2008

Upstream database and digital asset management in variable data printing

Franziska Frey

Nicholas Barzelay

Upstream Database and Digital Asset Management in Variable Data Printing

By

Franziska Frey, Ph.D.

McGhee Distinguished Professor

School of Print Media

Nicholas D. Barzelay, MPA, M.S. Print Media

Adjunct Professor

School of Print Media

Rochester Institute of Technology

A Research Monograph of the

Printing Industry Center at RIT

No. PICRM-2008-01

R·I·T Printing Industry Center

An Alfred P. Sloan Foundation Center

Upstream Database and Digital Asset Management in Variable Data Printing

By

Franziska Frey, Ph.D.
McGhee Distinguished Professor
School of Print Media
Rochester Institute of Technology

Nicholas D. Barzelay, MPA, M.S. Print Media
Adjunct Professor
School of Print Media
Rochester Institute of Technology

R·I·T Printing Industry Center
An Alfred P. Sloan Foundation Center

A Research Monograph of the
Printing Industry Center at RIT
Rochester, NY
January 2008

PICRM-2008-01

The research agenda of the Printing Industry Center at RIT and the publication of research findings are supported by the following organizations:

Table of Contents

Abstract

This study outlines the upstream database and digital asset management issues for variable data printing. The goal is to clarify what work environment and processes are needed during digital asset and data preparation. A literature review was conducted and complemented with the experiential experience of establishing and using a variable data preparation and testing platform.

Executive Summary

The purpose of this study is to understand the application of database and digital asset management tools, techniques, and skills to facilitate content preparation for variable data printing jobs. It does not address who is responsible for addressing the data and digital asset components in the overall workflow – the customer or the service supplier.

While variable data printing is widely recognized as an important adjunct to customer relationship management and targeted marketing, complexities with data and digital content preparation have limited the actual application of the technology. Digital assets and data involve processes closely associated with information technology. Lack of information technology savvy is reflected in how data is handled, what tools are used, and what skills are fostered to work the tools.

A review of available literature clearly demonstrates the difficulties encountered in comprehending and assembling the needed operational knowledge from the wide variety of topics involved in variable data printing. Each topic is a specialty by itself, particularly in the information technology area. This is a barrier to successful knowledge acquisition by printing companies.

Establishing an experimental variable data preparation platform confirmed that it is time-consuming to establish a working model. A number of different tools are available that have oftentimes overlapping functionality – a fact that may confuse their proper application.

A preliminary list of skill sets needed to adequately support VDP solution delivery was put together. This list will serve as the basis for future research into these skill sets and the ways to deploy them in the printing industry.

Introduction

The introduction of digital technology into the printing industry is predicated on a cross-technology, cross-media solution with the potential for delivering ancillary non-printing services to the print customer. In his book on the state of the printing industry, Romano (2005) alludes to this product and service mix. Cost (2005) discusses the integrated technologies within a historical context, as he describes the changes overtaking the printing industry:

> The evolution of the print medium as a service to human communications from the middle of the 15th century to its present is a story of the gradual disintegration of the production process into narrowly focused crafts that each took a share of the final product. We are at the beginning of a radical redistribution of the power enabled by digital technology and the Internet that sets the stage for the emergence of an entirely new industry. This new industry seeks to reintegrate the production and distribution of print with the businesses that are served by it. The industry will be dominated by companies that provide the technologies that will enable this to happen.

Among the integrated technologies are those that drive VDP (Variable Data Printing) -- data and digital content management, their storage structures, access methods, and processing techniques. The implication is that the digital printer, to be profitable, will need to provide not only the final print product but also the integrated technologies in the form of services. Separate researchers – Romano (2005), Sorce and Pletka (2005), and Frey and Christensen (2005) – have individually come to this conclusion.

This research monograph is intended to clarify what capabilities may be required to prepare text, images, and data for VDP pre-press activities. It aims to present a conceptual framework for using information and digital assets, and to explain what issues may be encountered in preparing them for print.

Although the technologies and tools currently exist – as do the professionals with the expertise to use them to fullest advantage – the continuing issue seems to be their deployment by digital printers. The key issue is the mix of workflow, tools, and skills employed by digital printing establishments. Without a conceptual understanding of digital content and database technologies, decisions about workflow, tools, and skill sets become difficult, and the probability of obtaining optimal results less likely.

Problem

The advantages of customer relationship management (CRM) supported by VDP are well documented. Findings from a Broudy and Romano white paper (1999 to 2000) show the significant revenue gains achievable using data-driven print: 574% gains for monochrome and 1,540% gains for full color.

In a series of articles for "OnDemandJournal.com" linking VDP and CRM, Heidi Tolliver-Nigro (2005) stated: "The foundation of a good CRM program is data on customer transactions, interactions, and behaviors. Many companies already collect this information and have more than they know what to do with it." She continued, "The trick is turning this information into something they can use."

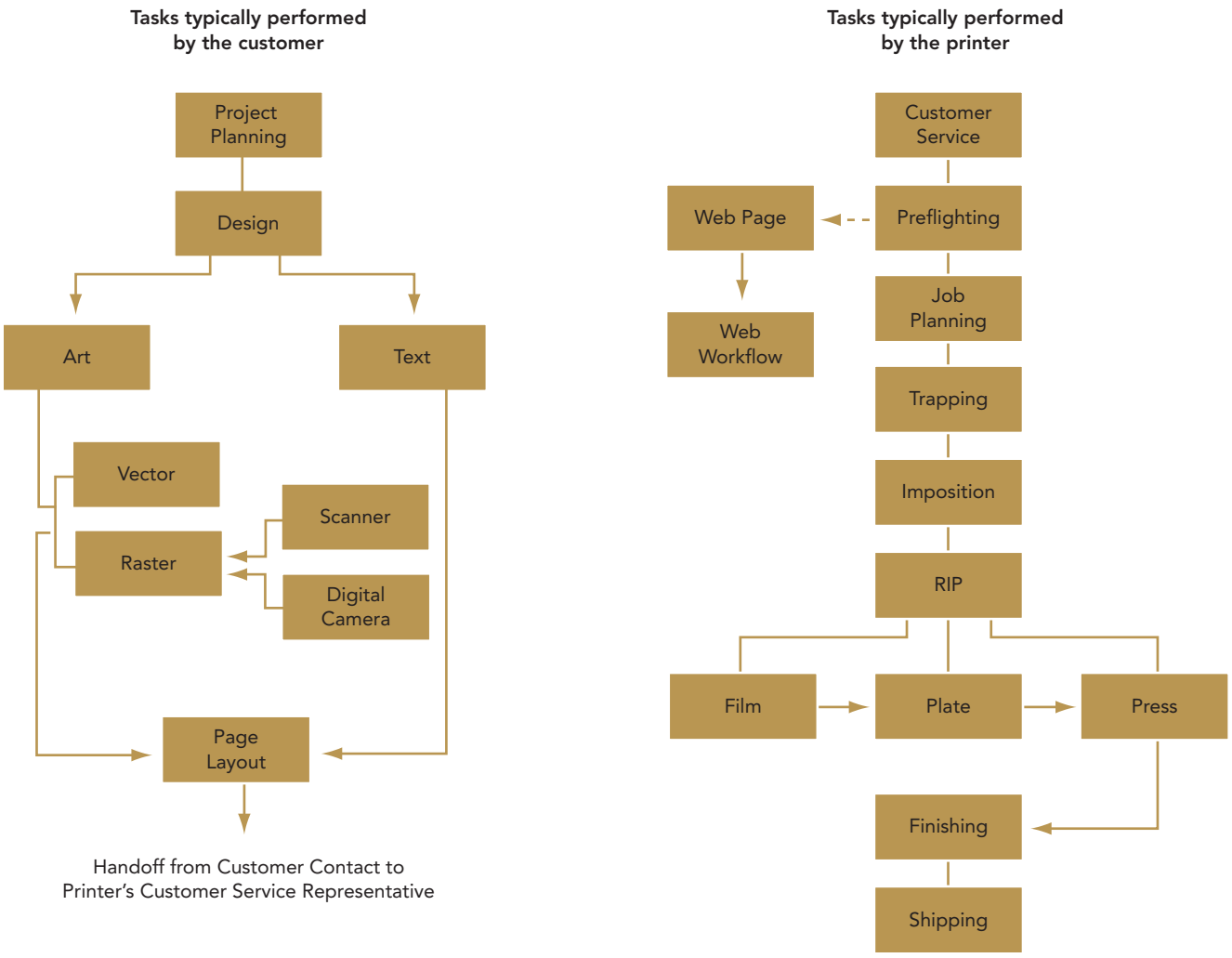
In their book, "Data Driven Print" (2006), Sorce and Pletka follow similar logic, stating, "Data quality can make or break VDP because of the need for data cleansing, scrubbing, and updating" (p. 61). Quality and delivery of data and other digital assets remain problems that are not readily addressed because the technologies involved are not related to printing. Correspondingly, the people addressing these issues may not have the level of knowledge in digital asset management (DAM) or information technology (IT) to adequately identify solutions.

In his book, "Status of Printing in the United States 2005," Romano (2005) stated, "There is still much to be done to educate marketers and 'creatives' about the value of CRM and targeted direct mail" (p. 81). Later in the text, he said, "The future of direct mail relies on the ability to appropriately cross-reference consumer information and use this to drive the content of mailing." He continued, "Marketing and transaction data are separate and need to be merged to get a complete picture of the customer" (p. 122). Romano also pointed out that among successful digital-only printers, "There is more emphasis on complex variable jobs where most of the revenue comes from database and programming services rather than print" (p. 79).

Multiple experts, including Romano, Sorce, and Frey, point to the need to tackle data and digital content with appropriate technologies. An implied need is the ability to understand information architecture and related processes. The crux of the matter is that the merging of differently purposed data structures into a meaningful data set is not a simple exercise in spreadsheet manipulation. IT in business practice goes well beyond the use of simple flat files and spreadsheets: it includes the relational databases, XML-based structures, and older hierarchical mainframe database technologies that may still be used by some companies.

Figure 1 shows a typical print production workflow that exemplifies this issue.

Figure 1. Customer/Printer workflow (Myers, 2006)



This model divides the print workflow into customer responsibilities and printer responsibilities. While digital content is implied, no digital content or database elements appear on the chart in either area of responsibility. It appears that commercial printing does not have to concern itself with the nuances of complex data structures inherent in multifaceted VDP applications, but rather only flat files (a spreadsheet or tab-delimited text file) that arrive at the right moment, are assumed to be clean, and are ready to use. Another interesting aspect of the chart is the fact that the customer is also not shown as dealing with data issues.

A chart by the IPA Workflow Team (2006) shows database management feeding into content definition (as part of the design concept activity), but there is no connection between digital assets and data. The data itself is shown going directly to the press production run without any review or quality check by the printing organization. This demonstrates data variability of mail merge with in-line addressing on a high-speed ink jet print head, but it does not address other types of personalization. Figure 2 shows a diagram modeled after the IPA chart.

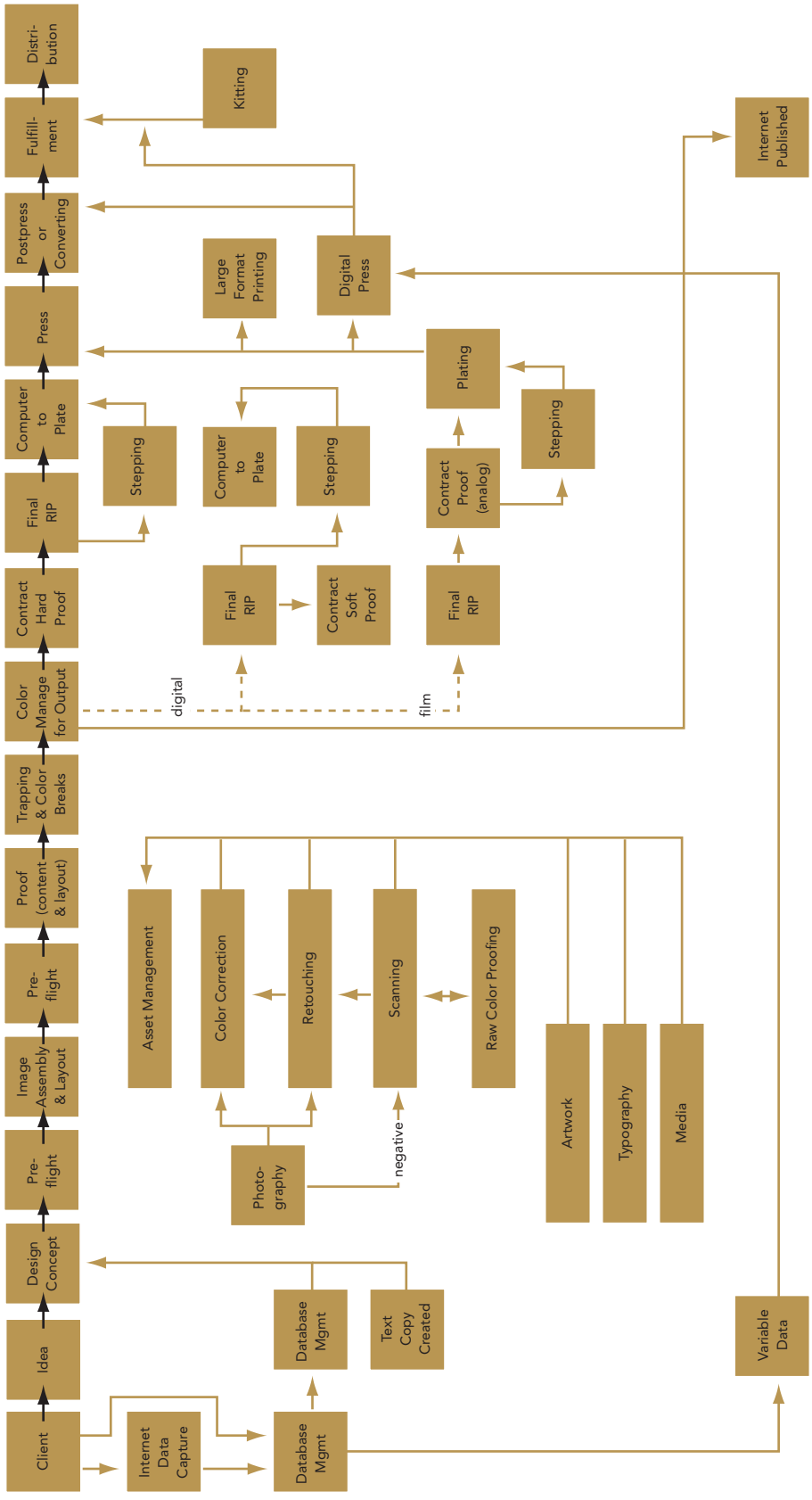


Figure 2. Graphic workflow – Part 1 (based on the IPA Workflow)

Beyond digital storage issues is the matter of providing ancillary database and digital content services to customers (Romano, 2005, p. 99). Sorce and Pletka (2006, p. 16) say it another way: "If the print provider wants to get a piece of the action, the printer must understand who the decision makers are, what their role will be in the promotional planning, what the marketing objectives are for the campaign, and how the print will be able to deliver the firm's objectives." This means that the printer will have to do a solution-based consultative sale with executive management, not a typical commodity sale and final product hand-off with a print buyer.

Background and Significance

Digital printing has become a major component in the toolbox available to commercial printers. It now encompasses varying electrophotographic processes and high-speed ink jet, all with sheet and web capabilities. Both benefiting from and driving these technical advances is the demand for increasingly sophisticated data-driven printing, as well as increased competition among printers to deliver it. If data and digital content are the driving elements in VDP, the printing industry does not necessarily reflect this.

In a survey of 103 printers of various types, Frey and Christensen (2005) found VDP and related DAM technologies to be considered growth areas. VDP ranked highly in comparison with other types of printing in the digital arena:

- Major applications for digital (p. 24) – marketing and promotional 24%, direct mail 20%, manuals and documents 18%, and others at lesser percentages.
- Among growth leaders (p. 25) – transactional financial documents 19%, signage 19%, marketing and promotional 16%, followed by book production, business communications, and direct mail.
- Future growth-related printing jobs (p. 26) – direct mail 24%, marketing and promotional 22%, and transactional 14%.
- Demand for DAM and VDP (p. 39) – 85% and 90%, respectively.

These are significant growth figures. If digital printing is an important competitive component in the printing industry, and if that component is deriving its most significant growth from various types of VDP and DAM, the expectation would be that expertise in those areas would similarly have high value.

However, when the survey (Frey and Christensen, 2005) covered such considerations as professional levels, education, and training, a completely different picture emerged (p. 45). Formal technical training (rated desirable at 26.7%) and experience or a bachelor's degree (25.7%) were both followed by "will train" (19.8%), and high school graduation (15.8%). These proportions seem reasonable until "will train" is viewed in conjunction with the survey question on training topics (p. 51). Of the training topics most mentioned, respondents' replies were in these proportions:

- Spreadsheet use (essentially use of text-based flat files) at 40.6%.
- Variable data applications (design layout using existing data) at 33.7%.
- Digital asset and database setup and administration (VDP and DAM essentials) at only 5.9%.

The significance here is the idea that the power of VDP is in the digital content and data. In an interview with Christensen (R. Parrett, personal communication, May 2006), Roger Parrett of Kodak Versamark said it this way: "Many companies claim to be strong in VDP, but in reality are not using anywhere near the capabilities available." He continued, "If they were more skilled, they could get more business." This presents both an internal problem and a sales problem. The internal knowledge, once acquired, must also be used to train customers so that they can understand the capabilities of VDP and know what to request.

Frey and Christensen (2005) found evidence to support Parrett's statement. According to Frey and Christensen, "on average, the companies that did hire DAM, VDP, and/or IT administration were larger than companies that did not hire in these areas." These companies "also experienced a higher revenue growth than companies that did not hire in these areas."

Delimitations

The primary purpose of this monograph is to illuminate the existing IT and DAM concepts for industry management and to show how these concepts can be put to work. This report does not purport to break new ground, but rather to encourage the use of existing tools and techniques more closely associated with IT than printing. However, it is not intended to be a technical manual.

The workflow diagrams and discussion in the latter part of the monograph are conceptual. Additional research and testing is needed to validate or revise these theories and concepts.

Given the potential for multiple solutions in a data- and digital-content-centric environment, no single tool set is promoted. In this regard, there is no intention to evaluate or compare various manufacturers' printing platforms.

While networking and system architecture form the underlying platform for VDP, it is not the intent of this report to delve into their subtleties. This paper solely focuses on hardware and networking as concepts that relate to the acquisition, manipulation, and delivery of data and digital content.

Background Theory

Effective VDP is based on three key elements: the physical computing system (the system model), information structure (the information model), and design configurations for different VDP types (basic VDP models).

System Model

The basic system architecture supporting VDP (based on the author's IT experience) is multi-layered, as shown in Figure 3. Primary layers, from the bottom up, concern connectivity, hardware, system software, and application software. Figure 4 demonstrates this in a physical configuration.

Figure 3. Basic system architecture

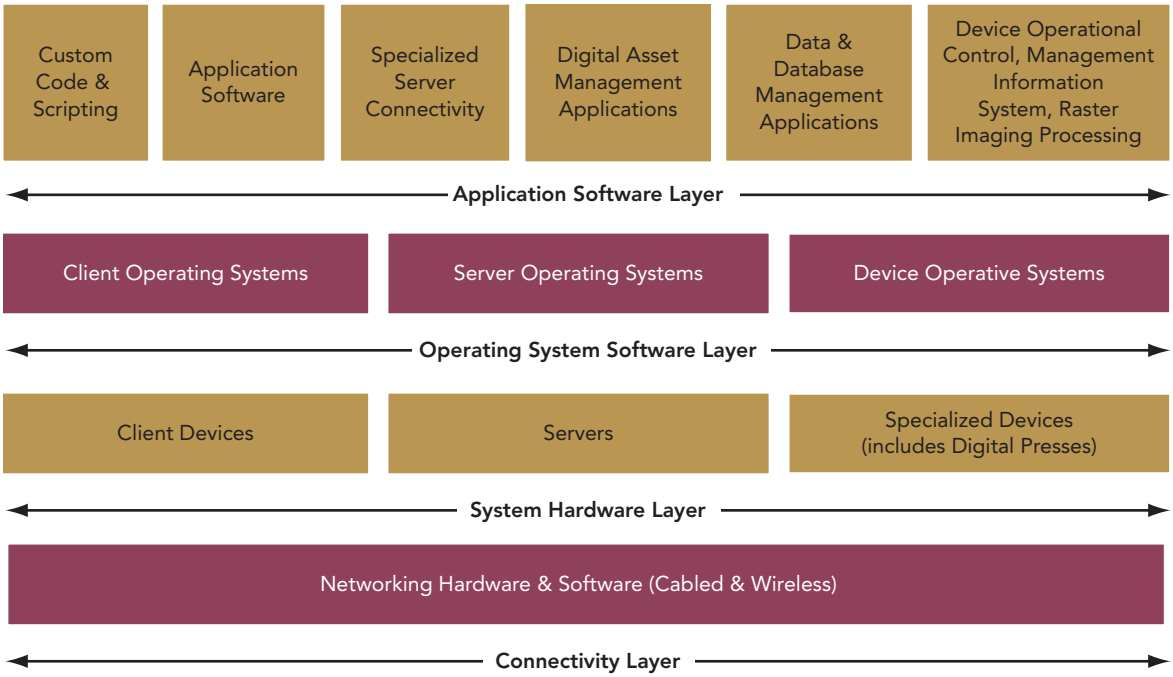


Figure 4. Basic physical system rendition (based on a configuration at RIT PAL)

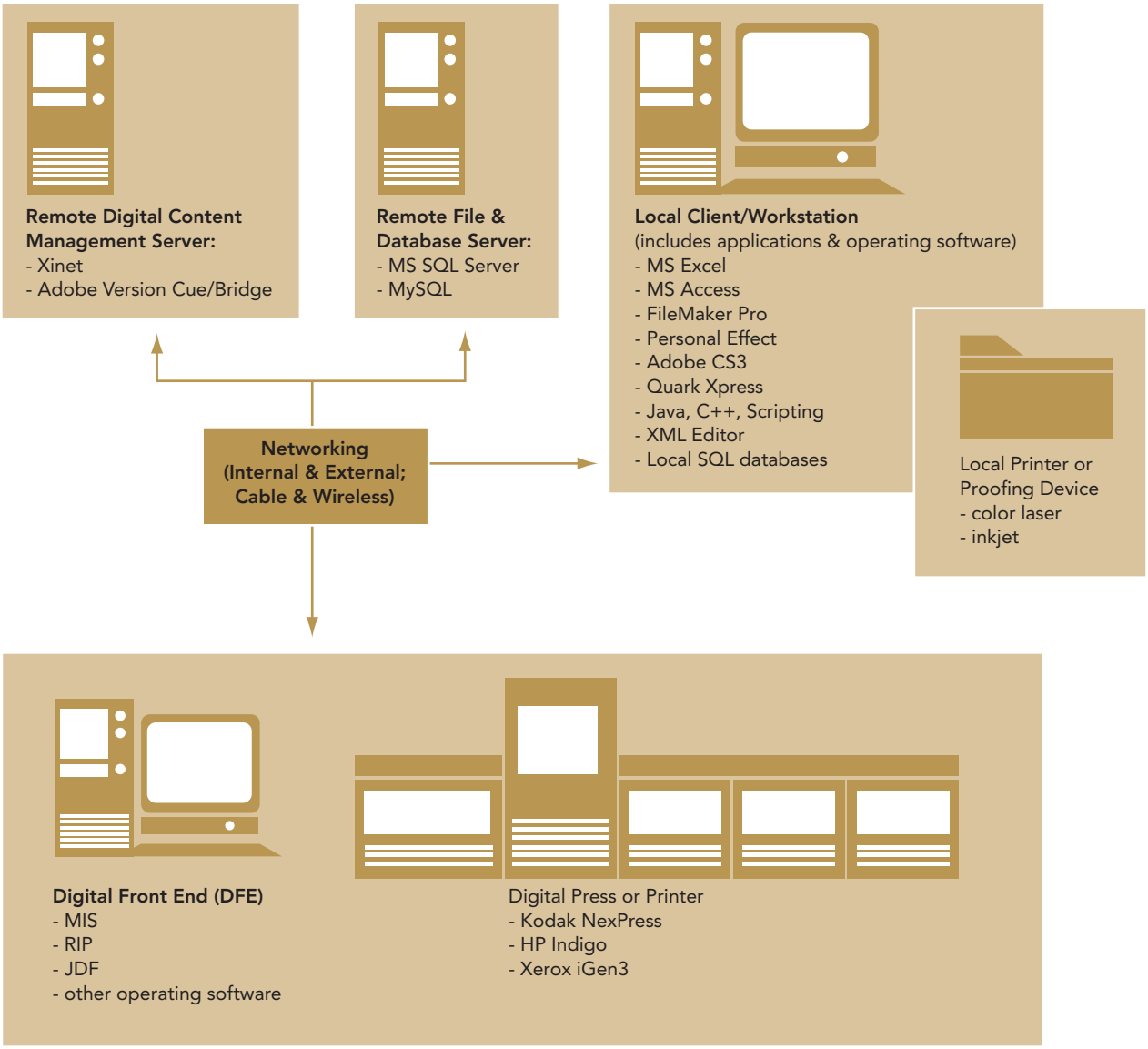


Figure 4 is modeled after the system configuration at the RIT Printing Applications Laboratory (PAL) that was used in research related to this monograph. The diagram does not imply system size.

Data and digital content sources may include mainframe systems, minicomputers, or servers of varying power. Client-server architecture is generally multi-tiered, spreading selective functionality across different platforms or processors. A discussion of client-server architectures can be found in books such as Liang's (2000) *Rapid Java Applications Development using JBuilder 3* or *Developing Java Enterprise Applications* (Asbury & Weiner, 1999).

DAM capabilities can reside on a different server than the database management system, though this is not a necessity. The two content sources feed a large storage-capacity workstation, which may be used for both data preparation and document design. Potentially, a single workstation, as shown in Figure 5, may actually represent multiple workstations organized to divide labor.

The VDP system is facilitated by a local area network (LAN), which requires networking skills in addition to application and data handling skills. Networking is limited to connecting the proofing printer, the workstation, and the digital front end (DFE) of the digital press or high-speed ink jet printer.

Customer-supplied data may eliminate some complexity, because the customer is responsible for assembling data and content. The customer, not the printer, has to deal with IT infrastructure. But this means that the printer has to assume the data is clean, organized, and usable – a potential risk for both the customer and the printer. The provided data and content may or may not reflect the customer's business or campaign objectives. It may be nothing but a mailing list of unknown quality from some commercial source. Whatever the source and whatever the content, the immediate need is to understand it, determine its level of quality, and then deal with it. This means that it is necessary to have some capabilities and facilities for examining the data and other content (digital images and potentially text), which, according to Sorce and Pletka (2006), can be very time-consuming.

Information Model

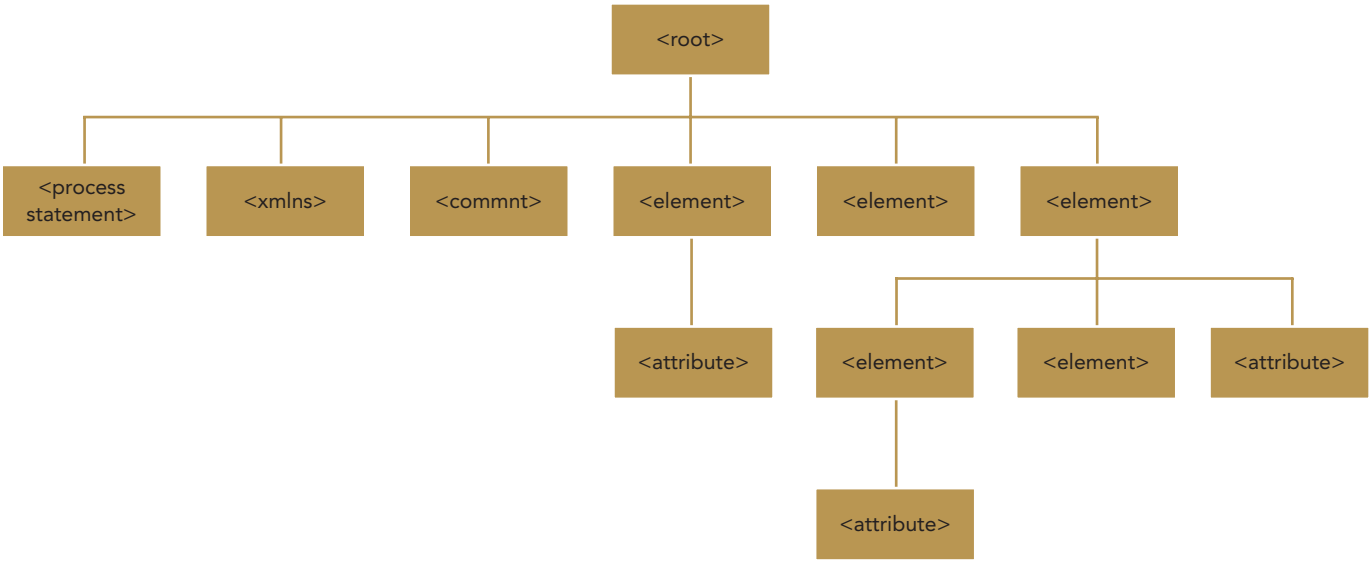
Information may consist of data in the form of facts or figures, digital content such as text, images, and photographs, or metadata (data about data), as aptly discussed by Boiko (2005) in his book on content management.

Information modeling is building information structures by categorizing, grouping, and organizing data elements for effective and meaningful accessibility – usually in the form of a tree of some type. It can be as common as preparing a report, or as esoteric as creating an object-relational database with XML (eXtensible Markup Language) components. Classic examples of information structures familiar to everyone are report outlines, books, and the common paper filing cabinet, shown in Figure 5. What is important is the idea that information structure is really a set of nested containers: the filing cabinet, the file folders, and the documents in each folder (which are also containers).

Figure 5. Basic physical system rendition (based on a configuration at RIT PAL)

Documents, as XML containers, are organized in a nested structure beginning at the document level, and then subdivided into smaller interrelated elements, identified by start and end tags (for example, <name> Smith </name>) to form the internal document containers called "elements", "attributes", or other nomenclature depending on their use within a document. As an example, Figure 6 shows an XML document tree structure of basic components.

Figure 6. Basic XML document structural components

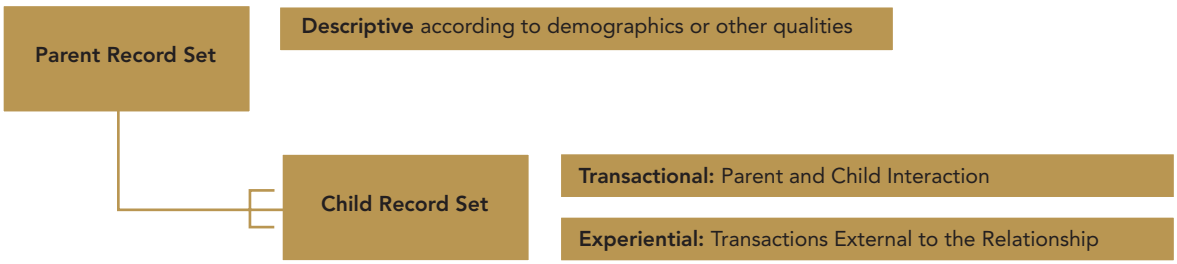


The root element represents the document as a whole. Different elements provide processing instructions, commentary, and content. The container elements may themselves contain elements. This same structure is similarly found in the storage directory approach of various "Windows" and Unix systems.

Content stored in a database uses a similar structural paradigm: nested containers of categorized and grouped data organized in parent-child relationships. Most books on database management systems (DBMS) go into considerable detail on table, row, column structure, and the relationships between tables. Examples of such publications are: "Oracle 8: The Complete Reference" by Koch and Loney (1997) at the high end of database systems; "Access Database Design and Programming" by Roman (1997) on the desktop; or, more recently, Forta (2006) on MySQL or Coffey and Prosser (2005) on FileMaker Pro.

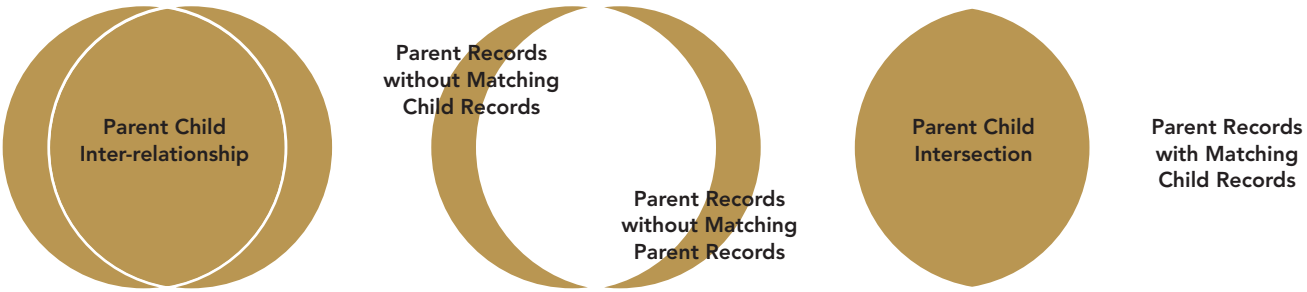
Most manuals on DBMS commonly use the entity relationship (ER) model depicted in Figure 7 to demonstrate the parent/child relationship between tables. Figure 7 shows the structural view of a basic database relationship.

Figure 7. Basic one-to-many relationship between tables



In their discussion on data mining, Westphal and Blaxton (1998) suggest another database view of table interrelationships using Venn diagramming to demonstrate their relative contents. Figure 8 takes this approach by converting the ER diagram in Figure 8 into a Venn diagram. Understanding these two views is critical when working with data for VDP.

Figure 8. One-to-many relationship as a function of content



In the logical view (Figure 7), the parent is a high-level descriptive container for the child, providing contact, demographic, and financial information. The child provides historical information on what the parent has done in a given context. Information may be transactional, meaning that the parent has engaged directly with the recording organization, or experiential, in that a parent has historical activity but not necessarily with the recording organization.

The content view of the parent-child relationship (Figure 8) is the match between parent and child populations. The match is demonstrated using set theory described by Bishir and Drewes (1970) in their book, "Mathematics in the Social Sciences." The Venn diagram shows the degree of fit between parent and child populations as an intersection of the two. Those portions not intersecting are parent records without transactional children and transactional children that cannot be associated with any descriptive parents. The intersection is critical, because it delineates what portion of the entire data pool is usable at a particular level of effectiveness for a given VDP marketing campaign.

Using the example in Figure 9, it is easy to imagine the mix of database content between the respective tables in this simple ER structure.

Figure 9. Retail database example (based on Figure 7)

Consider a retail grocer with multiple locations. Each store has customers, which may or may not be "customers of record" (defined as having their information in a database as a parent record). Assuming that all stores record all transactions (as shown in Figure 8), it is easy to understand how mismatched or orphaned data can develop.

Physically joining the information and system models together into a business system creates a further complication in that programs which manage and use content usually reside in different software environments. These application environments may be installed on a single powerful computer or on multiple computers. If not on a single machine, a network makes the physical connection between machines, but the environments must also talk to each other programmatically.

The methods of communication between database systems and applications are native (direct connections), open database connectivity (ODBC), Java database connectivity (JDBC), and intercommunication between connection types with an ODBC/JDBC Bridge. Asbury and Weiner (1999) and Liang (2000) describe these connections in more detail.

Conversations between the source environment where the information resides and the destination environment where it is used usually require one of two types of connections – a request language like Structured Query Language (SQL) for DBMS or a native request protocol for digital asset management systems (DAMS). The combination of system and information models creates the basis for the third level of modeling: application modeling that describes different VDP approaches.

VDP Models

In their book, "Data-Driven Print: Strategy and Implementation" (2006), Sorce and Pletka list six types of variable printing: versioning, mail merge, personalized printing, transaction printing, Internet-on-demand, and fully customized communications. For each of these types, some application architecture changes are required to deliver the right information at the proper moment in the correct sequence of events. Figures 10 through 15, which are based on Sorce and Pletka's VDP types (2006), demonstrate the various models.

One constant is the need to assemble, cleanse, and organize the data to facilitate a match with appropriate content. The other constant is the need to design necessary business rules to automate the data-content-design match through the entire job. The implication of this is that, no matter what type of VDP solution, the required skill set remains the same.

Figure 10 shows a versioning example where mail order catalogs are prepared for different areas of the country. Variability would be in the catalog content.

Figure 10. Versioning

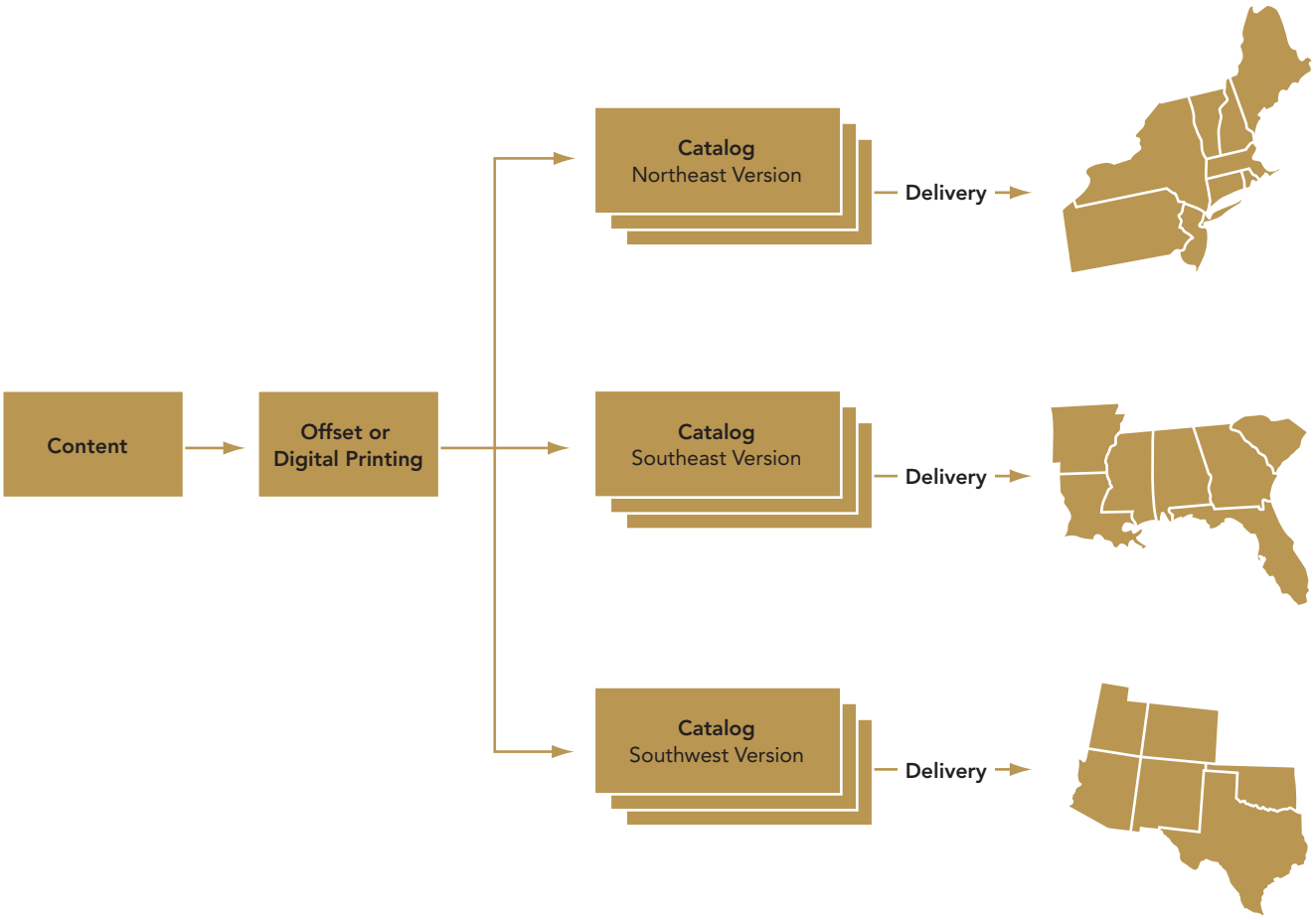


Figure 11 is an example of a mail campaign using mail merge. The campaign letters were printed on an offset press and addressed in a second run on a digital press. Addressing could also have been done in-line with a high-speed ink jet print head.

Figure 11. Mail merge

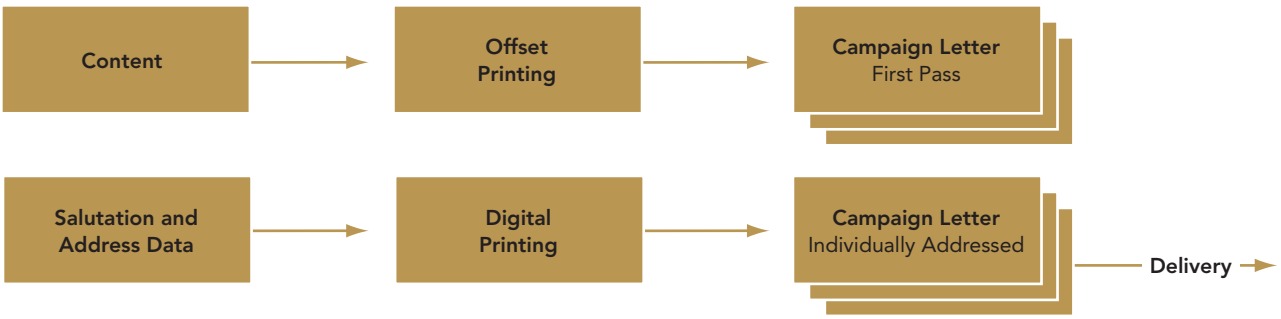


Figure 12 demonstrates personalized printing. In this example, individual store advertisements with coupons are prepared for each customer based on their buying habits.

Figure 12. Personalized printing

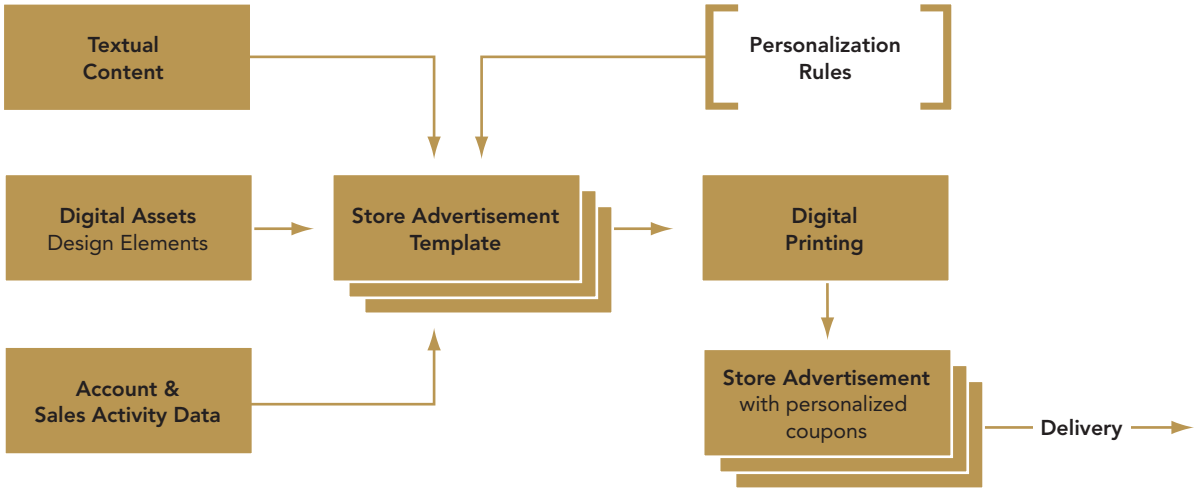


Figure 13 is a transactional printing example. Individual monthly investment reports are prepared based on month-end status and the past month's account activity for each investor.

Figure 13. Transaction printing

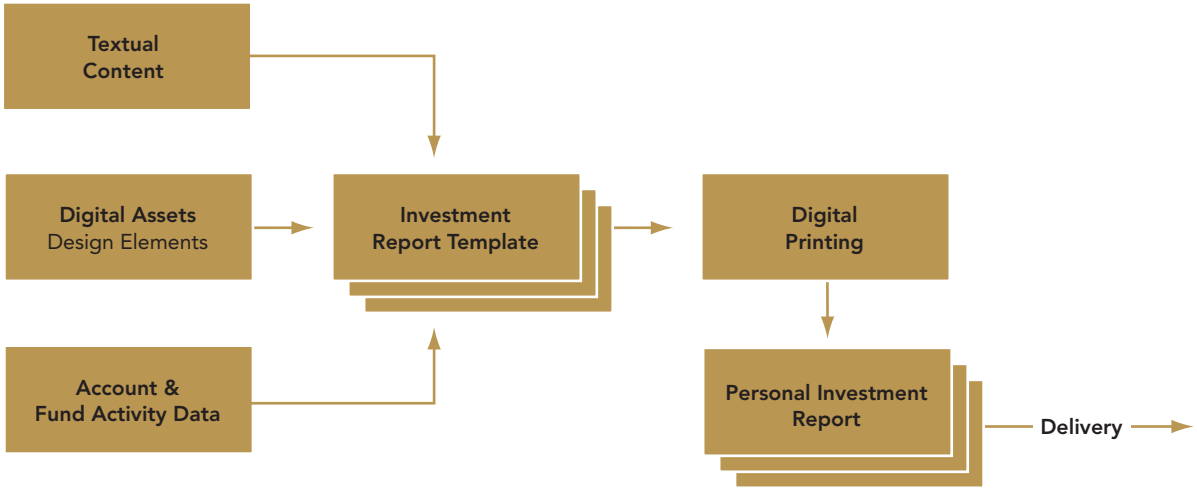
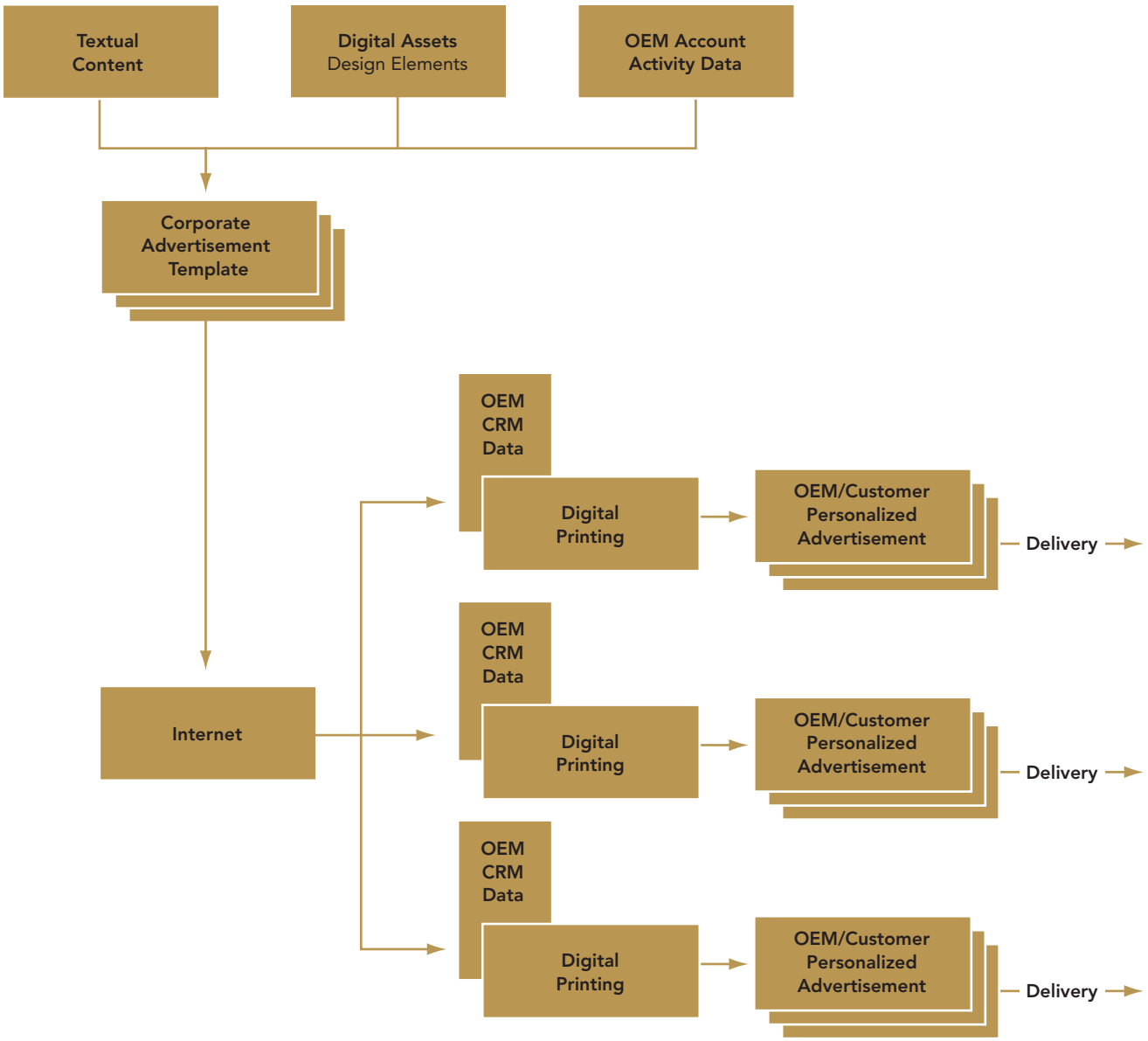


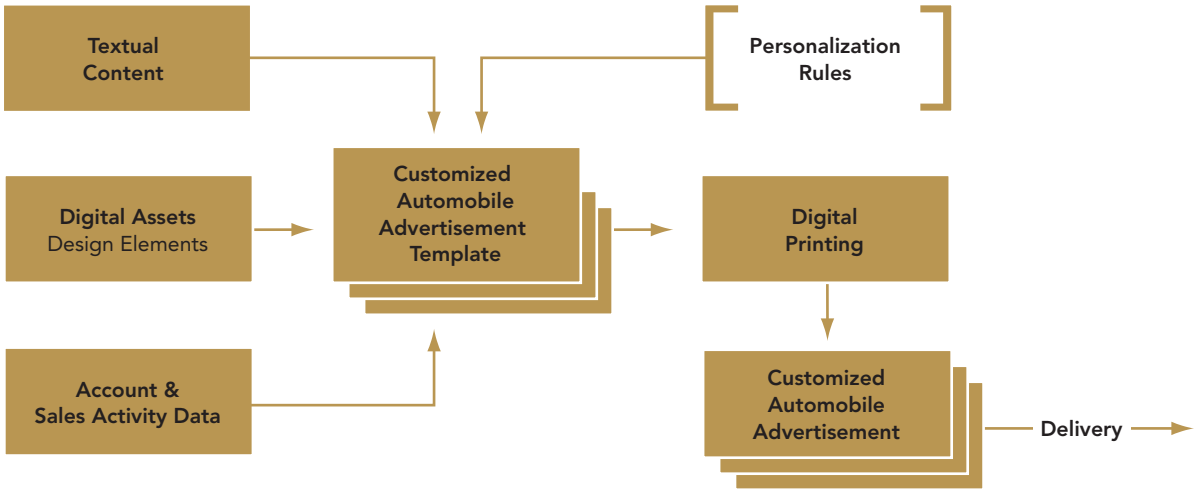
Figure 14 shows a model of Internet-on-demand. There are several approaches to this type of solution. One approach is to use the Internet as the request medium for a common print-on-demand (POD) solution. As portrayed here, the Internet is the distribution vehicle for a standard advertisement prepared at corporate offices. The advertisement is sent over the Internet to product distributors who customize the document to reflect their distributorship and, potentially, tailor it for their individual customers as well.

Figure 14. Internet-On-Demand



The most complex type of VDP is fully customized communications, shown in Figure 15. In this example, an automobile dealer sends out advertisements about the new model year to current customers. The content reflects the age of their current automobile, its color, model, accessories, and remaining lease or payment amount. The advertisement contains an offer for a new car purchase, showing pictures of the new vehicle with interior and exterior pictures that match the customer's tastes.

Figure 15. Fully customized communications



With an understanding of the basic architectural models (systems, information, and applications) and how they interact, the next step is to examine a cross-section of available literature and develop an appreciation of the problems in the VDP knowledge base.

Note that some of the literature is older. This is important, because it underlines the point that many key technology components critical to VDP have been around for a number of years. Most of the data and imaging technology needed for successful VDP operations is not cutting edge, and is presented in the literature with considerable experiential knowledge. It is primarily VDP itself that has not reached maturity.

Literature Review

This literature review is focused on preparation activities that take place prior to the press hand-off. The reviewed material covers both background literature and technical subjects.

In viewing this list of topics, keep in mind the Westphal and Blaxton (1998) discussion on different types of knowledge (pp. 60-66). Westphal and Blaxton distinguish between knowing how to do something and knowing factual information about something. They call the former "procedural knowledge" and the latter "declarative knowledge." They also distinguish between actual knowledge (what information actually exists) and meta-knowledge (what information may exist).

The importance of distinguishing knowledge types in reviewing VDP-related literature and expert opinion is that VDP discussions regularly mix declarative, procedural, actual, and meta-knowledge. For instance, knowing about an activity may be confused with knowing how to do it. If knowing how is not based on actual expertise, related declaratory knowledge (which might be valuable) may also contain inaccuracies.

While considerable literature relates to various facets of VDP, much of the literature does not mention VDP, so the reader has to make (or be able to make) the connection between the technology discussed and VDP. For example, in database technology material, the reader will have to connect database and VDP knowledge themselves because no connection is mentioned.

This literature review gives a high-level view of VDP in the printing industry, IT, DAM, and technical skills. Conclusions will be offered at the end of the review.

VDP and the Printing Industry

Commenting on the status of the printing industry, Romano (2005, p. 79) says, "There is a lot of talk about variable data printing, but the money is mostly in static work." He also says, "Among digital-only companies, there is more emphasis on complex variable data printing with most of the revenue coming from database services and programming, not printing per se" (p. 79). In other words, data and digital content are drivers for printing, and money-makers by themselves.

Romano predicts that new workflows will include support for databases (p. 103). He then proceeds to the heart of the matter, explaining that transaction data and marketing data are separate but need to be merged to provide a complete customer picture (p. 122).

These views are repeated by Tolliver-Nigro (2005) in her three-part series titled, "Better Together: VDP and CRM." Another article on CRM by Winer (2001), although not specifically mentioning VDP, discusses the necessary components of targeted marketing:

- A database of customer activity
- Analysis of the database
- Decisions about which customers to target
- Tools for targeting those customers
- Techniques to build a relationship with the targeted customers
- Privacy issues
- Metrics for measuring CRM success

Winer then discusses database content, explaining that a database needs to provide customer contacts, customer transactions, descriptive information, and responses to marketing efforts, while recording all this information over time.

Among the many experts on the printing industry and CRM, Romano, Tolliver-Nigro, and Winer introduced the linkage from marketing and CRM to VDP. This provided an opportunity for others to move the discussion directly to VDP procedural information and know-how.

In "Marketing Through Personalized Printing and Variable Data," Chapman (n.d.) comments, "If you are going to build a document based on a database, the quality of the information that appears in the document is only as good as the database itself." Kanonik (2004) tells printers that they need to practice what they preach regarding variable data. He lists a set of steps for developing their own use of VDP, and suggests that using VDP themselves will provide the success stories they can use to sell VDP to others. Kanonik (2004) concludes with some tips. However, in both articles actual know-how is not obviously evident.

In two whitepapers on Xerox Freeflow solutions, Davis (2004, 2004a) states that there is a major problem in bridging systems and organizations when generating promotional material. In discussing the particulars of the solutions, he states that Freeflow can plug into "any database." He goes on to list the following databases: dBase, Microsoft FoxPro, Microsoft Excel, Microsoft Access, Paradox, Delimited text files, and ODBC. Based on common industry knowledge in IT, these are all desktop data sources, some of which are not considered databases at all. Missing are heavy database solutions such as open source MySQL, Microsoft's SQL Server, Oracle from Oracle Corporation, and DB2 from IBM. In addition, making an ODBC connection only means that two application environments can communicate – it says nothing about how they converse or what storage paradigms they use.

A CAPV/InfoTrends article (March, 2005) entitled "Production Workflow Solution Product Compendium: Variable Data Design and Production Software" includes comments on database content and the use of conditional logic. In their book on data-driven print, Sorce and Pletka (2006) carry the technical discussion further with high-level details. In particular, they discuss business logic and basic (Microsoft Visual Basic) programmatic constructs needed to control the variability of a document according to a data stream. Christensen (2006) diagrams in detail the entire VDP job flow in an unpublished paper.

An important, though not surprising, detail from Christensen's paper is that procedural knowledge best comes from face-to-face interviews with experts and practitioners. The caveat, however, is the need to know who the expert is, who employs that expert, and whether the interviewee has an agenda. The tools and technologies a particular company promotes provide definite flavor to an interview or presentation. For example, a Throckmorton presentation (2006) primarily promotes and explains VDP as a marketing tool. When the presentation turns to the question of how to accomplish VDP, Throckmorton advocates XMPie's toolset. But in discussing how non-technical people can comfortably use XMPie, he says, "Everything can be done with XMPie." Whether true or not, the effect is to pad over the details needed to prove know-how.

The transcript of an interview by Christensen with an Oće representative (personal communication, 2006) provides valuable information on the steps and tasks of acquiring and preparing data for VDP. The tool of choice in this case was "Document Designer Advanced." In a similar interview with a Kodak Versamark representative, Parrett

(personal communication, May 2006), the selected tool was "Composer." Parrett has an IT background and is able to discuss database connectivity, response time, and issues associated with business logic. Parrett's advice is to minimize the amount of logic the printer has to deal with by using "Composer" to automate processes by building C, C++, and Java functions into the workflow. However, building C, C++, and Java functionality is not something the average printer is going to be able to do without considerable technical assistance.

A joint interview conducted by Christensen and the author with Meehan (B. Meehan, personal communication, August 2006), an independent IT and VDP consultant, gave a different view from the one-tool-does-all approach. Meehan advocates selecting from a variety of tools, depending on what problems exist. He has found that some of the most versatile tools are scripting languages, such as Javascript or Perl. Meehan also points out that "Composer" (mentioned by Parrett), is a version of "FusionPro." This point is important because it raises the question of tool lineage (ancestry) and casts further doubt on the one-tool-does-all approach. In fact, according to an RIT Printing Industry Center (PIC) study done by Sorce and Pletka (2004), respondents mentioned 24 different tools used to directly assist in VDP production. The study did not cover tools that played an indirect or "assisting" role in the VDP production process.

Reading what the tool manufacturer has to say about its software, though marketing-oriented, provides important detail that may be overlooked by some tool advocates. For example, XMPie sales material (2005, 2005b) on VDP offers two classes of tools. uDirect is used with single-table or flat file data sources. It cannot deal with the full relational aspects of a multi-table DBMS. However, it is simple to use with Adobe InDesign. PersonalEffect, the other application, provides full database connectivity via ODBC into the major data sources in business use: Oracle, SQL Server, DB2, MySQL, and a number of others.

Although procedural detail is one weak point in VDP discussion, a related and often overlooked topic is identifying the skills needed to do the work, particularly data acquisition and preparation. The small amount of discussion on this topic partially explains what Frey and Christensen (2005) noted about skills in their survey – that the diminished and sometimes simplistic approach to tools and skills belies the potential business importance of VDP in printing.

Information Technology

Information technology is a very broad topic by itself, covering multiple hardware and software components, diverse programming languages and technologies, and a variety of tools, methods, and techniques. Of particular interest in this discussion is the use of database technology to provide source data for VDP. At the same time, VDP and marketing issues are generally not a topic of concern in information technology (IT). One exception is the area of knowledge mining, which, in a business context, usually relates to marketing and, by implication, personalization.

The topic of database technology is very complex, with books and articles written on individual subtopics and different DBMS. Besides design and structure, subjects range from relational algorithms and object storage to recovery and integrity and performance tuning, to name a few. When considering the database in VDP, only basic concepts need be addressed. Roman (1997) provides a good conceptual framework at the desktop level, and Koch and Loney (1997) provide a discussion of concepts from the perspective of an enterprise-capable DBMS. These concepts address database structure, data quality, and concern for table field content, record uniqueness, referential integrity in related records, and processes involved in dataset acquisition and preparation.

Articles such as the one on database performance philosophy in "15 Seconds Weekly Newsletter" (Bostrup, 2003), or the collection of articles at the Web site SQL-Server-Performance.Com (2006), make important points to regard if data request and retrieval response times are affecting a VDP job. Even an article on client-server architecture and its effect on database throughput (Nichol, 1997) may be relevant in limited cases. However, the primary need is to be able to deal with database structure and content, leaving other considerations to a database administrator.

Topics needed for VDP can be termed "database basics." For high-end relational database management systems (RDBMS) which are most commonly used in operational settings, a new edition of a book such as "Oracle 8: The Complete Reference" (Koch and Loney, 1997), will provide an essential reference. For desktop databases, a newer edition of Roman's book, "Access Database Design & Programming" (1997), will give the reader key insights into database construction and use. One realization from reading Roman's book is that Microsoft Access is a simple but powerful tool that is much more appropriate for working with VDP-destined data than Microsoft Excel, which is primarily a grid manipulation and calculation application.

An important issue is the process of getting the data from the source into the hands of the VDP preparer. In a RIT PIC study by Sorce and Pellow (2003) directed at advertising agencies, marketing executives were polled on customer database and CRM use. The respondents stated that almost all their customers had some sort of customer database, but only 21% employed CRM. They went on to say that the biggest obstacles in recommending personalization strategies to their customers were price and the lack of a suitable database. In another study, Pellow, Sorce, Frey, and Banis (2003) found that there was no consistency in the software used to manage the source data. They identified 160 different data-related software applications in use. This underscores the problems of acquiring source data and moving it to a work environment for cleansing and preparation, as well as the need to simplify the number of software tools needed to do the job.

There are a number of popular database systems of differing levels of performance and popularity. Another offering from Microsoft (besides Access) is an upward scalable RDBMS called SQL Server. A good introduction to this system is a book by George and Delano (2006) entitled "SQL Server 2005 Express Edition Starter Kit." An open-source competitor to SQL Server is MySQL, which has gained considerable popularity in busi-

ness for supporting Web applications. Forta (2006) has written a good introduction to MySQL, but additional references may be helpful. Another easy-to-use desktop database with some client-server capabilities is FileMaker Pro, noted for its ability to handle digital images at desktop scale and its cross-platform capability (Microsoft Windows and Apple OS X). "FileMaker Pro: The Missing Manual" (Coffey and Prosser, 2005) is a good place to begin as it is quite comprehensive in its coverage.

An idea that emerged from discussions between the author, Christensen, and Meehan (B. Meehan, personal communication, August 2006), was that the trouble with using databases in VDP concerns basic subject matter. While not esoteric, it has its own complexities and subtleties. The number of different database systems and their operating system platforms further add to the range of subject matter that needs to be covered. When it comes to skill levels, the practitioner not only has to have some rudimentary knowledge, but also has to know enough to realize knowledge gaps, as well as where to find answers. In fact, knowing where to find answers may be more important than actually having a given answer.

Data may be dispersed among several data sources or media. In progressing from a variety of data containers to a prepared data set for VDP, one approach is to retrieve desired data from a data warehouse or data mart. Although understanding data origins and associated business processes is necessary, issues of data preparation may be mitigated to some degree because data may have already been prepared prior to leaving the warehouse or mart.

Knowledge and Data Mining

In his book, "Web Warehousing and Knowledge Management," Mattison (1999) presents an introduction to the data warehousing and knowledge topic. He starts by saying, "Data plus context plus application to specific business objectives equals business knowledge" (p.30). The data warehouse contains the essentials of business knowledge. He defines the warehouse as "any of a large variety of computer system initiatives whose primary purpose is to extract information out of legacy systems and make it usable to business people in the support of their efforts to reduce costs and improve revenues" (p.139). The distinction between warehouse and mart, according to Mattison, is that the warehouse is enterprise-wide, whereas the mart is departmental (p.148). The process of populating the data warehouse, as described by Mattison (p.155), is very similar to what would be used in VDP preparation: extract data, format it, validate it, merge it, purge problematic elements, and then stage it.

In her book, "Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management," Rud (2001) explains, "Because consumers have become more discriminating, there is a need for target marketing in order to direct an offer to a target audience" – in essence, VDP, though not mentioned by name. She continues by discussing types of targeting questions, types of analysis techniques, and different modeling methodologies. The results of this discussion provide the knowl-

edge to make business decisions on what audience groups get what level of customized, personalized, marketing communications.

Discussions of data warehousing and data mining techniques are predicated on arguments made by Westphal and Blaxton (1998), who discuss in some detail the issues involved in preparing data for analysis and application. They state, "Data cleanup may take eighty percent of a data mining effort" (p.105). They continue by listing the types of problems a data analyst is likely to encounter (pp.105-108):

- Inconsistent data
- Incomplete data
- Erroneous data
- Missing data
- Spelling and transposition errors
- Incomplete formats
- Improperly completed field entries
- Deliberate improper coding
- Ambiguous naming
- Naming variations

These conditions may also be encountered in preparing for VDP. Indeed, eighty percent of a VDP job may well be data preparation, whether data comes from a customer or a commercial supplier. Issues initially confronting a prospective user are a good introduction to the related issues of programmatically manipulating data to correct problems and match it with other sources of content through DAMS (Digital Asset Management Systems).

Programming

Programming concerns three processes: connecting to the data source for data acquisition purposes, cleaning and preparing the acquired data, and applying business rules in the form of program logic to facilitate desired data and content variability in the target documents. In their interviews, both Meehan (personal communication, August 2006) and Parrett (personal communication, May 2006) talked about several different programming and scripting languages they used in preparing for VDP. From their perspective, no one language is best. Selecting a particular language depends solely on the type of programming problem encountered.

There is an abundant supply of manuals, articles, and books for all the different program languages in use. For any given language, it is common to find that no single book tells everything a programmer needs to know. One reason is that there are so many differ-

ent technologies involved in creating a single solution. Another is that changing technologies quickly render books obsolete. Meehan mentioned this when talking about the number of books on the same subject he currently has and the number of books he has regularly discarded.

Many books are timeless – 'keepers' – while others have built-in obsolescence. An example of the former is Flanagan's book, "Java Examples in a Nutshell" (1997), which provides basic Java program constructs and solutions. An example of the latter, Liang's "Rapid Java Application Development Using jBuilder 3" (2000), would ultimately be a throwaway because it is tied to a particular version of a software development package that has become obsolete. However, Liang does provide a thorough discussion of JDBC-ODBC database connectivity and rapid development techniques, creating a possible reason for reprieve. "Developing Java Enterprise Applications" by Asbury and Weiner (1999) is an older book, but some of its material may be considered strategic and it is therefore valuable for specific problems. As stated before, the VDP developer has to know what to use, where to look, and also has to be cognizant of copyright dates and editions. It is not merely a matter of picking out the most recent edition. The problem being addressed may not be on a state-of-the-art system, and may therefore require older documentation.

Digital Asset Management

Besides data, the other element of variability in VDP is content. In his book, "Content Management Bible, 2nd Edition", Boiko (2005, p.4) makes the distinction between data and content: "Data consists of small snippets of computer information that have much of the human meaning squeezed out; content is also information but it retains its human meaning and context."

Content is most appropriately stored in a content management system (CMS) or DAMS. Experts differentiate the two in subtle and sometimes overlapping ways. One group (Geser et al., 2002) distinguishes the two according to the kind of organization using the technologies – institutional as opposed to business organizations. For the purposes of this discussion, DAMS will be the preferred appellation.

Davis and Walter (2003, p.9) define DAMS as "characterized by a shared resource architecture, the integration of content publishing with core business and supply-chain processes, and multi-channel rich-media content delivery." McCord (2002) describes DAMS this way:

- Digital content and associated metadata
- Collections or packages of digital assets
- Use of a wide range of digital formats
- Linkage to technical infrastructure

- Access and use rights, permissions, and processing rules
- Overall administration and control

Writing for the company GISTICS, Moon and Warwick (2004) explain DAMS as a collection of storage capabilities, security, and business rules. They depict the DAMS infrastructure as a series of management processes: document management, Web content management, fixed-form content management, collaboration content management, and content distribution management. The objective is to provide consistent centralized DAM to facilitate multi-purposed use.

In a series of articles for "DigiCULT Thematic Issue 2," Geser et al. (2002) define digital assets by example:

- Media assets, such as photographs
- Editorial assets (text)
- Informational assets as whole units, such as catalogs, marketing materials, educational curricula, and sound files

To facilitate management and usage, Boiko (2005) structures assets housed in DAMS according to purpose, type, and scope (p.27). He then proposes that functionality is also digital content that has been segmented, encapsulated, and shared (p.31). This speaks directly to the idea of stored programming objects and code snippets. After speaking about organization, Boiko proceeds into a discussion of matching content with data to deliver personalized content within a standard context. He describes the personalization (p.733) as "the process of matching the data that is collected about a user to the metadata used to tag the content." For printed documents, this technique is called VDP.

The ability to use content for multiple purposes, such as the personalization scenario described by Boiko, is made possible by technologies such as eXtensible Markup Language (XML), which separate content from format. Therefore, the same content, or extracts from the content, can be displayed in different styles and on different media. In his presentation, "XML for Multi-channel Publishing," Hinderliter (2005) explains how this is accomplished and why it is such a cost-effective publishing approach. In their book "XML Problem-Design-Solution" (2006), Amiano et al. take the reader through a progression of problems where content is multi-purposed to produce different print and electronic documents. Wagner and Mansfield (2003) discuss similar processes as they explain the different components of XML technology.

Digital Imaging

The content of DAMS is a subject by itself with its own array of complexities. A discussion about images in the context of VDP is needed because it is the other variable in the VDP workflow. Tools such as uDirect and PersonalEffect from XMPie (2006) attest to this. Digital images can vary along with the data, since the data determines the

image used. The selected images have to be prepared and then mated to the data using programming logic.

Image preparation is complex because of the number of image types that may be encountered and because of the types of corrections, adjustments, and enhancements that may be required to make an image suitable for print. Though he does not mention VDP, Gregory Baxes (1994) discusses (primarily) photographic image correction from a scientific perspective. In his discussion of imaging, John Russ (2002) includes a considerable range of photographic types and images, including ones produced by different electron microscopy technologies.

From a graphic arts perspective, authors have concentrated on how to prepare photographs and graphics for use in production output. Subjects may range from the basic use of standard industry tools, such as Adobe Photoshop and Adobe Acrobat and Distiller, to managing color when acquiring an image, preparing it, and finally using it in a press run. Basic Photoshop how-to books are exemplified by Adobe (2005) in "Adobe Photoshop CS2 Classroom in a Book" and McClelland (2005) in "Adobe Photoshop CS2 One-On-One." Both books cover how to prepare documents for print, addressing such topics as color profiles and settings for Adobe Portable Document Files (PDF).

In their books from New Riders Publishing, Haynes and Crumpler (2002) and London and Grossman (2002) go beyond the basics of photographic manipulation, yet stay away from the scientific perspective. Fraser et al. (2005) cover the nature of color and then relate it to source and destination device preparation and imaging workflow processes. In a discussion of VDP with the author, Smith (D. Smith, personal communication, 2006), the president of a digital printing company, explains that some VDP requires verification of output. This means that print-ready images (potentially PDF documents) must be tagged with an index (bar code) and that code simultaneously recorded in a database used in job production. As the physical documents are processed through printing and finishing, they are checked off against the database to verify correct production.

Applying various graphics and data technologies plays a role in producing VDP solutions. A professional level of technical skill and business insight are essential to apply the full capabilities of VDP and accrue optimal benefits for the print customer and the print solution provider. The next section will discuss these skills.

VDP Skills

The VuePoint conference on VDP (April, 2006) tackled the issue of personnel requirements. Above all, the presenters believed that leadership is needed to build VDP expertise in an organization, along with marketing that understands VDP to sell the concept to the organization's customers. According to Parrett (R. Parrett, personal communication, May 2006), if customers do not know what VDP can do, they are not going to know what to ask for. VuePoint (2006) seems to agree, giving specific recommendations

on what skills (titles) should be found on a VDP team: designer, database guru, variable data merge coordinator, and production employees.

Software and solution vendors take a different view. The Christensen interview (personal communication, 2006) with an Oće representative revealed that "mainframe databases have been migrated to client-server because it is more cost effective." The representative also explained that software capabilities take the place of high-level computer skills. In a marketing piece, XMPie (2005) took a similar approach, presenting the idea that the software can "bring it all together" and address all the issues that may be encountered.

Meehan, a practicing VDP consultant, took a somewhat different path when interviewed (B. Meehan, personal communication, August 2006). He said that the key skill is "the ability to translate business logic into code." He went on to say that "multiple language skills and the ability to determine which language is best for creating a given solution" is the important factor. Smith (D. Smith, personal communication, 2006) agrees with this statement. He points out that VDP is not just an end in itself, is not just marketing- or communications-related, but is also a tool for solving business problems that are not necessarily associated with communicating ideas. An analytical understanding of business operations and issues is very important.

During another interview, Parrett (R. Parrett, personal communication, May 2006), also a practitioner, put it in business terms: "More skill means the ability to get more business." He went on to explain: "The shorter [time] it takes to generate a job, the greater the return on investment. A computer science degree is really needed compared to somebody right out of high school. The computer science degree will turn around a job in one day, while the other guy may take a week. So you could hire four high school people or one [person with a] computer science degree to get similar turnaround."

Parrett further carried the need for technical skills to the customer: "A technical service representative facilitates a successful connection to the customer database management system. If you can't help the customer, you may lose the job or you may lose the customer. There are printers who turn away business because they don't know how to do the job." Speaking from considerable IT and VDP experience, he continued: "It's not the equipment; it's the lack of know-how. Hire very talented people – that is defining! Pay to get the level of skill and education needed. Instead of hiring more sales people to bring in more jobs in order to be profitable, maybe it is only a matter of using what you have more efficiently."

In his unpublished paper on implementing VDP, Christensen (2006) categorizes the skills needed according to VDP type:

- Versioning – Not necessary, but some VDP software and data preparation skills will optimize productivity.

- Mail Merge – Some VDP and software skills are necessary; some data preparation skills are necessary as well, and scripting language skills might come in handy.
- Personalized Printing – Extensive DBMS, SQL, and scripting language skills are necessary; some VDP layout and software skills are necessary.
- Transaction Printing – Extensive DBMS, SQL, and scripting language skills are necessary; some VDP layout and software skills are necessary.
- Internet-On-Demand – Extensive VDP layout and software skills are necessary. Extensive Java or C++ programming skills are needed for systems integration. Some DBMS, SQL, and scripting skills are necessary.
- Fully Customized Communications – Extensive VDP layout and software skills are necessary. Extensive Java or C++ programming skills are needed for systems integration. Some DBMS, SQL, and scripting skills are necessary.

Overall, there is considerable disagreement over the kinds of skills needed in a VDP environment. The strongest difference in viewpoint is shown between vendors and practitioners.

Another interesting point is that DAM skills were not mentioned at all, though digital content plays a vital role in the production of marketing literature, whether done in print or in electronic display. While digital imaging and design skills have been an integral part of the graphic arts and are closely associated with printing, skills related to the management of digital images and their use in VDP applications do not yet seem to be a topic of discussion.

Conclusions

The literature demonstrates the as-yet-unrealized opportunities offered by VDP. At the same time, there is a considerable gap between declarative and procedural knowledge that, so far, has not been brought under the umbrella of VDP.

Articles on VDP that mention database connections do so in the most superficial terms. This presents a problem for the neophyte, who may not know enough to put VDP and database together accurately, or may not realize the necessity of reading non-VDP literature to learn about VDP workflows.

While IT and DAM facilitate VDP, a bridge has not been erected across relevant technologies in the literature found. The result is that each technology must be examined separately; this reduces the likelihood that VDP knowledge will be widely disseminated or discussed in depth across its entire breadth.

Research Trials

Given the comments in the prior sections regarding the use of declaratory knowledge without an experiential basis, the obvious next step in this study was to run research trials. The nature of these trials was the establishment and, later, the operation of a VDP testing environment. This step was based upon comments by Sierra and Bates (2005) in their introductory book on the Java programming language, which stated, "a significant amount of detailed analysis and design occurs in creating the test cases prior to developing an application." A considerable amount of procedural knowledge is gained in this way.

The research trials addressed three questions with regard to processes followed, tools used, and skills required:

1. What techniques are needed to prepare data for VDP jobs?
2. What techniques are needed to prepare digital content (images, art, and text) for VDP jobs?
3. How are data and digital content brought together?

In addressing these questions, trial activity followed three steps: integration (hardware and software component selection, installation, interconnection, and tuning), data and content preparation (creation of standard test cases using a variety of data and content storage containers), and data and digital content (text and image) coordination. The significance of this approach from an industry perspective is that businesses implementing digital printing and VDP face a similar series of steps.

Trial Results

The results from the research trials provide a three-point perspective: insights gained from establishing the VDP development and test facility, insights from planning and operating the test facility for VDP production output, and insights into some of the nuances of data-driven workflows.

VDP System Installation and Integration

Implementation of a VDP-capable digital printing system is not a simple matter. The idea that (with the exception of the digital press) the rest of the system components can be implemented with a PC (personal computer) and some off-the-shelf software package is unrealistic.

It took considerable time to assemble the components of the system platform implemented for the research trials, hook them together, and get them to work smoothly.

Implementation continues to be an iterative process as software versions evolve and differences between declarative or "advertised" knowledge are found to vary from knowledge gained experientially, as not everything works as purported.

Taking into account the fact that an existing and experienced IT department at RIT supports the research system, the networking client-server infrastructure already exists, and the digital presses are already installed and in use in established production centers, various installation issues persist.

This part of the implementation concerns the system platform (networking, hardware components, and operating system software) and does not take into account application software critical in managing data and digital content. It does not address questions about data and digital content storage approaches, programmatic tools, VDP design and delivery tools, what tools actually work well together, and what tool combinations can be substituted for others.

Therefore, a shop that is installing, or has just installed, a VDP-capable system for the first time in an environment without the advantages outlined above faces considerable difficulties. It is understandable that some businesses may drop digital printing or decide to forego the digital "opportunity" altogether.

VDP Application Planning and System Operations

Planning and preparation for continued research looks to remain complex, although accessible data and content already exist in test cases. Design and analysis questions posed for initial test case development continue to be relevant when looking at future work. An ad hoc approach appears unworkable.

The trial experience demonstrated that the complexity and the potential business benefits associated with VDP necessitate planning that spans the entire production workflow, starting at the point of project inception. Planning questions (shown in Table 1) are grouped into three broad categories: questions relating to project design and processes, questions specific to data handling, and questions specific to digital asset handling. These questions need to be addressed before starting any development.

Table 1. Initial Planning Questions

Project Design	
What are the project objectives, what will it accomplish, and what are the desired results?	
What decisions have been made regarding project document recipients, given available data?	
What will be the approach for multiple recipients at the same location?	
Data	Digital Content
Has a preliminary data analysis been made?	Has a preliminary digital asset inventory been made?
What data is needed to support the project?	What digital assets are needed to support the project?
How is the data going to be used?	How are the digital assets going to be used?
What data is currently available?	What digital assets are currently available?
How will the data be managed during the project?	How will the digital assets be managed during the project?
Will data be acquired externally, and if so, from where, for how much, and when?	Will digital assets be acquired externally, and if so, from where, for how much, and when?
Data & Content Processing	
Data	Digital Content
Where is the data coming from, and will it be a single or multiple source?	Where are the digital assets coming from?
How will the data be organized – as single or multiple data sets?	How are the digital assets going to be acquired or organized?
What will the data look like?	What will the digital assets look like?
What is the quality of the data?	What is the quality of the digital assets?
Has any preparation been done on the data?	What are the source and target color profiles?
What will intermediate data storage be?	What will intermediate data storage be?
	How will the digital assets be used (as information or decoratively)?

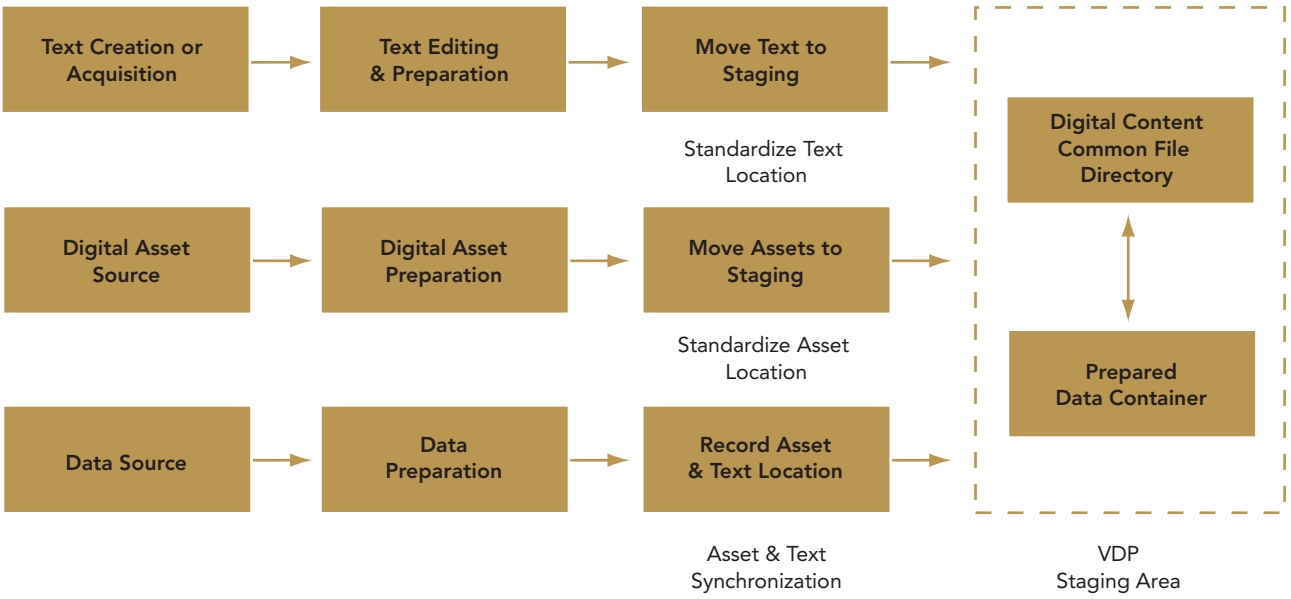
Data-Driven Workflows

The workflows discussed in this section form a conceptual framework that will require additional research and discussion to prove or disprove their validity and continued workability.

Once the VDP test development portion of the trials began, it was obvious that the traditional sequential workflow associated with offset printing would not suffice. This is due to the potential number of variable components that make up a particular design, and it is also due to the fact that those components would most likely require iterative adjustments to fit them together and choreograph them for production.

Digital assets (data, text, and images) are prepared in separate workflows and then brought together into a common production staging storage area as shown in Figure 16.

Figure 16. Data and digital asset preparation overview

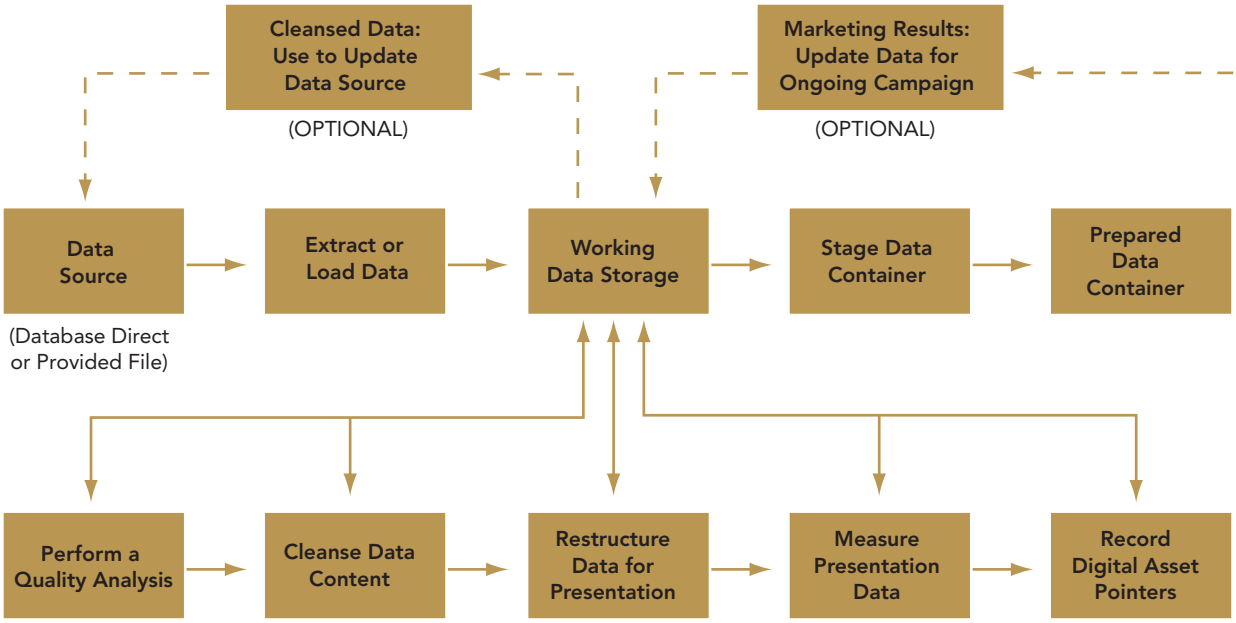


The results of each workflow are synchronized and integrated using one or a combination of approaches: data field pointers to digital assets, references to digital assets, stored uniform resource identifiers (URI's), the insertion of the digital asset into the database as a binary large object (BLOB), and actual placement into the document design for static components that have also gone through the preparation process.

Data Preparation

As a number of authors have said, data drives VDP quality and affects results. Given the potential scope of problems and applied solutions, and considering the likely large number of records or rows in the dataset, a considerable amount of functionality is needed to manipulate and structure the data for successful VDP production. The data preparation workflow is illustrated in Figure 17.

Figure 17. Data preparation workflow



This is not a workflow to delay until data and document design are merged, even when considering the capabilities of FusionPro, as explained in its reference guides (Donaschuk, 2006, and Printable Technologies, 2006), or XMPie PersonalEffect, as described in its reference guides (XMPie, Inc., 2005 and 2006).

The activities that compromise the data preparation workflow are significant. Rud's (2001), and Westfall and Blaxton's (1998) discussion on data mining, as well as Mattison's (1999) discussion on data warehousing and knowledge management, provide a list of considerations and issues when acquiring and preparing data for further use. Data acquisition processes involve more than just getting the data:

- Capture or load
- Transform, organize, and store in a standard storage format
- Eliminate superfluous data
- Administer or manage, providing security and recoverability

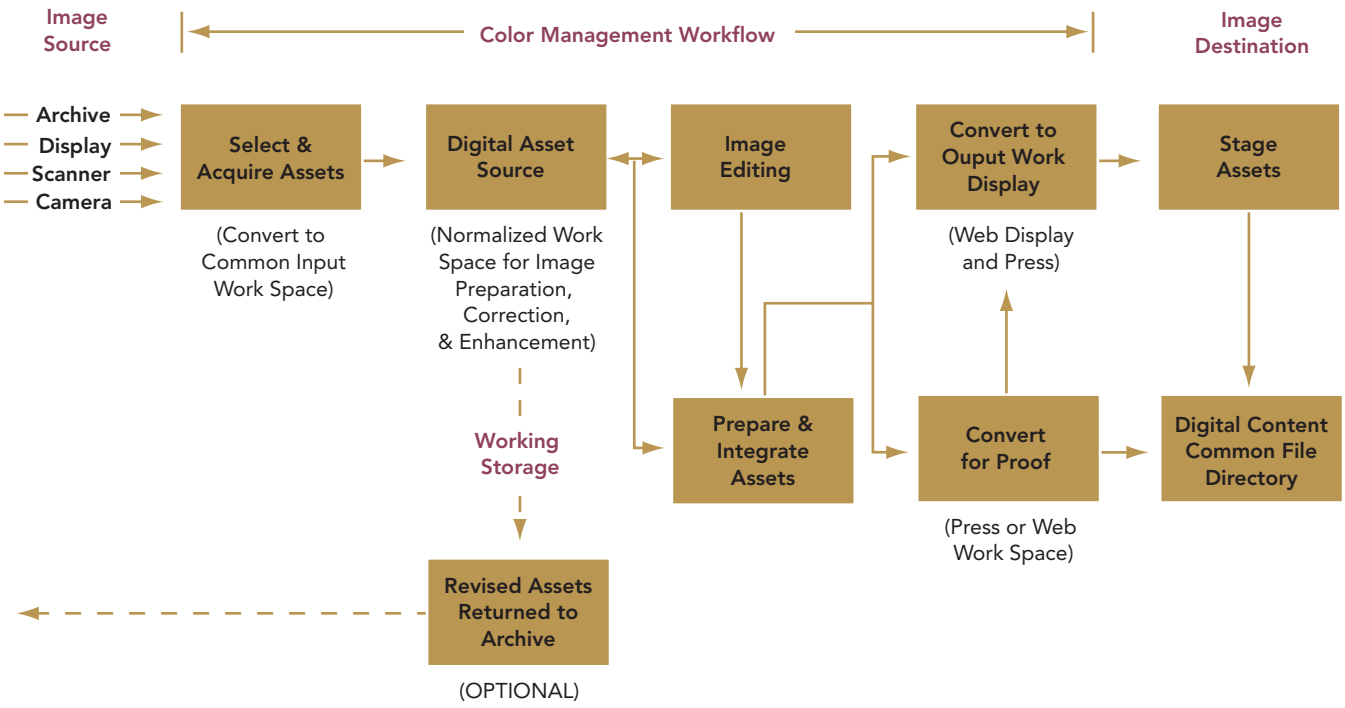
The objective is to ultimately work with a single set of data, but the data may come from multiple sources or reside in multiple datasets that present integration issues. Incoming data may present various difficulties: multiple data structures, schemas, and formats from multiple platforms and operating systems at different security levels, provided from multiple locations on a variety of transport media in multiple spoken languages. Integration may include use of matched data from each set, use of all records from the main data set plus matches from other data sets, or use of the data of one set where no matches are found in the other data sets. The approach depends on project objectives.

Digital Image Preparation

Digital assets are usually prepared in a design, pre-media, prepress sequence. VDP presents an additional issue because images are manipulated and placed on documents according to business logic based on a data stream. Therefore, digital images should be prepared iteratively and in parallel with data.

Digital image preparation, illustrated in Figure 18, can be very complex in terms of the types of issues encountered and the "repair work" potentially needed. The workflow is similar to that for data, but is further complicated by color management requirements to ensure desired colors at press-time.

Figure 18. Digital image preparation workflow

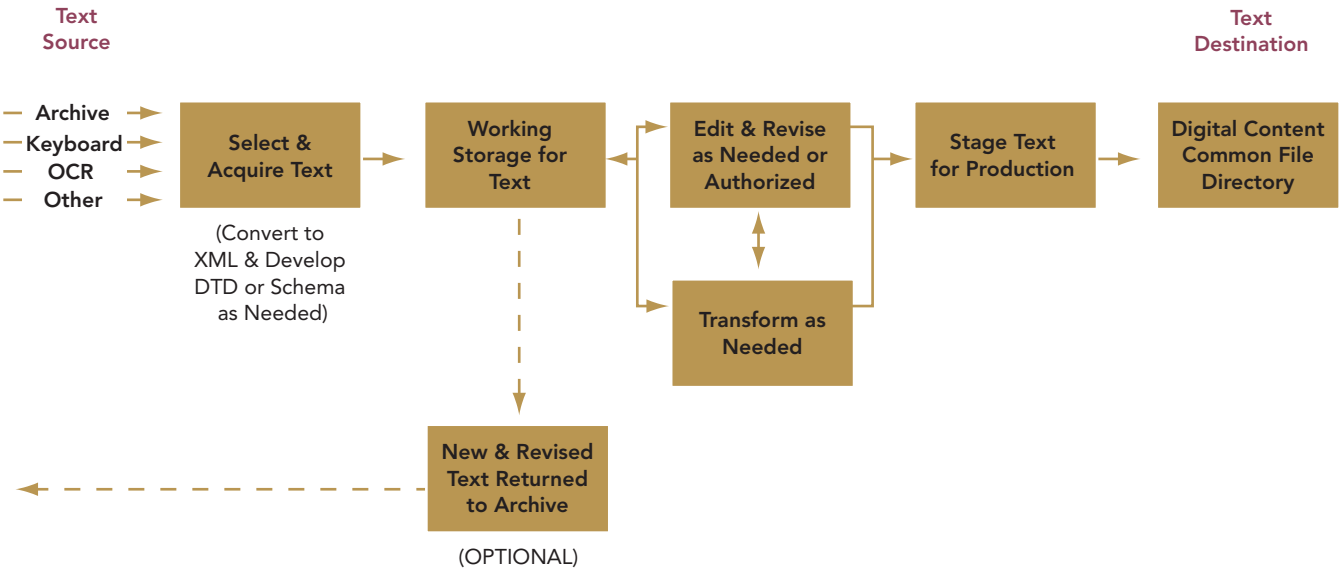


Digital imaging preparation involves a compilation of issues and remedies as described by Adobe (2005), Baxes (1994), and Russ (2002), and also involves the color management concerns presented by Fraser and Bunting (2005). Problems requiring digital content preparation may have a variety of causes, but usually relate to embedded color profiles, embedded fonts, copyrighted fonts and images, file formats, file sizes, and assumptions about appropriate color space – particularly in regard to the generation of Adobe Portable Document Files. Images that have already been processed as a result of artistic work may still require corrective action to make them usable and/or the desired results achievable.

Digital Text Preparation

Further complicating content preparation is the need to deal with the digital text assets, which may have their own unique, processing requirements such as XML-based content. A text-based workflow is pictured in Figure 19.

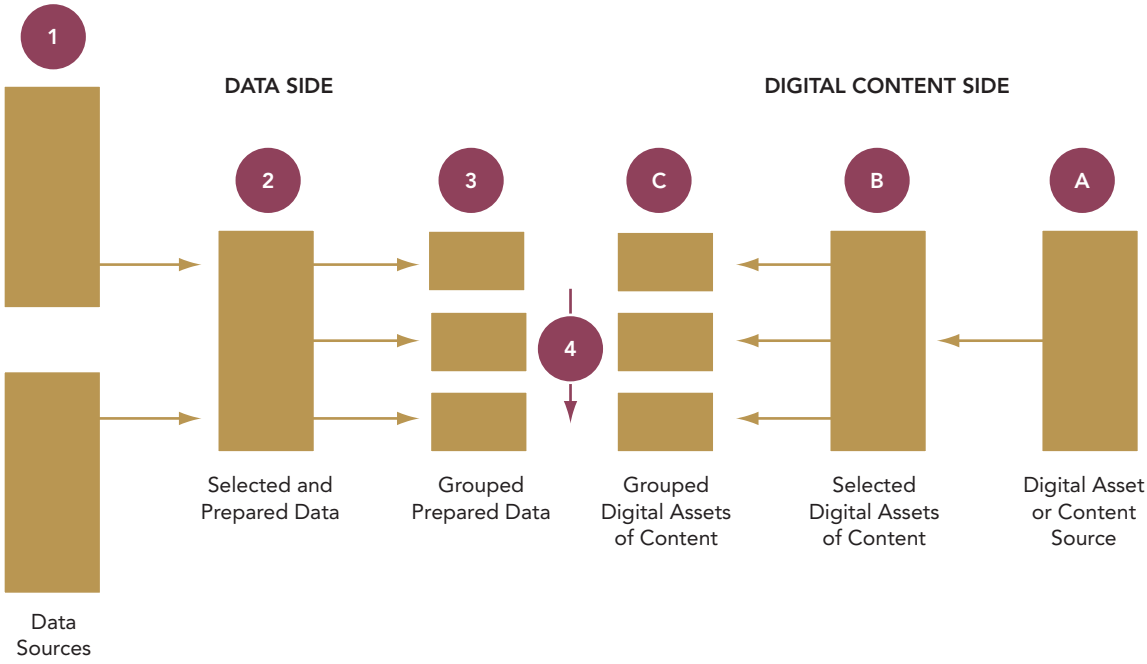
Figure 19. Text preparation workflow



Digital asset preparation that addresses textual content issues is relatively simple, although it can be time-consuming when reviewing text and making necessary corrections or when structured and tagged text components such as XML are involved. Editorial and technical functions that check accuracy, appropriateness, and wording, spelling, grammar, and punctuation, formatting and capitalization, and compliance with XML structure and process conventions are all necessary.

Once data and digital assets have been prepared and moved into a production staging area for data and digital assets, they need to be synchronized. If they were already synchronized during a prior step, then the synchronization needs to be validated. The choreography of data and digital assets (shown in Figure 20) can be complex. Potential benefits need to be researched and quantified.

Figure 20. Data & digital content choreography



1. Data is selected from multiple sources.
2. Data is merged and cleansed, preparing it for ultimate design application.
3. After analysis, data is grouped into logical levels based on postal, RIP, and response optimization strategies.
4. Matched data groups with digital asset and content groups are ready for use in document design and print submission.

A. Digital images and textual content are selected.
B. Digital images and textual content are groomed and edited to meet design requirements.
C. Digital images and textual content are grouped to match associated data elements.

A lot of time and effort, representing a potentially large expense, may have been expended on data, text, and digital image cleanup and preparation. Therefore, reusing the data and digital assets may be economically beneficial for both the customer and the printer. The VDP production data and digital assets can be used to cleanse or augment existing source databases and archives. They can also be used in ongoing communications projects as a basis for campaign management and automation.

Conclusions

This monograph is based upon the concept of supporting information from the literature about VDP with experiential information. The basis for developing procedural knowledge was a set of three questions: what techniques are needed to prepare data for VDP jobs, what techniques are needed to prepare digital content for VDP jobs, and how are data and digital content brought together?

The answers to these questions come to conclusions that differ from the seemingly common notion that VDP service suppliers can simply use a combination of spreadsheets and graphics tools to prepare and combine content into effective communication pieces that cross a variety of media and delivery mechanisms.

Instead, VDP is very complex, requires sophisticated tools and professional skills, and is expensive and time-consuming to successfully establish and operate in order to produce effective quality communications with a production process that is documented, repeatable, predictable, measurable, traceable, and that results in outcomes that are reusable.

A VDP solution requires the interplay of topical knowledge about business and marketing, information content (data and images), and printing applications across the entire print production workflow from inception to delivery. All these topics come into play at conception, and continue to operate together iteratively over the life of the project. Design concept, knowledge acquisition from data, and graphics modify each other as a design emerges.

It does not matter who is responsible for addressing the data and digital asset components in the overall workflow, be it the customer or the service supplier. It only matters that the work is done in conjunction at the proper time. This requirement presents an opportunity for the service supplier to expand the range of services offered, the duration of service delivery, and the degree to which the customer relationship can be enhanced.

A number of tools are available for handling data. Based on this research, most tools employed in VDP have some overlapping functionality – a fact that may confuse their proper application. It is important to choose the right tool for the given job. Experts each have their favorite tools. The decision to select one tool over another does not rest solely on the type of job to be performed, but also on the intent and functionality of each tool being considered.

Constraining the number of tools involved and using the right tools in the right sequence not only streamlines VDP processing, it makes finding the right skill combination simpler. It also simplifies the need to keep technical experts up-to-date with rapidly advancing technologies. The following is a preliminary list of skills needed to adequately support VDP solution delivery:

- Information, knowledge, and application architecture skills
- Networking and systems integration skills
- Business and requirements analysis skills
- Imaging, color management, and DAM skills
- Text manipulation processing skills (including XML and Perl)
- Desktop or server database administration, design, and SQL programming skills
- General object-oriented design and programming skills using compiled and interpreted languages
- Web design and cross-media development skills (including XHTML, CSS, and scripting)

This list can serve as the basis for future research into necessary skill sets and the ways to deploy them in the printing industry.

References

Adobe. (2005). *Adobe Photoshop CS2 classroom in a book*. Berkeley, CA: Peachpit Press.

Amiano, M., D'Cruz, C., Ethier, K., & Thomas, M. (2006). *XML problem-design-solution*. Indianapolis, IN: Wiley Publishing, Inc.

Asbury, S., & Weiner, S. (1999). *Developing Java enterprise applications*. New York, NY: John Wiley & Sons, Inc.

Baxes, G. (1994). *Digital image processing principles and applications*. New York, NY: John Wiley & Sons, Inc.

Bishir, J., & Drewes, D. (1970). *Mathematics in the behavioral and social sciences*. New York, NY: Harcourt, Brace & World, Inc.

Boiko, B. (2005). *Content management bible* (2nd ed.). Indianapolis, IA: Wiley Publishing, Inc.

Bostrup, T. (2003). Database performance philosophy. *15 Seconds Weekly Newsletter*. Retrieved July 7, 2006, from http://www.15seconds.com/issue/040115.htm

Broudy, D., & Romano, F. (1999-2000). *Untitled*. Digital Printing Council White Paper.

InfoTrends/CAP Ventures. (March, 2005). *Production workflow solution product compendium: variable data design and production software*. Retrieved June 28, 2006, from http://www.capv.com/eprise/main/store/pdf/VDP Compendium.pdf

Chapman, K. (n.d.). Marketing through personalized printing & variable data. *Digital Print Production Coordination*, Oregon State University. Retrieved June 28, 2006, from http://printmail.oregonstate.edu/copy/pdf/Direct Marketing.pdf

Christensen, H. (2006). *Implementing variable data printing*. Unpublished manuscript, Rochester Institute of Technology.

Coffey, G., & Prosser, S. (2005). *FileMaker Pro: The missing manual*. Sebastopol, CA: O'Reilly Media, Inc.

Cost, F. (2005). *The new medium of print*. Rochester, NY: RIT Cary Graphic Arts Press.

Davis, D. (2004). *Freeflow variable information workflow.* [Prepared by INTERQUEST, Ltd.]. Stamford, CN: Xerox Corporation. Retrieved June 28, 2006, from http://www.xerox.com/go/xerox/

Davis, D. (2004a). *Workflow for the new business of printing.* [Prepared by INTERQUEST, Ltd.]. Stamford, CN: Xerox Corporation. Retrieved June 28, 2006, from http://a1851.g.akamaitech.net/f/1851/2896/24h/casheA.xerox.com/Freeflow_white_paper_GBR.pdf

Davis, M., & Walter, M. (2003). The next wave: Revolution in process & content. *Seybold Report, 3(15)*, 3-15.

Donaschuk, R. (2006). *FusionPro rules editor reference guide*. Solana Beach, CA: Printable Technologies, Inc.

DuBoise, P. (2006). *MySQL cookbook*. Sebastopol, CA: O'Reilly Media, Inc.

Hewlett-Packard. (2004). *HP Indigo press 3050*. Retrieved June 28, 2006, from http://h30267.www3.hp.com/country/us/en/products/digital_presses/3050_DS_Low.pdf

Hewlett-Packard. (2005). *HP Indigo press 1050*. Retrieved June 28, 2006, from http://h30267.www3.hp.com/country/us/en/products/digital_presses/1050_DS_Low.pdf

Hewlett-Packard. (2006). *HP Indigo press 3250*. Retrieved June 28, 2006, from http://h30267.www3.hp.com/country/us/en/products/digital_presses/3250_DS_Low.pdf

Hewlett-Packard. (2006a). *HP Indigo press 5000*. Retrieved June 28, 2006, from http://h30267.www3.hp.com/country/us/en/products/digital_presses/5000_DS_Low.pdf

Flanagan, D. (2004). *Java examples in a nutshell* (3rd ed.). Sebastopol, CA: O'Reilly Media, Inc.

Flanagan, D. (2005). *Java in a nutshell* (5th ed.). Sebastopol, CA: O'Reilly Media, Inc.

Forta, B. (2006). *MySQL crash course*. Indianapolis, IA: SAMS Publishing.

Fraser, B., Murphy, C., & Bunting, F. (2005). *Real world color management* (2nd ed.). Berkeley, CA: Peachpit Press.

Frey, F., & Christensen, H. (November, 2005). *New skills for DAM and variable data printing services – Is the printing industry ready?* (PICRM-2005-08). Rochester, New York: Printing Industry Center at RIT.

George, R., & Delano, L. (2006). *SQL server 2005 express edition starter kit*. Indianapolis, IA: Wiley Publishing, Inc.

Geser, G., Kantor, N., van Kasteren, J., Moon, M., Ross, S., & Steemson, M. (2002). Digital asset management systems for the cultural and scientific heritage sector. *DigiCULT Thematic Issue 2*. Retrieved June 28, 2006, from http://www.digicult.info/downloads/Thematic Issue 2.pdf

Haynes, B., & Crumpler, W. (2002). *Photoshop 7 artistry*. Indianapolis, IN: New Riders Publishing.

Hinderliter, H. (2005). *XML for multi–channel publishing*. [Powerpoint presentation]. Presented at Rochester Institute of Technology. Rochester, New York.

IPA Workflow TEAM. (2006). *Graphic workflow process*. The Association of Graphic Solution Providers.

Kanonik, F. (2004). Variable data printing: Practicing what you preach. *DigitalPrintInfo IPA Bulletin*. Retrieved June 28, 2006, from http://www.ipa.org/pdf/var-data-ball04.pdf

Koch, G., & Loney, K. (1997). *Oracle 8: The complete reference*. Berkeley, CA: Osborne/McGraw–Hill.

Kodak. (2006). *Kodak NexPress 2100 plus digital production colour press.* Retrieved June 28, 2006, from http://graphics.kodak.com/global/products/printers_presses/production_color/

Kodak. (2006a). *Kodak NexPress 2500 plus digital production colour press.* Retrieved June 28, 2006, from http://graphics.kodak.com/global/products/printers_presses/production_color/

Kofler, M. (2005). *The definitive guide to MySQL 5* (3rd ed.). Berkeley, CA: Apress.

Lehman, E. (2006). *Data driven print teaching supplements*. Rochester, NY: Rochester Institute of Technology, School of Print Media.

Liang, Y. (2000). *Rapid Java application development using JBuilder 3*. Upper Saddle River, NJ: Prentice-Hall, Inc.

London, S. & Grossman, R. (2002). *Photoshop 7 magic*. Indianapolis, IN: New Riders Publishing.

Mattison, R. (1999). *Web warehousing and knowledge management*. New York, NY: McGraw Hill.

McClelland, D. (2005). *Adobe Photoshop CS2 one-on-one*. Sebastopol, CA: O'Reilly Media, Inc.

McCord, A. (2002). *Overview of digital asset management systems*. EDUCAUSE Evolving Technologies Committee, University of Michigan. Retrieved June 28, 2006, from http://www.educause.edu/ir/library/pdf/DEC0203.pdf

Meehan, B. (August, 2006) *Personal interview*. Rochester Institute of Technology, Rochester, New York.

Moon, M., & Warwick, J. (2004). *Business case for dynamic imaging in media service portals*. Oakland, CA: GISTICS Incorporated.

Myers, C., III. (April, 2006). Untitled slide presentation. *Cross Media Workflow II.* PowerPoint classroom presentation, Rochester Institute of Technology, Rochester, New York.

Nichol, S. (1997). *3-Tier for improved database throughput*. Retrieved July 7, 2006, from http://www.scottnichol.com/3-tier1.htm

Oċe. (2006). *Interview with a representative of Oċe North America*. [Henrik H. Christensen, Interviewer]. Rochester Institute of Technology, Rochester, New York.

Parrett, R. (May, 2006). *Interview with Roger Parrett of Kodak Versamark.* [Henrik H. Christensen, Interviewer]. Rochester Institute of Technology, Rochester, New York.

Pellow, B., Sorce, P., Frey, F., & Banis, H. (2003). *Marketing communication demand creation: Marketing executive study* (PICRM -2003-06). Rochester, NY: Printing Industry Center at RIT.

Printable Technologies, Inc. (2006). *FusionPro user's guide*. Solana Beach, CA: Printable Technologies, Inc.

Reese, G. (2003). *MySQL pocket reference*. Sebastopol, CA: O'Reilly Media, Inc.

Roman, S. (1997). *Access database design & programming*. Sebastopol, CA: O'Reilly & Associates, Inc.

Romano, F. (2005). S*tatus of printing in the United States 2005.* Nashua, NH: PennWell Corporation.

Romano, F. (2006). Distinguished professor Frank Romano discusses "the state of printing." *PUB Quarterly, Fall Quarter 2006*. Rochester, NY: Rochester Institute of Technology.

Rud, O. (2001). *Data mining cookbook: Modeling data for marketing, risk, and customer relationship management.* NY: John Wiley & Sons, Inc.

Russ, J. (2002). *The image processing handbook* (4th ed.). Boca Raton, FL: CRC Press LLC.

Sierra, K., & Bates, B. (2005). *Head first Java* (2nd ed.). Sebastopol, CA: O'Reilly Media, Inc.

Smith, D. (2006). *Interview on variable data print production.* [Interview by author]. Rochester, NY: Merlin International, Inc.

Sorce, P., & Pellow, B. (2003). *Demand for customized communications by advertising agencies and marketing executives* (PICRM-2003-08). Rochester, NY: Printing Industry Center at RIT.

Sorce, P., & Pletka, M. (2004). *Digital printing success models: Validation study (2004)* (PICRM-2004-06). Rochester, NY: Printing Industry Center at RIT.

Sorce, P., & Pletka, M. (2006). *Data–driven print: Strategy and implementation*. Rochester, NY: RIT Cary Graphic Arts Press.

SQL-Server-Performance.Com. (2006). *SQL server performance tuning articles*. Retrieved July 7, 2006, from http://www.sql-server-performance.com/articles_performance.asp

Throckmorton, B. (2006). *Presentation on Raine Media*. Presented at Rochester Institute of Technology, Rochester, New York.

Tolliver–Nigro, H. (2005). Better together: VDP and CRM [parts 1, 2, and 3]. *OnDemandJournal.com*. [PDF]. Retrieved from Rochester Institute of Technology online course content October, 2005.

Vanslette, J. (2006). *Personal interview on digital printing*. Rochester, NY: Rochester Institute of Technology, Printing Application Lab.

VuePoint. (April, 2006). *Variable data printing.* VuePoint 2006 Conference, Walt Disney World Resort, Orlando, Florida.

Wagner, R., & Mansfield, R. (2003). *XML all-in-one desk reference for dummies*. New York, NY: Wiley Publishing, Inc.

Westphal, C. & Blaxton, T. (1998). *Data mining solutions: methods and tools for solving real–world problems.* New York, NY: John Wiley & Sons, Inc.

Winer, R. (2001). *Customer relationship management: a framework, research direction, and the future.* Retrieved June 28, 2006, from http://groups.haas.berkeley.edu/fcsuit/PDF-papers/CRM paper.pdf

Xerox. (2006). *Xerox iGen3 90 digital production press.* Xerox Corporation. Retrieved June 28, 2006, from http://www.xerox.com/go/xrx/equipment/product_details/FILE_PROD_DocuColor_iGen3_90_Specs

Xerox. (2006a). *Xerox iGen3 110 digital production press*. Xerox Corporation. Retrieved June 28, 2006, from http://www.xerox.com/go/xrx/equipment/product_details/FILE_PROD_DocuColor_iGen3_Specs

Xinet. (2006). *Xinet solutions* (Web site). Retrieved December 4, 2006, from www.xinet.com

XMPie, Inc. (2005). *PersonalEffect modular dynamic publishing solutions*. Retrieved June 28, 2006, from http://www.xmpie.com/

XMPie, Inc. (2005a). *Proof set viewer user guide*. Netanya, Israel: XMPie, Ltd.

XMPie, Inc. (2005b). *uDirect professional.* Retrieved June 28, 2006, from http://www.xmpie.com/

XMPie, Inc. (2006c). *uPlan user guide 2.0*. Netanya, Israel: XMPie, Inc.

Rochester Institute of Technology
College of Imaging Arts and Sciences
55 Lomb Memorial Drive
Rochester, NY 14623
Phone: (585) 475-2733
http://print.rit.edu