

DATA "MINING" VERSUS DATA "OURS"ING

By Len Silverston

Published July 2007

How can we truly develop integrated information and information systems? Data management groups have struggled with the realization of this goal. Many of the people in organizations that I have witnessed have said (and I paraphrase):

"We need to understand and implement best practices from other organization who have succeeded."

"We need top level management commitment."

"We need business buy-in."

"We need to demonstrate value."

"We need proper incentives in place to motivate people and organizations to share and manage information on an enterprise wide basis."

It seems like these are all intelligent and appropriate statements that are important in developing more integrated information.

I want to share a hypothesis with you - one of the biggest issues in developing integrated information is data "mining" and what is really needed, at the core of this issue, is data "ours"ing. You may be thinking of data mining as in, "the process of automatically searching large volumes of data for patterns" ¹. I am not referring to this definition, but a new alternative definition for data mining meaning, "acting in a manner where this data is mine" As an example, John did not share company information as he was "data mining", declaring that the customer contact information was "his" own personal information that he collected and thus declared "this data in mine – not the company's or anyone else's".

Data "mining" is a root cause of not being able to integrate data – data silos come from people silos.

While there is much written on the need for data integration, it also seems that people and organizations have had a great deal of trouble in integrating data. I have given talks to thousands of data management professionals representing hundreds of organizations and when I ask the question, "Who has successfully integrated their data?", very few people claim huge successes. Why is this?

According to dictionary.com the definition of integrate is “to bring together or incorporate (parts) into a whole” and the definition of “disintegrate” it is “to separate into parts or lose intactness or solidness; break up; deteriorate”.

When people and/or departments within an enterprise act too often in a manner where “this data is mine” and thus are “data mining”, then people and organizations move towards separation and disintegration. When people move towards disintegration, data becomes disintegrated. Data silos come from people silos.

Consider one common example that occurs in many organizations. A sales person maintains data regarding their customer contacts. Perhaps there is an enterprise-wide customer contact database to facilitate sharing and synchronization to enable data consistency, cross-selling, collaboration, and more effective sales and service.

Enter data “mining”. The salesperson may think “I understand the benefits of an enterprise-wide customer database but this customer data is mine!”. I even brought in some of this data from a previous employer and this is how I make my livelihood for my family. I am doing a good thing by providing to my family and by protecting my personal customer contact information. If I share it with others who I don’t even know then they could mess it up, misinterpret it or misuse it.

This type of thinking makes a customer data integration effort very difficult and I believe that it is a core underlying issue that results in lost power for the enterprise. If we each think separately, our data will be separate. I have worked in many organizations where there are over 100 sources of inconsistent customer information and after many years of efforts, they have still been unable to integrate customer information across their enterprise.

This type of thinking happens in many different circumstances. Here are a couple, but this list could go on indefinitely.

- A department not wanting to share “their” department’s data into an enterprise data warehouse, master data management system, or enterprise data management effort. Sound familiar?
- Government agencies not wanting to share critical information about counterterrorism.

This illustrates that lives are sometimes at stake if we don’t appropriately share. For example, in the September 11th incident, there were two hijackers that were on the FBI’s most wanted list and another two hijackers who had expired visas and the airlines did not have access to this information. There may have been regulations, privacy, and security considerations that hindered sharing. But could we have shared this data more effectively and were there motivations of not wanting to share?

I want to stress that there are many good reasons and situations where people should not share data freely even in the above scenarios or, for example, in a human resources department that is responsible for securing sensitive information.

Be the change you want to see in the world

What can we do? Mahatma Gandhi said "Be the change you want to see in the world". If we each share more and consider data to be "ours" as opposed to "mine" more often, then I believe we will move towards data integration as opposed to data disintegration.

Here is an exercise that you can do and that you can share with others to do:

1. Identify one of your most valuable pieces of data.
2. Ask the question "Do you completely own it?"
3. How willing are you to share it and under what conditions will you share it?

The point of this exercise is to experience the feeling of data "mining" and get a clearer understanding of how data separation occurs, our underlying motivations, enabling us to act wisely and influence an appropriate level of data sharing.

So I invite you to conduct this exercise with me. What is one of the most valuable pieces of information that you own? Look at this data that you declare that you "own" and that is "your" data. For example, I have spent decades of my life working on a repository of re-usable data models that my company publishes and licenses called the Universal Data Model Repository. I would consider this one of the most valuable pieces of data that I own. What type of data is valuable to you? Is it personal data, intellectual property, corporate data, confidential data or some other type of data?

Do you completely own it? I am not saying that you don't - it may very well be your data. However, is this data really "yours" and not "ours"? Are you absolutely sure that this is "your" data or is it "our" data? Could it be that others own it, for example, your employer? Could it be that no one really owns it? For example, is the Universal Data Model Repository of template data models something that I, Len Silverston, own? Or is it something that my organization owns? Under intellectual property law, my organization has copyrights and trademarks, so there may be legal ownership. However, others have contributed many ideas to these template models for which I am so grateful. So at a deep level, are the ideas contributed by others owned by me? I must admit, that right now as I write, I feel a certain amount of discomfort (and data mining). Hey, I spent a great deal of my career on this repository of models! I own this. I have paid for this by paying others and with my time. So, data "mining" arises. I do have certain legal rights

regarding this data and at one level there is ownership by my company, Universal Data Models, LLC. On another level, I cannot personally claim ownership to many of the ideas and if I spend too much energy towards these being “my” ideas and “my” data, then they would not be as shared and integrated into the industry where they can be of greater benefit.

How willing are you to share “your” data or information? A thought that runs through my mind is that, ooh, the more I freely share “my” data the more I am at risk of losing power! For example, I may have to trust someone or some organization not to copy these models indiscriminately!

Some people have said to me, don’t publish your ideas so easy, it is a competitive advantage to offer these through my consulting practice. My personal experience has been that sharing these models through books and publications has been so gratifying to me in so many ways and has benefited so many people and organizations around the world. Sometimes I charge for sharing this work and sometimes I don’t. Data is an asset. So when to charge and when not? What do you need in order to share your valuable data?

The benefits of information sharing and data “ours”ing can be realized and cultures can be changed to produce extremely positive outcomes. For example, a Fortune Magazine article told the story of how Seagate Technologies returned to financial health based on a change in culture from being “divided into vertical silos” to “putting group genius to work”, a buzz phrase from Matt Taylor, who developed a collaboration process called DesignShop that was used by Seagate and many other firms to help change culture via Capgemini’s Accelerated Solution Environment.²

As a consultant, I have been involved in a few remarkable examples of the power of sharing information and data “ours”ing. For example, in a large financial services organization, the healthy environment that was developed enabled an enterprise wide system where there was appropriate sharing of client information throughout the enterprise, as well as with the clients, resulting in much better service levels.

Invitation to Share

Perhaps the next time someone is reluctant to share their data in a data warehouse, master data management or data management effort, we can better understand the behavior and influence more powerful and collaborative behavior. I invite us to share more and be an example of moving towards data integration versus data disintegration. I invite us to see data more as ours and less as “mine”. I invite us to move from people silos towards people integration and thus move towards data integration. I also invite us to share ideas about this topic with each other. I welcome any feedback from you and any thoughts regarding how we can move from data “mining” to data “ours”ing. Thank you.

¹ From **Wikipedia**

² Siekman, Philip, "Seagate's Three-Day Revolution" Fortune Industrial Management and Technology, Feb, 19th 2001.

Published in **Results Through Integrity** and **Gumption**