption patterns and load variation. This can help utilities to plan their supply and reduce wastage. Our survey found that 73% of energy and utilities organizations are running pilots/proofs-of-concept (POC) of this use case, while 16% have a fully-scaled ecosystem initiative.

Bidgely, an India-based start-up focused on energy consumption, uses data collected from more than 50bn meter readings to form an accurate picture of consumption patterns, down to appliance level. This data is shared with utilities to derive more precise customer-usage insights and plan demand.

PassivSystems – a UK-based smart-energy platform company – has partnered with Sheffield University and not-for-profit Open Climate Fix to make more accurate predictions of solar-energy generation. This helps balance the demand and supply of electricity, reduce costs for the grid and the consumer, and better utilize renewable energy, including bringing environmental benefits.[27]

Sharing aggregate mobility data for services such as smart cities, parking, and mapping

As the most popular use case for the automotive sector, this also applies across sectors, including insurance and public-sector firms. We found that 81% of automotive organizations in our survey are running pilots/POCs of this use case while 7% have a fully-scaled ecosystem initiative.

Avis Budget Group – a leading global provider of mobility solutions, partnered with automotive data-services platform, Otonomo, to derive insights from its large, connected vehicle fleet, which includes different brands and models. The connected fleet generated an estimated 7bn road miles of data in 2020. Analysis of this data has the potential to generate insights into streamlining Avis's operations, reducing costs, and improving customer experience with solutions such as predictive maintenance. JD Power is one of the first customers of this ecosystem.[28] The partnership also has the ambition to serve several smart-city initiatives.

Sharing online purchase and transaction data

Although banking data usually comes with huge regulatory constraints, it can provide useful insights for business and retail customers alike. Sharing online-purchase and transaction data consequently emerges as the most popular use case for the banking sector with 45% of organizations implementing it at full scale and 42% running pilots.

Visa enables participating issuers, such as Bank of America Merrill Lynch and Citi, to provide their US commercial customers with a comprehensive view of all purchases made on Amazon Business using Visa cards. This process, which can be time-consuming to track and record for businesses, is made seamless through a collaboration between Visa, Amazon, and participating issuers.[29]

A central team led by senior leadership conducts a thorough due diligence on data before we enter into any new data partnership. The team has resources and well-defined processes to conduct research on the use, value, relevance, and quality of available data sources.

- Safiya Ansari,
Data analytics manager for engineering & program management at Faurecia

How are organizations preparing for the new European data strategy? In discussion with the **European Commission's Daniele Rizzi and Jiri Pilar**

The European strategy for data announced in February 2020 aims to make Europe a leader in a data-driven society. As part of this strategy, the EU will create a single market for data, making more data available for use across the economy and in society in general.[30] The strategy will also create "data spaces" to ease the flow and usage of data across sectors (see Figure 15). However, our survey shows that nearly half of organizations are yet to put in place the structures and measures that would ease its adoption. For example, 61% have no team or measures to monetize data assets and to incentivize customer data sharing, and nearly half (45%) have no internal processes to ensure data interoperability within and across sectors.

We spoke with **Daniele Rizzi, principal administrator and policy officer at the European Commission, and Jiri Pilar, Legal and Policy Officer at the European Commission,** to understand the strategy and regulation in more detail and the impact on organizations and individuals.

Q. What is the key objective of the European data strategy 2020?

A. The data strategy, in conjunction with the digital strategy and AI strategy for Europe, aims to enable reuse of personal and organizational data in Europe for data-powered innovation. It intends to facilitate trustworthy data sharing between individuals and private- and public-sector organizations, in line with earlier regulations such as GDPR, and to improve Europe's competitiveness in the data economy.
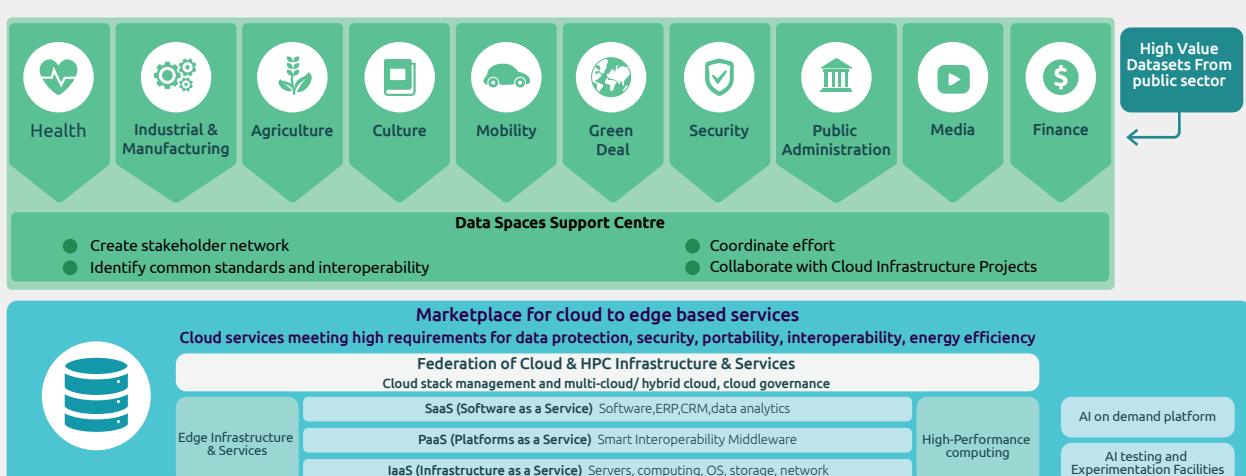
Q. While regulation is often seen as restricting innovation, how does this strategy aim to foster innovation?

A. The strategy aims to spur innovation by stimulating competition and balancing market power between various players, such as data generators, aggregators, processors, and so on. It would do so by setting clear and fair rules for access and reuse of data and pooling European data in key sectors, with common and interoperable data spaces. With Covid-19, we have all seen how more and better-quality data leads to better results for the economy and society in general.

Q. How will the strategy impact the private sector with the formation of new data spaces?
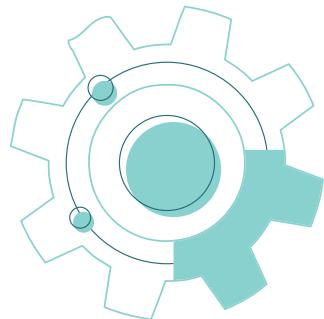
A. We believe that the private sector will benefit greatly from the clarity of rules, interoperability, and standards around data sharing within and across sectors. And we hope they will see more incentives to share data more openly for private and public use. Additionally, it will facilitate cross-border exchange of data and insights, where currently they only exist within certain countries or groups of companies.

Figure 15. Common European data spaces as outlined in the European data strategy

Source: CIGREF, French GAIA-X Hub: Les financements europeens & nationaux pour la donnee," March 2021.

YET, THE FULL POTENTIAL OF DATA ECOSYSTEMS REMAINS UNTAPPED

Many organizations engage in simple data sharing in ecosystems with low levels of collaboration
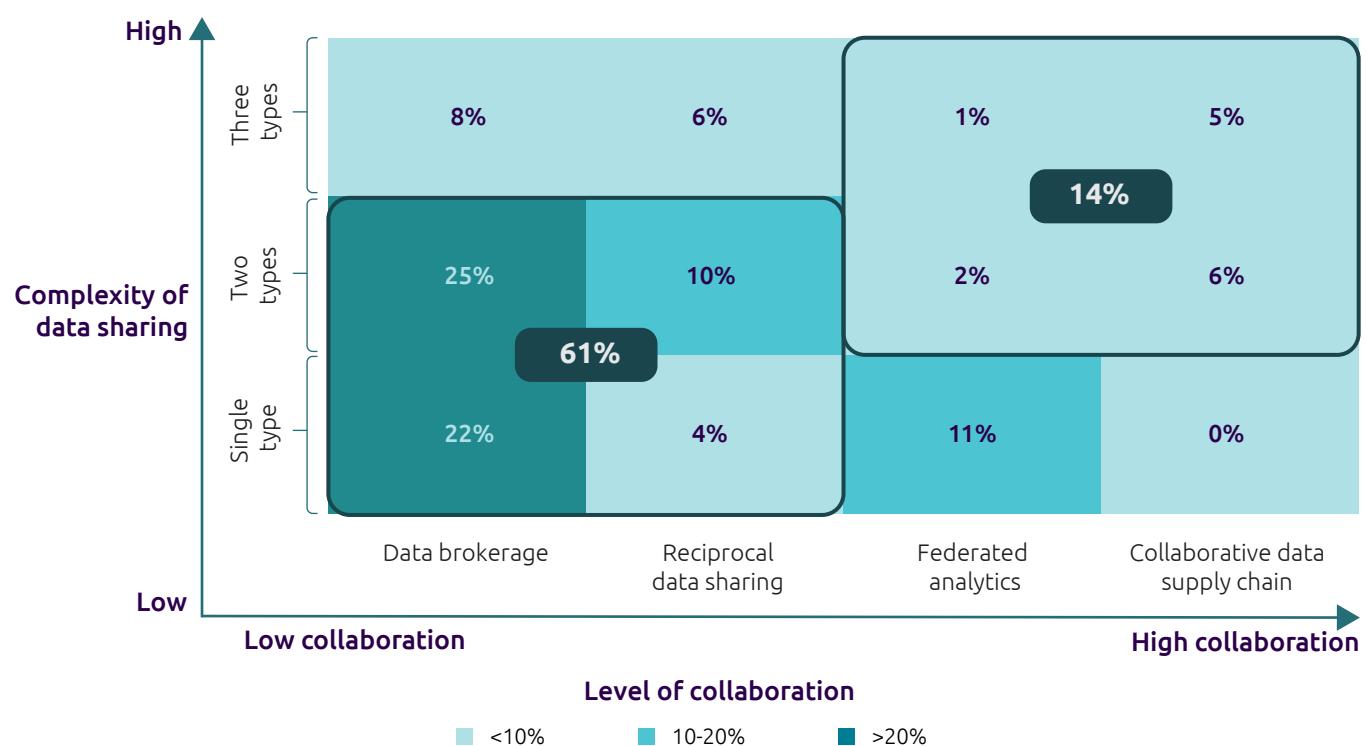
When we analyzed organizations' various modes for engaging in data ecosystems, we found that there is significant disparity in the adoption of different models. We analyzed the adoption of various models on two dimensions:

1. **Complexity types of data sharing:** We classified the complexity of data sharing in three categories, based

on whether they used one, two or all three of the types below:
– Direct data sharing (via file transfers, APIs, etc.)
– Aggregator-based data sharing (where a broker/ aggregator collects data from multiple data sources and shares it with a client organization)
– Data sharing via exchanges or marketplaces (such as AWS, Snowflake, or sectoral data exchanges, such as Verisk).

Figure 16. Organizations primarily engage in simpler data sharing and low collaboration models of data ecosystems

Percentages within boxes represent the share of organizations that adopt the given collaboration model and data sharing complexity

Source: Capgemini Data Ecosystem Survey; N=698 organizations, April-May 2021. Capgemini Research Institute analysis.

2. **Collaboration model:** We used the same classification of primary collaboration models in data ecosystems that we highlighted at the beginning of the report:

  – Data brokerage
  – Reciprocal data sharing
  – Federated analytics
  – Collaborative data supply chain.
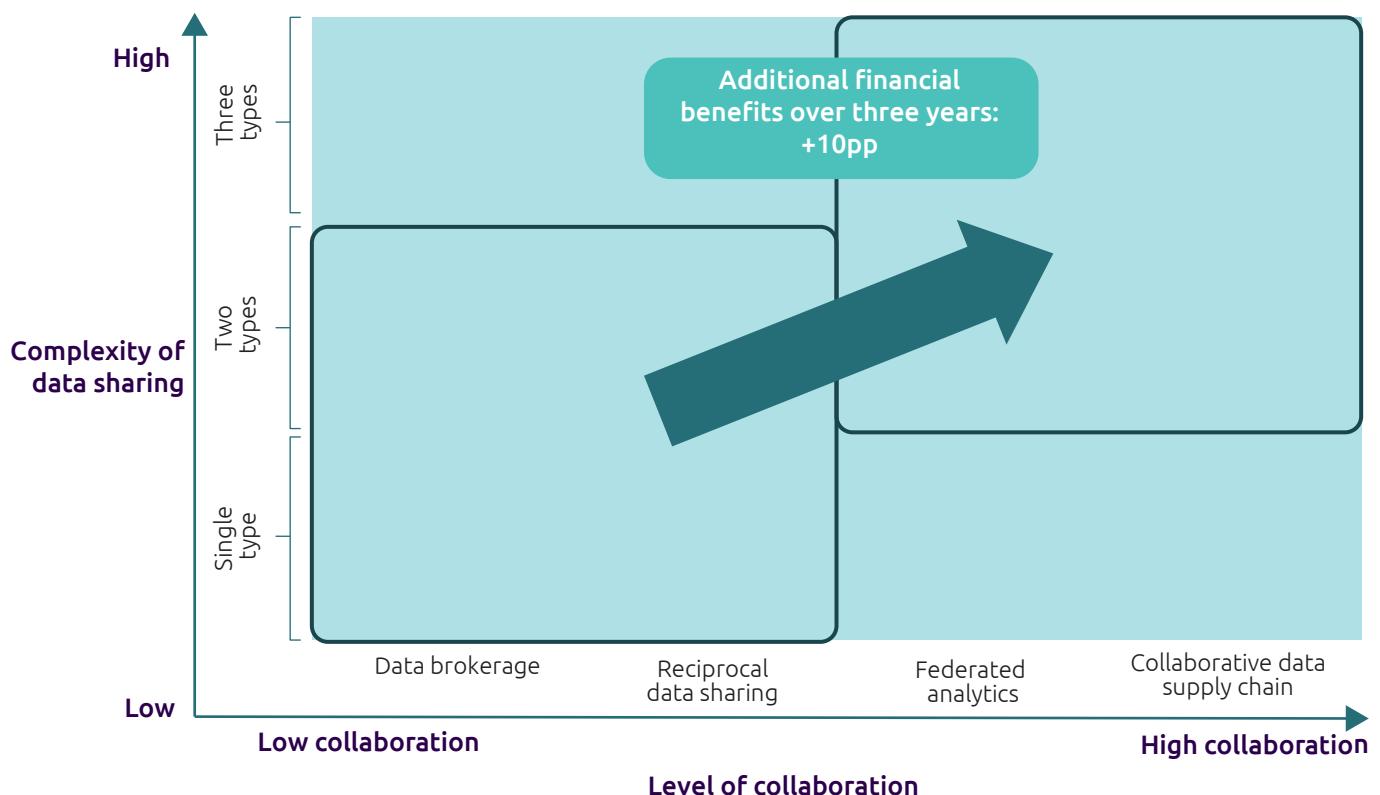
Our analysis of surveyed organizations on this framework revealed that a majority of organizations (61%) are low-complexity/low collaboration – i.e., using just one or two of the data sharing types and only focused on data brokerage and reciprocal models of collaboration (see Figure 16). These correspond to the four models in the bottom-right of the framework. Only 14% of organizations have adopted the most collaborative ecosystem models and complex types of data sharing.

Data ecosystem models with higher benefits remain largely untapped

To understand whether there are untapped benefits from data ecosystems, we analyzed the financial benefits associated with the various models using the same methodology as we did in chapter 1.

We found that organizations that are involved in more collaborative models and complex data sharing (14% of organizations) gain a financial advantage exceeding 10 percentage points over and above their peers who rely on simpler data sharing and low collaboration models (see Figure 17). This financial benefit includes total additional benefit considering the potential of generating new revenue, cost savings, and productivity improvement by data sharing in ecosystems. It implies that, over the next three years an organization in more collaborative and complex ecosystem model will generate 10 percentage points higher financial benefits (new revenue, reduced costs, and improved productivity) than that in a less complex and collaborative model, compared to today's levels, all else remaining the same. It highlights the competitive advantage inherent in more collaborative data ecosystems.

Figure 17. Data ecosystems with greater collaboration and complexity of data sharing offer significantly higher benefits over others

Source: Capgemini Data Ecosystem Survey; N=698 organizations, April-May 2021. Capgemini Research Institute analysis.

Various challenges hinder large scale adoption of data ecosystem initiatives

Organizations pointed to a range of hurdles that stand in the way of greater uptake of data ecosystem initiatives. As Figure 18 shows, these range from cybersecurity threats to unclear return on investment. Yet, as a majority of organizations face these top challenges, it is still not easy for a large share of organizations to embark on data sharing in ecosystems on a large scale.

Figure 18. Top challenges faced by a majority of organizations – by category

		Share of organizations citing it as a top challenge
Governance, risk, and compliance	Cybersecurity threats	62%
	Intellectual property and sensitivity of data	57%
Financial	Lack of investment in talent and skills	56%
	Unclear return on investments in a data ecosystem	29%
Trust	Consumer resistance towards sharing data	56%
	Lack of trust in external data	46%
Technological	Lack of data sharing platforms to control access rights	56%
	Poor data quality (including lack of data availability)	53%

Source: Capgemini Data Ecosystem Survey; N=750 organizations, April-May 2021

+10pp

The competitive financial advantage to organizations engaging in more collaborative and complex data sharing, over and above their peers who rely on simpler data sharing and low collaboration models.

Preserving privacy in data ecosystems

When it comes to data sharing, a significant area of concern is safeguarding the privacy of end-consumers or citizens:

- A 2020 study found that 74% of Americans are concerned about personal data privacy and 96% want companies to do more to protect consumer data privacy.[31]

- Multiple regulatory frameworks have come into effect around the world in recent years – such as the EU's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) – which aim to give people a level of control over their data to ensure trustworthy data sharing and usage.

- Despite these efforts, breaches and fines are increasing. According a recent report, GDPR data breach notifications were up 19% in 2020 over 2019.[32] And the EU regulators imposed €158.5m of fines in 2020, up 39% over the previous year-and-a-half since the law went into effect.

Here, we outline some principles, practices, and techniques that organizations can use to balance data sharing in ecosystems with preservation of privacy:

GDPR's data protection principles. Article 5 of the GDPR laid down seven data protection and accountability principles that can go a long way in ensuring a high level of data privacy[33] : Lawfulness, fairness, and transparency; purpose limitation; data minimization; accuracy; storage limitation; integrity and confidentiality; accountability.
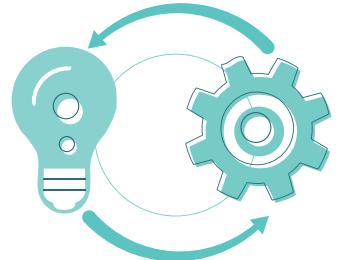
Anonymizing/pseudonymizing data. This is a process of masking or removing personal information from data records in a way that no record can be traced back to its individual data subject. Anonymized data can be safely shared with external and internal audiences and can be useful for a range of purposes, from trend analysis to customer segmentation for marketing and sales. However, anonymizing techniques need to be fairly robust – even when identifiers are removed from data, it may be possible for attackers to de-anonymize the data by cross-referencing with other data sources, thus revealing sensitive information.[34]

Differential privacy. This is a property of datasets and algorithms that allows you to publicly share information about the patterns displayed by groups while keeping information about individuals in the data set private. Apple uses a technique based on this concept called "local differential privacy" to glean insight into groups of users without learning about specific individuals.[35] By ensuring privacy during data collection itself, data is not even revealed to an organization's internal users, which creates robust levels of privacy.

Federated learning. A technique of machine learning, it allows algorithms to run analysis on data located at decentralized servers without exchanging or moving data. Data may contain sensitive information which can be used to identify the user. By avoiding the need to send this data to the cloud, this technique preserves user privacy while also providing a personalized experience. Federated learning finds application in use cases where data sensitivity is paramount. NVIDIA's Clara is a federated learning framework that allows different organizations and hospitals to securely collaborate, train, and contribute to a global model.[36]
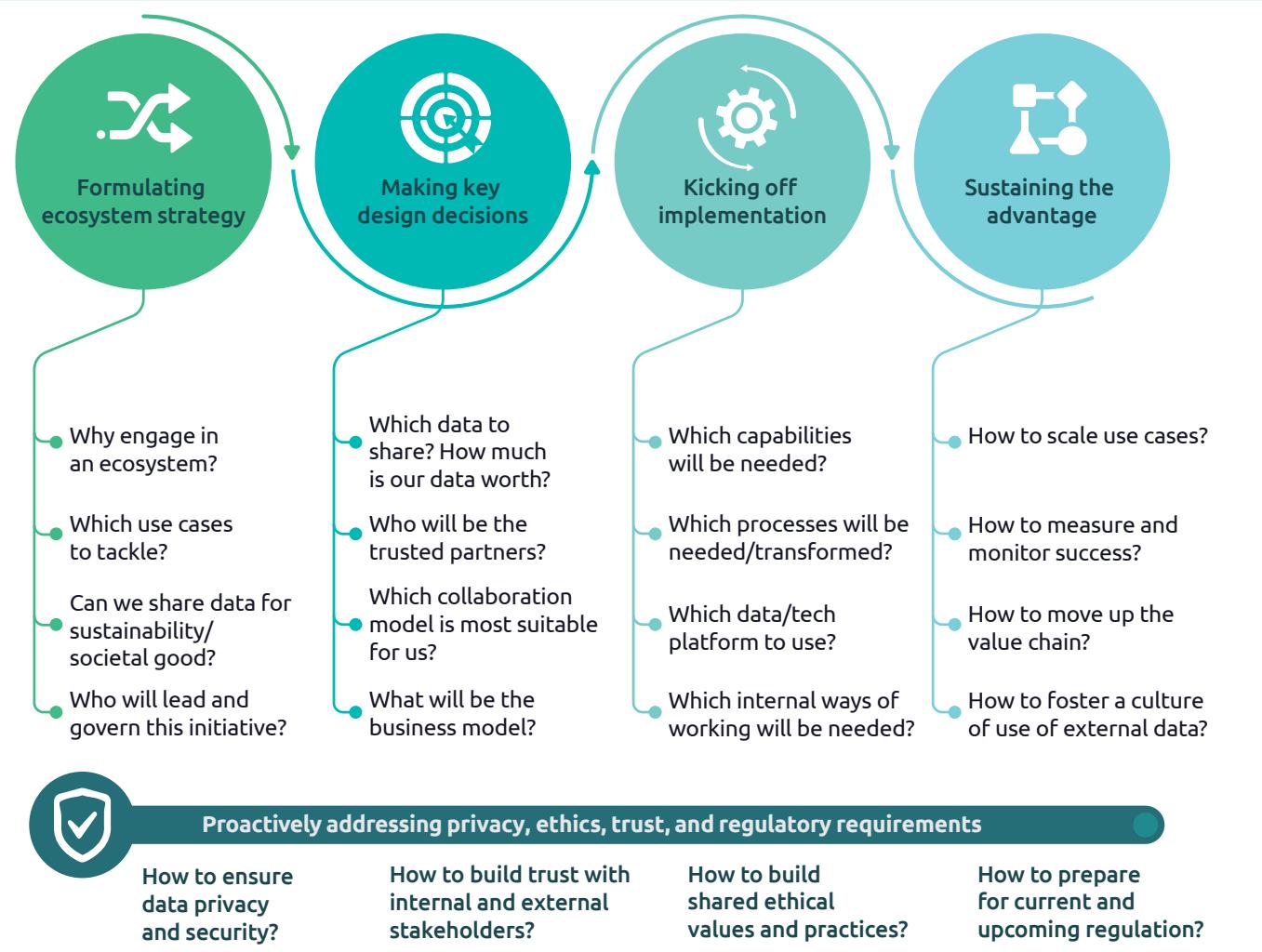
Homomorphic encryption. This is a special type of encryption technique that allows data processing without first decrypting the data. The results of processing are encrypted and can only be revealed by the owner of a secret decryption key. Microsoft has developed Microsoft SEAL – a code library that makes homomorphic encryption available in an easy-to-use form.[37] IBM recently completed field trials of fully homomorphic encryption.[38]

HOW ORGANIZATIONS CAN DESIGN, IMPLEMENT AND SUSTAIN A WINNING DATA ECOSYSTEM STRATEGY

As noted in the previous chapter, data ecosystems offer a host of advantages. To help organizations navigate the route to building a sustained advantage from data ecosystems, we have developed a roadmap based on our experiences with clients, discussions with industry executives, and insights gleaned from this research (see Figure 19).

Figure 19. A roadmap to build a sustained advantage with data ecosystems and key questions to answer at each stage

Formulating ecosystem strategy
- Why engage in an ecosystem?
- Which use cases to tackle?
- Can we share data for sustainability/ societal good?
- Who will lead and govern this initiative?

Making key design decisions
- Which data to share? How much is our data worth?
- Who will be the trusted partners?
- Which collaboration model is most suitable for us?
- What will be the business model?

Kicking off implementation
- Which capabilities will be needed?
- Which processes will be needed/transformed?
- Which data/tech platform to use?
- Which internal ways of working will be needed?

Sustaining the advantage
- How to scale use cases?
- How to measure and monitor success?
- How to move up the value chain?
- How to foster a culture of use of external data?

Proactively addressing privacy, ethics, trust, and regulatory requirements

How to ensure data privacy and security?

How to build trust with internal and external stakeholders?

How to build shared ethical values and practices?

How to prepare for current and upcoming regulation?

Source: Capgemini Research Institute analysis

I. Formulating a data ecosystem strategy

a. Clearly articulating the purpose of engaging in a data ecosystem. At the outset, organizations must clarify the purpose of engaging in data ecosystems. At E.ON, the European electric utility company, Gavin Roberts – VP, head of customer insight, data analytics and CRM – told us: *"For us, it always comes down to two things: How do we improve customer satisfaction? And second, how do we increase operational effectiveness and acquisition or retention of customers?"* Sometimes, sharing data internally with other parts of the organization can itself serve the purpose. In this scenario, it would be prudent to partner internally first and look for external partners later, if the need arises. Due to organizational size and complexity, the data ecosystem strategy may need to be made at CxO or function/department levels. In these instances, it is important to ensure that the purpose and strategy for ecosystem partnerships are aligned with the overall purpose and strategy of the organization.

b. Identifying data-sharing use cases that create value – for your organizations as well as for ecosystem partners. Organizations must carefully select use cases to achieve the objectives of your engagement with data ecosystems. Given that your organization will likely be involved in multiple ecosystems, it is important to consider which ones are likely to lead to quick value creation in the short term. For instance, automotive OEMs can significantly improve customer satisfaction through better data collaboration with parts suppliers and need not always require access to third-party data.

The most common use cases that cut across sectors and value chains (presented in chapter 2, above) are good places to start. As their value has been proven over time, they are likely to be beneficial to new participants. The appendix has a full list of use cases across sectors that we analyzed in this research. Sometimes use cases linked to sustainability benefits (reduced CO_2 emissions etc.) - or improving supply chain and operational efficiencies - are good places to start ecosystem collaborations. Organizations can start experimenting with these use cases to fine tune their collaboration strategy and proceed to scale it to other use cases. If your organization hosts an ecosystem (or is central in organizing it), it is important that selected use cases create significant value not just for a few participants in the ecosystem, but for all participants.

The head of the 'Big Data & IoT' value proposition at the data unit of a large telco outlined how important it is to customize the value proposition of the use case depending on the specific role of the ecosystem partner, saying: *"Every partner has a different need in terms of insights they wish to derive. To cater for that, we customize our data-based products to share the insights most relevant to the business of the partner. Retailers, for instance, see most value in footfall data clubbed together with credit-card consumption data, while transport and energy clients of the same data product find insights from weather data to be most valuable."*

c. Exploring opportunities to contribute to societal good. Business drivers are the usual starting points, but organizations are increasingly sharing data for sustainability and social-responsibility reasons. Our survey found that, when it comes to creating societal and sustainability impact, a majority of organizations engage in data ecosystems for sustainable development from an environmental, social, and governance (ESG) perspective (73%) and enabling social inclusion for marginalized sections of society (65%). The global head of data engineering at a multinational optical retailer told us: *"Societal good is one of the major drivers for our external data sharing. For instance, we share data with academic institutions and universities for research on optical health. It helps us understand the optical health conditions of citizens and suggest measures to improve them."*

"Data is much more than an asset to us and we are organised to maximize its potential through data sharing. We are witnessing a strong regulatory push in the EU to establish smoother data sharing systems in the financial services sector. The robust and comprehensivedata governance structures we have put in place while protecting sensitive customer data is in line with this evolving regulation and contributes to smoother data exchange with our ecosystem partners but also unlock greater benefits for our clients."

- Christina Poirson,
Group Chief Data Officer at Société Générale

d. Assigning a business leader and a team responsible for external data sharing and its governance. In over half of organizations, organizations delegate responsibility for leading external data-sharing initiatives to CIO or CDO:

- CIO: 29%

- Chief data/analytics officer: 24%.

Having a single leader entrusted with the overall responsibility of leading organization's data ecosystem initiatives ensures that there is clear ownership, oversight, and support to organization-wide external data sharing initiatives. The leader must appoint a team to work very closely with business teams involved in individual ecosystem initiatives to reduce overlaps in external partnerships, reduce redundancies of spending and cross-pollination of ideas and best practices across individual initiatives. A majority of organizations (51%) currently do not have a team dedicated to their data ecosystem initiatives. Rebecca Gan, a Senior Manager in Data Science at Cox Automotive – a US-based provider of software and data-based services to automotive dealers, manufacturers, and shoppers – told us: *"We have a dedicated, centralized team that looks into external data sharing and reports to the CDO. Having this centralized team helps us harness synergies between various data-sharing agreements that exist at a group level, while maintaining the agility needed for quickly responding to market demands."*

II. Making key design decisions pertaining to the ecosystem

Organizations must carefully consider all key characteristics of various ecosystem models, weighing their pros and cons and suitability for the organization, before starting or scaling up engagements. Whether or not your organization is orchestrating or engaging in an ecosystem of partners, you have certain decisions to make related to a range of areas: for instance, the type of data shared, partners and their roles, and ecosystem model.

a. Building a strong foundation of data management to align with use of external data. As the use of external data adds a whole new dimension to data handling practices of an organization, internal data management must be aligned appropriately. This step must begin with a careful assessment of the current state of data at your organization. It involves:

Conducting data discovery and profiling which maps all data existing within the organization and its characteristics.

Data cataloguing and data quality checks to finally build a complete picture of the data "real-estate" of your organization.

Devise a clear data governance approach to address – ownership/stewardship of external data, its quality, security and privacy, and transparency. More on this in the next section on implementation.

b. Decisions on data and insights. Key design decisions related to data and insights include::

- Whether external data can prove valuable for your use case and how much will you be willing to pay for it.

- Which data types are feasible to be shared externally or sourced from external sources, given the consent of data subjects and the applicable regulation.

- Whether the data is available in an acceptable and sharable form, complete with its data model, definitions, and explanations necessary for understanding the data.

- If you're planning to monetize access to your organization's proprietary data or insights, how do you attach a value or price to the data assets? Our survey found that 16% of organizations are planning to monetize access to data and 32% are planning to monetize insights within the next two years. However, only 43% of organizations have the capabilities needed to assess the value of their data assets. If monetization is the aim:

  – Consider monetizing only anonymized and zero-party data – the data that customers have willingly shared with your organization

  – There has to be an optimal pricing model for it. Forty-seven percent of organizations selling access to data and insights rely on a fixed/license fee. Only about 15% of organizations have been using a subscription-based or usage-based pricing model, which can be more profitable in the longer run.

- How often is data products shared – one shot, monthly, daily, several times a day, micro-batch, or real-time.

Safiya Ansari, data analytics manager for engineering & program management at Faurecia, told us that a team taking complete ownership of data-related decisions helps them, saying: *"A central team led by senior leadership conducts a thorough due diligence on data before we enter into any new data partnership. The team has resources and well-defined processes to conduct research on the use, value, relevance, and quality of available data sources. They also consider the value this data partnership will bring and how it is likely to evolve over time. There are also mandatory checks around data security, privacy, and ethics-related concerns."*

c.Decisions on the ecosystem model, and the role of your organization and partners. As explained above, there are four broad collaboration models of engaging in data ecosystems. Figure 20 provides a decision tree to identify the most suitable model for your organization's needs. Another key decision in this context is who to partner with and what role will be played by partners. If access to external data is crucial for your use case, consider whether the use case serves customer use cases, innovation, or improves operational effectiveness. Further, carefully consider the role your organization will play in the ecosystem and that played by other partners. Is this an ecosystem that your organization can orchestrate? If not, can a third-party organization
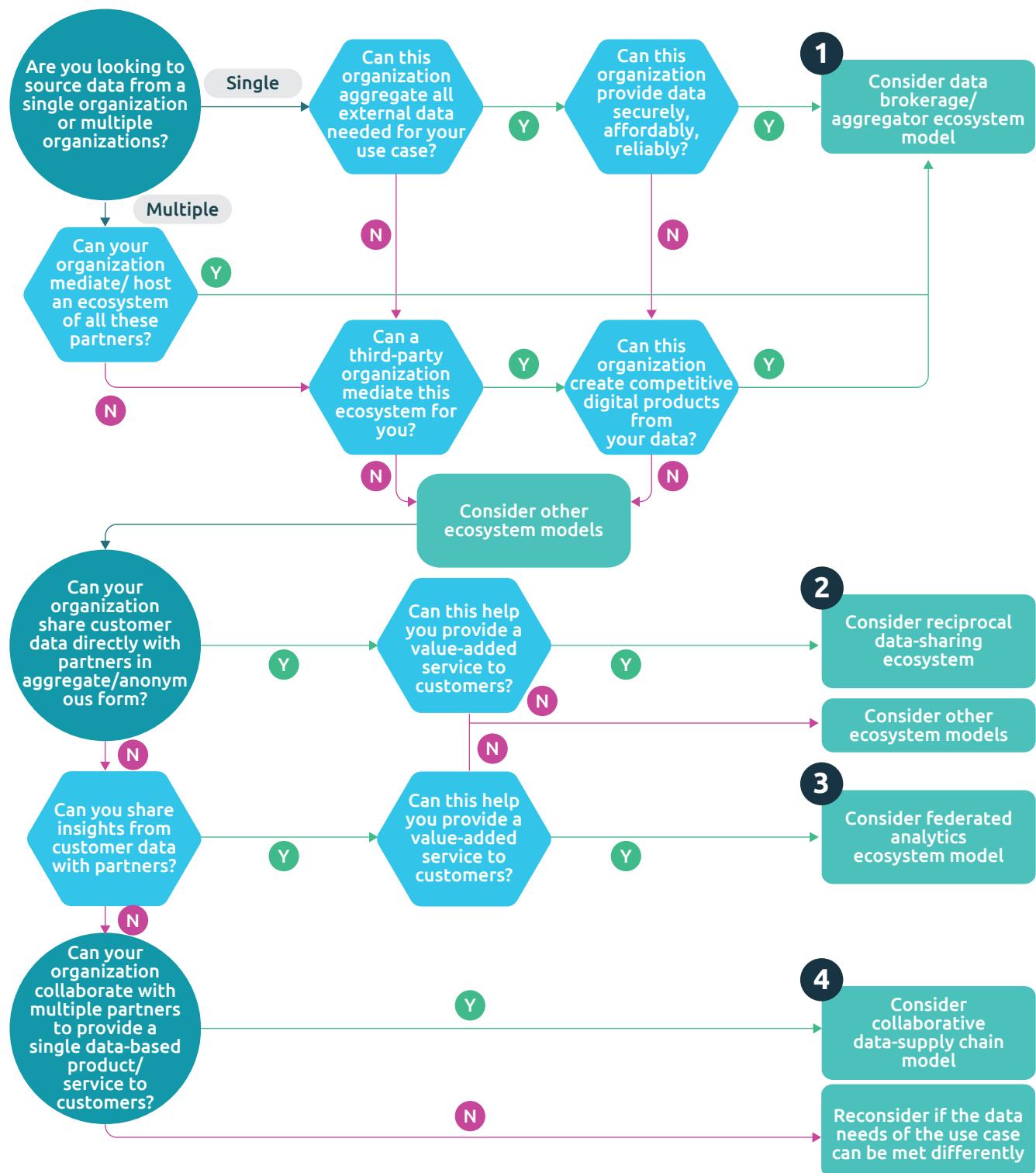
perform this role for you? Often, system integrators can act as a neutral third-party mediator in situations where it is not possible for organizations to share data directly owing to regulatory reasons or confidentiality constraints. In addition, they can provide consultative, technology-implementation skills, and managed services. Other key decisions to be made include:

- What are the service level agreements (SLAs)? What happens if SLAs are not met?

- What scope of use and related liability is borne by each party for use of these data products?

d. Decisions on the form of value exchange. To create a win-win partnership between all ecosystem partners, it is crucial to consider what the ecosystem partners will share in exchange for data and insights coming from either clients, end-consumers, citizens, or other partners. Often, this value can be monetary in nature if an affordable and acceptable price can be attached to data or insights. As seen previously, there are other forms of value exchange: for instance, improvement in quality, customer experience, and value-added services, among others.

43%
The share of organizations that have the capabilities for assessing the value of their data assets.

Figure 20. Decision tree to decide your organization's primary ecosystem collaboration model

Source: Capgemini Research Institute analysis.
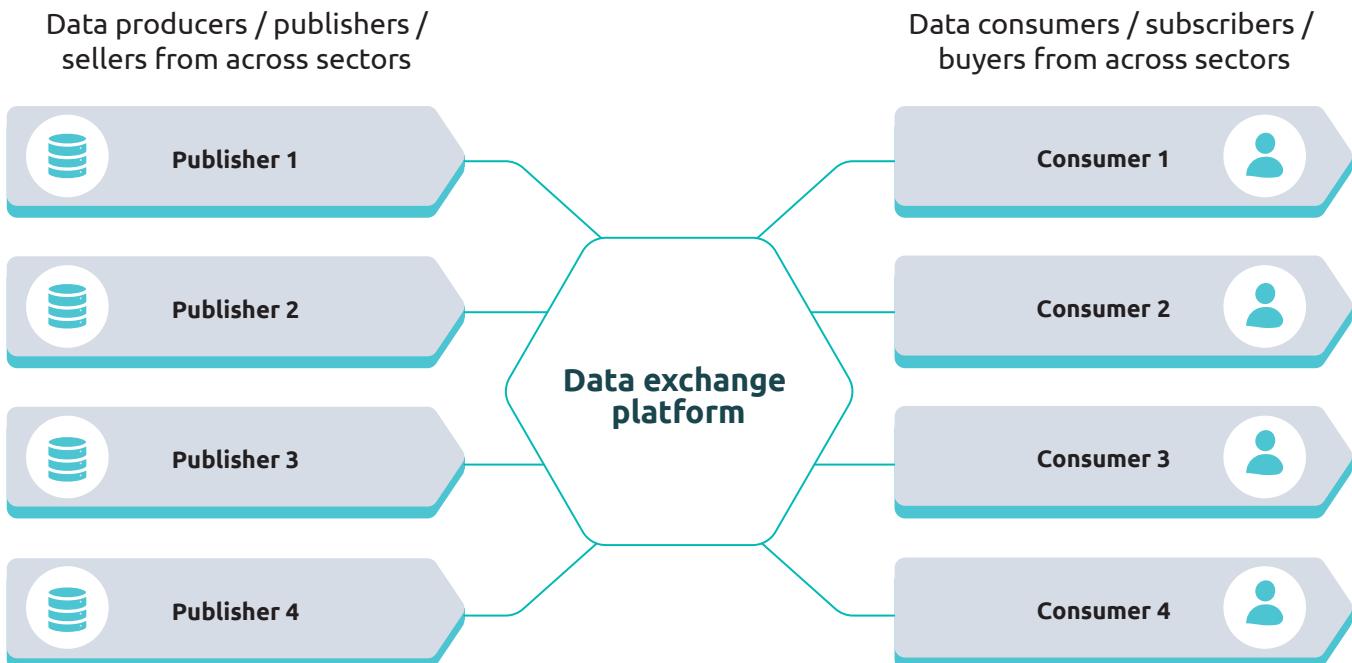
III. Devising a clear implementation plan and kicking it off

a. Considering suitable technology platforms.
A majority of organizations (56%) cite "lack of data-sharing platforms to control access rights" as the top technology challenge to embarking upon and scaling their data-sharing initiatives. It is therefore prudent to devise a tech-platform strategy before kicking off a data ecosystem. Broadly speaking, organizations have three data-sharing types and corresponding technology options to consider for their ecosystems (as explained in the first chapter):

- Direct data sharing (supplying and consuming) via file transfer, APIs, or cloud-based services, between two organizations

- Data mediator-driven sharing by a central, third-party organization allowing data collection from multiple sources and supplying to one or more clients/user organizations

- Data exchanges or marketplaces that allow multiple organizations to share data without the need for a central mediator (see Figure 21). These exchanges can be:

Figure 21. Data exchange or marketplace-driven sharing

Data producers / publishers / sellers from across sectors

Data consumers / subscribers / buyers from across sectors
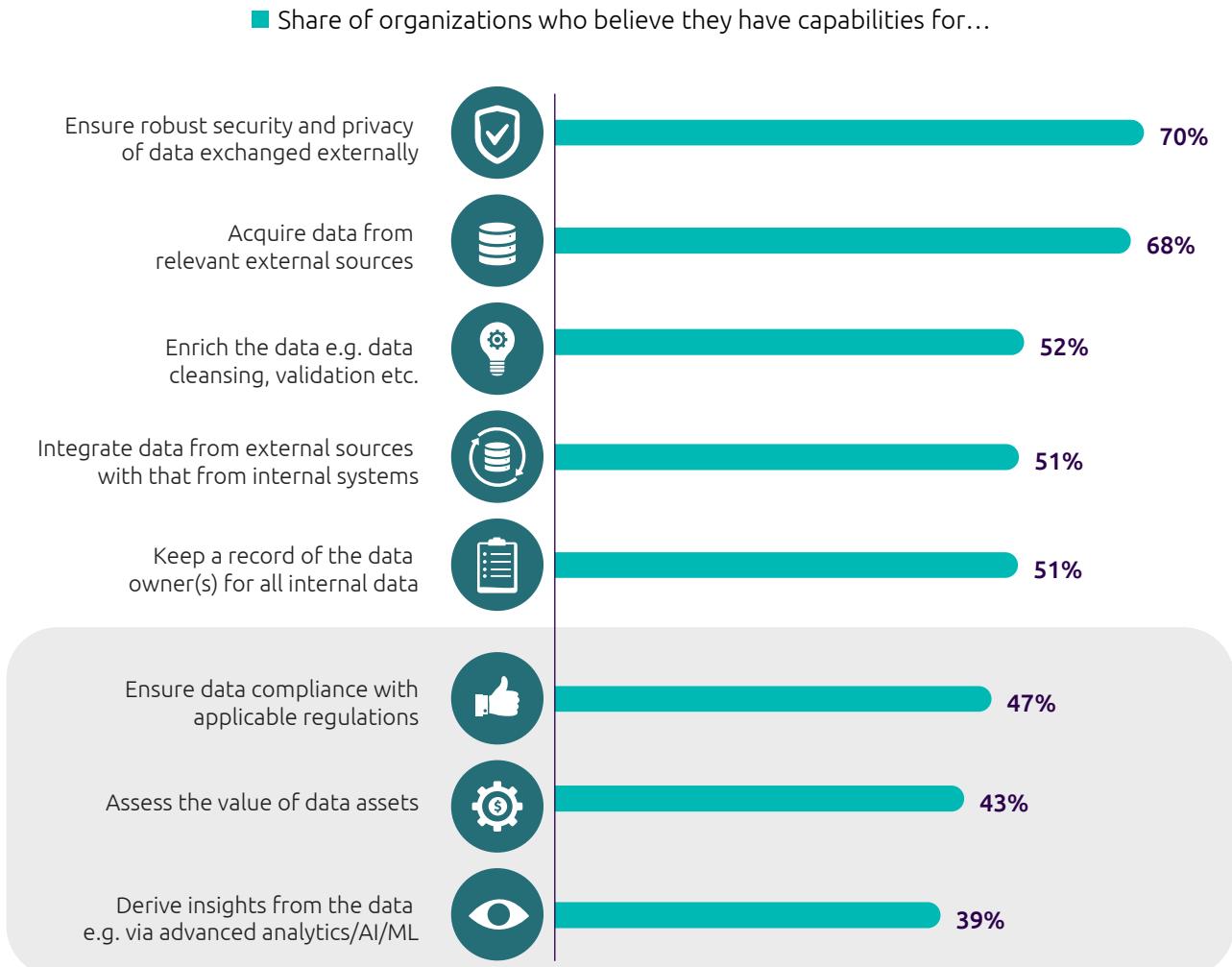
Source: Capgemini Research Institute analysis.

1. Public (i.e., open to all organizations), such as those owned by tech firms AWS Data Exchange, Azure Data Share, Snowflake, Dawex, and Narrative

2. Sectoral (i.e., relevant to a particular sector, such as Verisk for the insurance industry; Crux for financial services; and Terbine for machine-generated data)

3. Private (i.e., owned and orchestrated by an organization for its own set of partners, such as in the case of Airbus' Skywise program).

b. Addressing data capability gaps. Engaging with data ecosystems and making optimal use of the data requires capabilities that sometimes do not exist in organizations. To deal with this shortfall, organizations need either to build them in-house or source them from partners from within or outside the ecosystem. As Figure 22 shows, a majority of organizations lack the skills for deriving insights from data via advanced analytics/ML/AI (61%); assessing the value of data assets (57%); and ensuring compliance with regulations (53%).

Figure 22. Large capability gaps exist in deriving insights, assessing data value, and ensuring compliance

■ Share of organizations who believe they have capabilities for…

Ensure robust security and privacy of data exchanged externally — **70%**

Acquire data from relevant external sources — **68%**

Enrich the data e.g. data cleansing, validation etc. — **52%**

Integrate data from external sources with that from internal systems — **51%**

Keep a record of the data owner(s) for all internal data — **51%**

Ensure data compliance with applicable regulations — **47%**

Assess the value of data assets — **43%**

Derive insights from the data e.g. via advanced analytics/AI/ML — **39%**

Source: Capgemini Data Ecosystem Survey, N=750 organizations, April-May 2021.

c. Reassessing existing data processes. Availability of data from new external sources can fundamentally affect the internal processes that handle data flow within an organization:

- New processes and roles will be needed to collect, store, and manage external data coming from ecosystem partners. Interfaces with partners need to be defined with clear roles and responsibilities for supplying or sourcing data of a pre-agreed nature and quality. New processes may also be needed to verify and improve the quality of incoming data. Starbuck's Megan Brown told us: *"We have been working on a multi-year plan for how we quality-check data [for] data differences or data errors. We are also building more automated systems, such as auto-anomaly detection, which can flag issues with the quality of incoming data."*

- Some existing processes need to be transformed to better utilize external data from the ecosystem in conjunction with internal data. In addition, data protection needs to be strengthened, along with ensuring regulatory requirements; for instance, data deletion and portability as part of GDPR.

- Some processes may need to be made redundant; greater availability of external data would provide deeper levels of insight.

If data monetization is involved, sales teams will be needed to own the end-to-end sales process of data-based products and services. Finance/CFO teams will also need to be involved to ensure proper accounting of costs and revenues, and the required auditability.

IV. Sustaining the advantage.

Once the groundwork for implementation has been laid, the next step is to run a pilot of data-sharing use cases to prove the ROI from sourcing data from the ecosystem. As a senior data executive at a large US bank told us: *"We extensively test the data sourced from external partners. It is essential to determine the ROI we're making for investing in the acquisition of this data. This constant testing and measurement are essential to make sure use cases, such as personalization, are really enhancing the customer experience."*

a. The importance of data governance. In addition, organizations need to set up proper data governance, clearly defining the internal priority of various data ecosystem initiatives among other data-related initiatives. There is also a need to set up key roles, such as data owners, data stewards, and data-protection officers, to ensure responsibility and accountability of data at each step. Société Générale's Christina Poirson emphasizes the importance of the bank's governance structure, saying: *"Data is much more than an asset to us and we are organised to maximize its potential through data sharing. We are witnessing a strong regulatory push in the EU to establish smoother data sharing systems in the financial services sector. The robust and comprehensive data governance structures we have put in place while protecting sensitive customer data is in line with this evolving regulation and contributes to smoother data exchange with our ecosystem partners but also unlock greater benefits for our clients."*

"Over time, we have grown into cross-selling and upselling using insights derived from external data [...] That's a revenue stream that will pay dividends over five or maybe even ten years vis-à-vis just a one-time product sale."

- Gavin Roberts,
VP, head of customer insight, data analytics and CRM at E.ON

b. Moving up the data and insights value chain. As value creation and exchange are industrialized in the ecosystem, there should be an attempt to enhance this over time to keep the ecosystem relevant and sustain the advantage for every player in the system. E.ON's Gavin Roberts says: *"Over time, we have grown into cross-selling and upselling using insights derived from external data. For instance, while reselling a solar panel, we would try selling a maintenance contract or a power-supply contact. That's a revenue stream that will pay dividends over five or maybe even ten years vis-à-vis just a one-time product sale."*

V. Proactively addressing privacy, ethics, trust, and regulatory requirements.

Organizations are not well equipped to handle ethical concerns related to data sharing in data ecosystems. However, it is absolutely crucial to have clarity on such ethical requirements before embarking on new or existing data ecosystems. Our survey found that:

a. Nearly half of organizations (49%) have processes in place to define/secure ethical standards in use of data

b. Less than half (47%) monitor and guide all data partners working on data for potential ethical issues (e.g., having diverse data sets, unbiased data sets)

c. Less than half (47%) conduct frequent audits of data partners to ensure that the ethical guidelines are being followed.

Organizations would need to build these guidelines and processes where they are not in place, to create a foundation of trust and understanding between all ecosystem partners. As Megan Brown from Starbucks told us: *"We have conducted user experience research and found that distrust in data or insights stems from confusion and lack of transparency in definitions and interpretations of insights. Learning from this exercise helped us launch data literacy programs to acquaint business users with various ways data are analyzed and how they need to be interpreted. Alongside those initiatives, we have been building better and more reliable data products and we've seen the trust in data and insights has grown significantly from these efforts."*

Dawex's Didier Navez told us how trust needs to be built at two important levels: *"At first, you need a trusted environment where organizations can come together to collaborate with faith in the onboarding and vetting of participants and their respective roles, and second, data exchange has to happen in a trusted manner, in compliance with regulations under mutually agreed terms and conditions, including transparency in data sharing and pricing."*
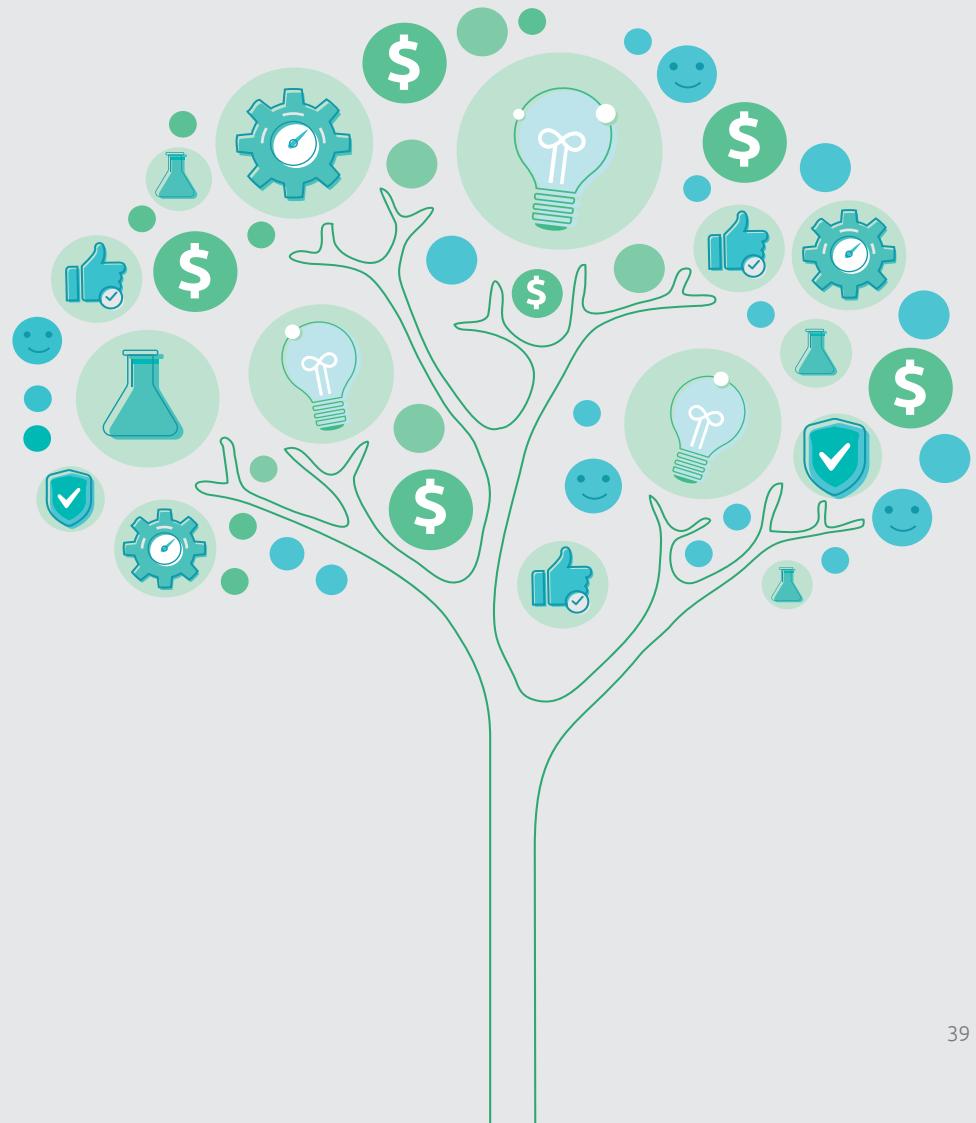
If one ecosystem partner's ethical guidelines and charter are well defined, it can be discussed and potentially adopted as a shared set of policies and acceptable set of practices for all partners, while engaging in the ecosystem. Having this shared set of principles will be a bedrock of ecosystem partnership and trust.

47%

Share of organizations that conduct frequent audits of data partners to ensure that the ethical guidelines are being followed.

CONCLUSION

Data ecosystems have delivered significant improvements in customer satisfaction (15%), productivity (14%), and cost reduction (11%), among others. Organizations significantly engaging with data ecosystems will drive a significant financial advantage – up to 9% of annual revenue in total (for an organization with annual revenue of $10 billion) compared to their less involved peers in the next five years. Front-runners will have a significant competitive advantage and late movers will face the risk of disruption and disintermediation from their better-equipped competitors. To join the winning pack of 'data sharing masters', organizations must urgently devise a strong data ecosystem strategy, and make key design decisions as early as possible. A small scale implementation with a view on data enablers and behaviors, followed by scaling to a higher level, while proactively addressing ethics, trust, and regulatory requirements will ensure lasting benefits.

APPENDIX

Research Methodology
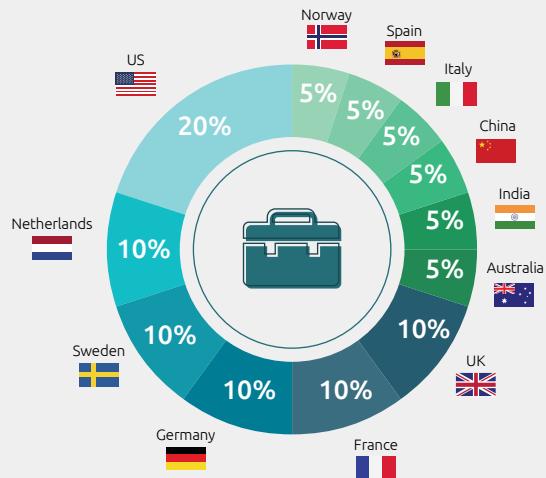
We surveyed 750 senior executives from organizations across the US, Europe, and APAC who are currently engaged in external data sharing.
In addition, we also conducted in-depth interviews with more than 30 industry executives and academicians.

Disclaimer: The views of the European Commission officers have been included as subject matter experts in their individual capacity and do not reflect the views of the European Commission nor an endorsement of Capgemini's work.
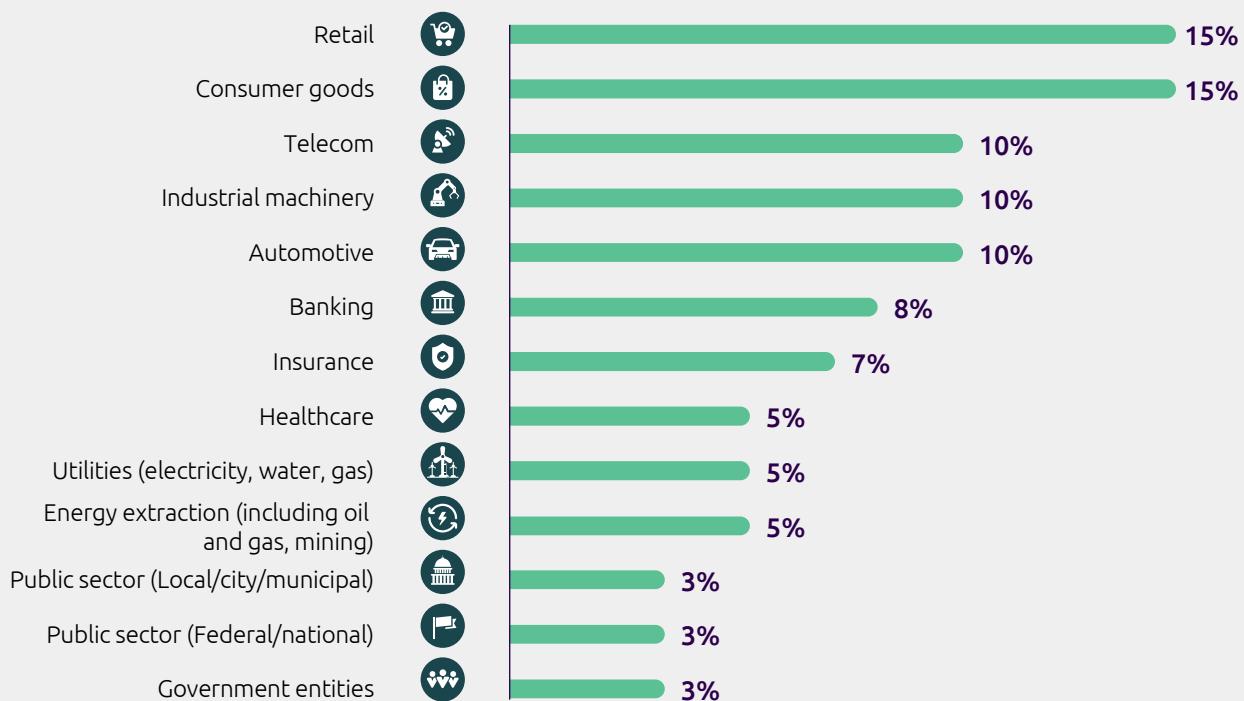
Executives by country of residence

In which country do you primarily work in?

US 20%
Netherlands 10%
Sweden 10%
Germany 10%
France 10%
UK 10%
Australia 5%
India 5%
China 5%
Italy 5%
Spain 5%
Norway 5%
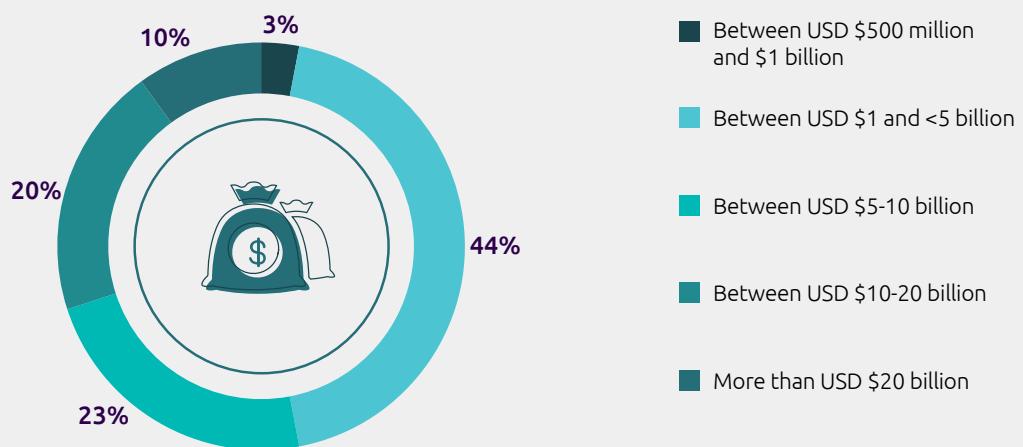
Organizations by sector of operation

In which of these sectors does your organization primarily operate?

Sector	
Retail	15%
Consumer goods	15%
Telecom	10%
Industrial machinery	10%
Automotive	10%
Banking	8%
Insurance	7%
Healthcare	5%
Utilities (electricity, water, gas)	5%
Energy extraction (including oil and gas, mining)	5%
Public sector (Local/city/municipal)	3%
Public sector (Federal/national)	3%
Government entities	3%
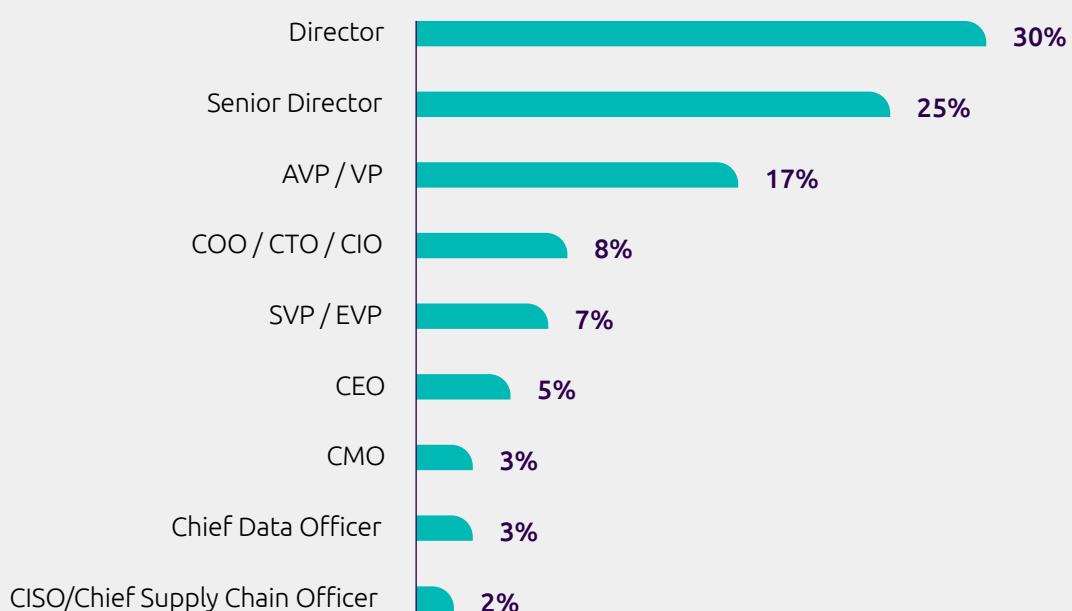
Organizations by annual revenue

What was your group's global revenue/budgets (for Public Sector organizations) in the last financial year?

- ■ Between USD $500 million and $1 billion
- ■ Between USD $1 and <5 billion
- ■ Between USD $5-10 billion
- ■ Between USD $10-20 billion
- ■ More than USD $20 billion

3%
10%
20%
44%
23%

Which of the following best describes your current job title?

Job Title	Percentage
Director	30%
Senior Director	25%
AVP / VP	17%
COO / CTO / CIO	8%
SVP / EVP	7%
CEO	5%
CMO	3%
Chief Data Officer	3%
CISO/Chief Supply Chain Officer	2%

ECONOMIC CONTRIBUTION OF DATA ECOSYSTEMS ON ORGANIZATION

In order to analyze the overall impact of data ecosystems on the financials of an organization, we estimated the potential benefits on three main financial areas over 5 years: new revenue opportunities from new business, productivity improvements and cost savings enabled by data sharing in data ecosystems.
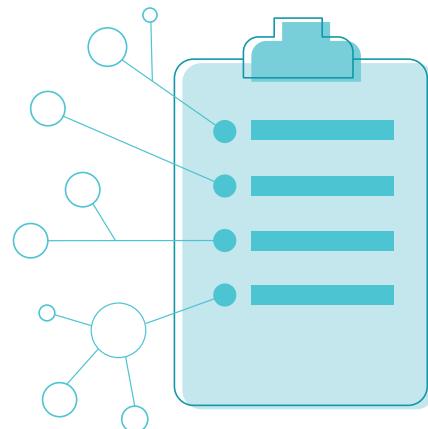
Our analysis is based on the characteristics of the organizations from our survey – namely, average revenue, average improvement in productivity gains/cost benefit as reported by organizations and percentage of organizations gaining these benefits. We then try and compute the impact of this on an average organization's revenue. We use the following assumptions based on current market trends:

1. The EU study on data market[39] which estimates that the data economy will constitute 4% of the EU GDP by 2025 in a baseline scenario. From this, we estimate that the data sharing in ecosystems can potentially add 4% of additional revenue to an average organization in an optimistic scenario. We estimate this to result equally – 2% each – from new revenue potential and productivity improvement potential, and potential cost savings to impact 4% of the overall cost of operation.

2. From our discussions with industry experts and clients, we estimate that in a conservative scenario, the top line impact will be limited to 1% of organizational revenue. We estimate this 1% to equally come from new revenue opportunities (0.5%) and productivity improvements (0.5%), and potential cost savings to impact 1% of the overall cost of operation.

3. For our cost and productivity analysis, we estimate the impact on operating expenses. Historic 20-year (2001-2020) data from S&P 500 shows an average operating margin (EBIT) of 14.3%. We utilize this along with the conservative and optimistic estimates above to derive the potential impact on the overall cost of operation of the organization.

LIST OF EXTERNAL DATA SOURCES

Following is the list of external data sources that we had analyzed in our
data-powered enterprise study published earlier:

1. Supplier data

2. Data from distributors/ retailers

3. Data from platform providers

4. Anonymous consumer data (example cookies etc.)

5. Personal Identifiable Information (PII) of consumers

6. Consumer usage data

7. Social media listening data

8. Data from blogs/product reviews etc.

9. Publicly available competitor data

10. Analyst/Industry reports/ reports from agencies such as D&B etc.

11. Data from hyperscalers like Google, Amazon, Facebook etc.

12. Proprietary datasets from data aggregators like Nielsen, Experian etc.

13. Open data (Govt and other public sector data)

14. Other external data (such as weather, traffic etc.)

EMPOWER YOUR BUSINESS LEVERAGING DATA & AI WITH CAPGEMINI

Perform AI

Activate data. Augment intelligence. Amplify outcomes.

Data promises infinite possibilities. Now, more than ever, those tasked with driving their industry forward, need to activate its full potential. To become a Data Master, you must be at the helm, ready to harness data in critical dimensions, to outperform cohorts on financial performance and make crucial decisions from a qualitative base. A data master is a high-performer that has harnessed the full suite of data capabilities in their organization. They have fine-tuned their processes that turn insights into action, and beyond.
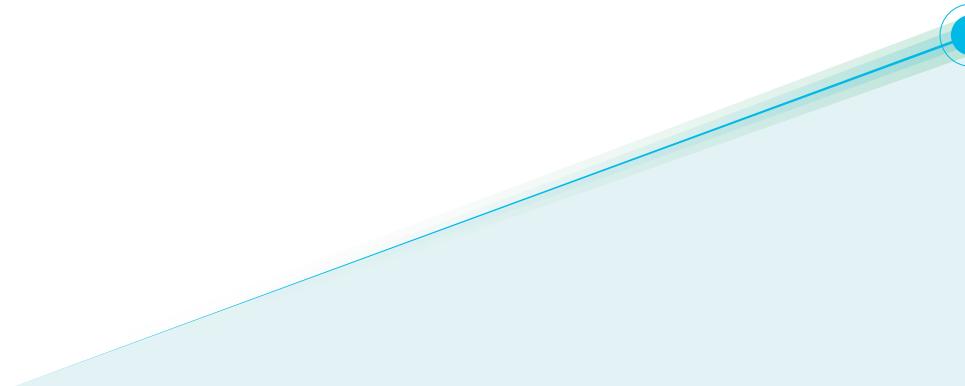
Data masters significantly outperform their peers, generating +70% higher revenue per employee and up to 22% higher profitability generate and lead in top-line benefits increasing sales of traditional or new products and services1 .

+70% higher revenue per employee

up to 22% higher profitability

We help our clients become Data Masters by managing data as a strategic asset

Perform AI is a **portfolio of Data and Artificial Intelligence services** to help organizations become data masters, take business control of their data as a strategic asset and transform their business to be a leader in the data economy.

Our seamless approach, that connects business strategy and data operations, enable growth, performance, and innovation with trusted outcomes. We empower some of the largest data & AI transformations at scale in the market – bringing together Data & AI strategy for CxOs. Together, we can improve your organization's business and cultural change, data engineering and data science skills, data & AI engineering industrialization and intelligent process automation, and execute at scale.

Speed up your business outcomes with trusted Data & AI at scale

Our teams deliver high-performing, tangible business outcomes, across industries and sectors:

Increased Sales 25%

Organizations successfully scaling AI have seen more than a 25% increase in sales of their products and services.

Cost Savings €100m

A major international retailer used our powered solution to optimize sales forecasting which directly drove €100m of inventory costs savings.

Fraud detection10X ROI

€100m ROIA European government agency has delivered a 10x return on investment thanks to our AI-powered fraud detection systems

Welcome to the era of Data Commerce

Unleash the potential of your data by connecting it to the right ecosystem. Our recent report shows, that financial benefits of data sharing in ecosystems represent a 9% increase in revenue for a typical organization.

There's an ocean of business opportunities to create new intelligent experiences, products and services leveraging data from your ecosystem. To do this, organizations must start with laying the foundation of trust, ethics, and regulatory compliance, formulating a robust ecosystem strategy, before kicking off the implementation at a small scale and then scaling it up with a clear roadmap.

Activate the full potential of data ecosystems to drive net new value for your business.

We help organizations navigate the route to building a sustained competitive advantage from data ecosystems in the cloud. Perform AI value proposition is Business Focused & Industry Specific and structured around two main pillars. with data ecosystems

Our approach

We build industry specific point of views to increase business collaboration

Business Model & Monetization
We support you in setting the strategy for data and insights sharing. We design pragmatic and progressive roadmap through all design phases: new business models, industry benchmarks and potential market and partners selection, new services and data products embedding user and partners experience. In addition, to fit your market, we design the industry specific strategy of data sharing with a consistent roadmap.

Governance & Operating Models
In order to support the governance of the partners and secure the trust between stakeholders, we set up operating models and design rules for efficient collaboration. To secure the privacy of data and ensure its quality, we set up foundations of the data governance.

And deploy at scale leveraging our partners network and technologies enablers

Security & Privacy
Our technologies enablers allow seamless and secured collaboration through: Homomorphic Encryption and Secure computing, ensuring collaborative working without data access - Differential Privacy enabling access to PII and protected data without risking value leakage - Data Centric Security enabling security models to follow the data and be controlled even when shared.

Data Sharing Platforms
We provide the environment to accelerate in federated analytics and foster new types of collaboration and business models, with data sharing platforms: Data Marketplaces to publish and contractually manage data sharing - Data Mesh architectures and Federated learning capabilities that allow access to federated data and provisioning into purpose driven collaborative environment.
Ultimately, to speed up scaling, we leverage hyperscalers foundations as key enabler of data execution and collaboration

Get the right team to scale Data & AI
With 25,000 Data & AI at scale practitioners, working for more than 800 clients worldwide, supported by nine Centers of Excellence in AI in all regions, Capgemini's capabilities are unmatched.

We fulfill your ambition, from the right strategy with industrialization, to operationalization and scale in mind. We ensure you take the right approach to finding actionable, trusted insights for your teams, as well as including them in the transformation journey. And we're capable to build the right solutions for your business and implement them with the right continuous deployment and operations to maximize their positive impact.

"We help you shift gears, scale trusted Data & AI solutions for the best business outcomes, and Get The Future You Want."

- Anne-Laure Thieullent
*Vice President, Artificial Intelligence &
Analytics / Perform AI Group Offer Leader*

Capgemini Perform AI

How do we look holistically at Data & AI transformation

CP & Retail
Augmented operations and immersive customer engagement

Public Sector
Infusing AI to better serve citizens and governments

Manuf, Auto & Life Sciences
Augmented operations for operational efficiency & risk reduction

Telco Media Tech
Deep Customer Engagement and Operational Excellence

Energy Utilities & Chemicals
Sustainability powered by Data & AI and Operational Excellence

Financial Services & Insurance
Right Product, Right Price

Data Activation

| Customer First | Intelligent Industry | Enterprise Management |

Data Advantage

Data Partnerships & Data disruptors

Data Enablers

Strategy

AI Activate

Transformation Target and foundations for execution of AI @Scale

AI Reimagine

Disrupt, Invent and Implement "the next" business, "AI First"

Build, Deploy, Manage, Operate

AI, Analytics Data Science & 890

Scalable, Trusted AI and Augmented BI & Data Viz solutions

Intelligent Process Automation

Data & AI powered Frictionless Enterprise

AI & Data Engineering

Hybrid Cloud Technology platforms for Trusted Data & AI @Scale transformation

---

Delimited Impact

Human In Control

Data Trust

Continual Validation

USE CASES ANALYZED

We included 37 use cases in our research. These are:

Sector/Functions	Use case
Retail	Monetizing transaction and sales data created in-house
	Sharing consumer transaction and loyalty data
	Share consumer behavior data with a range of partners across sectors
	Tracking advertising efficiency and targeting
Consumer products	Sharing sales data by brands
	Sharing social media and third party data with brands
	Aggregating and sharing farm produce data with other growers and government
	Tracking advertising efficiency and targeting
Financial services	Monetizing transaction and sales data created in-house
	Sharing online purchases and transaction data
	Partnerships with peers for seamless flow of customer data between organizations
	Sharing data for compliance/regulatory purposes e.g. credit risk data
Automotive	Sharing operational data of vehicles such as maintenance log, fuel efficiency on an aggregate level
	Sharing anonymized and aggregated telematic data of vehicles with city governments, service providers, etc
	Sharing vehicle usage data with insurance and finance organizations
	Sharing sales data by brands and types
	"Industrial cloud" that combines data from all plants and makes it available across the supply chain partners and integration partners
	Sharing aggregate mobility data for services such as smart cities, parking, mapping etc.
Energy and utilities	Sharing data with peers to define best practices to reduce risks at sites
	Sharing real-time energy/water supply and consumption data to improve demand forecasts
	Sharing aggregate and anonymized customer data
	Sharing data between plant and facility managers, system integrators, and tech providers for more efficient operations
Healthcare	Sharing data/access to data of patients for quicker diagnosis
	Sharing data/access to data to mutualize health and patient history
	Sharing data/access to data to enable research and development
Telecom	Monetizing transaction and sales data created in-house
	Sharing call revenue with CP brands for targeted promotion
	Sharing aggregate traffic data with local directory services providers and merchants
	Sharing consumer transaction and loyalty data
	Share consumer behavior data with a range of partners across sectors
	Tracking advertising efficiency and targeting
Manufacturing & supply chain (consumer goods, automotive, industrial machinery, healthcare)	End-to-end monitoring of goods e.g. monitoring goods' location, ETA, ambient conditions for sensitive components
	Sharing operational data – asset performance, breakdown information, maintenance and efficiency among others to predict occurrences of failure in advance and take speedy corrective actions
	Inbound logistics management e.g. cargo data, scheduling of supplies, delivery times, etc. for improved forecasting and production planning
	Production/warehousing facilities management by sharing data on utilization of facilities and individual machinery
Innovation (all sectors)	Sharing data and insights on new research and development of new products
Sustainable operations (all sectors)	Sharing data on sustainable operations – best practices, freshwater consumption, leakage, spillovers etc. with suppliers/distributors for an overall sustainable supply chain

REFERENCES

1. Capgemini Research Institute, "The data-powered enterprise: Why organizations must strengthen their data mastery," November 2020.
2. Data.europa.eu website; The European University Institute, "European data portal," accessed April 2021.
3. Mayo Clinic, "Mayo Clinic launches its first Platform initiative," January 2020.
4. Financial Times, "Pharma groups combine to promote drug discovery with AI," June 2019.
5. Statista, "Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2024," accessed May 2021.
6. The European Data Market Monitoring Tool, "Final Study Report," June 2020.
7. Cisco Annual Internet Report, "By The Numbers Projecting the future of digital transformation (2018—2023)," Accessed July 2021.
8. GSMA, "IoT Revenue: State of the Market 2020," August 2020.
9. Company websites.
10. Definition by the Data Sovereignty Now community, https://datasovereigntynow.org/
11. European Commission, "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - A European strategy for data", February 2020.
12. European Commission, "Building a data economy," March 2021 update.
13. Company website, accessed May 2021
14. NVIDIA, "Federated Learning powered by NVIDIA Clara," December 2019.
15. OpenGroup, "What is the Open Footprint Forum?", accessed July 2021.
16. The Open Data Institute, "The value of sharing data to drive open innovation: A case study in the manufacturing sector," March 2020.
17. The Open Data Institute, "The value of sharing data for benchmarking and insights: A case study in the shipping industry," March 2020.
18. European Commission, "The European data market study update, March 2021 update," accessed June 2021.
19. Openbanking.org.uk, "Three years since PSD2 marked the start of Open Banking, the UK has built a world-leading ecosystem," January 2021.
20. Harvard Business Review, "3 Lessons from Chinese Firms on Effective Digital Collaboration," August 2020.
21. Quartz, "Employee-sharing schemes are softening the blow for China's pandemic-hit businesses," April 2020.
22. Vodafone, "Correct use of telco data can help in this crisis," March 2020.
23. e-Estonia, "e-governance," Accessed July 2021.
24. Think Big, "Planning the location of a retail site with the data," February 2021.
25. The Open Data Institute, "The value of sharing data in supply chain optimization: A case study in the aerospace sector," March 2020.
26. EBI, "Open data sharing accelerates COVID-19 research," October 2020.
27. The Open Data Institute, "The value of sharing data to address sector challenges: A case study in the energy sector," March 2020.
28. Otonomo, "Avis Budget Group Partners with Otonomo to Unlock the Potential of Its Connected Car Data," July 2019.
29. Visa, "Visa Launches Enhanced Transaction Data for Amazon Business Customers in the U.S.," October 2017.
30. European Commission, "A European strategy for data, 9th March 2021 update," accessed May 2021.
31. Consumer Reports, "Consumer Reports Study Finds Marketplace Demand for Privacy and Security," October 2020.
32. CPO Magazine, "European Regulators Have Imposed £245.3 Million in GDPR Fines To Date; 39% More Issued in 2020," January 2021.
33. Gdpr-info.eu, "Art. 5 GDPR: Principles relating to processing of personal data," Accessed June 2021.
34. IEEE Explore, "Robust De-anonymization of Large Sparse Datasets," May 2008.
35. Apple, "Differential Privacy Overview," accessed June 2021.
36. NVIDIA, "Federated Learning powered by NVIDIA Clara," December 2019.
37. Microsoft Research, "Homomorphic Encryption," Established March 2016.
38. Ars Technica, "IBM completes successful field trials on Fully Homomorphic Encryption," July 2020.
39. European Commission, "The European data market study update, March 2021 update," accessed June 2021.

ABOUT THE AUTHORS

Zhiwei Jiang
CEO, Insights and Data Global Business Line
zhiwei.jiang@capgemini.com

Zhiwei is the CEO of I&D Global Business Line for the last two years. Before that, Zhiwei was the Head of Insights & Data for Financial Services Business Unit. Prior to Capgemini, he spent 20 years in banking. His passion has been in data privacy, security, sustainability, and now more into AI ethics and social impact.

Anne-Laure Thieullent
Managing Director, Artificial Intelligence & Analytics Group Offer Leader
annelaure.thieullent@capgemini.com

Anne-Laure leads Perform AI, Capgemini's Artificial intelligence & Analytics group offer. She works with clients to scale Artificial Intelligence, so it infuses everything they do, and they become AI-driven, data-centric and innovative. With 20 years of experience in big data, analytics and AI systems, from design to production roll-out, her passion is to bring companies what they need to transform themselves into intelligent enterprises. Anne-Laure is committed to guiding clients to increase activating their data, while cultivating the values of trust, privacy and fairness.

Steve Jones
EVP, Chief Data Architect
steve.g.jones@capgemini.com

Steve Jones is the Chief Data Architect for Capgemini's Group Portfolio, with over a decade in Big Data and Cloud he's been at the front of helping our clients adopt new technologies to drive new business value and become more data driven. Steve was one of the first people to propose using mobile applications and analytics to help track and manage individuals carbon footprint and has continued to be at the front of Capgemini's CSR and Diversity initiatives.

Valérie PERHIRIN
Managing Director, Insight driven enterprise Capgemini Invent
valerie.perhirin@capgemini.com

Valérie leads the data, analytics and artificial Intelligence consulting activities. She advisesCapgemini Invent clients to design their datastrategy and designs roadmap and execution plan. She helps our clients to drive their data powered transformation in order to reach operational excellence and leverage data and AI to invent new sources of revenue. She has over 19 years of experience in data centric transformation powered by data, analytics and AI.

Marie-Caroline Baerd
Executive Vice President, AI, Capgemini Invent
marie-caroline.baerd@capgemini.com

Marie-Caroline leads Capgemini Invent's Artificial Intelligence offer, developing thought leadership, new frameworks and solutions, promoting a "Trusted AI" approach. She has a track record of more than 20 years in business management consulting, and is applying her experience in helping organizations leverage the full potential of AI & Data to innovate, enhance their revenues and efficiency, supporting their transformation journey from awareness and AI / Data strategy to implementation at scale within the organization.

Padmashree Shagrithaya
Global Head – AI Analytics & Data Science, at Capgemini
padmashree.shagrithaya@capgemini.com

Padmashree leads a team of 2500 data scientists, AI architects, data storytellers, and business intelligence experts. She and her team are responsible for helping clients across industries unlock the value of data investments through AI & analytics. In a diversified career spanning over 27 years, Padmashree has crafted and led multiple large & complex transformation programs delivering strong business outcomes for many clients, leveraging data, technology, Machine Learning, and AI. She has deep expertise in structuring, incubating, and scaling Analytics & AI Hubs.

Gianfranco Cecconi
Director, International Data Policy and Data Sharing Acceleration Lead, Capgemini Invent
gianfranco.cecconi@capgemini.com

Gianfranco leads Capgemini Invent's work in international data policy in Europe. His team runs for the European Union its two main programmes for the promotion and support of open data and data sharing: Data.Europa. EU and the Support Centre for Data Sharing, respectively. More recently, leveraging the ongoing momentum to accelerate data sharing in the Union, his focus is on advising Capgemini's public and private sector clients worldwide to transform, to make the best of the opportunities arising from the new data models and ecosystems.

Léo Isaac-Dognin
Director, Citizen Services, Capgemini Invent
leo.isaac-dognin@capgemini.com

Léo Isaac-Dognin is director for data, cloud and cybersecurity within Capgemini Invent's global public sector unit. He leads Capgemini Invent's engagement with policymakers on national and European strategies for artificial intelligence, data, as well as trusted cloud and edge computing. Leo is also Adjunct Lecturer at Sciences Po Paris, and co-author of *Cyber, la guerre permanente* (2018).

Jerome Buvat
Global Head of Research and Head of Capgemini Research Institute
jerome.buvat@capgemini.com

Jerome is head of the Capgemini Research Institute. He works closely with industry leaders and academics to help organizations understand the nature and impact of digital disruption.

Amol Khadikar
Senior Manager, Capgemini Research Institute
amol.khadikar@capgemini.com

Amol is a senior manager at the Capgemini Research Institute. He leads research projects on key frontiers such as artificial intelligence, sustainability, and the future of work to help clients devise and implement data-driven strategies.

Yashwardhan Khemka
Manager, Capgemini Research Institute
yashwardhan.khemka@capgemini.com

Yash is a manager at the Capgemini Research Institute. He likes to follow disruption fueled by technology across sectors.

Shahul Nath
Senior Consultant, Capgemini Research Institute
shahul.nath@capgemini.com

Shahul is a senior consultant at the Capgemini Research Institute. He keenly follows disruptive technologies and its impact on industries and society.

About the Capgemini Research Institute

The Capgemini Research Institute is Capgemini's in-house think tank on all things digital. The Institute publishes research on the impact of digital technologies on large traditional businesses. The team draws on the worldwide network of Capgemini experts and works closely with academic and technology partners. The Institute has dedicated research centers in India, Singapore, the United Kingdom, and the United States. It was recently ranked number one in the world for the quality of its research by independent analysts.

Visit us at www.capgemini.com/research-institute/

For more information, please contact:

Anne-Laure Thieullent
Managing Director, Artificial Intelligence &
Analytics Group Offer Leader
annelaure.thieullent@capgemini.com

Valérie PERHIRIN
Managing Director, Insight driven enterprise
Capgemini Invent
valerie.perhirin@capgemini.com

Steve Jones
EVP, Chief Data Architect
steve.g.jones@capgemini.com

Global

Zhiwei Jiang
CEO, Insights and Data Global
Business Line
zhiwei.jiang@capgemini.com

Ron Tolido
CTIO, Insights & Data
ron.tolido@capgemini.com

Anne-Laure Thieullent
Vice President, Artificial Intelligence
and Analytics Group Offer Leader
annelaure.thieullent@capgemini.com

Steve Jones
EVP, Chief Data Architect
steve.g.jones@capgemini.com

Valérie PERHIRIN
Managing Director,
Insight driven enterprise
Capgemini Invent
valerie.perhirin@capgemini.com

APAC
Frank Windoloski
APAC Insights & Data Leader
frank.windoloski@capgemini.com

Dr. Jing Yuan Zhao
AI Leader
jing-yuan.zhao@capgemini.com

Europe
Robert Engels
AI Leader
robert.engels@capgemini.com

France
Marion Gardais
VP Presales & Solutioning
marion.gardais@capgemini.com

Germany
Fabian Schladitz
AI Centre of Excellence Leader
fabian.schladitz@capgemini.com

India
Padmashree Shagrithaya
AI Centre of Excellence Leader
padmashree.shagrithaya@capgemini.com

Italy
Massimo Ippoliti
Vice President
massimo.ippoliti@capgemini.com

North America
Jerry Kurtz
NA Insights & Data Leader
jerry.kurtz@capgemini.com

Ramana Bhandaru
FS Insights & Data Leader
ramana.bhandaru@capgemini.com

Dan Simion
Vice President, Insights and Data leader
dan.simion@capgemini.com
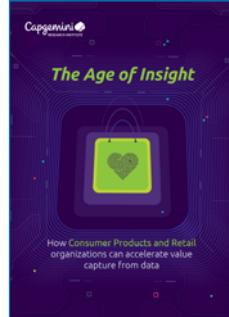
United Kingdom
Joanne Peplow
AI Centre of Excellence Leader
joanne.peplow@capgemini.com

Discover more about our recent research on digital transformation
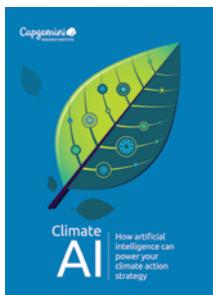
The data-powered enterprise: Why organizations must strengthen their data mastery

The Age of insight
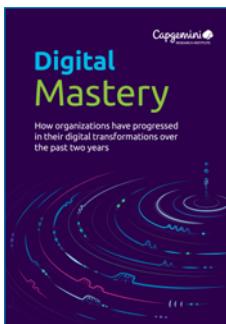

Accelerating the 5G industrial Revolution
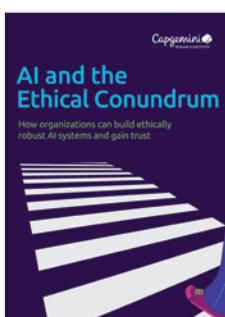
**Climate AI: How artificial intelligence can power your climate action strategy**
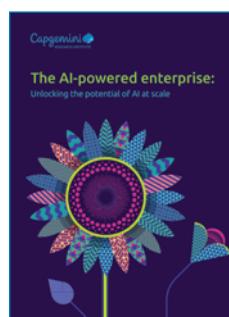
What's the big idea? Why most innovations fail to scale and what to do about it

Digital Mastery

AI and the Ethical Conundrum: How organizations can build ethically robust AI systems and gain trust

**The AI-powered enterprise: Unlocking the potential of AI at scale**
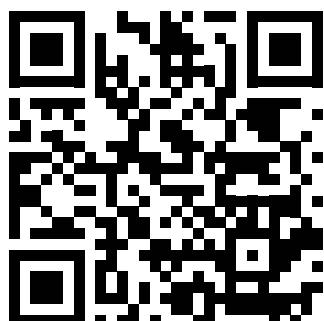
Accelerating the 5G industrial Revolution

Subscribe to the latest research from the Capgemini Research Institute

Receive copies of our reports by scanning the QR code or visiting

https://www.capgemini.com/Research-Institute/

Capgemini Research Institute   ⊗

Fields marked with an * are required

First Name*

Last Name*

Email*

☐ By submitting this form, I understand that my data will be processed by Capgemini as indicated above and described in the Terms of use.*

Submit

About
Capgemini

Capgemini is a global leader in partnering with companies to transform and manage their business by harnessing the power of technology. The Group is guided everyday by its purpose of unleashing human energy through technology for an inclusive and sustainable future. It is a responsible and diverse organization of 270,000 team members in nearly 50 countries. With its strong 50 year heritage and deep industry expertise, Capgemini is trusted by its clients to address the entire breadth of their business needs, from strategy and design to operations, fuelled by the fast evolving and innovative world of cloud, data, AI, connectivity, software, digital engineering and platforms. The Group reported in 2020 global revenues of €16 billion.

Get the Future You Want | www.capgemini.com

About
the Capgemini Research Institute

The Capgemini Research Institute is Capgemini's in-house think-tank on all things digital. The Institute publishes research on the impact of digital technologies on large traditional businesses. The team draws on the worldwide network of Capgemini experts and works closely with academic and technology partners. The Institute has dedicated research centers in India, Singapore, the United Kingdom and the United States.

Visit us at: www.capgemini.com/ researchinstitute/

MACS_20200901_SS