

Know Your Data

Pour the foundation before you build the house

Jason Keith, CIA, CISA | Partner, Saltmarsh

The Dream

[Money Ball](#)

Today's conversation

1

The opportunity

The vault we
forgot

2

The foundation

Name it, tag
it, own it

3

The warning signs

Flying Blind

4

Getting it Right

What clean
data buys
you

5

The first move

One “junk
drawer” at a
time

Jamie Dimon “We have large amounts of extraordinarily rich data that, together with AI, can fuel better insights and help us improve how we manage risk and serve our customers” 2023 annual letter to shareholders.

A customer call scenario.

Riverside Manufacturing has used us for 22 years: The CFO has a question about a wire.

The reality

Banker puts them on hold to check the core
We make a call to treasury management team
Deposit ops gets involved looking for an old wire instructions form
It takes 40 minutes and requires an in person visit

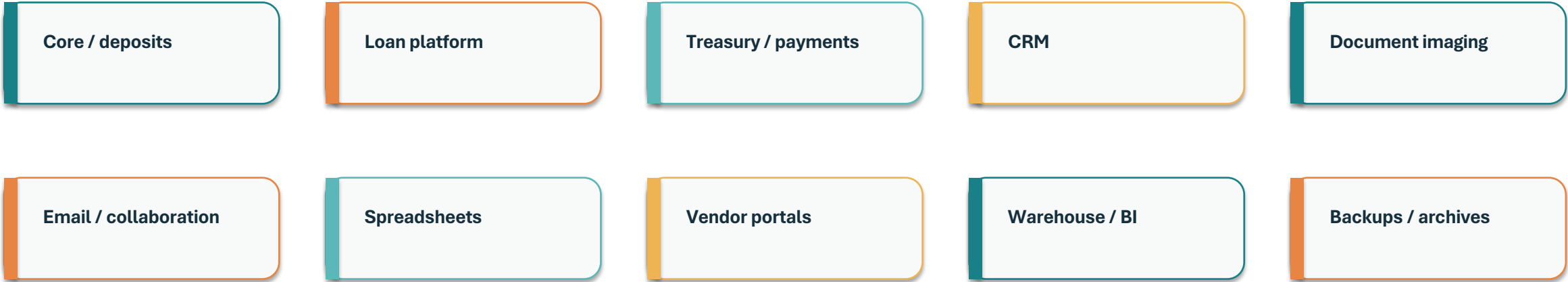
What the customer expects

“You know us. This should be easier”

Example Sofi - loan refinance

The knowledge is scattered

One relationship can be divided among systems, documents, people, vendors, and archives.



We may have the information. Do we have a clear relationship view and a clear source of truth.

What the bank loses when it cannot use its own information

Customer confidence

The customer repeats information and questions whether the bank understands the relationship.

Employee capacity

Bankers, operations, and analysts spend time searching, re-keying, and reconciling.

Decision quality

Teams work from incomplete, stale, or conflicting records.

Modernization speed

Every automation or AI project pauses to resolve basic data questions.

The missed opportunity is larger than efficiency. The bank loses the ability to turn relationship knowledge into timely service and informed decisions.

Downloaded from <http://ajph.org/> on November 10, 2014



What does “know your data” actually mean?

We should be able to answer five questions before we automate, analyze, share, or pass to AI.

- What is it?
- Where does it live?
- How sensitive is it?
- Who owns and may use it?
- Can we trust it for this purpose?

Classification is the discipline of attaching these answers to information before the bank uses it.

Classification should change what the bank does with the data

Public	Published rates, branch hours	May support public-facing tools.
Internal	Procedures, training, operating guidance	May support approved employee knowledge tools.
Confidential	Internal financials, vendor contracts, strategy	Limited employee and approved-vendor use.
Restricted	Tax returns, account numbers, loan files, credentials	Strong access, approved environments, strict vendor / AI controls.
Customer-impacting decision data	Credit inputs, fraud signals, dispute evidence	Defined source, monitoring, exception handling, human review.

The Bank's Underused Asset: Trusted Relationship Data

A community Bank's advantage is not simply its data. It is the ability to responsibly turn it into better service, better decisions, and deeper relationships. We build relationships.

Raw information

- Deposit activity
- Loan performance
- Treasury usage
- Financial documents
- Service interactions
- Related entities

Trusted relationship insight

- Resolved customer and entity links
- Current, sourced exposure
- Approved document status
- Context for needs and risk

Outcomes

- Faster response
- Better relationship management
- Lower manual work
- More relevant conversations
- Safer automation and AI

Data is not the asset by itself. Our personal relationship leveraged by data is the x factor.

Value the Use Case

Do not try assign a dollar value to “the bank’s data.” Measure the value by a specific use case.

Decision velocity (speed)

Hours saved × loaded cost
Fewer stronger customer touches
Shorter service cycle

Revenue + retention (growth)

Retained deposits/renewal rate
Incremental contribution margin
from expanded relationships

Risk + control (protection)

Expected loss avoided
False positives reduced
Rework and remediation avoided

Option value (flexibility)

Number of future uses enabled
Reusable relationships, definitions,
and permission controls

A long drive (automation bias)

Sabine Moreau, Belgian Woman, Drives 900 Miles Off 90-Mile Route Because Of GPS Error

Woman Drives 900 Miles Off Route Because Of GPS Error



By Ryan Grenoble

Jan 15, 2013, 06:03 PM EST | Updated Dec 6, 2017

Amazon: when historical data teaches the wrong lesson

Amazon discontinued a recruiting tool after it learned bad patterns from historical resumes.

What it received

Historical resumes shaped by prior hiring patterns and a predominantly male technical workforce.

What it learned

Patterns associated with prior applicants not a reliable definition of capability or job fit.

The lesson

The data may have been real. It was not fit for the question the model was asked to answer.

For banks: prior approvals, losses, overdrafts, or fraud alerts may describe history. They may not define the right future decision rule.

Source: U.S. House hearing record (2019) — [The Future of Work: Protecting Workers' Civil Rights in the Digital Age](#).

A report can run successfully and still be wrong

A data failure often looks like business as usual until we are forced to ask: “How did we get here?”

What happened

JP Morgan London Whale Spreadsheet Error. Manual copy and paste error cost. \$6 Billion in fines and over 900 million in regulatory fines.

The conclusion

A report needs a source, definition, refresh time, owner, and validation before it is dependable.

The most dangerous report may be the one that runs successfully but cannot be trusted.

“Which total commercial loans is correct?”

All four answers may be explainable.

Core system

\$247M
Close-of-business

Credit admin

\$500M
Includes participations

Warehouse

\$248M
Prior-night refresh

Board package

\$250M
Rounding

Centralizing confusion does not create clarity.

FDTA **NOT** a community bank mandate but a valuable perspective.

The agencies' June 11, 2026: A rule for financial regulators. The direction may be highly relevant.

Common identifiers

A customer, entity, product, or instrument should not become five different records in five systems.

Machine-readable data

Information should be usable by people, reporting tools, workflows, and approved AI, not trapped in PDFs and re-keying.

Semantic meaning

The institution should be able to explain what a field means, its source, update timing, and limitations.

FDTA's practical message: data becomes more valuable when meaning travels with it not merely when it is stored somewhere central.

Source: OCC Bulletin 2026-25, June 11, 2026 — FDTA joint final rule.

Section 1033 Open Banking Rules

- 1 Customer permissioned data movement
- 2 Frictionless movement of data and money
- 3 Accelerates Fintech model

As with everything else in banking things are speeding up. If we don't master our data someone else will.

What good data makes possible

BankUnited

Bounded, owned policy content can become a reliable employee knowledge tool.

Lloyds Athena

Curated internal knowledge can reduce search time and improve frontline service.

Relationship view

Connected customer, account, document, and service data can reduce friction across the bank.

Analytics and AI

Approved, fit-for-purpose data supports better automation without asking AI to guess.

The common pattern: a defined use case, a bounded and governed data set, and a visible human or control boundary.

The best first AI use case is often internal knowledge

BankUnited: SAVI

About 400 policy and procedure documents supported an internal GenAI assistant. BankUnited reported 95% accuracy and responses in less than 10 seconds.

Lesson: Begin where content is owned, current, approved, and easier to verify.

Lloyds: Athena

Athena searches about 13,000 internal articles. Lloyds reported average search time declined from 59 seconds to about 20 seconds.

Lesson: AI can help employees find trusted institutional knowledge without replacing judgment.

The first AI win may not be a customer decision. It may be helping an employee find the current, approved answer while the customer waits.

Sources: AWS Case Study — BankUnited SAVI; Lloyds Banking Group — Athena knowledge hub.

AI-ready data is:

- **Authoritative: The Source is understood**
- **Current: Fresh Not Stale**
- **Connected: Context Comes With It**
- **Permissioned: Access is Clear**
- **Fit for purpose: Right Data, Right Question**

A practical start

Crawl

Choose a high-value data domains. Map systems, owners, copies, definitions, and vendors.

Walk

Fix one failure point: unclear metric, unresolved identity, excessive access, or uncontrolled content.

Run

Prove one improvement: reconciled report, AI-ready knowledge set, narrowed access role, or approved use case.

Don't get bogged down with a 100-page policy and a committee that meets for a year. Start small with one data set, one known risk, and one result the bank can measure.

Three questions to take back to the bank

Can you find it?

If asked to show everywhere one customer's data lives: could you do it?, how long would it take? and would you be sure you had it all?

Can you trust it?

Would you bet your job on your board reporting or the answer to question 1?

Who owns it?

Can you define who owns a data set, assign responsibility?

Our mindset needs to shift from protecting data to using the right data, for the right purpose, with the right control before asking AI to use it.

Thank You!

Jason Keith, CIA, CISA | Partner

Jason.keith@saltmarshadvisors.com