A PRACTICAL GUIDE TO RESOLVING

DATA DEFINITION
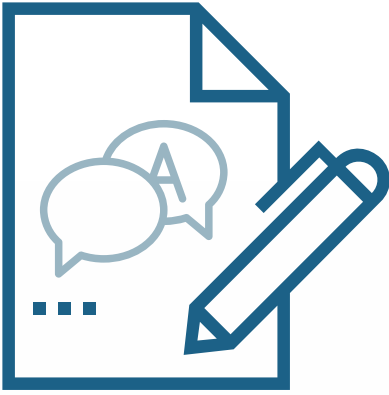
CONFLICTS

Principles, Process and Progress

FIRST SAN FRANCISCO
PARTNERS

by Gretchen Burnham
First San Francisco Partners

What's the data mean?

Why does our organization care about it?

What is "good" for the data?

Business Data Definitions are *Everything*

Definitions drive clarity of business purpose. They're about how we come together to create a common vocabulary, a common language.

It's not just about the data dictionary or business glossary that's probably in an Excel file somewhere. It's actually about how your organization thinks about itself.
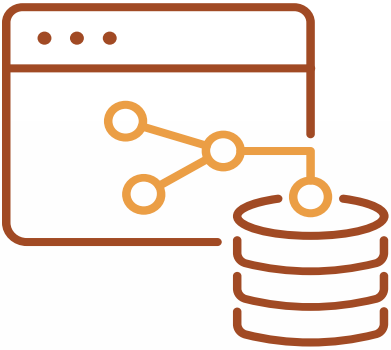
Before we can begin to effectively manage and govern data, we first have to understand why our organization needs it and how it flows through our business. Before data can be controlled, measured or optimized, it needs to be defined.

Data is a tool by which we make aspects of our business machine-readable for computers.

Data is a mechanism to reflect an organization's reality in its systems.

So data is actually about the business.

The reality of our business and how we make our computers interact with that reality — that's what all this data talk should be about.

Who are our customers?

How do our customers interact with our different lines of business?

What products or services do we sell?

Who is financially responsible for which products or services?

How do we organize what we sell?

What do we need to understand about our products and services?

The answers to these questions are the business decisions that get reflected in our data.

Historically, systems were built within operational silos.

Each system encoded the definition that was active within that silo/context.

Now, data is expected to flow seamlessly throughout the organizations' operational systems, as well as to combine flawlessly for reporting and analytics.

Data that was previously fine within its own data stream is now shown to be in conflict with similar data from another stream.

Why Are There Data Conflicts? It's Complicated

Many organizations started out with systems that were built within operational siloes.

When the data had to interoperate between different functional areas, it did so in controlled ways with batch processes and ETL with thousands of lines of code.

If there were different definitions, they could be handled within those defined hand-off points.

Now, data is expected to flow seamlessly throughout our organizations' operational systems. And we expect the data to combine flawlessly for reporting and analytics.

Data that was previously fine within its own data stream is now in conflict with similar data in another part of the organization.

This is where the big thorny data problems begin. There can be fundamental disagreements in your organization about what certain data is supposed to mean, based on the varied operational perspectives. *And it's hard work!*

But there's real value in bringing people to consensus and solving data conflicts.

## Icky Data Conflicts and Data Governance's Role

Data lives at the intersection of people, processes and systems.

To truly resolve a conflict or misalignment, it's very likely that there will need to be a change — usually to the processes and/or the systems involved. (You can't wordsmith your way out of a definition conflict in any meaningful way.)

Data governance teams, as well as related groups such as Information Architecture, are uniquely placed to work on definition conflicts. Sometimes they are not brought to the table because they are not seen as a direct stakeholder. But there is an opportunity to be neutral facilitators with deep expertise in how data flows and works.

We've seen governance organizations shy away from this because they understand they don't have the authority to make any one change. But in reality, they have an opportunity to influence all of the players and to enlist the authority of the organization's executives.

Be willing to dive into the icky stuff. That *is* the value of data governance.
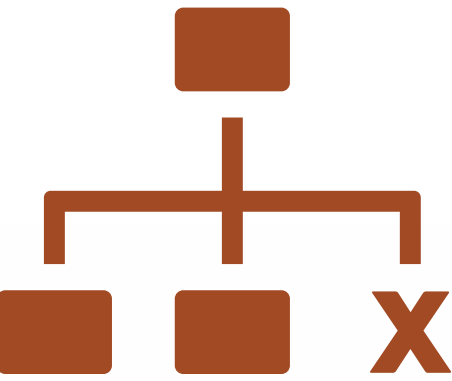
Conflicts that show up in data reflect conflicts or lack of alignment in the organization

Do any of these phrases sound similar to what's said in your organization about your data?

*"That definition is close, but for **my** department, we mean X."*

"There are two types of this thing, which is sort of a hybrid of the other two."

"That's not what the term means **at all**. That term means Y."
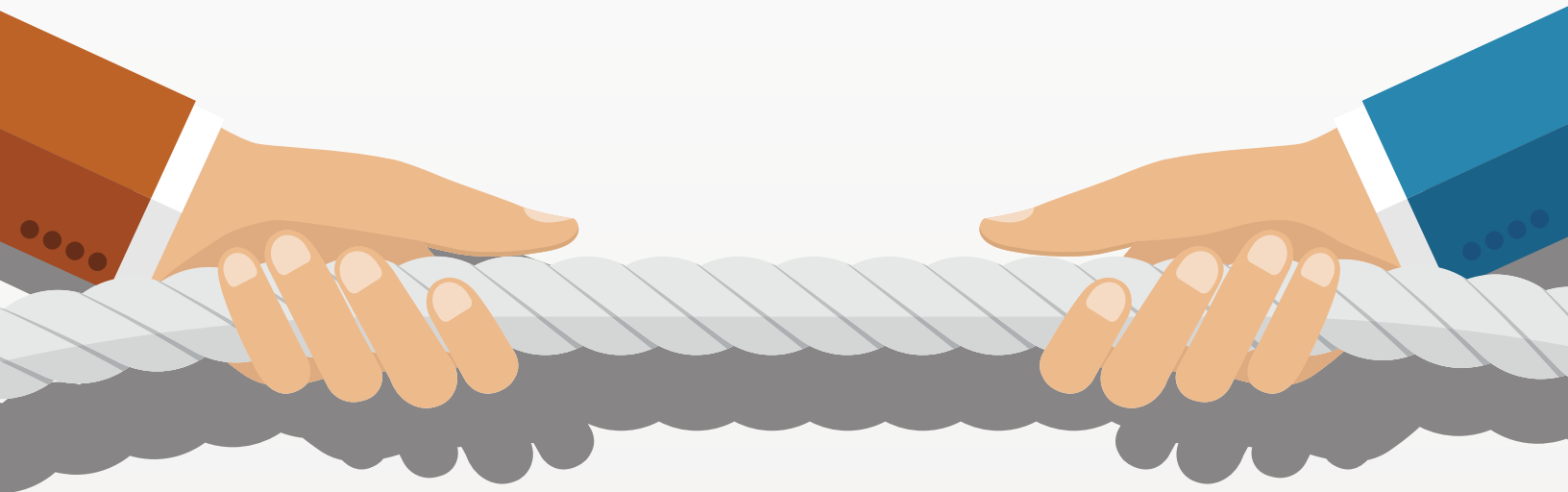
When data conflicts occur, they're reflecting the organization's different perspectives about the data.

1. Avoid Judgment

2. Focus on the Data

3. Question the Status Quo

4. Be Pragmatic

Conflict Facilitation: Four Guiding Principles

1. Avoid Judgment

Data doesn't have emotions. But people do.

Data definition conflicts are actually about people, their perspectives and the decisions people make on a day-to-day basis (how they do their business processes, etc.).

Data is at the heart of how many people do their jobs, which means that data is part of their professional self-identity.

No particular business usage of data is *wrong.* No employee is bad for having a different definition. Everyone is trying to get through their day and do their work — and they're using the data in whatever way is needed to serve that goal.

When we say things like, *That's junk data,* we ignore the fact that the junk data often had (or still has) a specific business purpose that we're not aware of. It might be misplaced and misnamed, but don't assume that it's bad.

It's probably safe to say that most everyone in the organization is trying to get through their day doing the best job they possibly can.

1. Avoid Judgment

2. Focus on the Data

3. Question the Status Quo

4. Be Pragmatic

Conflict Facilitation: Four Guiding Principles

2. Focus on the Data

Remember, data is a tool that we use to do our work. When working to resolve data conflicts, it's good to keep the focus on the fact we're trying to make sure everyone has the right tools for their jobs.

How we describe our intent can either make or break our efforts.

> I'm not here to judge you or your business goals.
>
> I'm here to talk about the data and how it all fits together and flows.
>
> We need the data to interoperate, and it's not doing that right now.
>
> Let's talk about how we can resolve this together.

Additionally, it can be helpful to look at the actual production data to see how well it conforms to the definition, rather than relying on "how it's supposed to work" knowledge.

By keeping a focus on the data, we can cut across the business processes and systems that are part of the conflict.

1. Avoid Judgment

2. Focus on the Data

3. Question the Status Quo

4. Be Pragmatic

Conflict Facilitation: Four Guiding Principles

3. Question the Status Quo

Business vocabulary (and the data associated with it) is the cornerstone of an organization's culture. There are legitimate business purposes for why we describe things the way we do.

However, those reasons may no longer exist. Yet, we often continue to enforce legacy constraints that are no longer needed.

"

Our data is how we reflect our reality in our systems.

If some of our reality has changed, should some of our data change, too?

Let's talk about how we can let go of constraints we no longer need.

This doesn't mean the old way is wrong. But we need to be willing to consider a new way.

1. Avoid Judgment

2. Focus on the Data

3. Question the Status Quo

4. Be Pragmatic

Conflict Facilitation: Four Guiding Principles

4. Be Pragmatic

The goal is to identify a resolution that will work for the organization from an enterprise perspective.

Sometimes we need to make a decision that isn't the 100% theoretical ideal, but it will solve the issue and allow the organization to move forward.

We're striving for progress, not perfection. And the progress we can make will be acknowledged and maybe even celebrated!
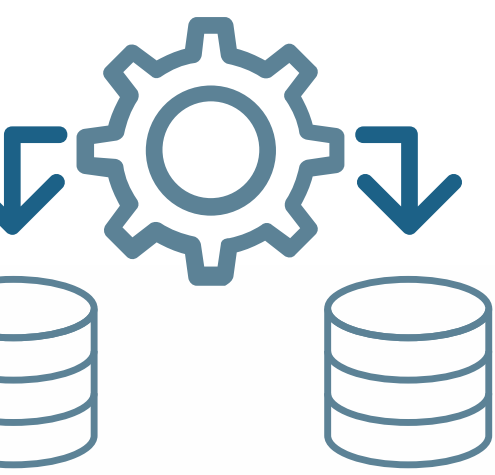
Data Conflict Resolution: A Practical Approach

Before tackling data definition conflicts, your organization needs a data governance framework that addresses some key questions.

> Who gets to make decisions around people, process, technology and data?

> What level of consensus is required in the organization?

> How are decisions recorded and communicated?

Also, stakeholder buy-in is required... because conflict resolution is a team sport!

> Do the relevant parties agree to participate?

> Is anyone able to do an end-run around the forum?

Determining the implementation pathway is also essential.

> How will your forum's decision become real?

> What process and/or system changes will be needed?

1. Gather Stakeholders

Ensure all the relevant people are in the room.

2. Analyze the Conflict

Break down the definition, identify the specific conflicts.

3. Formulate Resolution Options

Propose solutions, identify impacts, make recommendation.

4. Approve and Implement Resolution

Record decision in glossary, ensure any impacts are handled.

Data Definition Conflict Resolution Process

Step 1: Gather Stakeholders

Bringing together stakeholders is the absolute foundation for your success. Even if you've been working with them (for example, on a business glossary), it's important to say, Let's stop and take a look at all the stakeholders we need.

Questions to address:

- Who are all of the producers and consumers of the data?

- How does data get created and brought into your organization?

- Who owns the business processes?

- Who are the primary consumers of the data, both internally and externally?

- What stakeholders need to be brought into your data conflict efforts because they create or use the data?

- Are there other stakeholders? (Think broadly, for example, Information Security, Legal, Privacy, Compliance and Human Resources.)

- Who gets to make the final decision to implement changes around the information?

Cast the net wide to include people outside the key stakeholders. You need to communicate, at some level, with a broader group to truly resolve data definition conflicts.

1. Gather Stakeholders

Ensure all the relevant people are in the room.

2. Analyze the Conflict

Break down the definition, identify the specific conflicts.

3. Formulate Resolution Options

Propose solutions, identify impacts, make recommendation.

4. Approve and Implement Resolution

Record decision in glossary, ensure any impacts are handled.

Data Definition Conflict Resolution Process

Step 2: Analyze the Conflict

Data definition conflicts often come to us like big balls of yarn — a tangled mess — that no one wants to touch.

It's important to work together to break things apart, for example, in a whiteboard session or working line by line through glossary definitions.

Collectively, identify points of agreement, specific points of conflict and the cause of conflict — for example:

- Different contexts for similar data

- Overloaded data fields

- Name conflicts

We'll examine these causes in more detail shortly.

Get agreement early on so you can move forward. And if you have to get very granular — "Can we agree at the top level that a customer is someone who gives us money?" — that's okay.

1. Gather Stakeholders

Ensure all the relevant people are in the room.

2. Analyze the Conflict

Break down the definition, identify the specific conflicts.

3. Formulate Resolution Options

Propose solutions, identify impacts, make recommendation.

4. Approve and Implement Resolution

Record decision in glossary, ensure any impacts are handled.

Data Definition Conflict Resolution Process

Step 3: Formulate Resolution Options

Next, you'll start to formulate a plan of action for resolving data definition conflicts.

> What **should** the data be in order to meet all of the various business needs? What are the options?

Then, identify the impact of each solution.

> Will this affect a business or technical process?

> Will it require a system change?

Identify the expected benefit of each solution.

> If we make this change, how will we know it works?

> What does success look like for us?

Get consensus from the cross-functional team on the impact and benefits of the plan.

1. Gather Stakeholders

Ensure all the relevant people are in the room.

2. Analyze the Conflict

Break down the definition, identify the specific conflicts.

3. Formulate Resolution Options

Propose solutions, identify impacts, make recommendation.

4. Approve and Implement Resolution

Record decision in glossary, ensure any impacts are handled.

Data Definition Conflict Resolution Process

## Step 4: Approve and Implement Resolution

Use your organization's data governance framework for making decisions around data.

> Who in the organization is required to make these changes happen?

> How will we track and measure the implementation and adoption of the solution?

This is a great opportunity for the data governance team to demonstrate its value.

You're facilitating a cross-functional group of people to solve a problem for the organization and seeing the benefit. Now that's powerful!
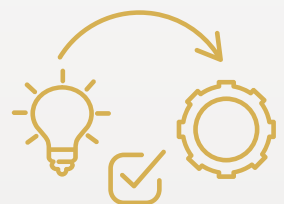
Business Terms and Definitions

There are two fundamental aspects of business data definitions to cover before we dive into the specific kinds of definition conflicts.

Sometimes the issue isn't the definition — *it's the name of the business term* — and here's why this can be problematic:

- People tend to name things *before* defining them. Then once they finish the definition, the name might not be quite right. Yet, the name sticks.

- The name and the definition are related, of course, but it can help to think of them as separate.

- Business term names don't have to be set in stone. Sometimes the answer isn't to change the definition but to adjust the name instead.

Data definitions are often named before they're defined, which can lead to problems.
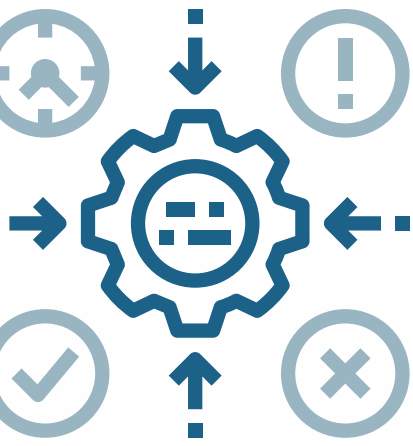
Business definitions can exist at two different levels:

- The "higher" level is that of a **business concept**. Business concepts are system-agnostic and can be true across many different functions and implementations.

- The "lower" level is that of a **data element** — an implementable data field, such as a column in a relational database, or a field in a file layout. Most data elements are related to a business concept.

A business concept might be implemented in different ways across different columns in many systems. For example, "Mailing Address" is a single business concept which is often captured in at least five data elements ("Address Line 1," "Address Line 2," "City," "State" and "ZIP Code").

Different data elements related to the same business concept can have different rules and formats.

When you get consensus on the business concept-level first, this can reduce the amount of churn in trying to deal with conflicts at the data element level.
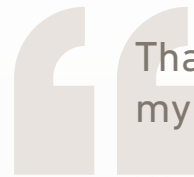
Business Glossary Basics: Different Contexts for the Same Data

This is a very common pattern — that various stakeholders have slightly different meanings for essentially the same thing. For example, what encompasses an "active" product is different for the product design team than it is for the marketing team.

What this might sound like in your organization:

> That definition is close, but for my department, we mean X.
>
> We used to do it that way, but we changed how we did that two years ago. Oh, has your department not changed?

Possible resolutions include consolidating across contexts — or keeping all context, but with clear naming and distinguishing the differences.

Here's an example from the world of financial services: "AUM" (assets under management). AUM is being calculated by the client reporting, finance and portfolio management areas, and all are getting different results.

First, define AUM as a core concept. Then define Client AUM specifically, as well as other contexts. Include the business calculations for each context.

Overloaded Data Fields

This pattern is often tied to having changed business processes without changing the data systems over time, squeezing new functionality into legacy databases. So a single field takes on more and more different meanings until there is no single definition possible.

What this might sound like in your organization:

> This field used to mean X, but two years ago we started to use it to mean Y.
>
> There are two types of this thing. Except for this other thing, which is sort of a hybrid of the other two. Oh, and except for this other weird condition that only happens twice a year.

Break definitions apart into separate terms, even if those terms only apply in some circumstances.

If a field is holding two different business concepts, be sure to document how to distinguish between the two in the data itself.

Document legacy and current definitions and determine how to tell the difference between the two in the data.
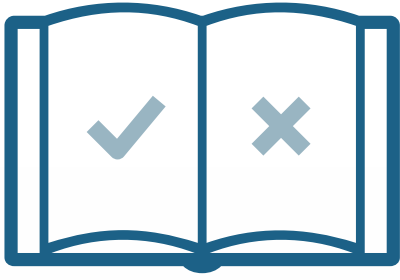
Data Naming Conflicts

This pattern is a bit less common than the others. To truly resolve it will require some organizational willpower to change the business vocabulary.

What this might sound like in your organization:

> "That's not what that term means **at all**. That term means X.

In this pattern, the same term name has two completely different definitions. Your options include changing the name of one of the terms or allowing both terms to use the same name but with a qualifier that clearly calls out the difference.

The most effective way to make such a change may be to update the screen label for the systems where that data is seen by users.

Completely changing the business vocabulary in your organization takes time and can be difficult. Patience can be a virtue.

Implementing Recommendations
Top Takeaways

- Resolve the conflict at the source(s), if at all possible.

- Update screen labels to reflect new data names.

- Update the originating business processes.

- Update the source data structures, if needed.

If changing the originating systems or processes is not possible, you could:

- Document and publish standard rules for parsing the data downstream.

- Transform the data in the data warehouse.

- Know that your organization is incurring data debt.

Go Forth and Resolve Data Conflicts

The resolution of data definition conflicts is a high-value activity for an organization.

Data governance is uniquely qualified to facilitate resolution activities. With effective communication and a "we can do this!" mindset, your data governance team can guide the organization through its conflicting data definitions, creating a common language for the enabling broad data understanding and literacy.

Author Gretchen Burnham

Gretchen Burnham, CDMP, is an experienced data management consultant with years of experience in business metadata management, data governance and stewardship and product master data management.

At First San Francisco Partners, Gretchen's focus is on delivering business value through data management and governance best practices.

info@firstsanfranciscopartners.com

www.firstsanfranciscopartners.com

1-888-612-9879

About FSFP

First San Francisco Partners is a leading business advisory and information management consultancy dedicated to helping companies leverage their data to improve strategic decision-making, reduce risk, create operational efficiencies and fuel unprecedented business success.

Our services span data governance, data quality, data architecture, master data management, analytics and big data.