Methodology and Technology Services

Google™

www.ies.aust.
Web   com

| **Home** | **Courses** | **Certification** | **Projects** | **Papers** | **EA Blog** | **Online Store** | **Contact Us** |

Home
Courses
Certification
Projects
Papers
TEN Archive
EA Blog
Contact Us
Search
Links
Online Store

XML AND CORPORATE PORTALS

Printable PDF Version

Clive Finkelstein

Extract from "*Building Corporate Portals with XML*"
by Clive Finkelstein and Peter Aiken,
McGraw-Hill (Sep 1999) [ISBN: 0-07-913705-9]

"I know the data is there, but I can't get the information I need!" How many times have you heard this cry from management? But you are not alone; the same cry has been expressed in most languages around the world. It is a common problem: the data is in the computer, but cannot be located readily; or it is not in a format that is suitable for use by management. So what do you do?

Corporate Portals, also called Enterprise Portals (EPs) or Enterprise Information Portals (EIPs), are based on Data Warehousing technologies, using Metadata and the Extensible Markup Language (XML) to integrate both structured and unstructured data throughout an enterprise. Metadata, XML and EPs will be vital elements of the 21st century enterprise. This White Paper briefly introduces basic concepts of metadata, XML and Enterprise Portals. It is an extract from: "*Building Corporate Portals with XML*" by Clive Finkelstein and Peter Aiken, published by McGraw-Hill in September 1999 [ISBN: 0-07-913705-9].

Structured data exists in databases and data files that are used by current and older operational systems in an enterprise. We call these older systems legacy systems; we call the data they use legacy data. In most enterprises, structured data comprises only 10% of the data, information and knowledge resources of the business; the other 90% exists as unstructured data in textual documents, reports or email, or as graphics and images, or in audio or video formats. These unstructured data sources are not easily accessible to Data Warehouses, but EPs use metadata and XML to integrate both structured and unstructured data seamlessly, for easy access throughout the enterprise.

1. What is Metadata?

IT staff in most enterprises have a common problem. How can they convince managers to plan, budget and apply resources for metadata management? What is metadata and why is it important? What technologies are involved? Internet and Intranet technologies are part of the answer and will get the immediate attention of management. XML is the other technology. The following analogy may help you outline to management the important role that metadata takes in an enterprise.

Every country is now interconnected in a vast, global telephone network. We are now able to telephone anywhere in the world. We can phone a number, and the telephone assigned to that number would ring in Russia, or China, or in Outer Mongolia. But when it is answered, we may not understand the person at the other end. They may speak a different language. So we can be connected, but what is said has no meaning. We cannot share information.

Today, we also use a computer and the World Wide Web. We enter a web site address into a browser on our desktop machine -- a unique address in words that is analogous to a telephone number. We can then be connected immediately to a computer assigned to that address and attached to the Internet anywhere in the world. That computer sends a web page based on the address we have supplied, to be displayed in our browser. This is typically in English, but may be in another language. We are connected, but like the telephone analogy -- if it is in another language, what is said has no meaning. We cannot share information.

Now consider the reason why it is difficult for some of the systems used in an organization to communicate with and share information with other systems. Technically, the programs in each system are able to be interconnected and so can communicate with other programs. But they use different terms to refer to the same data that needs to be shared. For example, an accounting system may use the term "customer" to refer to a person or organization that buys products or services. Another system may refer to the same person or organization as a "client". Sales may use the term "prospect". They all use different terminology -- different language -- to refer to the same data and information. But if they use the wrong language, again they cannot share information.

The problem is even worse. Consider terminology used in different parts of the business. Accountants use a "jargon" -- a technical language -- which is difficult for non-accountants to understand. So also the jargon used by engineers, or production people, or sales and marketing people, or managers is difficult for others to understand. They all speak a different "language". What is said has no meaning. They cannot easily share common information. In fact in some enterprises it is a miracle that people manage to communicate meaning at all!

Each organization has its own internal language, its own jargon, which has evolved over time so similar people can communicate meaning. As we saw above, there can be more than one language or jargon used in an organization. Metadata identifies an organization's own "language". Where different terms refer to the same thing, a common term is agreed for all to use. Then people can communicate more clearly. And systems and programs can intercommunicate with meaning. But without a clear definition and without common use of an organization's metadata, information cannot be shared effectively throughout the enterprise.

Previously each part of the business maintained its own version of "customer", or "client" or "prospect". They defined processes -- and assigned staff -- to add new customers, clients or prospects to their own files and databases. When common details about customers, clients or prospects changed, each redundant version of that data also had to be changed. It requires staff to make these changes. Yet these are all redundant processes making the same changes

to redundant data versions. This is enormously expensive in time and people. It is also quite unnecessary.

The importance of metadata can now be seen. Metadata defines the common language used within an enterprise so that all people, systems and programs can communicate precisely. Confusion disappears. Common data is shared. And enormous cost savings are made. For it means that redundant processes (used to maintain redundant data versions up-to-date) are eliminated, as the redundant data versions are integrated into a common data version for all to share.

---

2. What is XML?

Much effort has earlier gone into the definition and implementation of Electronic Data Interchange (EDI) standards to address the problem of intercommunication between dissimilar systems and databases. EDI has now been widely used for business-to-business commerce for many years. It works well, but it is quite complex and very expensive. As a result, it is cost-justifiable generally only for large corporations.

Once an organization's metadata is defined and documented, all programs can use it to communicate. EDI was the mechanism that was used previously. But now this intercommunication has become much easier.

Extensible Markup Language (XML) is a new Internet technology that has been developed to address this problem. XML can be used to document the metadata used by one system so that it can be integrated with the metadata used by other systems. This is analogous to language dictionaries that are used throughout the world, so that people from different countries can communicate. Legacy files and other databases can now be integrated more readily. Systems throughout the business can now coordinate their activities more effectively as a direct result of XML and management support for metadata. We discuss XML fully in Chapter 11 of this book: "*Building Corporate Portals with XML*".

XML now provides the capability that was previously only available to large organizations through the use of EDI. XML allows the metadata used by each program and database to be published as the language to be used for this intercommunication. But distinct from EDI, XML is simple to use and inexpensive to implement for both small and large organizations. Because of this simplicity, we like to think of XML as:

"XML is EDI for the Rest of Us"

XML will become a major part of the application development mainstream. It provides a bridge between structured and unstructured data, delivered via XML then converted to HTML for display in web browsers. Together with metadata, XML is a key component in the design, development and deployment of Enterprise Portals. We discuss XML for Business Reengineering in Chapter 12, and for Systems Reengineering in Chapter 13, of this book: "Building Corporate Portals using XML". We show how Data Warehouses evolve into Enterprise Portals in Chapter 15 of the book.

---

3. How Is Metadata Used with XML?

Metadata is used to define the structure of an XML document or file. Metadata is published in a Document Type Definition (DTD) file for reference by other systems. A DTD file defines the structure of an XML file or document. It is analogous to the Database Definition Language (DDL) file that is used to define the structure of a database, but with a different syntax.

An example of an XML document identifying data retrieved from a PERSON database is illustrated in Figure 1. This includes metadata markup tags (surrounded by < … >, such as <person_name>) that provide various details about a person. From this, we can see that it is easy to find specific contact information in <contact_details>, such as <email>, <phone>, <fax> and <mobile> (cell phone) numbers. Although we have not shown it here, the DTD also specifies whether certain tags must exist or are optional, and whether some tags can exist more than once -- such as multiple <phone> and <mobile> tags below.

```
<PERSON person_id="p1100" sex="M">
  <person_name>
    <given_name>Clive</given_name>
    <surname>Finkelstein</surname>
  </person_name>
  <company>
    Information Engineering Services Pty Ltd
  </company>
  <country>Australia</country>
  <contact_details>
    <email>cfink@ies.aust.com</email>
    <phone>+61-8-9402-8300</phone>
    <phone>(08) 9309-6163</phone>
    <fax>+61-8-9402-8322</fax>
    <mobile>+61-411-472-375</mobile>
    <mobile>0411-472-375</mobile>
  </contact_details>
</PERSON>
```

Figure 1: An example of an XML document with metadata tags (surrounded by < … >) identifying the meaning of following data

Metadata that is used by various industries, communities or bodies can be used with XML to define markup vocabularies. The World Wide Web Consortium (W3C) has developed a standard framework that can be used to define these vocabularies. This is called the Resource Description Framework (RDF). It is a model for metadata applications that support XML. RDF was initiated by the W3C to build standards for XML applications so that they can inter-operate and intercommunicate more easily, avoiding the communication problems that we discussed earlier.

With XML, many applications that were difficult to implement before -- often due to metadata differences -- now become possible. For example, an organization can define the unique metadata used by each supplier's legacy inventory systems. This enables the organization to place orders via the Internet directly with those suppliers' systems, for automatic fulfillment of product orders.

XML is enabling technology to integrate structured and unstructured data for next generation E-Commerce and EDI applications. Web sites will evolve to use XML, with far greater power and flexibility than offered by HTML. Netscape Communicator 5.0 and Microsoft Internet Explorer 5.0 browsers both support XML. Most productivity tools and office suites will support XML. For example, Microsoft Office 2000 uses XML to maintain the internal formats and styles used by

Word, Excel and PowerPoint when converted to HTML, so that those HTML documents can later be opened again by the same originating source products without losing relevant formatting detail. Business Intelligence and Knowledge Management tools will support XML. XML development tools are also being released so that XML applications can be developed more easily.

The acceptance of XML is progressing rapidly, as it offers a very simple -- yet extremely powerful -- way to intercommunicate between different databases and systems, both within and outside an organization. How well an organization accesses and uses its knowledge resources can determine its competitive advantage and future prosperity. Use and application of knowledge will become even more important in the competitive Armageddon of the Internet, in which we will all participate.

The tools are coming, but a greater task still remains to be completed. This is the definition of your own metadata, your common enterprise language for intercommunication, so that you can use these tools effectively. The definition of metadata depends on knowledge of data modeling, previously carried out by IT people. But this is not just a task for IT. As it is vitally dependent on business knowledge, it also requires the involvement of business experts. Not by interview, but by their active participation.

While data modeling has until now been a technical IT discipline, business data modeling is not. It can be learned by business people as well as IT staff. It is based on strategic business planning (discussed in Chapter 2), data modeling (Chapter 3), strategic modeling (Chapter 4) and decision early warning (Chapter 5). This uses Forward Engineering techniques to identify metadata, based on management information needs for the future. Metadata is also extracted from legacy databases and systems using Reverse Engineering, and is discussed in Chapters 6 - 10.

---

4. The Impact of Technology

One thing we are not short of today, is information. We are swimming in it! Our information comes from traditional printed sources such as books, magazines, newspapers, subscription reports and newsletters; from audio sources such as radio; from video sources such as free-to-air television or cable TV; from email and from word-of-mouth. The saving grace with these information sources -- apart from radio and free-to-air TV -- is that they are limited only to those who have subscribed to receive that information.

Not any more. Even today, and certainly more so in the future, each of these sources is moving to the Internet. They are offered as free services, where the cost of preparation is paid not by subscription but by advertising. Even word-of-mouth, previously a reliable source of information from people you knew personally and whose opinion you respected, has moved to the Internet in newsgroups and chat rooms -- but with opinions offered by people, perhaps in another country, who are totally unknown to you. Both accurate and inaccurate comment now circle the globe not at word-of-mouth speed, but at electronic speed.

Email is the killer application of the Internet; even more so of the corporate Intranet. Enormous knowledge is retained in corporate email archives -- much to the chagrin of Microsoft, with certain email messages used by government prosecutors in the Microsoft Antitrust trial as smoking guns to illustrate alleged abuses of monopoly power. Corporate email is a knowledge resource that is of great value, yet until now it has been largely inaccessible.

Text searches on the Internet by traditional search engines are largely ineffective; a simple query can return thousands of links containing the entered keywords or search phrase. Only a small fraction of these may be relevant, yet each link must be manually investigated to assess its content -- if relevancy ratings are not also provided.

The problem is no less severe with enterprises. We are inundated with information. To the credit of the Information Technology (IT) industry, at least this information is being organized and made more readily available through Data Warehouses. Most information in Data Warehouses is based on structured data sources as operational databases used by older legacy systems and relational databases. Data Warehouse products are also now becoming available that use Internet technologies. These valuable information tools can be used within an enterprise across the corporate Intranet. The information is thus more readily available.

We discussed earlier that structured data represents only 10% of the information and knowledge resource in most enterprises. The remaining 90% exists as unstructured data that has been largely inaccessible to Data Warehouses. Text documents, email messages, reports, graphics, images, audio and video files all are valuable sources of data, information and knowledge that have been untapped. They exist in physical formats that have been difficult to access by computer -- as if they were behind locked doors.

The technologies are now available to open these doors. XML is one technology, as we have briefly seen. XML enables structured and unstructured data sources to be integrated easily, where this was extremely difficult before. Organizations will develop new business processes and systems based on this integration, using Business Reengineering and Systems Reengineering methods. They will at last be able to break away from the business process constraints that have inhibited change in the past.

4.1 Process Technologies in The Industrial Age

Most organizations today still use processes based on principles that are no longer effective. They were designed using the process engineering "bible". Here is a short quiz: which book are we referring to? Who was the author? When was it published?

Was the process engineering bible written by Michael Hammer, acknowledged by many as the "Father" of Business Process Reengineering [Hammer 1990]? Was it [Hammer & Champy 1993]? No, it was before them …

Was it written by Ed Yourdon, Tom deMarco, Ken Orr or Gane and Sarson -- all giants of the Structured Software Engineering era, which was process-driven? No to all of these …

Was it written by Edwards Deming, regarded by many as the "Father" of the quality movement? No, not him …

What about Peter Drucker, considered the "Father" of management gurus? Not him, either …

Was it Henry Ford, the "Father" of the assembly line?

No, not him …Yet all of these giants have contributed in their separate ways to improve the design, operation and functioning of enterprises and of information systems. We owe them all

our thanks; we are in their debt. They contributed greatly to the theory and practice of management, of organization and process design, of systems design and development. We draw on their works many times throughout this book.

No, the process engineering bible was written long before each of these esteemed gentlemen.

We are in fact referring to "The Wealth of Nations" by Adam Smith, written around 1776, published most recently in [Smith 1910]. This has been the basis of most business processes used in enterprises today!

Expressing what he wrote, but in today's terminology, Adam Smith took complex processes and broke them down into simple steps. These were then carried out using the technology of his day -- a workforce that was largely illiterate. He showed that people could be trained to carry out these simple process steps, which they repeated endlessly. He then combined each of these steps in different ways to build complex processes. While we have greatly simplified above what he wrote and translated it into today's environment, essentially this was its impact. For these became the processes that fueled the Industrial Age.

Organizations grew as complex processes were built in this way. Manual technologies also used other technologies to supplement them. Mechanical technologies, electrical, electronic and other technologies lead to corresponding engineering disciplines: mechanical engineering, electrical engineering etc. Yet the basic principle behind all of these processes was the work done by Adam Smith.

Henry Ford made a great contribution, with the assembly line. But still essentially the same approach was being used to design processes. And as these processes were automated, they were implemented on computer in much the same way as the processes were carried out in the enterprise. The computer was used basically to do the same tasks, yet faster and more accurately.

The processes referred to relevant data. Each part of the enterprise maintained its own copy of the data that was required. As the processes were automated, the data was also automated. The same data was implemented often in different versions, redundantly. The Information Engineering (IE) methodology, developed from 1976, was designed to address this problem -- evolving in the mid 1980s into Enterprise Engineering (EE) [Finkelstein 1981a, 1981b, 1989, 1992].

By the late 1980s, the inhibiting factor in the effectiveness and operation of processes in many enterprises was seen to be due to this evolutionary approach to business process design. The Business Process Reengineering (BPR) revolution of the early 1990s began to address these problems. This was largely started by Michael Hammer in his landmark paper, provocatively titled: "Reengineering Work: Don't Automate, Obliterate!" [Hammer 1990].

XML and Enterprise Portals offer technologies that will progress these methods further. We discuss their impact on Business Reengineering and on Systems Reengineering in Chapters 12 and 13 of this book: "Building Corporate Portals using XML".

4.2 Data Technologies in the Information Age

Our focus in this book is on Data Warehouses and Enterprise Portals. Data Warehouses provide access to structured data as discussed earlier. We introduce Enterprise Portals here.

The term "Enterprise Information Portal" (EIP) we believe was first used in a report published by Merrill Lynch on November 16, 1998. A summary of this report is available from the [SageMaker] web site. The full report can be downloaded in Adobe Acrobat Portable Document Format (PDF) file from this same web site. The Merrill Lynch summary and report define EIPs as:

> *"Enterprise Information Portals are applications that enable companies to unlock internally and externally stored information, and provide users a single gateway to personalized information needed to make informed business decisions.*
>
> *Enterprise Information Portals (EIP) are an emerging market opportunity; an amalgamation of software applications that consolidate, manage, analyze and distribute information across and outside of an enterprise (including Business Intelligence, Content Management, Data Warehouse and Mart, and Data Management applications."* ... Merrill Lynch: Nov 16, 1998 [SageMaker] Web Site.

The Merrill Lynch report and summary highlight the emergence of Enterprise Information Portals as an investment opportunity for their clients and others. InfoWorld presented a summary of the report as a Front Page article of the January 25, 1999 issue. A copy of that article is available from the [InfoWorld] web site. A financial summary of the potential of the EIP market from the Merrill Lynch report was provided in the InfoWorld article. This is reproduced here as Figure 2. The summary states:

> *"We have conservatively estimated the 1998 total market opportunity of the EIP market at $4.4 billion. We anticipate that revenues could top $14.8 billion by 2002, approximately 36% CAGR* (Compound Annual Growth Rate) *for this sector."*

As Figure 2 illustrates, software is required for Content Management, which is projected to grow from a market worth $1.2 billion in 1998 to one worth $4.7 billion in 2002. Products in the Business Intelligence EIP market are expected to grow from $2.0 billion to $7.2 billion. The Data Warehouse and Data Mart EIP market is projected to grow from nearly $1 billion to $2.5 billion, while the Data Management market will grow from $184 million to $360 million. The total EIP market therefore was projected in the Merrill Lynch report to grow from $4.4 billion to $14.8 billion over the period 1998 to 2002.

Discussing the potential of the EIP market, the authors of the Merrill Lynch report believe it will "*eventually reach or exceed the investment opportunities provided by the Enterprise Resource Planning (ERP) market.*" They give three main reasons why:

"Enterprise Information Portals will emerge from a consolidation within and between the Business Intelligence, Content Management, Data Warehouse, Data Mart and Data Management markets:

EIP market starts to take off

Estimated total EIP market revenues (in millions)

Content management	
1998	$1,225
1999	$1,650
2000	$2,300
2001	$3,250
2002	$4,700

Business intelligence	
1998	$2,000

- **EIP systems provide companies with a competitive advantage:** *Corporate management is just realizing the competitive potential lying dormant in the information stored in its enterprise systems. … EIP applications combine, standardize, index, analyze and distribute targeted, relevant information that end users need to do their day-to-day jobs more efficiently and productively. The benefits include lowered costs, increased sales and better deployment of resources.*
- **EIP systems provide companies with a high return on investment (ROI):** *The emergence of 'packaged' EIP Applications are more attractive to customers because they are less expensive than customized systems, contain functionality that caters to specific industries, are easier to maintain and faster to deploy. … EIP products help companies cut costs and generate revenues.*
- **EIP systems provide access to all:** *The Internet provides the crucial inexpensive and reliable distribution channel that enables companies to make the power of information systems available to all users (employees, customers, suppliers). Distribution channels include the Internet, Intranet and Broadcasting. … Companies will need to use both "publish" (pull) and "subscribe" (push) mediums to ensure the right information is available or distributed to the right people at the right time."* They go on to say that they: *"envision the Enterprise Information Portal as a Browser-based system providing ubiquitous access to business related information in the same way that Internet content portals are the gateway to the wealth of content on the Web."*

Business intelligence	
1998	$2,000
1999	$2,700
2000	$3,700
2001	$5,180
2002	$7,250

Data warehouse and data marts	
1998	$992
1999	$1,235
2000	$1,562
2001	$1,993
2002	$2,555

Data management	
1998	$184
1999	$220
2000	$261
2001	$309
2002	$360

Total market	
1998	$4,401
1999	$5,805
2000	$7,823
2001	$10,732
2002	$14,865

SOURCE: MERRILL LYNCH & CO. INC.

Figure 2: *Enterprise Information Portal Market.*
Source: [InfoWorld] and [SageMaker] Web Sites.

The Merrill Lynch report and the InfoWorld Front Page article triggered a flurry of articles in other publications. Software companies in these markets scrambled to refocus their software development plans to deliver products for the new emerging market that had been identified.

4.3 Enterprise Information Portal Directions

The market potential had been identified, the software vendors had begun to develop products, but there was no clear definition of the EIP market apart from general directions in the Merrill Lynch report. And there was no technical guidance that would help software vendors and their

enterprise customers to build these Enterprise Information Portals.

The report also affected ourselves: your authors. We had been writing a book on Data Warehousing. Our purpose was to publish a book that would to help enterprises move their Data Warehouses and Data Marts to the Internet, Intranet and Extranet. We felt that this would provide benefit to the enterprises, their employees, customers, suppliers and business partners.

This was a difficult task to do, as another author who we respected had found. Richard Hackathorn had published "Web Farming for the Data Warehouse" [Hackathorn 1999]. He was writing this around the time when the groundswell of support for XML had begun to build up following its acceptance as a recommended standard by the W3C Committee in February 1998 [W3C].

As discussed earlier in this chapter, XML is an technology that enables many applications and databases to overcome the great constraints of legacy systems and databases that had evolved as redundant data versions. We saw it also as an important component to move Data Warehouses and Data Marts to the Web. The Merrill Lynch report identified the market potential that justified what was, until then, just a "gut feel" for us. It highlighted a glaring omission; the absence of clear technical direction on how to build for this new environment. As authors, we do not pretend to have all of the answers. But this is our field; having built many Data Warehouses, Data Marts, Web Sites and Electronic Commerce applications as consultants, instructors and webmasters over many years.

We will share our knowledge with you in this book. The three of us, together, will discuss problems and solutions. And there will be others after us who will add more, based on their experience. They will also write, or consult or teach: this new discipline will further evolve -- that is the nature of the Information Technology industry.

4.4 Enterprise Portal Terminology

A number of terms have emerged along with the growing interest in Enterprise Information Portals. Internet Content Portals such as NetCenter (Netscape), MyYahoo (Yahoo), MSN (Microsoft) and AOL became popular in 1998 as a central point that could be visited by millions on the Internet -- as a gateway or jumping-off point to other locations on the World Wide Web. Some of these are content providers; others are search engines. The terminology differs, but we feel a general term describing all of these is "Internet Portal". This is the term we will use in this book for reference to WWW consumer portals.

In the many articles that have appeared since publication of the Merrill Lynch report and the InfoWorld article, the terms "Enterprise Information Portal" (EIP), "Corporate Portal" (CP) and "Enterprise Portal" (EP) have been variously used.

"Enterprise Information Portal", being the first used, is the obvious term. But we find many articles are using "Enterprise Portal" and "Corporate Portal" as equivalent terms to refer to an EIP. This is a new field and the terminology has not settled yet. So we will use all three terms interchangeably in this book to refer to portals for all enterprises: large Corporations; Small or Medium Enterprises (SMEs); Federal, State or Local Government departments; and Defense departments.
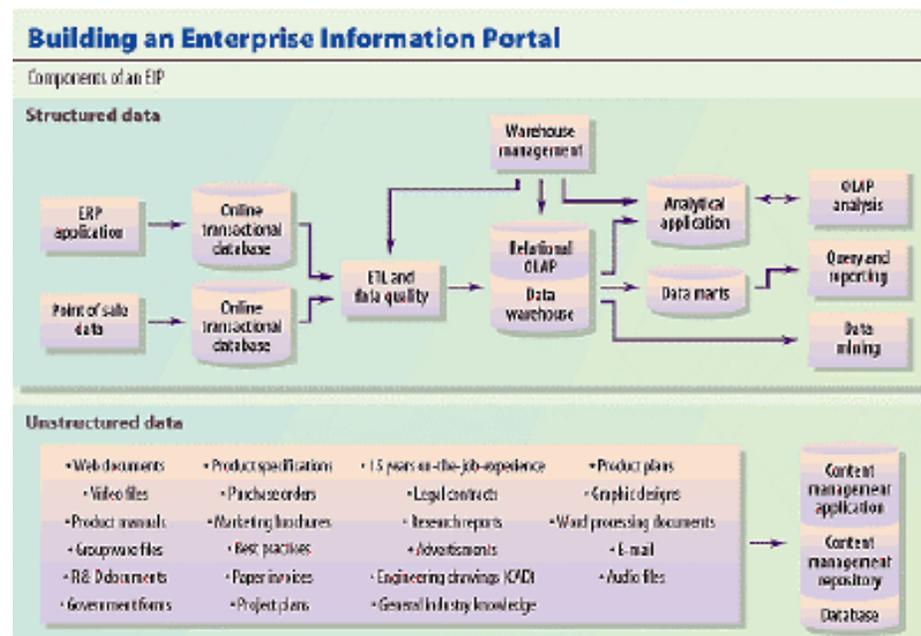
4.5 Enterprise Portal Concepts

We will introduce some of the basic concepts of an Enterprise Portal in this section, with related

concepts covered later in this chapter. The remainder of the book will progressively introduce you to the concepts and methods that can be used to build Enterprise Portals.

In Part I: Enterprise Portal Design, there is a great parallel with Data Warehouse design. Part II: Enterprise Portal Development also parallels Data Warehouse development. Our focus in these two parts is therefore mainly on Data Warehouses and Data Marts.

In Part III: Enterprise Portal Deployment we cover XML in Chapter 11. XML is an enabling technology that offers great benefit for Business Reengineering and Systems Reengineering. These are covered in Chapters 12 and 13. Many enterprises are struggling to move out from under the weight of legacy systems and processes that are not appropriate or responsive enough for the Information Age. Enterprise Portals and XML will enable these enterprises to transform themselves more effectively, without first having to throw all those legacy systems away and develop new systems at great cost. Chapter 14 addresses quality in these transformed enterprises.

Finally, in Chapter 15 we will return to discuss the central role of Enterprise Portals, summarizing the main points from the book.

Figure 3: *Enterprise Portal Concepts. Source: [InfoWorld] Web Site.*

The main concepts of Enterprise Portals are illustrated in Figure 3, from the InfoWorld article on the [InfoWorld] web site. The focus of Data Warehouses is Structured Data, shown in the top part of Figure 3. Source data is drawn from online transactional databases such as ERP applications, legacy files or other relational databases. Source data may also be point of sale data. This source data is first Extracted, Transformed and Loaded by ETL and data quality tools into Relational OLAP databases and/or the Data Warehouse. Data marts take subject area subsets from the Data Warehouse for query and reporting. Analytical applications carry out OLAP analysis using OLAP tools. Business Intelligence tools also provide analytical processing, such as EIS and DSS products. Data mining tools are used to drill down and analyze data in the warehouse. Warehouse management operates to manage the ETL and data quality stage, the Relational OLAP databases and Data Warehouse and the analytical applications.

The bottom part of Figure 3 lists Unstructured Data sources that are used by Enterprise Portals. In Chapter 11 we see how XML can use metadata tags to integrate unstructured data sources with the Structured Data sources above. These unstructured data sources are managed by a Content Management Repository as Content Management Applications and Database. While they are conceptual in Figure 3, we will see these referenced as XML databases later in the book. Enterprise Portals extend Data Warehouses to the Intranet and Internet. But unlike Data Warehouses which are data-driven, Enterprise Portals are also process-driven. They enable organizations to change their business processes and workflow practices in dramatic ways. We introduce some of these ways when we discuss reengineering in Chapters 12 and 13. We cover many more changes and opportunities in Chapter 15.

---

5. The Next Few Years ...

In discussing the move towards Corporate Portals over the coming years in "The Portal is the Desktop", Gerry Murray - Director of Knowledge Technologies research at International Data Corporation (IDC) [Murray 1999] - says:

> *"Corporate portals must connect us not only with everything we need, but* (also) *with everyone we need, and provide all the tools we need to work together. This means that groupware, e-mail, workflow, and desktop applications -- even critical business applications -- must all be accessible through the portal. Thus, the portal is the desktop, and your commute* (to work) *is just a phone call away."*

> *"This is a radical new way of computing. It's much more effective for companies than traditional approaches, since they can outsource the entire infrastructure as a monthly service."* He makes the point that: *"Corporate Portals will provide access to everything from infrastructure to the desktop, so portal vendors will be the Microsofts of the future."*

He discusses four stages in the evolution of Corporate Portals:

- Enterprise information portals, which connect people with information
- Enterprise collaborative portals, which provide collaborative computing capabilities of all kinds
- Enterprise expertise portals, which connect people with other people based on their abilities, expertise, and interests
- Enterprise knowledge portals, which combine all of the above to deliver personalized content based on what each user is actually doing.

He then goes on to describe a number of products that are starting to appear in each of these Corporate Portal evolution stages. His complete article is available on the Internet [Murray 1999].

5.1 Application Service Providers

We are beginning to see the early moves into the portal environment described above by Gerry Murray, with the emergence of Application Service Providers (ASPs). Early ASPs will typically also be Internet Service Providers (ISPs). They will not only provide ready access to the Internet, but also offer access to much of the software that you need from your desktop, as well as to other products such as Enterprise Resource Planning (ERP) systems from SAP and

others.

This will be the true realization of Network Computing. Not by using Java as a portable language as promoted by Sun and Oracle. But by outsourcing hardware, servers, networks and network management, software and software management, help desk, maintenance and other Total Costs of Ownership (TCO) to ASPs. This is a radical move that will transform desktop computing as we know it. It will provide ubiquitous computing through the Internet and the intranet. And with a move to wider bandwidths on the Internet -- with higher data rates available also through wireless computing via PDAs or mobile phones that access the Internet for email and browsing -- we will soon be able to work not just from the office, but from anywhere. In a few short years these ASPs will become Information Utilities for the future.

Seeing the potential threat to its desktop monopoly that is presented by Corporate Portals and by ASPs, Microsoft has decided that it will adopt a win -- win strategy by also becoming part of this ultimate move to Network Computing. The release of Internet Explorer 5.0 and Microsoft Office 2000 provided some support for this capability. With Office 2000, Microsoft Office Web Server extensions for Intranet servers within the enterprise support collaboration and other groupware applications. But Microsoft will also make these extensions available to ISPs to help them become ASPs. In the future, many of these ASPs will enter into license agreements with Microsoft; two initial ASP licensees were announced with the release of Office 2000. ASPs will be able to offer rental access to their customers so they can use Microsoft and other applications. These will be rented for a fixed monthly or annual fee, or on a pay-for-use basis. So Microsoft will benefit both ways -- not just from new product sales and upgrade sales as we have today, but also from license fees that are paid by ASPs to Microsoft.

With the use of XML and the emergence of Corporate Portals (Enterprise Portals) over the next few years, we will see radical changes in the way we use computers. The Internet and intranet will become more and more a part of our daily work lives. Instead of commuting by road, rail or bus to work, increasingly we will be able to telecommute from wherever we are via the Internet or intranet. The Corporate Portal will be our desktop, available anywhere we log-on to our personalized portal page. From there we will have access to all of the software, systems and other knowledge resources that we need to do our job -- with XML integrating these various data, information and knowledge resources seamlessly across the internet, intranet or extranet.

---

6. Author

Clive Finkelstein, acknowledged worldwide as the "Father" of Information Engineering, is Managing Director of Information Engineering Services Pty Ltd in Australia. He is the Chief Scientist of Visible Systems Corporation in the USA and is Managing Director of Visible Systems Australia Pty Ltd. He is a member of the International Advisory Board of DAMA International and has over 38 years' experience in the Computer Industry.

This paper is extracted from Chapters 1 and 15 of: *"Building Corporate Portals with XML"*, co-authored with Peter Aiken, published by McGraw-Hill (Sep 1999). An extract can be read online at http://svc004.bne009i.server-web.com/catalogue/visible/default.shtml. Click on the *Read Extract* link, below the image of the book front cover on the *Home* page.

He has published many books and papers throughout the world including the first publication on Information Engineering: a series of six InDepth articles in US ComputerWorld in May -

June 1981. He co-authored with James Martin the influential two-volume report titled: *"Information Engineering",* published by the Savant Institute in Nov 1981. He wrote two later IE books: "*An Introduction to Information Engineering",* Addison-Wesley (1989); and *"Information Engineering : Strategic Systems Development",* Addison-Wesley (1992). He has contributed Chapters and Forewords to books published by McGraw-Hill [*"Software Engineering Productivity Handbook"* (1992) and Foreword: *"Data Reverse Engineering: Slaying the Legacy Dragon",* Peter Aiken (1996)], and by Springer-Verlag [*"Handbook on Architecture of Information Systems"* (1998)].

His latest book is:

"Enterprise Architecture for Integration: Rapid Delivery Methods and Technologies", by Clive Finkelstein, Artech House, Norwood MA (March 2006)

- Read the Book Review

His current focus helps organizations to evolve from Data Warehouses and Data Marts to Corporate Portals (also called Enterprise Portals) using the Extensible Markup Language (XML). These provide a central gateway to the information and knowledge resources of an enterprise on its corporate Intranet and via the Internet. Enterprise Portal, XML and related technologies and products will rapidly become available over the next 2 – 5 years. Enterprise Portals will be the central computing focus and interface for most enterprises in the 21st century.

Clive writes a monthly column, "The Enterprise" for DM Review magazine in the USA and also publishes a free, quarterly technology newsletter via email: "The Enterprise Newsletter (TEN)". Past issues of TEN, and of the DM Review Enterprise column, are available from http://www.ies.aust.com/~ieinfo/articles.htm.

---

7. References

[Drucker 1974]	Drucker, P F, *"Management: Tasks, Responsibilities, Practices",* (1974), Harper & Row: New York, NY.
[Finkelstein 1981a]	Finkelstein C, *"Information Engineering",* a series of six InDepth articles published in US Computerworld, (May-June, 1981), IDG Communications, Framingham: MA
[Finkelstein 1981b]	Finkelstein C and Martin J, *"Information Engineering"*, two volume Technical Report, Savant Institute, (Nov, 1981), Carnforth, Lancs:UK
[Finkelstein 1989]	Finkelstein C, "*An Introduction to Information Engineering"*, Addison-Wesley, (1989), Sydney, Australia
[Finkelstein 1992]	Finkelstein C, *"Information Engineering : Strategic Systems Development",* Addison-Wesley, (1992), Sydney, Australia

[Hackathorn 1999]	Hackathorn, Richard (1999), *"Web Farming for the Data Warehouse",* Morgan Kaufman, ISBN: 1-55860-503-7. Includes use of XML for data sources from Internet (368 pages). See also the Web Farming web site - http://www.webfarming.com/
[Hammer, 1990]	Michael Hammer, *"Reengineering Work: Don't Automate, Obliterate",* Harvard Business Review, Cambridge: MA (Jul-Aug 1990).
[Hammer & Champy 1993]	Hammer M and Champy J, *"Reengineering the Corporation: A Manifesto for Business Revolution",* Harper Business, (1993).
[InfoWorld]	The InfoWorld front page article on the Merrill Lynch report is available from - http://www.infoworld.com/cgi-bin/displayStory.pl?/features/990125eip.htm
[Murray 1999]	Murray, G. (1999) *"The Portal is the Desktop",* International Data Corporation, Framingham: MA. This article is available from the Group Computing web site http://www.groupcomputing.com/Issues/1999/MayJune1999/mayjune1999.html
[SageMaker]	SageMaker web site - http://www.sagemaker.com/company/lynch.htm
[Smith 1910]	Smith A, *"The Wealth of Nations"*, London: Dent (1910)
[W3C]	WWW Consortium: XML, XSL and XLL Specifications - http://www.w3.org/

| Home | Courses | Papers | Services | Online Store |

| Home | Courses | Certification | Projects | Papers | TEN Archive | EA Blog | Online Store | Contact Us | [Search] |