

Fix Your F***ing Data Mess

Don't Let Poor Data Quality Undermine
Enterprise Design

Scott W. Ambler

Data Methodologist | Author

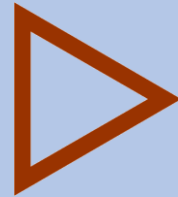
Ambyssoft Inc.



Scott Ambler



- scott@scottambler.com
- [linkedin.com/in/sambler/](https://www.linkedin.com/in/sambler/)
- [@scottambler.bsky.social](https://bsky.social/@scottambler)



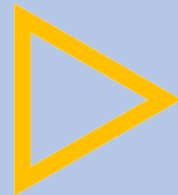
Ambysoft

Data Methodologist,
Board Advisor
[Ambysoft.com](https://www.ambysoft.com)



Agile
Data

Thought Leader
[AgileData.org](https://www.agiledata.org)



Agile
Modeling

Thought Leader
[AgileModeling.com](https://www.agilemodeling.com)



Disciplined
Agile

Co-Creator
pmi.org/disciplined-Agile

Agenda

- Today's takeaways
- Data debt and enterprise design
- Industry statistics
- Data is the new water
- Fixing your f***ing data mess
- Parting thoughts

Today's Takeaways

Data quality (DQ) is a critical factor for effective decision making and successful AI

DQ impacts the delivery of your operational services

The best place to address DQ issues is at the source

Definition: Data Debt

Technical debt is the accumulation of defects, quality issues (such as difficult to read code or low data quality), poor architecture, and poor design in existing solutions

Data debt (data technical debt) refers to quality challenges associated with legacy data sources, including both mission-critical sources of record as well as “big data” sources of insight.

Source: AgileData.org/essays/dataTechnicalDebt.html



Causes of Data Debt

- Business prioritizing time to market over quality concerns
- Manual data entry
- Multiple siloed sources
- Lack of input validation
- Inconsistent data collection methods
- Inconsistent business rules across applications
- Ineffective data management
- Weak data literacy



Why Should Enterprise Designers Care?

What good is a well-designed service if you don't have the required data?

What does it say about your organization if the data about your customers, products, and services isn't trustworthy?

How long will your brand survive the effects of poor service?

Are your systems really assets?

How can leadership guide your organization effectively?

The Impact of Data Debt

15-25%

of revenue lost due to bad data in most orgs (Thomas Redman, 2017)

\$12.9 million

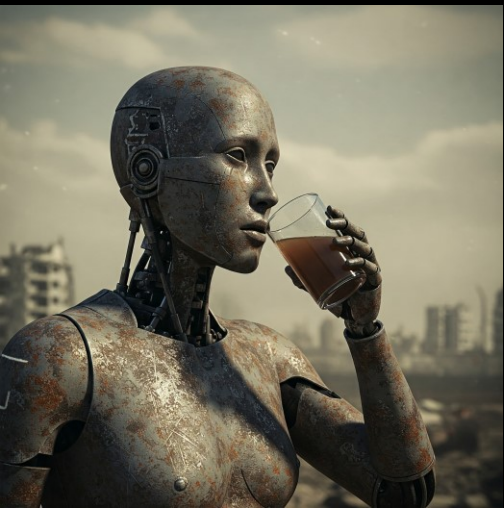
The average annual cost of poor-quality data (Gartner, 2021)

33%

of executives trust their data to derive value from it (Accenture, 2019)

46%

of organizations report that data quality and consistency is a challenge for their AI projects (K2View, 2024)



Source: AgileData.org/essays/impact-of-poor-data-quality.html

Data is the New Water



When water is dirty, we can:

1. Filter it just before we drink it
2. Filter it coming into our home
3. Filter it before it is dumped into our water supply
4. Filter it at the source
5. Clean and fix the actual source

Source: AgileData.org/essays/data-quality-metaphor.html

Data is the New Water: Fixing Your F***ing Data Mess

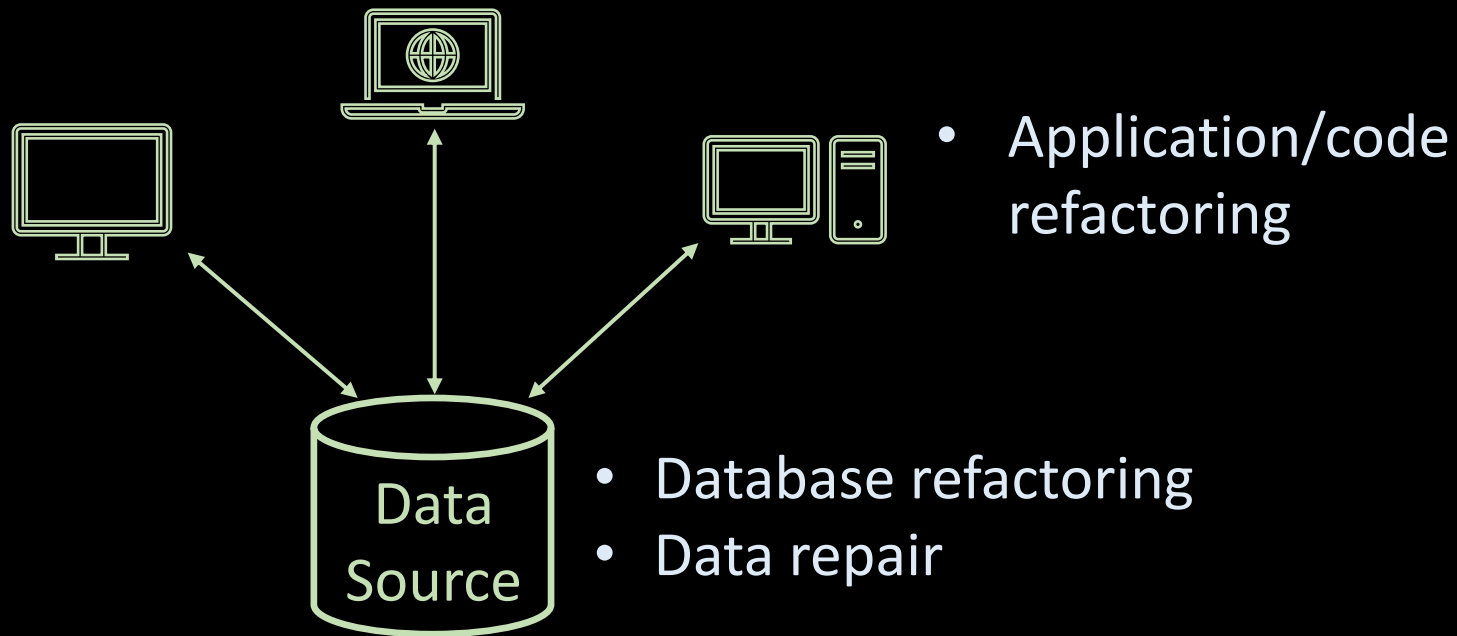


When data is dirty, we can:

1. Cleanse data at point of use for a function
2. Cleanse data at point of use for a system
3. Create a copy of the data and cleanse that
4. Wrap the data source with a data contract or API
5. Refactor/repair the data source and any systems updating it

Source: AgileData.org/essays/data-quality-metaphor.html

Fixing Your F***ing Data Mess: Technical Solutions



- AgileData.org/essays/databaseRefactoring.html
- AgileData.org/essays/data-repair.html
- Refactoring.com

Fixing Your F***ing Data Mess: Overcoming the Bullshit Excuses



We don't have the time → You're likely fixing the data already, why not do it once at the source?

We don't have the skills → Then invest the time to gain them

We have more important things to do → Do you?

My team doesn't own the data → Another team does, work with them

Our company doesn't own the data → Another organization does, work with them

Parting Thoughts

Data quality (DQ) is a critical factor for effective decision making and successful AI

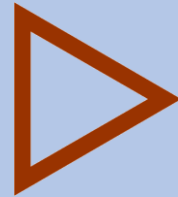
DQ impacts the delivery of your operational services

The best place to address DQ issues is at the source

Thank You!



- scott@scottambler.com
- [linkedin.com/in/sambler/](https://www.linkedin.com/in/sambler/)
- [@scottambler.bsky.social](https://bsky.social/@scottambler)



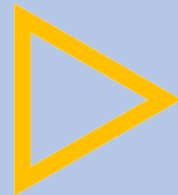
Ambysoft

Data Methodologist,
Board Advisor
[Ambysoft.com](https://www.Ambysoft.com)



Agile
Data

Thought Leader
[AgileData.org](https://www.AgileData.org)



Agile
Modeling

Thought Leader
[AgileModeling.com](https://www.AgileModeling.com)



Disciplined
Agile

Co-Creator
[pmi.org/disciplined-Agile](https://www.pmi.org/disciplined-Agile)