Information As Product

How to Deliver the Right Information, To the Right Person, At the Right Time.

Michael C. Daconta

Former Metadata Program Manager,
Dept. of Homeland Security
and
Chief, Enterprise Data Management
Oberon Associates, Inc.

Outskirts Press, Inc.
Denver, Colorado

Table of Contents

[Auto-generated]

List of Figures

List of Tables

Introduction

"The customer is the most important part of the production line"[1]
- Dr. W. Edwards Deming, Out of the Crisis

Whether for homeland security, regulatory compliance or patient medical records; the role of information in our society has reached a critical juncture. In the past five years, I have delved deeply into data design, enterprise data management and information sharing - along the way I have uncovered flaws and gained new insights into achieving information superiority. As a result of that process, this book provides practical and concrete techniques for organizations to reach higher levels of agility, effectiveness and efficiency by treating information just as they do tangible products. This is the central idea of this book:

Manufacture consumer-centric information products just as we do physical products.

Only after truly understanding the above, can you hope to deliver the right information to the right person at the right time; for those are applications of the more basic information production process. In other words, before you can deliver information, you must understand how information is constructed from data and metadata. Additionally, your organization must be willing to shift key processes from a producer/data orientation to a consumer/information orientation.

An information product is an enhanced, virtual analogue to a physical product just like the can of food in a grocery store as depicted in Figure 1.

[1] Deming, W. Edwards; Out of the Crisis; Published 1982 by the Massachusetts Institute of Technology; Pg 174.

Figure 1 A Physical Product with Packaging and Labeling[2]

A physical product, like the can of food in Figure 1, is produced, packaged, labeled and advertised to consumers to satisfy a specific need. Our data is analogous to the food inside the can – however, we don't go to the store and scoop out raw food from unmarked bins? So, why do we get our data that way, in a raw, unprocessed form?

The central point of this book is the shift away from raw data towards a concrete specification for information products. This understanding came about over many months in working with numerous data architects from across the federal

[2] Stock photos from imagebase with permission. http://imagebase.davidniblack.com/license.htm

government to define the Federal Enterprise Architecture's Data Reference Model[3]. As a group we exposed numerous fallacies, misunderstandings and flaws about data, metadata and information exchanges. You may be amazed to discover how little we know about creating information instead of just shuffling around bits of data. Our inability to reliably distinguish data from information breeds a lack of trust in IT from the business managers. In turn, such lack of trust holds back the progress and promise in the information revolution by turning phrases like "the paperless office" into jokes. The solution offered in this book is to define a repeatable process to create, assemble and deliver information products. This requires understanding how you "productize" data to create information. How you package it, label it and store it.

Organization of this Book

This book is the first in a multi-volume series on information production. This volume lays the foundation for all other volumes by defining the basic components and processes of information production. This book is organized into six chapters and book-ended with an introduction (this section) and a conclusion. Each chapter supports the others in a building block approach:

- Chapter One: Why Information Production – this chapter provides the core justification for formalizing the information production process. It also introduces and explains the 4C's of an Information Product: Consumer, Context, Catalog and Content.

- Chapter Two: Data and Information – this chapter clearly distinguishes data from information.

[3] Federal Enterprise Architecture Data Reference Model (DRM) V2.0; Published November 2005; Available at http://www.whitehouse.gov/omb/egov/a-5-drm.html

Additionally it distinguishes both from knowledge and wisdom. It closes by distinguishing information from information products.

- Chapter Three: The Information Product – this is the central chapter of the book that defines an information product both conceptually and technically. To do this we follow a physical product analogy. In line with that analogy we examine the importance of consumer profiles similar to those used in marketing physical products. The chapter also introduces a new concept of "Information MVC" and details its role in implementing information products.

- Chapter Four: The Information Catalog – this chapter describes how to create and populate a metadata catalog. Before exploring the catalog implementation, the concept of metadata is demystified. A new, unambiguous definition for metadata is constructed and explained in detail. The chapter closes by exploring the role of the metadata catalog in the information production process.

- Chapter Five: The Information Production Process – this chapter introduces the information supply chain as analogous to the physical product supply chain. Each value activity in the value chain is explored. The chapter closes with a case study on a real-world information product: the electronic mortgage.

- Chapter Six: Strategic Information Delivery – this chapter focuses on the application of information products to achieve "rightness" along three specific axes: information, person and time. To achieve this I

devised a “rightness pattern” composed of three interlocking components.

- Conclusion - this section completes the book by envisioning the result of a robust information production process: individual empowerment via real-time relevance. The chapter closes with a summary of the key “take-away” points

- Appendix A: Effective Metadata Design – this appendix is provided to enable technical managers to differentiate between good and bad metadata designs. It walks through seven methods for describing data and demonstrates each one via a case study. The chapter includes a detailed examination of formal taxonomies adapted from an article I wrote for xml.com. A successful metadata catalog will use all of the techniques described in this section.

- Appendix B: The Evolution of Data – this appendix describes the history and future of data representation and is adapted from an article I wrote for Enterprise Architect magazine. While quite technical in some areas, I have taken pains to expand and clarify the material for technical managers. The appendix also serves as an interesting indicator of the evolution of my own thought into data, information and knowledge representation.

After reading this book, it is my hope that senior executives, technical managers and technical implementers have a deep understanding of information and the tools necessary to reliably produce it. Treating information as tangible products is the linchpin to information superiority: moving beyond

information platitudes to the science of targeted information production and delivery.

Acknowledgements

The writing of this book has been an arduous process of discovery, research and experimentation. The title, theme and outline of the book have changed several times as I examined different aspects of the problem. In essence, I stumbled about as the proverbial blind-man examining the elephant. And, as this introduction has made clear, the research began before the writing. I am a stickler for having the proper prerequisites before taking on a job or a major project. In other words, I make it a priority to feel qualified to write before I write. My experience over the past five years is the source of all lessons learned in this book. Thus, the list of people I need to thank spans both the time of this writing and the time gaining the experience in which to write about. The list of people I am indebted to is long. If I perchance, forget to mention someone who has helped me along the way, I sincerely apologize. And now I'd like to sincerely thank the following people:

- My dear wife Lynne who has tirelessly supported me in every way possible. To join the Department of Homeland Security, I again uprooted the family from Arizona back to Virginia. Lynne and the kids - CJ, Gregory and Samantha - accepted this and worked hard to adjust rapidly. They are my strength, my joy, my love.

- My brother Joseph Craig Daconta, who always encourages and challenges me in our weekly discussions. And of course the rest of my brothers and sisters, parents, in-laws and relatives who are always rooting for me.

- Mary McCaffery, a colleague and friend at the Environmental Protection Agency, who has supported me as a professional sounding board, advisor and proofreader. Mary is a class act.

- My former colleagues at DHS especially Lee Holcomb (my boss), Martin Smith, Tarrazzia Martin, Merv Leavitt and Steve Cooper (former CIO of DHS). And to my current DHS customers who had the vision to launch an Enterprise Data Management program at TSA – Joe Peters, Chris Allen, Mike Karas, Kevin Lawson and Paul Worsham.

- My former colleagues in other federal agencies whom I worked with on the Federal Enterprise Architecture Data Reference Model and the National Information Exchange Model: first, special thanks to Jim Feagans, my chief collaborator on NIEM whom I am now proud to call my friend. To other federal colleagues to including Kim Nelson, Suzanne Acar, Bryan Aucoin, Richard Burk, Patrick McCreary, Clay Robinson, Glenda Hayes, Brand Niemann and Susan Turnbull.

- My colleagues at Oberon Associates, Inc. headquarters including Dave Young, Jodi Johnson, Everett Nelson, Anthony Iasso, Dale Johnson, Randy Brooks, John Healy, John Hibbert, Gail McGraw, Ken Diller, Bob Cotter, Maria and John Shugars, Sandy Corbett, and Ericka and Vanessa Speis. Special thanks to Dave and Jodi for actively supporting my voracious interests and writing schedule.

- My colleagues at Government Computer News whom I write a monthly column for; especially, Wyatt Kash,

and Joab Jackson. And to my other colleagues in the media to include Martha McGrath, Jason Miller and David Perera.

- My Enterprise Data Management Team at TSA to include David Dutton, Kelly Fahey, Michelle Stern, Cole Wilson, Peter Rush, Huma Munir, Alex Gandy, Van Pham, and Jon Renaut.

- My other professional friends to include Danny Proko, Kevin T. Smith, Leo Obrst, Charlie Sowell, Dwayne Jackson, Nicole Davis-Helmantoler, Darrell Gaskins, Craig Tanner, Mark Oliphant, Michael Tiemann and Felix Rausch.

Again, even though the list above is long, I am sure to be forgetting many people to whom I apologize. Thank you all for your assistance and support. Of course, the list is not complete without adding you, my reader, and to the previous readers of my other books and articles. Thank you for listening.

Feedback

This book is the start of an important dialogue on rigorous information production techniques that effectively support consumers. In light of the dot-com bubble bursting, we must be both sensitive to cost and clear of our benefits. I believe this book achieves that by delving beneath the vision to practical, concrete implementation approaches. I am very interested in your opinions on these important subjects and thus offer you a variety of methods to go beyond the contents of this book:

- Blog - this web log, located at http://iproduct.blogspot.com, will be the primary source of collaboration on the book. All readers will be able to post

commentary on all the sections of the book. The blog will work in conjunction with the website for more permanent and reference information on the book.

- Website - a set of pages on my personal website http://www.daconta.us/books/infoasproduct.html will cover general information about the book, errata, and projects related to the book.

- Email – you are welcome to contact me via email at mdaconta@aol.com. I look forward to your reporting of errors, feedback, constructive comments and suggestions for improvements. Please prefix the subject of your email with "[Information As Product Feedback]" so I don't confuse it with spam.

I encourage you to take advantage of these collaboration methods and join me in designing a more effective digital future.

Best Wishes,

Michael C. Daconta
Woodbridge, Virginia

Chapter One: Why Information Production

"The division of labour, however, so far as it can be introduced, occasions, in every art, a proportionate increase of the productive powers of labour."[4]

- Adam Smith, Wealth of Nations

Manufacturing began in the Stone Age with simple techniques for preparing food, creating utensils and clothing. In the Industrial Revolution in the late 18th century mass production forever changed the face of the world via the widespread introduction and use of specialization and machinery in the manufacturing process. Figure 2 depicts the steam engine invented by James Watt in 1765. This shift in the efficient use of labor was the foundation of modern society and the harbinger of revolutionary power shifts through the resulting redistribution of wealth.

Figure 2 A Watt Steam Engine in Madrid[5]

[4] Smith, Adam; Wealth of Nations; Published 1991 by Prometheus Books; Pg. 11

Just as physical product manufacturing brought huge gains in efficiency which enabled the production of high-quality, low-cost products; those same gains are desperately needed for organizations in the information age. By not having any reliable process for creating and managing information, we waste countless hours and resources inefficiently working, reworking, searching for and repurposing information. Information production can be a repeatable process of integrating and assembling disparate data into a useable information product which is then delivered to the right consumers in a community of interest. The key elements of such a reusable process are summarized by the "4C's of Information Production": Consumer-centric Content, Catalog and Context as depicted in Figure 3. Or put another way: An Information Product =

[Consumer-centric [[[Content], Catalog], Context]]]]

While the above looks like a formula, it is not. The brackets only serve to highlight how the various elements are bound to each other. Figure 3 graphically depicts and explains the 4C's of Information Products.

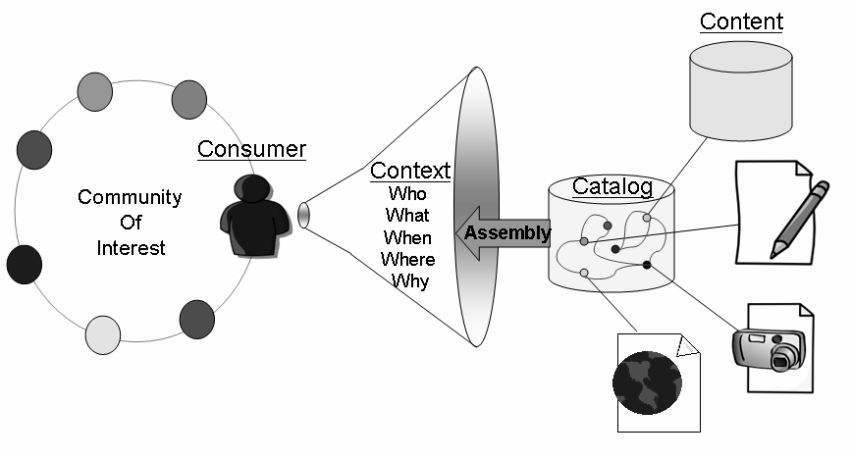

Figure 3 The 4C's of an Information Product[6]

[5] Wikipedia, http://en.wikipedia.org/wiki/Image:Maquina_vapor_Watt_ETSIIM.jpg
[6] Clipart icons reused are either public domain or drawn by the author. Public domain icons are available at http://clipart.nicubunu.ro

Figure 3 should be viewed from left to right. Let's walk through each key component of the diagram:

- The Consumer – All information products must begin with a known consumer or consumer segment. While this potentially halts the information production process until a consumer has identified a need, it also eliminates waste. Additionally, this requires the organization's leadership to understand what information each job function needs to perform its duties and how it may be more efficient or effective if provided additional information. For example, a program manager must monitor resource expenditures (labor, travel, indirect costs, etc.) against a budget and monitor project progress against a schedule. These two artifacts, a budget and schedule, are basic information products for the position; however, there are additional products that can be provided to improve productivity. In general, you find the information requirements for a position by examining four perspectives: Efficiency, Effectiveness, Teamwork and Autonomic Response. Though not always the case, organizations tend to focus on these in order along a maturity scale similar to Carnegie Mellon's Software Engineering Institute's Capability Maturity Model®[7] Let's examine each of these:
  - **Efficiency** – this is a measure of consumption which implements the saying, "do things right". By "doing things right" we mean performing them in the right way. The right way involves economy of movement and time (lowest possible consumption) to produce the desired outcomes. Information to promote efficiency involves the basics to perform the duty (at the operator level) and basic feedback measurements on the activities

[7] More information on the Capability Maturity Model ® Integration (CMMI) can be found at http://www.sei.cmu.edu/cmmi/general/general.html.

which makeup the work (at the management level). For example, a postal carrier needs to know where to pick up mail, where to pick up a carrier car and a route of addresses to deliver the sorted mail to. Feedback on this duty can be in the form of customer feedback, check-in/check-out times for the car and daily quantity of mail (in pieces) delivered. Efficiency improvements always lead to cost-savings via doing more with less.

- **Effectiveness** – this is a measure of strategic alignment which implements the saying, "do the right things". Doing the right things refers to the alignment and contribution of individual duties to organizational performance objectives. For example, in our postal example, the leveraging of the carriers to deliver stamps ordered from the online postal store increases the ease at which customers can use mail without incurring additional cost (by integrating the work into the normal mail delivery process). Another area of effectiveness would be for the postal service to track mail along its delivery path which will measure the speed of delivery and give customers additional assurance on the delivery progress of their mail. At the same time, this new feature could provide management with awareness of any bottlenecks in their national system. It is critical to distinguish effectiveness from efficiency measures. For example, an airport baggage screener's effectiveness is best measured by how many dangerous materials are prevented from entering a plane and not how many items are screened. Screening items is an activity (which can be made more efficient) while successfully preventing

dangerous materials which in turn protects passengers is an effectiveness outcome.

- **Teamwork** – this implements the saying that "a chain is only as strong as its weakest link." In other words, this looks at how an organization's people look to reach out and assist a larger "enterprise" in order to promote the operation of a larger system. Such "enterprises" can even cross organizations which create the "community of interest" as discussed in the book's introduction and as shown above in Figure 3. Teamwork requires collaboration and advanced information products that integrate data from various community members. To continue our postal example, the U.S. Postal inspection service is part of a larger law enforcement community against criminal's use of the mail system. This may require participation in active investigations with numerous other law enforcement officials for extended periods of time. Such work in an Enforcement Community of Interest requires advanced information products like "electronic case files" and "incident reports" that link and aggregate information from numerous sources.
- **Autonomic response** – the body's autonomic nervous system controls, more or less automatically, the basic functions of the body like heart-rate, breathing, gland secretion and other organs not subject to voluntary control. The ability for actions to occur without centralized, willful control can be duplicated in an organization by supplying enough information and training to "line" employees and short feedback loops to line managers. This principle is put to practice in the Toyota Production System whereby any employee

> can stop the line to fix a defective part. As Taiichi Ohno states, "In our production plant, an autonomic nerve means making judgments autonomously at the lowest possible level; for example, when to stop production, what sequence to follow in making parts, or when overtime is necessary to produce the required amount."[8] In the United Stated Defense Department this is one objective of "net-centric" warfare whereby small networked units use superior situational awareness and collaboration to decrease decision cycles.

So, understanding your information consumers from these four perspectives will enable you to create detailed information consumer profiles on each market segment. Those profiles explicitly define the general needs of your information customers and what they are looking to achieve or optimize. Now, let's move to the right in Figure 3 and see how we manufacture, package and deliver information products to satisfy those needs.

- The Context – this is the setting or background of any object (physical or digital) that gives it its unique characteristics and often clarifies the meaning of an object. Context development is a double-edged sword if it is not managed correctly because it is both critical to information discovery but can also be overworked or bogged down in debate. Context development can be overdone by attempting to dig too deep into the nature of a thing instead of constraining yourself with just enough context to provide utility to the consumer. The funnel in figure 3 depicts the universal aspects of context with the 5W's (Who, What, When, Where, and Why). Every business manager understands these and they are effective techniques to correlate information. Additionally the

[8] Ohno, Taiichi; Toyota Production System: Beyond Large-Scale Production; © 1978 by Taiichi Ohno; Translation published by Productivity Press; Pg 45.

figure represents context as a narrowing and focusing tool for the consumer. In other words, the mechanisms of context tailor and guide the information assembly process to target specific consumer segments. We will examine the detailed mechanisms to apply context in Chapter 4. Additionally, technical implementers will want to jump to Appendix A for a detailed explanation of best practice in designing information context.

- Assembly – between the customer's context and the catalog is the assembly process to manufacture information products from raw data. The assembly process will be introduced in Chapter 2, refined and expounded upon throughout the book, culminating in the last chapter. Successful assembly is a pull-based process that draws data backward from the content creator using the consumer's requirements, filters them against a "rightness methodology" (discussed in Chapter 6), and dynamically transforms them into a packaged, consumer-oriented product. We must apply to the information age, Taiichi Ohno's advice for the industrial age, "Today, the industrial world has been forced to master in earnest the multi-kind, small-quantity production system."[9] Heeding that same advice, we must 'master in earnest" an information production system. This will be a significant shift away from our current system of haphazard, one-off and manually intensive information "best-guesses" to a disciplined, responsive science of information product development.
- The Catalog – a repository or inventory of data assets available to satisfy consumer requirements. A data asset is a managed container of data like a database, website or content management system. The use of the term "asset" implies the data container represents value to the

[9] Ohno, Taichi; Toyota Production System: Beyond Large-Scale Production; © 1978 by Taiichi Ohno; Translation published by Productivity Press; Pg xiv.

organization and deserves its active management. A catalog of these data assets is sometimes referred to as a metadata registry/repository. Metadata and Metadata repositories will be discussed in detail in Chapter 4. The fidelity, accuracy and functionality of this catalog will directly and proportionally affect the information production process.

- The Content – this is the data and data structures that represent the raw collection of facts. For example, an image is a collection of pixels captured from a photograph or digital camera. A text document is a collection of an individual's thoughts on a particular topic. A database table is a collection of records where each record represents an entity instance (for example a record on a specific person). In our process, content becomes a part of an information product. This will require an understanding of how different parts of content satisfy a complex information requirement. Let's examine a simple example of a consumer's information requirement: a manager that seeks to fill a new position needs to provide answers on the position duties, salary range, team composition, minimum qualifications, preferred qualifications, availability of "in-house" candidates, urgency to fill, and numerous other judgement criteria. Content is data. Information is data in useable form. The key assertion of this book is: to reliably create information, you must treat it as a well-designed, concrete product.

Thus the 4C's of an Information Product describe the essence of a "pull-driven" process whereby a consumer within multiple different communities of interest can pull and assemble content (regardless of whoever produced it) by filtering the catalog entries through their unique situation. The goal here is to establish that clear line of sight from the consumer to the content. It is also to insure we have our

priorities straight by focusing on the information consumer and not the producer. While information producers may not like to hear that – it is an undeniable fact that the value of information cannot be measured in its production but only by its consumption. Thus every producer that wishes to produce worthwhile content is always focused on the consumer first. Now let us deeply understand what it means to label that content “data”, “information” or even “knowledge”.