

TWENTY FIVE Golden Rules to tame the CDI Beast

Mani Kumar Manda
Rhapsody Technologies, Inc.
mmanda@rhaptech.com

Introduction

The C-level executives are puzzled and rightfully so, as to why CDI projects are so complex, time consuming and expensive when the subject matter is simply the "CUSTOMER" data. Achieving nirvana for a robust CDI solution is far fetched given the current maturity level of CDI/MDM technologies. It is in this context that this presentation makes an attempt to provide a direction with TWENTY FIVE golden rules, distilled with years of experience to clear the path for any CDI implementation.

The Challenge

Customer data is dynamic – it changes over a period of time with out any involvement. About 17% of the business names change over a period of year according to Dun & Bradstreet. US Population will grow to 394 Million by 2050 (Census Bureau). Per USPS, about 14% of the population moves every year. The facts such as these indicate the dynamic nature of customer data. Most of these changes often are not communicated by Customers and businesses are left alone to find their own solutions and approaches to obtain more accurate data, some times from sources other than the Customer itself.

The nature of customer data also leads into complexities – Customer Name for example, illustrates the complexities involved in keeping up with it. Over a period of time, there are multiple names associated with the Customer firms. Customer may have a Legal Name, Doing Business As (DBA) Name, Preferred Name, Former (Previous) Name, Phonetic Name, etc. Often some legal entities have multiple DBA Names. The name alone can lead into capturing the same customer multiple times with varying names. This problem is further compounded by the fact that the applications varying standards in terms of case, abbreviations etc., even if an enterprise level standard is developed.

This situation is not that much different for persons. There are over 36 formats for representing the Customer names globally based on the components and sequence of these components. The key components being First Name, Middle Name, Last Name, Name Prefixes and Name Suffixes. Often in some cultures there is no concept of Last Name and in some cultures, the women take their husbands first name as their last name after their marriage. In many countries some names have common nick names (Bill for William in US), etc., factors complicate the management of customer data.

Address is yet another item that poses its own set of challenges. The components of address vary by country and formatting of an address is dependant upon country and cultures. Often times, the technology solutions that have been historically capturing customer data had inadequate flexibility resulting in capturing unrelated information as part of addresses. At many firms, it has been observed that addresses often contain company names, person names, and other information that is not truly address information.

Contacts are changing even faster. People get married and change their names, get transferred internally within an organization, move across geography and so on. All of this result in changes to contact information including their addresses, phone numbers, emails, etc. Phone numbers change not only in the context of contacts but also for companies often triggered by changes made by governing agencies or telecom companies in terms of area codes, etc.

All of the above mentioned issues are only a partial list and every entity must deal with them as they constantly change without their involvement. It is often said that a perfect and 100% accurate customer data becomes obsolete at least 20% at the end of one year if it is not maintained. Such is the beastly nature of the customer data. With this situation, managing customer data requires special tools, technologies, methodologies and procedures to deal with it in an effective manner. The problems such as these lead to the concepts of Master Data Management (MDM), Customer Data Integration (CDI) technologies that are trying to address the challenges associated with managing the Customer data.

Evolution of CDI Technologies

In spite of using CDI Technologies for over 5 years now, the technologies haven't matured yet. The Vendors products continue to evolve and the whole industry is evolving with more maturity added to the products. Some vendors are focusing on domain specific products such as Oracle with CDH and UCM for Customer data management, PIM for Product data management, Hyperion DRM for managing Chart of Accounts, etc. Some other vendors are trying to champion the platform that facilitates the management of multiple domains of Master Data Management. Both types of Vendors believe their approach is better over others. There are merits and down sides to both approaches.

To complicate matters, large Vendors are acquiring smaller vendors and assimilating their products into their existing product base. This often results in loosing the support for the implemented products.

The traditional Software Development Life Cycle (SDLC) methodologies, while useful in implementing CDI solutions, falls short of addressing unique challenges associated in dealing with Master Data. The current methodologies are being modified; additional steps are being added for the SDLC life cycle to increase the success rate of CDI initiatives.

The Governance has been predominantly a procedural task and Vendors products falls short in providing technological solutions to address the needs for data governance.

“TWENTY FIVE” Golden Rules

While best practices in managing MDM and CDI implementations are still evolving, the following Golden Rules have been identified as important things that must be paid attention during any CDI implementations.

They are:

1. Executive (CXO) Sponsorship
2. Data Governance Program
3. Stakeholders with Enforcement Authority
4. Strong Data Stewardship
5. Trading Community Model (TCM)
6. Design Considerations
7. Conceptual Customer Model
8. Canonical Designs
9. Define and Enforce Data Security
10. Data Privacy
11. Regulatory Compliance
12. DPL Compliance
13. Implementation Approach
14. Data Quality Metrics
15. Hybrid Project Implementation Approach
16. Go beyond the Identity Data, in stages
17. Data Profiling
18. Enrich Data
19. Data Standardization
20. Validate Addresses
21. Classify, Classify, Classify!!!
22. Analytical Use Cases
23. Certify Data
24. Latest Versions
25. Search UI with robust Parameters

The Rules are listed in random order and are not listed based on their significance or importance.

Rule #1 → Executive (CXO) Sponsorship

The importance of the need to have an Executive Support and Sponsorship can not be stressed any further than it has already been communicated by every published article by experts in MDM domain, Research Analysts from Gartner, Forrester, etc. Yet, often times a good number of projects has only minimal Support or Sponsorship and many of them took off without having any support at all or with half-hearted support. The initial stages of the project move smoothly, until the project gets to a point where decisions have to be made about what data attributes to include in CDI implementation, how to standardize, what classifications to choose from what sources, etc. This is when the customer data politics intensifies and each and every CDI

initiative that did not have strong CXO support will start to flutter. A strong executive support is the only one that will get you through hard times that are bound to surface due to customer data politics.

A Business Case can be made for ROI based on benefits the firm can realize through increased sales and profits, better customer services, increased customer satisfaction (upside potential) as well as avoid risks by increasing the compliance with regulatory requirements, cost efficiencies (downside risks). It is easier to sell the Business Case when ROI is based hard benefits such as increased sales and profits rather than soft benefits such as improving customer satisfaction. Many benefits obtained through CDI initiative are indirect and it is hard to quantify these benefits. Under turbulent times, such as the current state of economy, it is essential to high light all of the hard benefits to get buy in from executives but soft and indirect benefits should be identified as well.

Rule #2 → Data Governance Program

It is important to have corporate level Data Governance Program. Often times, Customer is the first domain to be tackled by enterprises and Data Governance Programs focus primarily on the Customer Domain. However, it is important that even if the current initiative is customer domain related, the Data Governance Program should be at the MDM level that should span multiple domains. Creating the Data Governance Framework early on will increase the success of the project, especially when the Governance Council may have to step in to resolve the disputes/problems resulted due to customer data politics. Without having a functioning Data Governance Program, the CDI/MDM initiative is bound to fail. Establishing the Data Governance framework and structure should begin as soon as CDI/MDM initiative is launched. The Governance Framework should be structured to include at 3 distinct groups with varying authority and responsibilities. The first of which is Governance Council that must include stake holders that have responsibility and authority to enforce whatever is required from CDI/MDM initiatives. The second is a Data Steward team that is tasked with responsibility to address Data Quality, Data Management, Meta Data procedures and also should establish required Roles and Responsibilities and Standards and Procedures. When problems arise, the Data Steward team should tap into Data Governance Council. The third team includes Data Librarians, Data Management Specialists, IT staff that actually implements and manage the technology.

Rule #3 → Stakeholders with Enforcement Authority

As part of Data Governance, we have talked about the need for Governance Council that is staffed with stakeholders that have responsibility and enforcement authority. Unless the scope of a CDI initiative is very limited, the customer data politics are bound to happen in any large organization, especially when multiple business units are involved. The only way to address and resolve these problem is to build a Governance Council with Senior Executives and possible Mid-level Managers that have authority to enforce standards established by Data Steward team or help resolve conflicts when they arise. Without having a strong Data Governance Council, even with strong CXO support, it would become difficult to succeed.

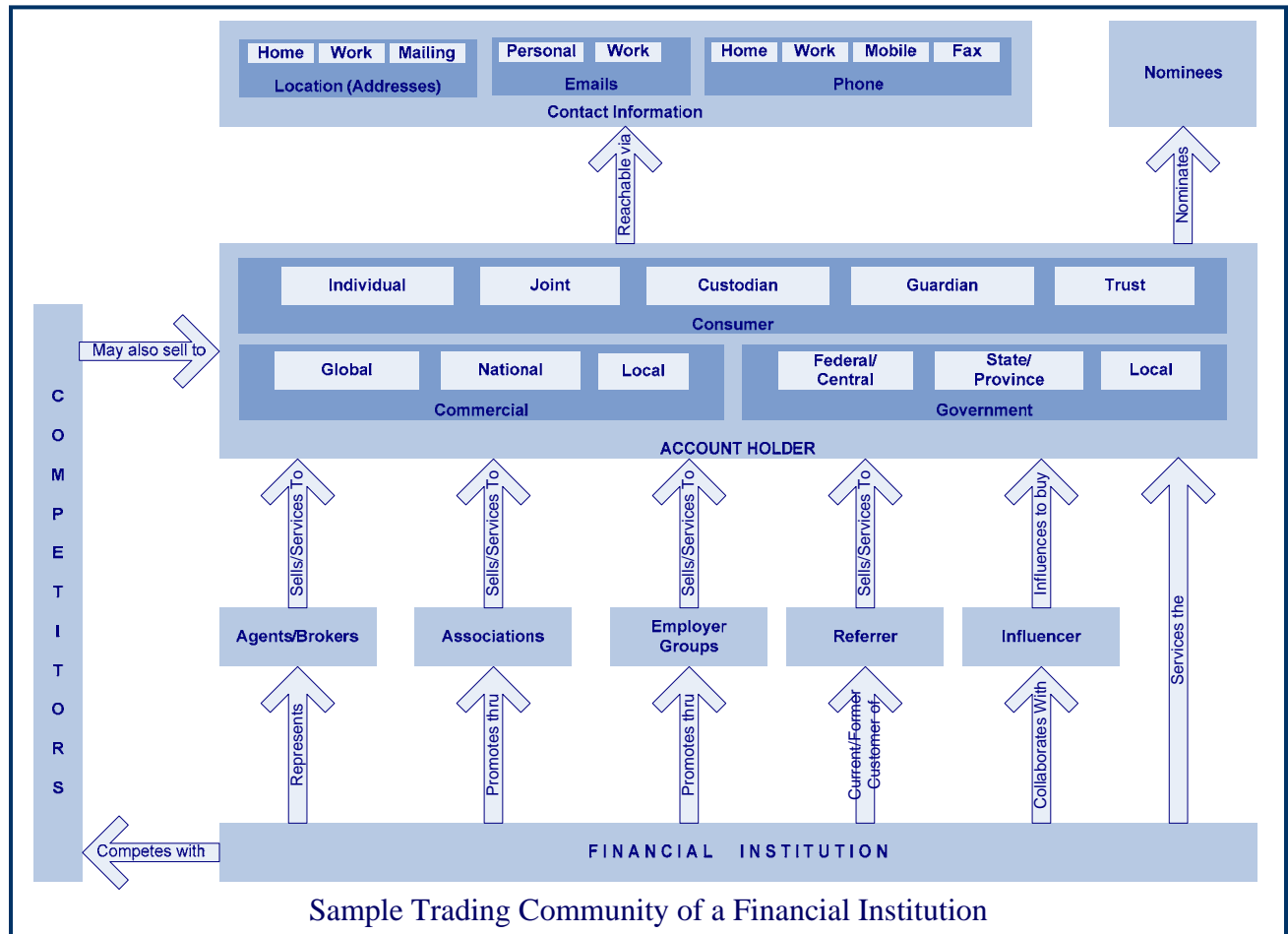
Rule #4 → Strong Data Stewardship

Data Stewards plays a key role in CDI/MDM initiatives. The Data Stewardship entails establishing not only Standards and Procedures, Roles and Responsibilities, but also addressing the Data Management aspects of the initiative. It is not essential that the same person address all of these aspects, rather depending upon size and complexity of the initiative, there will be a need for multiple Data Stewards who can split responsibilities. A Data Steward should be a person who understands the business, has strong Data Management capabilities and preferably has some exposure to the CDI/MDM technology chosen to be implemented. It is not possible to find a perfect Data Steward from the outside market who knows all three aspects. Hence we strongly recommend finding a Data Steward internally within the organization and training the person in the chosen technology for CDI/MDM initiative. Large enterprises often have an Architecture department that possibly may have persons who understand the business and have Data Management backgrounds. They can be targeted for the Data Steward Role. Also persons within an organization that have been historically tasked with managing customer data may be another set of targets. Rhapsody strongly recommends recruiting Data Stewards internally from the organization.

Rule #5 → Trading Community Model (TCM)

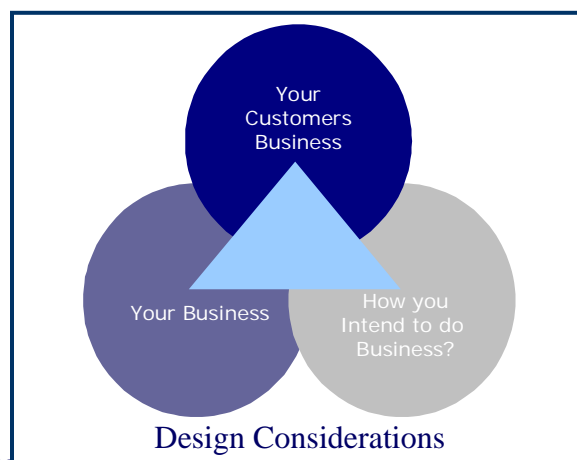
The Trading Community Model (TCM), a term coined by Rhapsody refers to a visual representation of people and entities that are involved in making commerce happen for your business and all the relationships between them. This representation can further highlight key classifications and attributes to better understand the business. The TCM picture forms the basis for making decisions in terms of what data to host in the CDI environment, how to classify the data and even if some of the data is available in any application. In our experience, it took 8 to 12 weeks to develop this diagram and often required interaction with 50 to 200 persons that span across the enterprise – various departments, business units and geographies to come up with the comprehensive portrait of the firm's business. Surprisingly, there was often no knowledge of previous attempts to document the business in this fashion. We also found that any where from 30 to 50 % of entities and relationships represented in the TCM diagram are not even stored in any database or application and some of this data possibly exists in documents (word, excel, power point), or lost in people's minds. Once TCM is developed, a decision has to be made what to capture in CDI environment vs. what not to capture. A cost benefit analysis is needed about especially if it pertains to data that were never captured.

The TCM diagram will also help in configuring the application at a later stage. A sample trading community model (scaled down version) for a financial institution is given below:



Rule #6 → Design Considerations

It is important to look at how business is conducted at your firm as well as at your customers. See the picture below:



The questions that are important are:

1. How does your Customer operate their business and how are they organized?
 - Has multiple Branches and Divisions
 - Global Player
 - Number of Business Units
 - Centralized vs. Decentralized
2. How do you want to do business with your Customer?
 - Treat Customer as one Entity vs. each branch or division or business unit as separate entity
 - At what level of the Customer do you want to track?
3. How is your business organized?
 - Global vs. Local
 - Number of business units (lines)
 - Profitability and Accountability tracking within the firm
 - How are Sales teams organized?
4. How much you want the D&B Data to influence?
 - Model along D&B Data (DUNS Number)
 - Capture customer entities regardless of biz relationship
 - How much to maintain?
5. Does Your Legacy System continue to exist?
 - Do Oracle Customer and Transactional Data need to be interfaced back to Legacy system?
 - Any other systems need Customer and Transactional Data besides Legacy system?

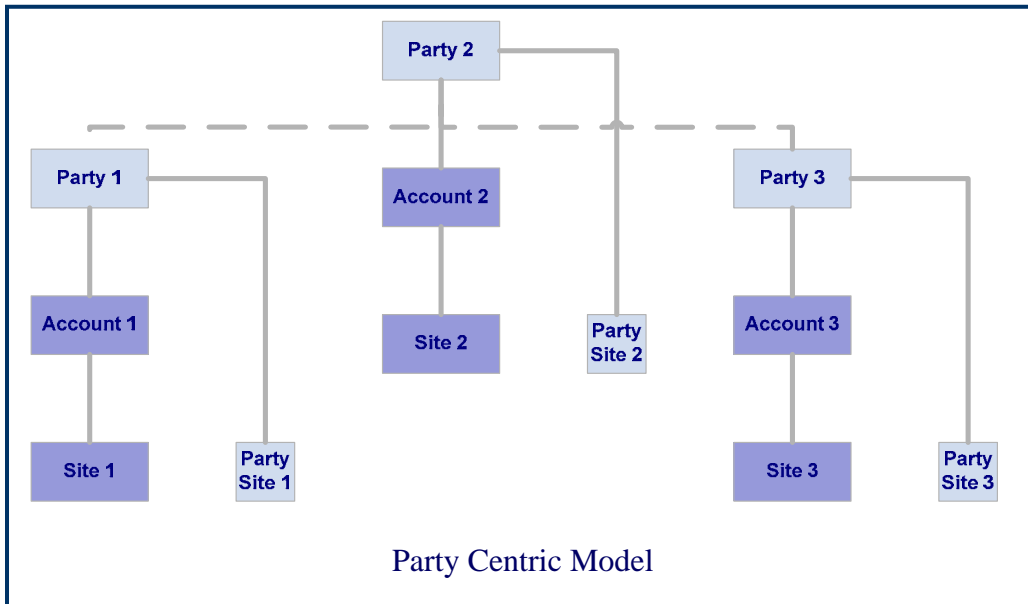
Answers to the first two questions will determine how many parties are created for each customer. And Answers to the second and third questions determines how many accounts need to be created. These aspects are further influenced by answers to the 4th and 5th questions.

Rule #7 → Conceptual Customer Model

With most of the CDI solutions separating the Identity data (party layer) and the business relationship data (Accounts) and with ability to relate customer locations to both layers, the customer modeling task becomes challenging and it is an important step to contemplate various models and their advantages and disadvantages. On one extreme is “Party Centric Model” and on the other extreme is “Site Centric Model”.

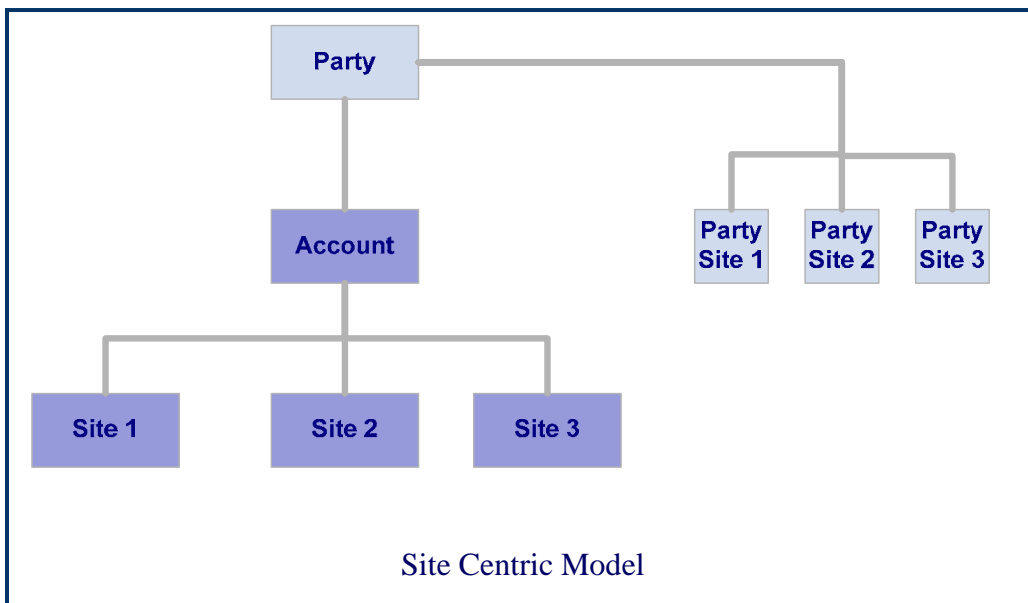
Party Centric Approach (aka Multiple Party Approach, Best Practices Model)

- is when a Party is created for every single site (location) of the Customer and party relationships are used to build the corporate hierarchy
- The focus is on creating parties and relating them to each other.

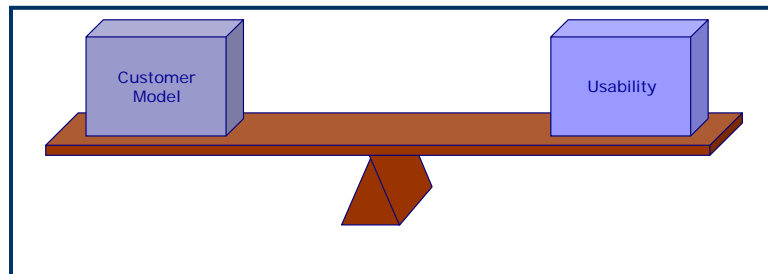


Site Centric Approach (aka Single Party Approach, AR Customer Model, and Legacy Customer Model)

- is when only one party is created for the Customer and all sites of the customer are created as party sites of this party
- The focus is on creating multiple sites



Most customers follow a model that is some where in between. From the perspective of CDI, a balance needs to be achieved between a model that is conducive for better data quality and the usability of the model.



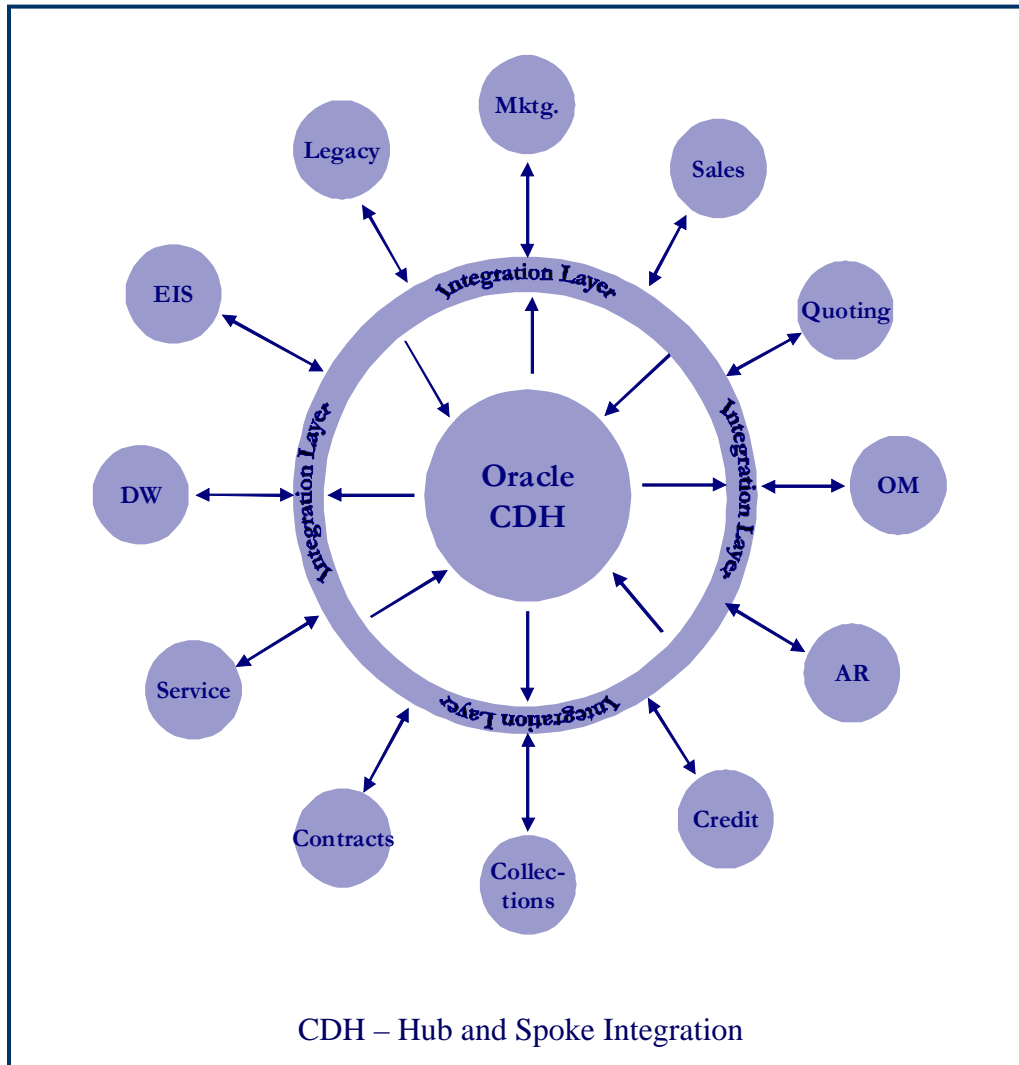
The usability factor gets further complicated and often conflicts with the objectives of data quality when a CDI initiative is targeting to use an existing operational application that might be catering to the needs in the areas of Sales, Service, Marketing, Finance, etc.

Rule #8 → Canonical Designs

The concept of CDI (or Customer Data Hub) inherently is conducive to adopt a Canonical definition for customer data integration between Hub and Spoke applications.

A canonical definition of an object is a horizontal language defined for that object (for example customer) which facilitates the building of integration between two applications using canonical definition as a medium. In this approach, source system data is mapped and translated into the canonical format by an integration object that is close to the source application and in turn translated by an integration object from the target application into a format conducive for target application.

The success factor of a CDI initiative is increased significantly if an open standards based canonical definition is adopted in the early stages of the project as a basis for the canonical definition of the customer domain. Open Applications Group (OAG) has developed a global open standard (known as OAGIS) for various objects including customer data. The AIA Foundation from Oracle uses standards established by OAG.



Once chosen, the open standards based canonical definition for customer data must further be reviewed in the context of enterprise needs and adjusted to accommodate business specific data elements to accommodate all business requirements. The functional and technical specifications for integrations should be developed based on the chosen canonical definition of the customer data.

Rule #9 → Define and Enforce Data Security

Data Security if not addressed can bring down the firm or cause severe damages in terms of public perception about the company. Recent regulations govern the various aspects of data security, including the security measures taken when data is transported in real-time or on an offline basis, reporting requirements if there any breaches, etc. Also there is a clear segregation

of duties warranted based on SOX, which require users be restricted to obtain only that functionality that is required based on their job roles.

In this context, whenever possible, one must utilize the native functionality of the technology that caters to fulfill this need. (Oracle CDH and eBusiness suite have a functionality called 'Data Sharing and Security (DSS)' that can be used to manage the access levels for users.). It is a not common practice to store credit card and bank account information in a CDI instance, however if the approach is to host this data, then the instance must comply with PCI standards for security of credit card information.

All regulatory requirements in all countries where your business operates must be identified earlier in the project during Analysis phase. As the functionality in the current vendor solutions is often limited or does not exist, identifying such a need at a later stage of the project, especially after a CDI solution from a specific vendor is chosen, is going to turn out expensive to customize if that solution doesn't provides a means to accomplish or meet the regulatory requirements.

Rule #10 → Data Privacy

Data Privacy pertains to protecting the Personally Identifiable Information (aka PII). PII is defined by Wikipedia as "In Information Security and Privacy perspective, a Personally Identifiable Information can be defined as any piece of information which can potentially be used to uniquely identify, contact, or locate a single person."

The examples of PII include Personal Information such as Full Name, Date of Birth (DOB), Sexual Orientation, Race, Health Information; National Identification Number (for example Social Security Number – SSN in United States); Contact Information such as Phone Numbers, E-Mail addresses, Street Addresses; Images, Financial Account Information (Credit Cards, Bank Accounts), etc.

Often times an element such as a first name that is common may not identify the person (non-identifying attribute) and the same is true for data elements such as location (another non-identifying attribute) when it stands alone. However, it is possible that when these two non-identifying attributes are revealed together (i.e., a common first name when combined with location) under certain circumstances this can uniquely identify a person and hence care must be taken in implementing policies and procedures as well as technology solutions in terms of compliance with privacy laws.

Here is a partial list¹ of some of the Privacy related Laws:

- United States
 - Federal Laws
 - Privacy Act of 2005
 - Attempts to strictly limit the display, purchase or sale of PII without the person's consent
 - Anti-Phishing Act of 2005
 - To prevent the acquiring of PII through phishing.
 - HIPAA – Health Insurance Portability and Accountability Act
 - Patient Information
 - Social Security Number Protect Act 2005
 - Limit the distribution of SSN
 - Identity Theft Protect Act 2005
 - Limit the distribution of SSN
 - Information Protection and Security Act of 2005
 - Consumer Privacy Protection Act of 2005
 - Wireless 411 Privacy Act
 - US 'Safe Harbor' Rules (EU Harmonization)
 - Title 18 of United States Code, Section 1028d(7)
 - State Laws
 - California – OPPA – Online Privacy Protection Act of 2003

For more comprehensive list of Privacy laws in other countries contact the Author.

Rule #11 → Regulatory Compliance

Regulatory Compliance refers to systems and departments at corporations and public agencies to ensure that personnel are aware of and take steps to comply with relevant laws and regulations. In general, compliance means conforming to a specification, standard or law that has been clearly defined.

There are Regulations that impact all businesses located in certain geography or fit a specific profile and there are regulations that impact businesses in certain industries. For example, in the United States, Sarbanes-Oxley Act (aka SOX) has been developed in 2002 in response to corporate scandals which defined significant tighter personal responsibility of corporate top management for the accuracy of reported financial statements. According to Section 404 of SOX, Organizations are required to take an active interest in the accuracy, consistency and timeliness of the data. Lately, the combination of above mentioned requirements of (two statements) SOX have been the drivers for many MDM initiatives.

In the United States, the Security and Exchange Commission (SEC) requires public companies to report risk exposure for bottom-line and top-line numbers in their quarterly reports filed with

¹ This is a partial list taken from Rhapsody's "CDH Functional Fundamentals" course material.

SEC. Any public firm can only comply with provision of the SEC regulations only when they capture and actively maintain complete corporate hierarchies of their customers. Without which, the numbers stated for risk exposure has a potential for under statement.

Gramm-Leach-Bliley Act (aka GLBA) is a regulation that governs financial institutions. This Act requires all financial institutions in United States to put policies and procedures that govern the collection, disclosure, and protection of consumers' non-public personal information; or personally identifiable information. This law also mandates, whether a financial institution discloses nonpublic information or not, there must be a policy in place to protect the information from foreseeable threats in security and data integrity.

The Federal Information Security Management Act of 2002 (aka FISMA) was meant to bolster computer and network security within the Federal Government and affiliated parties (such as government contractors) by mandating yearly audits.

The Health Insurance Portability and Accountability Act (aka HIPAA) requires all healthcare organizations that submit insurance claims to CMS to provide National Provider Identification (NPI) Number.

It is important to know all of the regulations associated with mastering data in a single instance across the enterprise. Due to this consolidation of data, the need to secure data and comply with various laws across different geographic jurisdictions is an essential part of the implementation.

Rule #12 → DPL Compliance

Denied Party List (aka DPL) is a list of persons and entities with whom no one should be doing business with.

According to the list² published by the Bureau of Industry and Security, the following persons or organizations (a partial list) with whom one must not transact with. They are:

ZHAN GAO 12731 MILL HEIGHTS, HERNDON, US	12/14/2005	03/05/2014	standard
Appropriate <i>Federal Register</i> Citations: 70 F.R. 75447 12/20/05			
ZHENG ZHENG C/O UNIVISION TECHNOLOGY, 764 VIOLET CIRCLE, NAPERVILLE, US	09/07/2006	09/07/2016	non standard
Appropriate <i>Federal Register</i> Citations: 71 F.R. 54467 9/15/06			
ZHONGDA JIN 1895 DOBBIN DRIVE, SUITE B, SAN JOSE, US	07/31/2001	07/31/2026	standard
Appropriate <i>Federal Register</i> Citations: 66 F.R. 40971 8/6/01			

With out having a well defined process, a business can not stop from transacting with the persons or entities listed above, if they try to purchase products or services from a business, a serious violation of US Department of Commerce.

² As per the list published as of 12th March 2008 at <http://www.bis.doc.gov/dpl/thedeniallist.asp> by Bureau of Industry and Security, US Department of Commerce.

Here is a partial list³ of federal agencies in United States that publishes such lists. Any business operating in United States are subject to all of these lists.

- Department of Treasury
 - Specially Designated Nationals and Blocked Persons (SDN)
 - Narcotics Trafficking Sanctions
 - Anti-Terrorism Sanctions
 - Non-proliferation Sanctions
- Bureau of Industry and Security
 - Denied Persons List
 - The Entities List
- FBI
 - Most wanted list
 - Issued Watch List
- US Department of State
 - Debarred Parties List
 - Designated Terrorist Organizations

Rule #13 → Implementation Approach

A dilemma often contemplated by many firms is the question of why not combine MDM initiatives in their existing operational instance? Why should they implement an MDM solution as a brand new application, especially if the operational application has the functionality to cater the need of MDM of the firm?

The answer is two fold. First of all, one these operational applications have been in use for years yet they have contributed to and/or failed to address the data quality problem which is typically addressed by MDM/CDI initiatives. Not all operational applications have the capability to address the tools and technologies offered by Vendors as part of MDM products. There are certainly exceptions – for example Oracle CDH and Oracle PIM Hub are part of Oracle eBusiness Suite of applications which are typically used by many large enterprises.

Another aspect is the problems that surface when these two initiatives are combined into one application although certainly some benefits will also emerge. The benefits are the reduced licensing costs as well as operating costs; however, the downside risks are much higher in this approach, since the combination of the MDM solution into an operational application often times prevent the application of patches that are required by MDM side and the implementation/upgrade/patch cycles are forced to cater to for the needs of operational requirements. At the end of the day, it is important to fulfill the orders, send invoices, collect and apply cash in an operation application. Nothing can stop that, not even the holiest MDM.

³ This is a partial list taken from Rhapsody's "CDH Functional Fundamentals" course material.

To complicate this further, the conceptual customer models (Discussed in Rule #7 above) required to maintain the data quality, a primary objective of MDM/CDI initiatives may conflict with the conceptual customer model required by operational application. This fact is clearly evident in Oracle eBusiness Suite of applications, which can also be used as a CDH instance or PIM Hub instance in addition to being an operational application.

What happens when there are multiple operational environments addressing different business units or geographies? Will only one of these operational environments be chosen to serve as a basis for MDM?

It is important to understand the ramifications of an approach for maintaining a separate environment for MDM vs. using an existing operational environment.

Rule #14 → Data Quality Metrics

The Data Quality management initiatives often take birth in reaction to acute problems traced to data quality problems (caused by failure of improper or lack of required policies and procedures, governance, or tools and technology) which are adversely affecting an organization. In some organizations that are more thoughtful, the data quality initiatives take birth in response to business case made for data quality improvement as a result of assessing how poor data quality has an impact on the business.

The data quality once initiated is an on-going effort. It can not be stopped; otherwise all the resources and time put in to the effort of increasing the data quality go down the drain. It is necessary that organizations addressing data quality should form the baseline of the current state of data quality so that one can identify the progress made with the effort expended in data quality and see if the efforts are going in right direction.

The other objective is to measure and publish the data quality metrics to instill the confidence in that data to business users. For all of this to happen, data quality metrics that are meaningful for the business must be defined, captured as baseline and periodically recalculated to correlate with the improvements made in data quality. Without having data quality metrics, it is very difficult to validate the effectiveness of the data quality programs that are often part of CDI/MDM initiatives.

All CDI/MDM initiatives must include the time, resources and money for defining Data Quality Metrics, measurement/monitoring procedures and reporting them to the business. This will facilitate the value proposition and the criticality of the CDI/MDM initiative to the business. This can also become a ROI factor for justifying the multi-million dollar expenditures that are often required for CDI/MDM projects.

Rule #15 → Hybrid Project Implementation Methodology

In large global organizations, it is not uncommon to have customer data in hundreds of applications. Imagine the challenge associated in bringing customer data (or any other MDM domain data) into a single CDI/MDM application such that the customer data is cleansed, deduped, optionally enriched forming the basis to provide a panoramic 360⁰ view of the Customer. Projects of this size often take years to complete and require tens of millions of dollars to achieve the nirvana of providing quality customer data that is accurate, dependable and available on a timely basis.

The challenge is holding on to the executive support and sponsorship as well as the necessary user participation during the implementation of a global enterprise wide CDI/MDM application that solves the impending problems.

Given the complexity involved in such projects, Rhapsody strongly recommends to take a hybrid⁴ project approach that combines the benefits of two project management methodologies – Waterfall Methodology, Spiral Methodology. Per Hybrid approach, the requirements and high-level design are performed at the global level (as per Waterfall Methodology), however the implementation is done in multiple and meaningful phases (as per Spiral Methodology). Each implementation (phase) built upon the earlier implementation (phase) of the project capitalizes on the functionality that was already implemented. Though this approach takes more time than going with a completely phased approach (pure Spiral Methodology), it takes significantly less time to provide benefits to the business though only parts of the business may benefit. Due to the benefits realized by parts of the business, the CDI/MDM initiatives will continue to hold the support of both Executives as well as Business Users thus increasing the success factor of the whole initiative.

Rule #16 → Go beyond the Identity Data, in Stages

A single view of the customer can only be accomplished by using the portions or all of the identity data of a party – some times by itself and sometimes in conjunction with the data associated with the business relationship. Hence a case can be made that the initial objectives of a CDI/MDM implementation are confined to the critical identity data that helps achieve a single view of the customer across an enterprise that spans across multiple business units, geographies, and applications. For some organizations, this approach may not work where an organization is trying to build operational applications such as web stores that need to know the business relationship information (Accounts) and are expecting that the CDI/MDM applications will provide this information requiring the CDI/MDM initiatives to go well and beyond the identity data.

On the path to maturity, an CDI/MDM initiative can bring in Accounts, enrichment data from content providers (such as number of beds, specialties, number of physicians, number of nurse practitioners from a third party provider like Verispan that can not be easily obtained during

⁴ This is an excerpt taken from the Rhapsody's Implementation methodology for CDI projects

normal course of doing business), segmentation data, automated validation of the existing data through content subscriptions, etc. However, doing all at once is going to increase the risk potential for failure. Another touch area in case of CDI initiatives is the ability as well as the need to build corporate hierarchies for all customer-base. Building corporate hierarchies incorporating the existing customer data is time consuming, and may even become resource intensive. May be it is prudent to build corporate hierarchies only for a top customers that bring in biggest portion of sales during the initial stage of the implementation. As the initiatives progresses, build corporate hierarchies for the entire customer base.

Rule #17 → Data Profiling

In our experience, we notice that many organizations begin CDI/MDM initiatives without actually looking at the profile of current customer data. The belief is that they are aware of the state of data quality of customer data and hence the initiation of current initiative. So the thinking goes that there is no need to further spend time and resources on profiling the data.

At one company that is embarked on CDI journey, a design decision was made to create the applicant information as person parties. This is based on a fact that they do business with individuals, or groups of people. What is not evident is that some of the applicants are Trusts and are possibly even organizations that are legally incorporated. Also, over a period of time the business relationship with this client can be assigned by an applicant to a Trust. A Trust is neither a person nor a legal entity. What if the trust is established such that at some point in time, it is transferred to an organization that is legally incorporated? Over a period of time the type of applicant has changed, however the design calls for creating the applicant as person party. The percentage of this scenario is very small but not identified by business users as well as business analysts. It will be too late to identify the problem during or after converting the entire data per the designed approach. Only data profiling can reveal the patterns of such behavior in the customer data that can form the basis for correct design of the conceptual customer model.

Another aspect of the data profiling is to identify the data elements that are relatively good and meaningful and can be brought over to CDI/MDM as oppose to data elements that are not so good or inconsistent and provides little business value, unless corrected.

The data profiling also identifies the level of standardization required during conversion for those data elements that must be brought into CDI/MDM application.

Data Profiling must be done in the early stages of the initiatives, especially prior to the design phase of the project to reap the benefits as well as prevent uncalled design decisions.

Rule #18 → Enrich Data

At a large global organization, a slide was presented for an audience who are Senior VPs, Business Unit Heads as well as Departmental Heads with a list of about 50 companies. The

audiences were asked if they have identified any of the companies listed on the slide. The answer was unequivocal no. When it was revealed that the list of companies on the slide actually belongs to the corporate hierarchy of the organization whose employees are in the audience, the audiences were shocked. The point is that if the employees of a large global organization could not identify the affiliates of their employer's corporate hierarchy, how can one expect a person tasked with entering customer data to identify corporate hierarchies? Here is where the content providers such as D&B come into play as enrichment providers.

As discussed before, for a manufacturing organization that builds diagnostic equipment primarily dealing with healthcare industry, information such as number of beds, types of specialties, specialties of physicians and their numbers of a hospital is important. Yet this information can not be easily gathered at the time of establishing a relationship with the customer. This is where content providers such as Verispan come in to play. There are many content providers serving different industries and you may already know about them. Some times it is even beneficial to source content from multiple vendors – for example corporate hierarchies from D&B, industry specific information from relevant vendors.

The content providers spent millions of dollars, huge number of resources to actively maintain the content they are providing. Regardless of the incompleteness of this data, it is much better to rely on content providers than spend time, resources and money to get to the same quality level of the data that can be purchased at a fraction of the cost from content providers.

Rule #19 → Data Standardization

Data standardization refers to a process of transforming the data in conformance with established standards and best practices for data quality. This is often done manually at many organizations, but the task is impossible to keep up with, at a global level if continued to be done manually. There are tools and technologies available that can help in standardizing the customer data. Data standardization increases the possibility of identifying duplicates thus increases the quality of data when the identified duplicates are resolved.

In case of Oracle, Oracle has a built in mechanism for data standardization only for the purposes of identifying duplicates, but the original data was never changed. However, it is not that difficult to use this existing framework to standardize the customer data as part of conversion.

Rhapsody strongly recommends performing the data standardization as part of the conversion or even prior to converting the customer data into a CDI/MDM instance.

Rule #20 → Validate Addresses

Address validation is a part validation of mailability of an address and part standardization of the address data. Many third party address validation service providers such as Trillium (owned by Harte-Hanks), First Logic (now owned by Business Objects), Group 1 Software.

Rhapsody strongly recommends the using of one of these vendor solutions and also suggests adopting a vendor solution that is pre-integrated with the chosen CDI/MDM application or technology platform.

Rule #21 → Classify, Classify, Classify!!

One of the causes for head-aches during CDI/MDM implementations is that the customer data politics resulting from the lobbying effort by individuals from various departments or business units to include their way of classifying the data in CDI/MDM applications. However, many CDI/MDM solutions have an ability to capture unlimited set of classifications, for example Oracle CDH.

Another classic problem often found during customer data profiling is that a single classification often turns out to be a set of classifications. At one firm, customers were classified based on following scheme – Large Customer, Minority Owned, Women Owned, etc. In reality, it is possible that a specific customer can be large customer which is owned by a minority woman satisfying all three different values of the classification. Analyze all classifications no matter how important they are to analyze for such scenarios.

At some implementations it has been observed that classifications are applied at incorrect entity level -- a common problem often noticed in Oracle eBusiness Suite implementations is the common classification of the customer based on industry that was applied to accounts (business relationships) rather than parties (the entities). So not only the classification should be important for business, it should be captured at appropriate level.

Rule #22 → Analytical Use Cases

Of late, analytical applications are driving the need for CDI/MDM solutions. Even when such is not the case, Rhapsody recommends to taken into consideration of analytical use cases as base requirements in designing a solution. For many organizations this may be as simple as capturing segmentation information form business intelligence applications.

Rule #23 → Certify Data

Not all but many of the CDI/MDM solutions provide functionality that lets you mark the validated customer data through certifications. The display of such certification information is very useful for the users when presented in search results.

Rule #24 → Latest Versions

The CDI/MDM technologies are still evolving and haven't reached maturity level where pace of changes has been slowed. Due to this vendors of CDI/MDM solutions often releases enhancement patches as well as bug fixes frequently. The new release of applications often provides new functionality that is worth upgrading for. (For example Oracle CDH R12 provides several enhancements that justify the upgrade, if you are on R11i version of Oracle CDH).

As a best practice, Rhapsody recommends the application of all latest patches as soon as the software from the Vendor is installed to reduce the disruptions at a later stage of the project.

Rule #25 → Search UI with Robust Parameters

One of the key benefits of CDI/MDM applications is to bring the customer data from multiple systems and de-dupe it. The resulting merged customer data now has Source System References (SSR) pointing to source system it hails from. It is conducive for the users to be able to identify the customer using any of the source system references in addition to other identifying attributes. Often times, the user may even want to filter the result set by source system. Vendor solutions such as Oracle CDH provides a mechanism through built-in DQM that lets you provide means to create search parameters that suits a specific user or business role.

Capitalize on such functionality from get going to provide meaningful results to the business, when CDI/MDM applications are opened up for selected list of business users for the purpose of querying.

Build several search profiles that suits differing needs of various users.

About the Author

Mani Kumar Manda is Founder of Rhapsody Technologies, Inc. and OAUG Customer Data Management Special Interest Group (OAUG CDM SIG). Mani has been a consultant for his entire career to mid-market and large customers and has been a recognized speaker in the areas of MDM and CDI with numerous presentations at OAUG, Open World, MDM Summit and several local GEO and SIG groups. Mani has developed a RHYTHM Methodology that significantly increases the success factor for any MDM/CDI implementations.

About Rhapsody Technologies, Inc.

Rhapsody Technologies, a pioneer in advisory, auditing, and implementation services for Master Data Management (MDM), Customer Data Integration (CDI), ERP and CRM solutions for mid to large customers. Over the years, Rhapsody had provided consulting services to Fortune companies such as ADP, GE, DoubleClick, Hewlett Packard, Motorola, etc., firms in their pursuit to solve business problems using enterprise applications such as Oracle CDH, Oracle eBusiness Suite, and PeopleSoft Applications. In addition to Consulting, advisory and auditing services for CDI, ERP and CRM, Rhapsody also provides 'MDM Awareness Seminar for CXOs', 'CDH Functional Fundamentals' training, and training for Data Stewards and Data Librarians.

Web sites

OAUG CDM SIG
Rhapsody Technologies, Inc.
For Latest Version of this Presentation

<http://groups.yahoo.com/group/cdmsig>
<http://www.rhaptech.com>
<http://www.rhaptech.com/resources.html>