



Road Map to Safe Data Monetization

THE GUIDE TO TRANSFORMING
EXISTING DATA ASSETS INTO
REVENUE

arm INSIGHT

Executive Summary

Data monetization is a big opportunity that is misunderstood and underutilized, especially in the financial services industry. When done correctly, data monetization is a high margin (85%+) revenue stream with minimal or no risk. When done incorrectly, data monetization drains budgets and resources, and exposes your data to security, compliance, and privacy issues.

Over the last six years, ARM Insight has helped 1,000+ institutions monetize their data. We've experienced our share of successes and mistakes along the way. One thing we've learned over the years is that the perceived fear of data monetization is higher than the real risk. From this experience, we've developed a simple, best-practice Data Monetization Road Map (see page 7) that illustrates how to safely monetize financial data, and of equal importance, what not to do.

Key Takeaways

- The data monetization opportunity is larger, safer, and shorter term (6 months) than you think.
- The following Data Monetization Road Map (“Road Map”) is a 4-step proven process that maximizes your data revenue while minimizing your risk, ensuring that your team is an excellent steward of data.
- There are three data types to monetize (raw, anonymous, and synthetic). Each data type has a different revenue potential and risk profile. Synthetic data is the “break-through” data set; it is the future of data, and you should be using it now.
- You should not attempt any artificial intelligence (AI) or machine learning (ML) unless your data is structured correctly (see step 2 on the Road Map).

“The perceived fear of data monetization is higher than the real risk.”

Audience:

The audience for this Road Map is Chief Executive Officers, Chief Data Officers and other C-Suite executives who are looking for ways to unlock the value and insights of financial data while ensuring that privacy is protected. ARM Insight would be happy to discuss the details of this road map and our data monetization experiences with these executives and their teams.

Road Map Methodology:

The Road Map is a best practice methodology (not a sales pitch or product) based on our experience with thousands of clients. We do not recommend a certain vendor, database, or software, but do recommend best practice steps that do need to be taken to safely drive business value from your data assets.

Table of Contents

Executive Summary	2
Data Monetization Opportunity & Landscape	4
Why Financial Institutions Are Slow to Monetize Data	5
The Road Map	7
Step 1: Embrace Compliance	8
Step 2: Refine the Data	9
Step 3: Three Data Types: Raw, Anonymous & Synthetic	10
Step 4: Your Data Monetization Options	11
Conclusion	12

Data Monetization Opportunity & Landscape

Data Monetization: A Huge Missed Opportunity

While still in its early stages, data monetization is already a fast-growing industry with large, global financial companies monetizing their data. However, most financial institutions have been slow to monetize their data due to a combination of a) misunderstanding of data monetization as an industry, b) perceived compliance complexity and c) lack of data knowledge. We will highlight each of these obstacles later, but let's first address the evolving opportunity.

When The Economist published, "The world's most valuable resource is no longer oil, but data,"¹ the article focused on Internet giants such as Google and Facebook. However, what stands out in the article is that data is already a billion-dollar industry. In fact, Allied Market Research (Figure 1) forecasts the global data monetization market to reach \$370.9 billion by 2023.² For individual institutions, McKinsey speculates that data monetization will generate between 5 and 10 percent of revenues.³

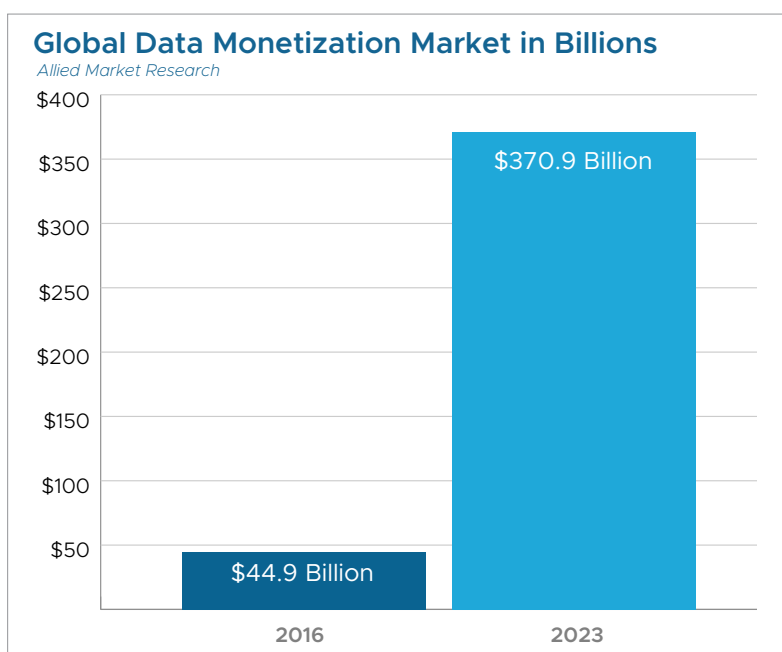


FIGURE 1

Some financial services companies have been externally monetizing data for years. BBVA, for example, operates an online marketplace for other companies to purchase anonymized transaction data from BBVA cards and point-of-sale terminals⁴. Visa, Mastercard, and American Express use data to better target customers through advertising and loyalty programs.^{5,6}

While there are safe and secure ways to drive new revenue streams externally with your data, not all data monetization activities relate to data being sold to other organizations. There are also myriad examples of financial services companies internally monetizing their data to drive greater operational efficiencies, build better analytics, or test new use cases through machine learning and AI.

Why Financial Institutions are Slow to Monetize Data

There are three main reasons why most financial institutions have avoided data monetization to date.


Misunderstanding the definition of Data Monetization


Data monetization is a broad industry term, with multiple paths an organization can take to drive business value or new revenue streams from their data assets. It's important to understand the whole data monetization landscape. Banking executives immediately think of selling financial data to external parties when they hear the term, "data monetization." Again, while there is a safe and secure way to participate in this largely untapped method of *external* data monetization, there are also other valuable data monetization paths that drive *internal* economic benefits from data assets (see step 4 in the Road Map).


Current Regulations Drive Perceived Compliance Complexity

Data privacy, regulatory compliance, reputational concerns, and headline risk have trained executives to just say "no" as it relates to data monetization. However, understanding the opportunities of safe data monetization allow both compliance and business executives to properly understand risk in context, making it more likely they'll say, "yes."

Financial institutions have high fiduciary standards as data stewards, driven not only by customer demand, but also by strict governmental and privacy regulations. Current industry hot buttons include:

 **Gramm-Leach-Bliley Act (GLBA)**
GLBA regulates the ability of financial institutions to disclose "nonpublic personal information" unless the individual has been offered an opportunity to opt-out of the disclosure or when certain exceptions apply.⁷

 **General Data Protection Regulation (GDPR)**
This European Union regulation broadly defines personal data protections to include information that identifies or is identifiable to a person.⁸

 **California Consumer Privacy Act (CCPA)**
The CCPA goes into effect in 2020 and requires data stewards to protect "information that identifies, relates to, describes, is capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household."⁹

Lack of Data Knowledge Leads to "No Clear Path" Mindset

Many bank executives just aren't sure where to start. Siloed data, a lack of internal resources, and general lack of knowledge around data transformation make it difficult to monetize data with confidence. In a 2018 survey of bank leaders by Accenture, nearly 40% of respondents cited complexity of their current state and legacy environment as the greatest barrier to digital transformation. In addition, 70% of respondents either agreed or strongly agreed that they hold hidden value within their operational data that have yet to be leveraged.¹⁰ Translation: data silos make it impossible to have visibility into the comprehensive data estate.

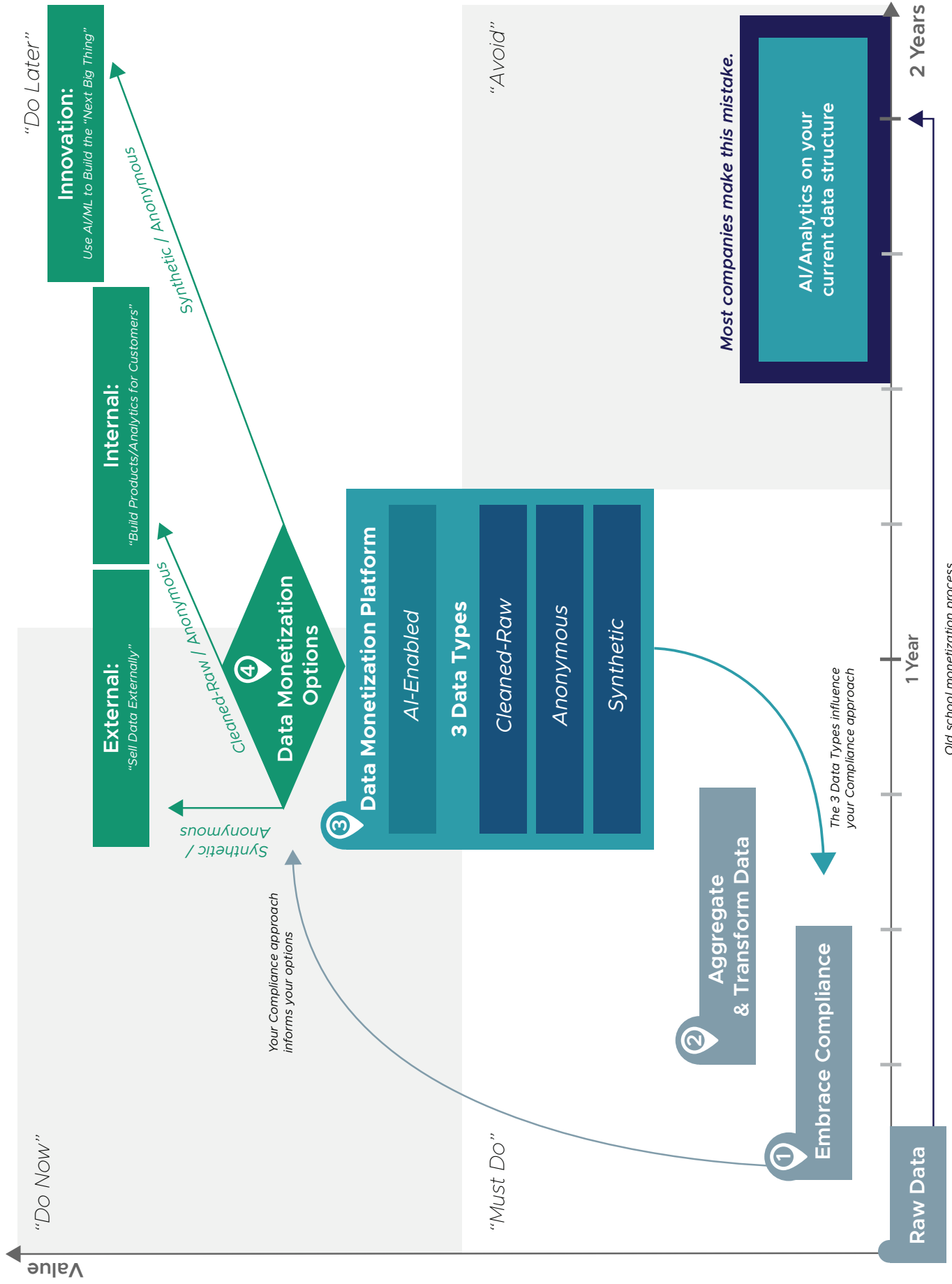
The good news is there is a clear path forward to true value and benefit from your data, all it takes is a little knowledge and understanding of safe data monetization.

The Road Map

Over the last six+ years, ARM Insight has monetized data for 1,000+ financial institutions, and 60+ billion transactions. From this experience, ARM Insight has built a Data Monetization Road Map. It is simple, proven, and illustrates the clear 4-step process to maximizing data value, while minimizing risk.

The 4-Steps

- 1. Embrace Compliance*
- 2. Refine the Data*
- 3. Understand the 3 Data Types:
Cleaned-Raw, Anonymous, and Synthetic*
- 4. Choose your Data Monetization Options*



① Step 1: Embrace Compliance

The Road Map ensures that your entire organization will be safe stewards of data. Your organization cannot afford a data misstep. Furthermore, every Facebook/Cambridge Analytica data mishap negatively affects the entire data ecosystem.

Regulation is here to stay, and more is coming—and common-sense regulation is good for the entire industry. Compliance (regulation, security, and privacy) is an absolute, not an option. Therefore, you don't have to like it or agree with it, but your data monetization approach must flexibly deal with it.

- **Don't be afraid of this (compliance) step**
Embrace it. It is not difficult or complex once you understand the three data types in step three and your data monetization options in step four.
- **Compliance AND three data types relationship**
There is a direct relationship between the three data types and compliance. Understanding this fact will greatly simplify your approach, minimize your risk, and accelerate your monetization.
- **More regulation is coming**
You must build your data structures to accommodate for new and changing regulations that are sure to come.
- **Compliance and data monetization can co-exist**
As you embrace compliance and understand the three data types, it becomes apparent that data monetization and data compliance go hand-in-hand.

Expect the U.S. to pass more GDPR-like regulation on data usage soon. It levels the playing field so that companies that are not good data stewards don't get a competitive advantage or ruin the data opportunity for everybody.

Let's face it, compliance executives confuse technologists and business executives. Alternatively, technologists confuse compliance executives and business executives. The result is usually an uneducated "no" to data monetization. The Road Map defines a common language (highlighted by the three data types) so compliance, technologists, and business executives can have a knowledgeable, pragmatic discussion that leads to an educated decision.

See Appendix Note 1: Customer Contract Compliance

2 Step 2: Refine the Data

If “data is the new oil,” you must refine (clean and format) the raw data to capture its value. Real data monetization and machine learning are simply not possible unless you refine data in a modern data warehouse.

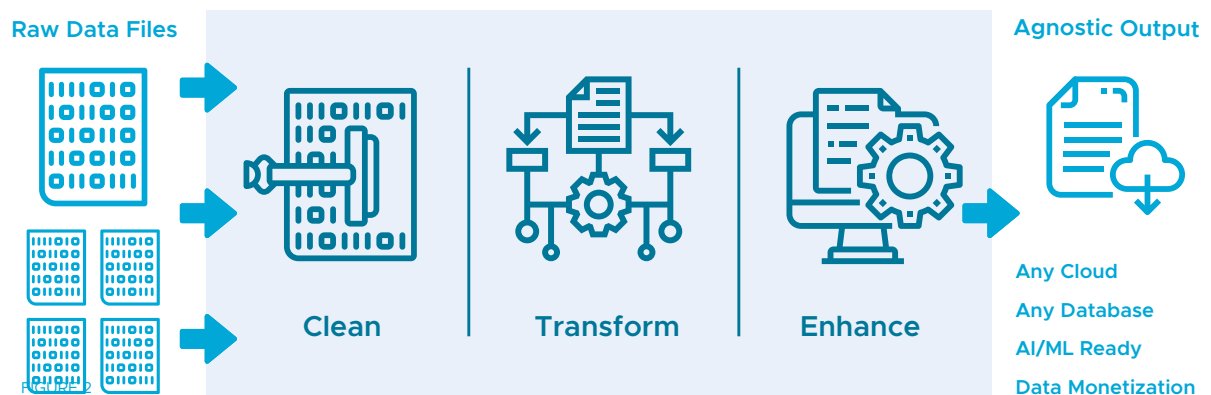
There are five steps to refining data:

- Aggregate raw data files from various siloed environments to be used as part of your data monetization strategy.
- Clean the data to allow for deeper analysis, while ensuring the integrity of the data is intact.
- Transform the data into the three data types based on your compliance approach in step one and your data monetization options in step four of the Road Map.
- Enhance the cleaned and transformed data: Tag merchants, tokenize customers, hash specific fields, etc.
- Output agnostically: Structure your output so that it can work with any PCI-compliant cloud and front-end tools to enable you to execute on your monetization options. (Note: Not necessary for all Financial Institutions (FIs), but we recommend it for medium-to-large FIs).

You must refine your data, or you will certainly fall into the lower right “Avoid” quadrant of the Road Map, wasting valuable time and resources, with minimum gain. Not even basic analytics (let alone AI/ML) are possible without refining data.

Regarding which modern data warehouses and which cloud to use, you have a number of solid options. ARM Insight has worked with all the major data warehouses and clouds, and in some cases we use multiple partners for the same client. You can be successful with any of the following: (Note: ARM Insight expects this list to grow over time)

- **Data Warehouses:** Snowflake, AWS RedShift, Google BigQuery, Microsoft Azure
- **Clouds:** AWS, Google, Microsoft



See Appendix Note 2: OnPrem vs. Cloud Battle, Note 3: “Avoid” Quadrant

3 Step 3: Understand the 3 Data Types

Once you have refined (cleaned and formatted) your data, **understanding the three data types is the most important aspect of the Road Map**. They directly influence all of the other steps.

Cleaned-Raw

All the raw fields for both the consumer (name, SSN, email, phone, etc.) and the associated transaction remain tagged to data. Cleaned-Raw data has the highest monetizable value but also the most risk.




Anonymous

Personal identifiable information (PII) is removed to create anonymous data, but the real transaction data remains unchanged. Anonymized data is low risk.

Synthetic

Synthetic data is fake data that cannot be reverse-engineered back to the original consumer (or banking institution), yet it retains all the statistical value of the original data set. Minor and random field changes are made to the original data set to completely protect the consumer identity and transaction. Synthetic data is the breakthrough data type and has no associated risk — it is fake data.

Data Type:

			
	Cleaned-Raw	Anonymous	Synthetic
Name	John Smith	Customer 123	Customer 123
Address	1234 Main St.	Removed	Removed
City, State, Zip	Los Angeles, CA 97007	Los Angeles, CA 97007	Census Block
DOB/SSN	12/25/80 123-45-6789	12/25/80 77C67BF61841	11/22/80
TRX Date	01/19/19 08:45:23	01/19/19 08:45:23	01/17/19 09:15:13
TRX Amount	\$13.56	\$13.56	\$14.96
Risk	High	Low	None
Monetization Option	Internal	External, Internal, Innovation	External, Innovation

See Appendix Note 3: Synthetic Data

4 Step 4: Choose Your Data Monetization Options

When banks and other financial institutions consider data monetization, there are three types of revenue streams to choose from: external, internal, and innovation. It's important to include your compliance team fully as you embark on this phase of data monetization. The opportunity-versus-risk profile of each option will inform which data type (cleaned-raw, anonymous, or synthetic) should be used with which data monetization option (external, internal, or innovative).

External Monetization: Selling data (usually synthetic) to other third parties, likely outside of your industry

- **Data Types:** Synthetic and Anonymous
- **Time:** 6-8 months
- **Risk:** Very low to none for synthetic data, low risk for anonymous data. ARM Insight recommends starting with synthetic data to protect the privacy of the consumer and comply with data regulations.
- **Use Case – Retail Competition:** Retailers usually have a solid grasp of their own data. However, they know less about their competition's. Retailers are willing to pay for a market share analysis or competitive insight (using synthetic data) of average transactions (over time) or percentage of online vs off-line transactions.

Internal Monetization: Building new products or analytics for your existing customers (consumer or business) for internal use.

- **Data Types:** All three – Synthetic, Anonymous, Cleaned-raw
- **Time:** 8-12 months
- **Risk:** Very low for synthetic data, high for cleaned-raw data (make sure you understand your compliance and customer contracts well).
- **Use Case 1 - Customer Insights:** Most FIs are not great at analytics. Using anonymous data, FIs can build and potentially monetize a variety of analytics, including spending for small business users, customer segmentation for marketing, fraud detection trends, customer loan likelihood, etc.
- **Use Case 2 – Security:** Usually, the single biggest security data risk is employee misuse/abuse of data. In most organizations, many employees/analysts have direct access to raw PII data to do their jobs, and this should stop unless it is absolutely critical that they have PII access. Synthetic data will vastly reduce this risk.

Innovation Monetization: Using machine learning or AI to build the “next big thing” (i.e., creating innovation labs or fintech incubators that test new ideas and develop proof-of-concepts) or deliver personalized banking experiences.

- **Data Types:** Synthetic, Anonymous
- **Time:** 16-24 months
- **Risk:** Very low for synthetic data. High for cleaned-raw data (make sure you understand your compliance and customer contracts well).
- **Use Cases:** Chat bots, virtual banking assistants, real-time credit limits and scoring, personalized or dynamic credit limits and enhanced payments fraud detection, with many more ideas in the “lab”.

Conclusion and Next Steps

After reviewing the Road Map, you should be very confident in the following:

- Data Monetization is a big opportunity, that can be safely executed, ensuring that you are great stewards of the data.
- The Road Map is a proven, 4-step process to safely monetize your data. Assuming you have sufficient data, your team should start using the Road Map now.
- Start synthetic data now: It is the future of data monetization due to the following benefits:

Security: It solves your biggest security risk. Your internal employees should not have unchecked access to raw PII data.

Revenue: It can bring revenue (at 85% margin) within six months.

Compliance: It greatly simplifies all your compliance (regulatory, privacy, security, and customer) needs.

Cloud: It is the fastest, cheapest way to either migrate to the cloud or enhance your current cloud migration.

If you have any questions or clarifications, contact ARM Insight. ARM Insight is happy clarify any points of the Road Map methodology with your team.

Contact us at info@arminight.com, 503-546-2288 or visit www.arminight.com.

To learn more about the privacy benefits of synthetic data as well as its compliance implications for GLBA, CCPA and GDPR, read the legal opinion, **“Enhance Privacy Through the Use of Synthetic Data.”**¹¹

About the Author

Randy Koch is the CEO of ARM Insight. ARM Insight is a financial data company that helps monetize data for 1,000+ companies, by leveraging the latest technologies in big data, analytics and artificial intelligence/machine learning.

Randy has 20+ years of experience in building software, analytics and data companies across multiple industries: Financial, Healthcare, Security, and Public Sector (both federal and state).

Appendix 1

Note 1 - Customer Contract Compliance

- Separate from regulatory and privacy compliance, you must also review your existing consumer/customer contracts. You can be 100% compliant with all state and federal regulations, but you may have a client contract that, for example, “prohibits the Bank from selling or reusing Anonymous Data in any form.” Again, the three data types will also make this task much easier and clearer for your team.

Note 2 - OnPrem vs. Cloud battle

- Don't let your team have dogma regarding OnPrem vs cloud. You should be quickly migrating to the cloud (private or public), even if it's in small steps. The cloud databases are simply superior: better, faster, cheaper, and scaling flexibility that is incredible. They are secure and the major cloud providers will continue to heavily invest in this space. That being said, you can make OnPrem work, but it is a much tougher task. Eventually, you should be migrating to the cloud. You don't have to migrate your entire infrastructure at once; but you should start. You should pick a data monetization use case (Road Map step 4), and go.
- Total Cost of Ownership (TCO): The TCO with cloud providers is lower, mainly due to two main drivers:

Lower Maintenance Costs: Database hardware and software maintenance is 90% cheaper in the cloud. With OnPrem, you (assuming a mid or large-size FI) need a small army of DBAs (data base administrators). Cloud providers not only perform this function, they do it better, with better tools.

Scale Flexibility: Scaling is easy and “pay by the drink.” Let's say you need to load 10 years of data (80 Terabytes). You can simply “spin-up” these extra nodes (servers) for two days, load the data, and then shut them down, and only have to pay for the two days. Contrast this with the OnPrem model of buying four servers, spending labor on installation and data loading, and now you're stuck with the servers for years.

- Financial Note: OnPrem does have the advantage of amortizing the hardware over a number of years. However, this benefit is overshadowed by the cloud benefits. Additionally, ARM Insight anticipates that the accounting rules will change to become more cloud-friendly over the next three years.

Note 3 - “Avoid” Quadrant

- The biggest mistake that most companies make is attempting to monetize data or build machine learning on legacy technologies or silos. You are better off doing nothing than attempting this route. While it is “technically” possible to go this route, it will likely drain your technical resources, drain your budget, and become obsolete in 1-2 years.

Note 4 - Synthetic Data is the Future:

Synthetic data is the future of data monetization and data security

- **Data Security:** Currently thousands of internal employees have access to raw data for a variety of reasons. This exposes you to unnecessary risk and must change. In most use cases, these same employees can do their jobs with synthetic data, eliminating a huge security risk.

Healthcare example: When a breakthrough pharmaceutical or medicine is emerging, testing is usually done on cells, prototypes, rats, etc (i.e. synthetic data), not humans (raw data)

- **Data Monetization:** Synthetic data is where you achieve your fastest revenue (within six months), at no risk.
- **“Old school pushback”:** You may hear some pushback from your team along the lines of, “Duplicating data is too expensive,” (i.e., creating synthetic data in addition to raw data). There are three reasons why this is no longer the case: 1) Storage is cheap and will continue to become cheaper. 2) It is much safer to have your internal team analyzing synthetic data than raw data. 3) The revenue and reduced risk from synthetic and anonymous data will far outweigh additional storage costs (if any).
- **The Algorithm:** Synthetic data mimics the real data while removing the identifiable characteristics of the individual, banking institution, and transaction. Because synthetic data maintains the overall data accuracy, it can be safely used by your internal teams, and safely monetized.

Note 5 – Pragmatic Machine Learning (ML):

- No doubt that AI/ML is a fast-growing topic and is on the mind of many executives. Based on the Road Map, we recommend the following path:

Structured Data 1st, ML 2nd: Machine learning is a waste of money unless your data is structured first (step 2) so that the machine can do its job.

Identify the Problem/Question: ML won't be effective unless you give it a specific problem or question to address. Don't let your team fall in love with ML for ML's sake. ML is a great tool, not the solution. You must ask 1 of 2 questions: “What problem am I trying to solve?” or “What specific question am I trying to answer?”

Note 6 - 85%+ margin revenue

- The advantage of data revenue is that the margin is usually high, especially through external monetization with synthetic data, where we routinely see 85% margin.

The Road Map to Safe Data Monetization: The Guide to Transforming Existing Data Assets Into Revenue may be updated periodically. While reasonable efforts are made to keep the content accurate and up-to-date, ARM INSIGHT MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND, EXPRESS OR IMPLIED REGARDING THE CORRECTNESS, ACCURACY, COMPLETENESS, ADEQUACY, OR RELIABILITY OF OR THE USE OF RESULTS THAT MAY BE GENERATED FROM THE USE OF THE INFORMATION OR THAT THE CONTENT WILL SATISFY YOUR REQUIREMENTS OR EXPECTATIONS.

Components of the content original to and the compilation produced by ARM INSIGHT is the property of ARM INSIGHT and cannot be reproduced without its prior permission.

Sources

1. [The world's most valuable resource is no longer oil, but data](#)
2. [Data Monetization Market by End User \(BFSI, E-commerce & Retail, Telecommunication & IT, Manufacturing, Healthcare, Energy & Utilities, and Others\) - Global Opportunity Analysis and Industry Forecast, 2017-2023](#)
3. [Monetizing data: A new source of value in payments](#)
4. [Don't let that data you're sitting on go to waste](#)
5. [Visa and Mastercard earnings: More than just payments at play](#)
6. [Mastercard, AmEx And Envestnet Profit From \\$400M Business Of Selling Transaction Data](#)
7. 12 C.F.R. § 1016.3(p).
8. GDPR Article 4(1) (“‘personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.”)
9. [Cal. Civ. Code § 1798.140\(o\)](#)
10. [2018 North America Banking Operations Survey](#)
11. ARM Insight “Enhance Security Through the Use of Synthetic Data”. To request a copy, visit www.arminsight.com