Arbortext

Enterprise Publishing Software

"Create Once and Deliver Many: An Overview of XML and Content Management"

Jeff Deskins

Agenda

- What We're Trying to Solve
- What is XML
- What is Content Management
- Demystifying It All

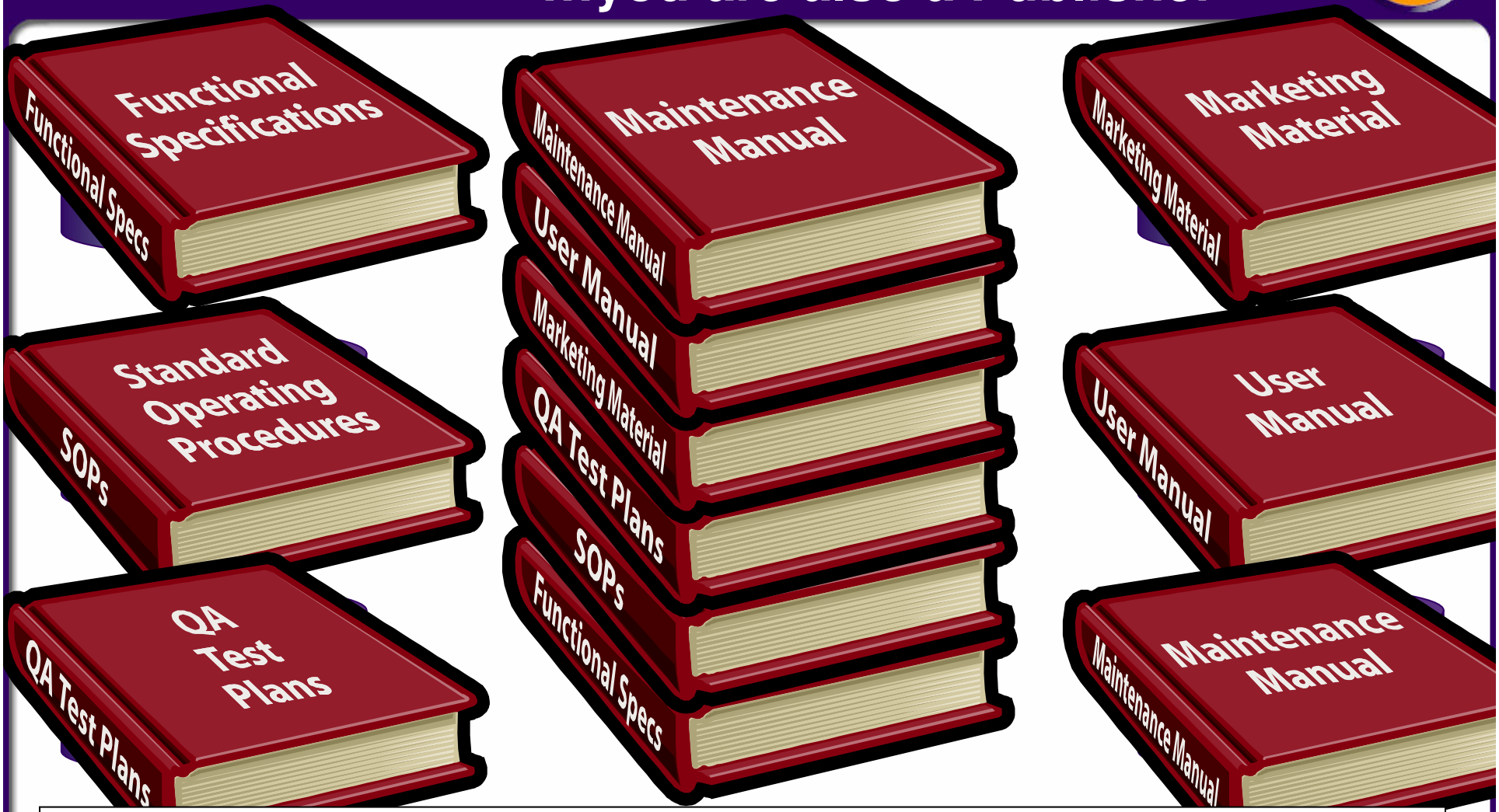
Did you know…

Businesses can publish hundreds of thousands of pages of information each year!

- Aventis Medical Information Services
  - Over 120,000 pages of medical inquiry responses per year
- US Airways
  - Over 500,000 pages in flight manuals alone
- BNA (Bureau of National Affairs)
  - Million of pages of research and report content each year
- Caterpillar
  - Over 5 million pages with 800 new pages added daily
- KBC Bank & Insurance Group
  - Over 12.5 million pages per year!

Arbortext

Regardless of what business you are in…
…you are also a Publisher

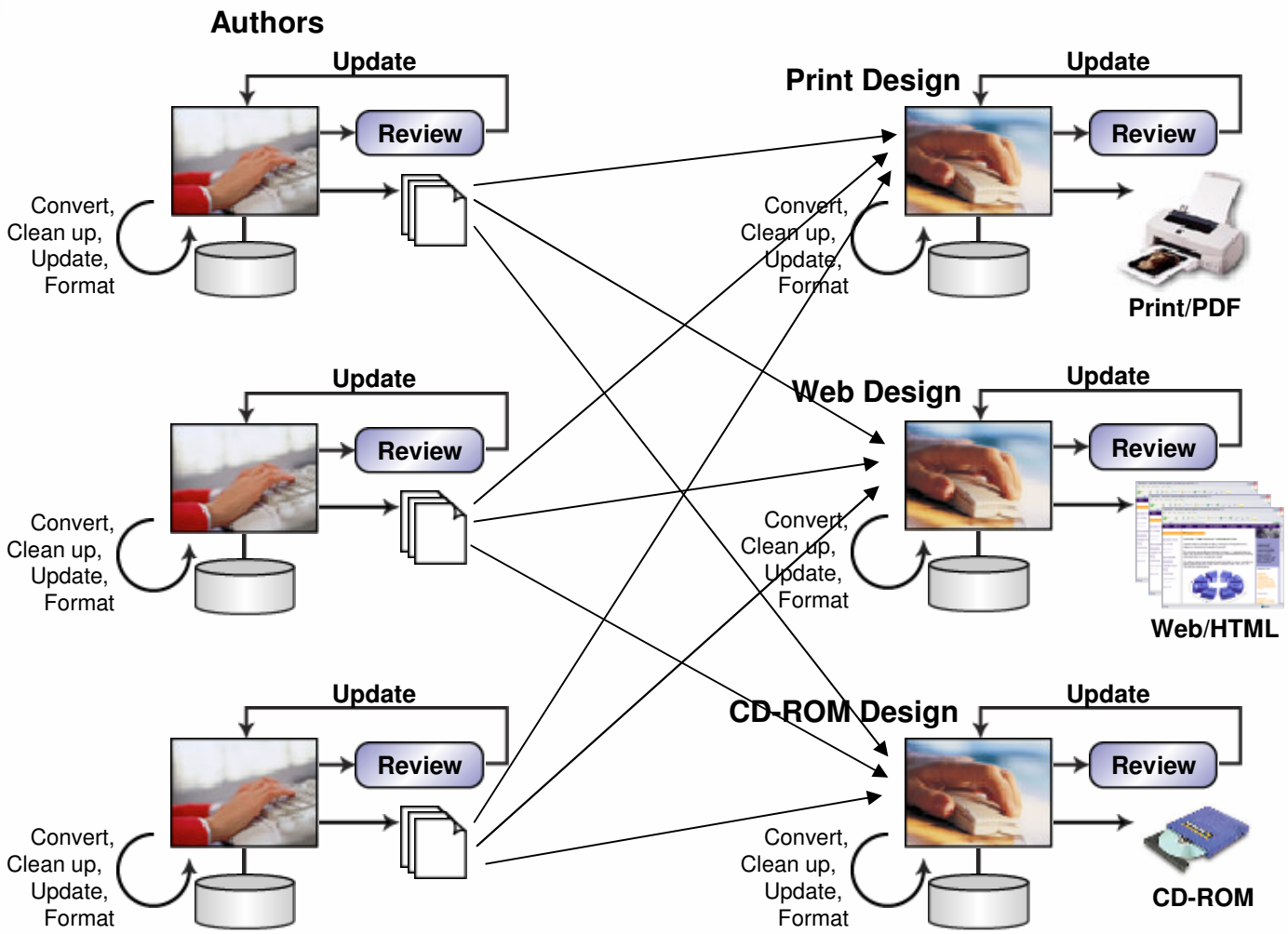


You create, maintain and publish information for every aspect of your business.

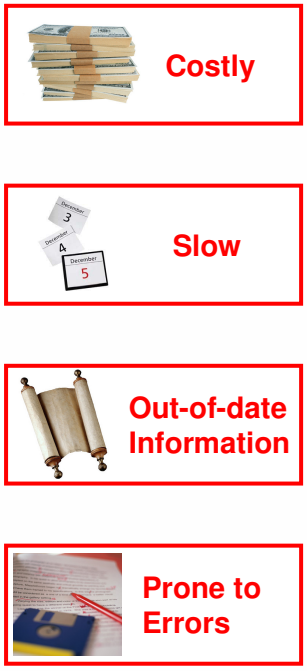
It’s about more than quantity…

- Accuracy, consistency and relevance is critical
- Frequent updates are cost-prohibitive
- Resulting information is error-prone, inconsistent and out-of-date
- Primary source of internal and external communication

Traditional Publishing Process



=



How can you improve your publishing process?

Enter XML

- Extensible Markup Language
- A Web language for:
  - Publishing information for human consumption
  - Facilitating data interchange between computers
- A Standard from the World Wide Web Consortium (W3C)

XML - Bird's Eye View

- **It's a** language
  - **With rules on how you** tag **information (embedding descriptions, or** metadata**)**
  - **You can use it to pre-define** tag sets **for use by whole classes of documents**
  - **You can use it to pre-define rules (a** data model**) for how to properly use a tag set**
- **XML can embody any kind of data**
  - **Hierarchical, tabular**
  - **Blurs the distinction between *documents* and *data***

What Does XML Look Like?

```
  <article pub="IBD" date="19990303">
<author name="Susan E. Fisher"/>
<title>XML Will Make Web Easy as XYZ</title>
<p>The data treasure chest known as the
World Wide Web finally might get a standard
way of labeling the stuff it holds.</p>
  ...
</article>
```

Basic XML "parts of speech"

- Elements **are like nouns**
  - **They name a piece of information**
  - **Elements can nest inside each other**

Quote element is wholly inside **title** element

```
<title>XML Will Make Web Easy as
<quote>XYZ</quote></title>
```

- Attributes **are like adjectives**
  - **They further describe a piece of information**

An attribute has a name (e.g., **pub**) and a value (e.g., **IBD**)

```
<article pub="IBD" date="19990303">
```

Arbortext

What is a DTD?

- "Document Type Definition"
- Specifies the "data model":
  - Which tags are permitted
  - Which tag structures are permitted
  - What content is permitted
- Enforces consistency
  - "Helps the pieces fit"

DTD and Documents - Applied

- **It's like starting a conversation with "I'm going to be speaking in English now"**
- **Then, English rules (like subject-verb-object) must apply**
- **This is called a** valid document

```
  <!DOCTYPE article SYSTEM
"http://www.myorg.org/dtds/article.dtd">
<article pub="IBD" date="19990303">
  ...
</article>
```

The DTD is accessed with a URL

The model is encoded formally as a **DTD**

Comparing Word Processing to XML

Word Processing

XML Markup

Centered, 24pt, Bold Book Antiqua

Centered, 20pt Arial

Left Justified, 20pt, Book Antiqua Bold

Body Text, 10pt Book Antiqua

Indent 2X, 20pt Book Antiqua Italic

Arboractin-D

Arborex sodium

Extended release tablets

60 and 120 mg

Description.

This is Arboraction-d (arborex sodium) in extended release tablets in 60 and 120 mg strengths.

Clinical Pharmacology

This is Arboraction-d (arborex sodium) in extended release tablets in 60 and 120 mg strengths.

Indications and Usage

For the treatment of high blood pressure. 1-2 Tablets a day as directed. Not for use by children under 18.

Contraindications

Pregnancy
Nursing Mothers

Drug Name

Generic Name

Formulation

Strength

Description Section

Paragraph

Clinical Pharmacology Section

Paragraph

Indications Section

Paragraph

Contraindications Section

Contraindications List

What's XML good for?

- ePublishing
  - Web (intranet, extranet, Internet)
  - CD-ROM
  - Print
  - Wireless
- eCommerce
  - Electronic commerce (business-to-consumer)
  - Electronic Data Interchange – EDI (business-to-business)
- Software applications
  - Data transfer between databases
  - Application integration

Benefits of XML

- **Makes structure explicit**
  - <ingredient> lactose </ingredient>
  - Context-related searches
- **Separates content from format**
  - Enables reuse of information
  - Information delivery is requirement specific (content & format)
- **Enables automated content creation and delivery**
- **Supports emerging regulatory standards**
  - Enables application interoperability and faster cycle time
  - eCTD, ebXML, SAML, Election Markup Language (EML)

Other Uses of XML

- Databases (Oracle XDB)
- Web Services
- Products
  - BEA Web Portal
  - Macromedia Flex
  - Apache Tomcat
- RSS (news aggregators and syndication)
- PODcasting

All the X's

- **XML Related Standards:**
  - Modeling Rules – DTDs & schemas
  - Stylesheets – XSL & XSLT
  - Linking – XLink & XPointer
  - SOAP
  - SVG
- **Visit www.w3c.org**

Content Management

- CMS – DMS – Records Management
  - Content Management System
  - Document Management System
  - File or Records Management
- So Many Acronyms, So Many Players
- What's the Difference?
- What Are Your Needs?

Differences

- **Document Management**
  - Documents are treated as “blobs” except
  - Compound Document Manager required for managed reuse
  - Full-text indexing for searches
- **Content Management**
  - Parses content
  - Stores elements as individual objects
    - Version, Reuse at the element level
    - Manage access at fragment level (minimum revisable unit, Locking unit, etc)
  - Deeper Searches with Elements, Attributes, Text

Differences

- **Web Content Management**
  - **Manages web content for staging, editing, and delivery to web sites**
- **Records/File Management**
  - **Manages scanned forms, documents, reports**
  - **Apply metadata to enable search. No full text search.**

Examples of Each

- **Document Management**
  - **Documentum, Oracle iFS, OpenText**
- **Content Management**
  - **Astoria, Oracle XDB, Documentum**
- **Web Content Management**
  - **Vignette, Interwoven, Sharepoint**
- **Records Management**
  - **FileNet**

Basic Management Functionality

- Store Object
  - in database
  - in controlled file system w/ database pointer
- Library Services
  - Cabinets, directories (folders)
  - Access Control
    - Delete, Write, Version, Read, Properties, None
  - Versioning
    - Simple, branching, labeled, time based
  - Content Indexing for search performance

Basic Management Functionality

- **Properties on Object**
  - **Owner, creation/revision date, etc**
  - **Searchable (Indexed)**
  - **Extensible**
    - **Category, Usage, Profile, Keywords, Workflow Status, etc.**

Questions to Ask When Selecting

- Do you plan on managing whole documents, document objects, individual elements?
- What DMS/CMS features do you need?
- On what criteria are virtual documents built?
- Will data for certain elements be provided from/validated against external data repositories (e.g., part number elements should be populated with data from an external relational database)?
- Workflow and Document Lifecycle tools?

Content Reuse

- **Written once and reused many times**
  - **Increased consistency**
  - **Reduced development and maintenance costs**
  - **Rapid reconfiguration**
  - **Translation**

Managing Reuse

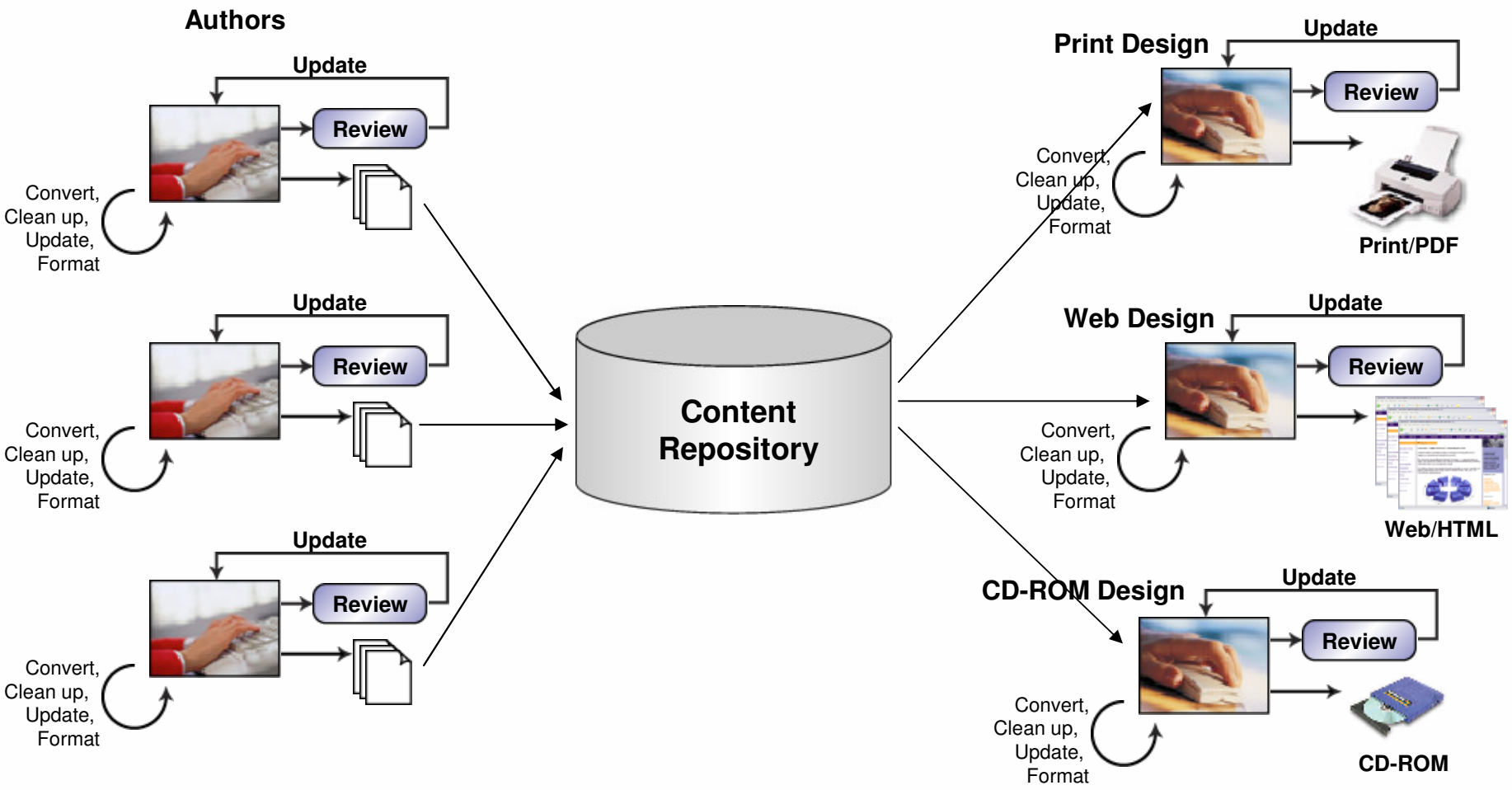
Basic Requirements for Managing Reuse

- Where Used - what are the books that use this chapter?
- Uses What - what are the chapters used in this book?
- Referential Integrity - don't allow an object to be deleted if it has parents

• Additional Requirements for Managing Reuse

- Version specific links (version binding)
- Late binding
- Dynamic links
- Freezes, snapshots

Content Management – Partial Solution



Slide 27

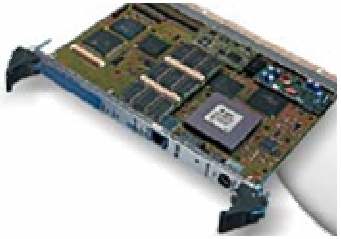
tgb11 This slide should only be used if the prospect/customer already has CMS or considering a CMS to solve these problems.
Tonua Brown, 9/2/2004

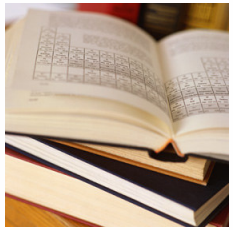
Content Management Resources

- **CMS Watch: http://www.cmswatch.com/**
- **The Gilbane Report: http://www.gilbane.com/**

XML and CMS by Industry

- Aerospace/Airlines
- Automotive
- Electronics/Telecom
- Finance, Insurance
- Government
- Heavy machinery
- Pharmaceuticals, Medical Equipment
- Reference & Legal Publishers

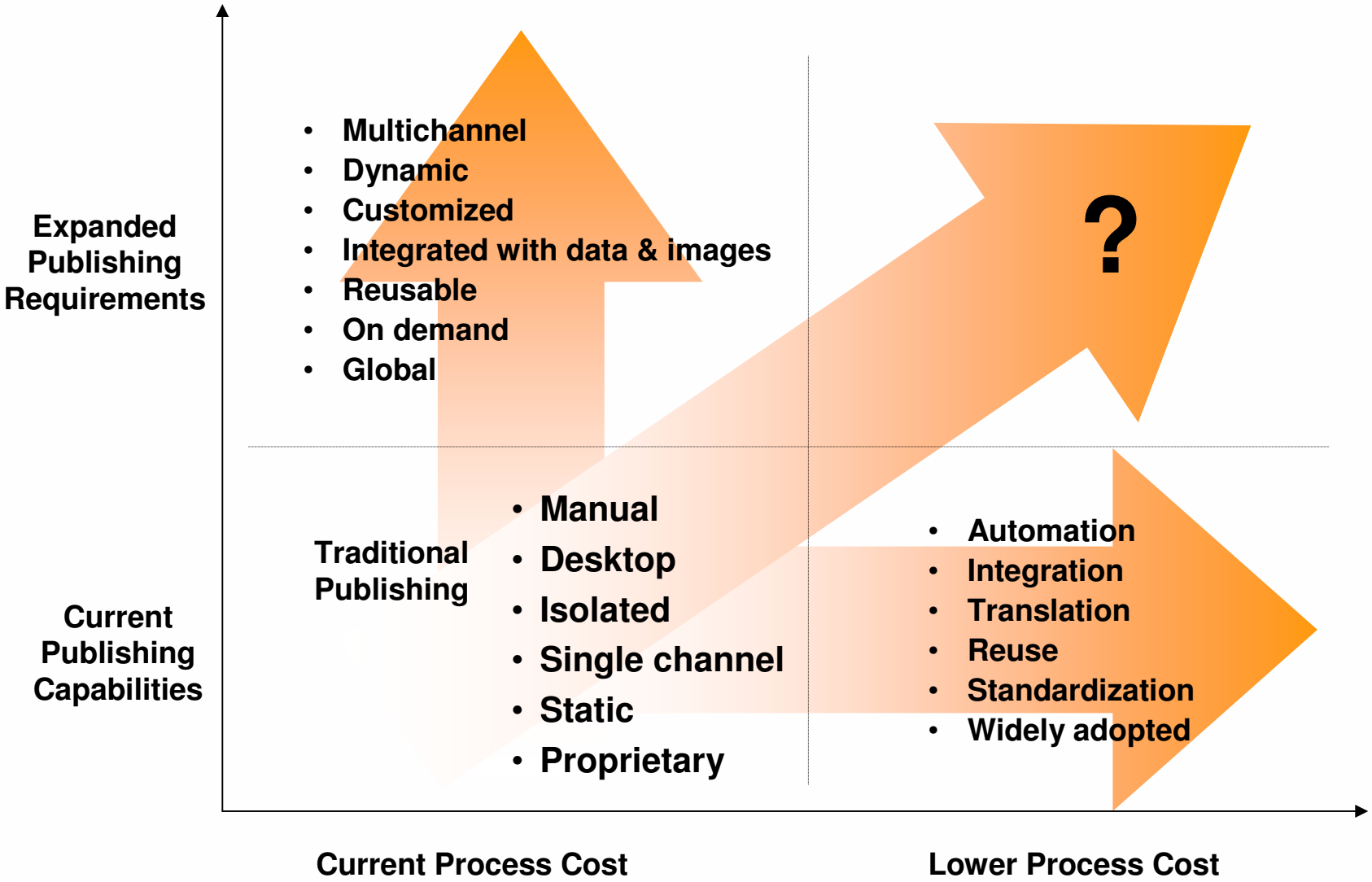




Arbortext

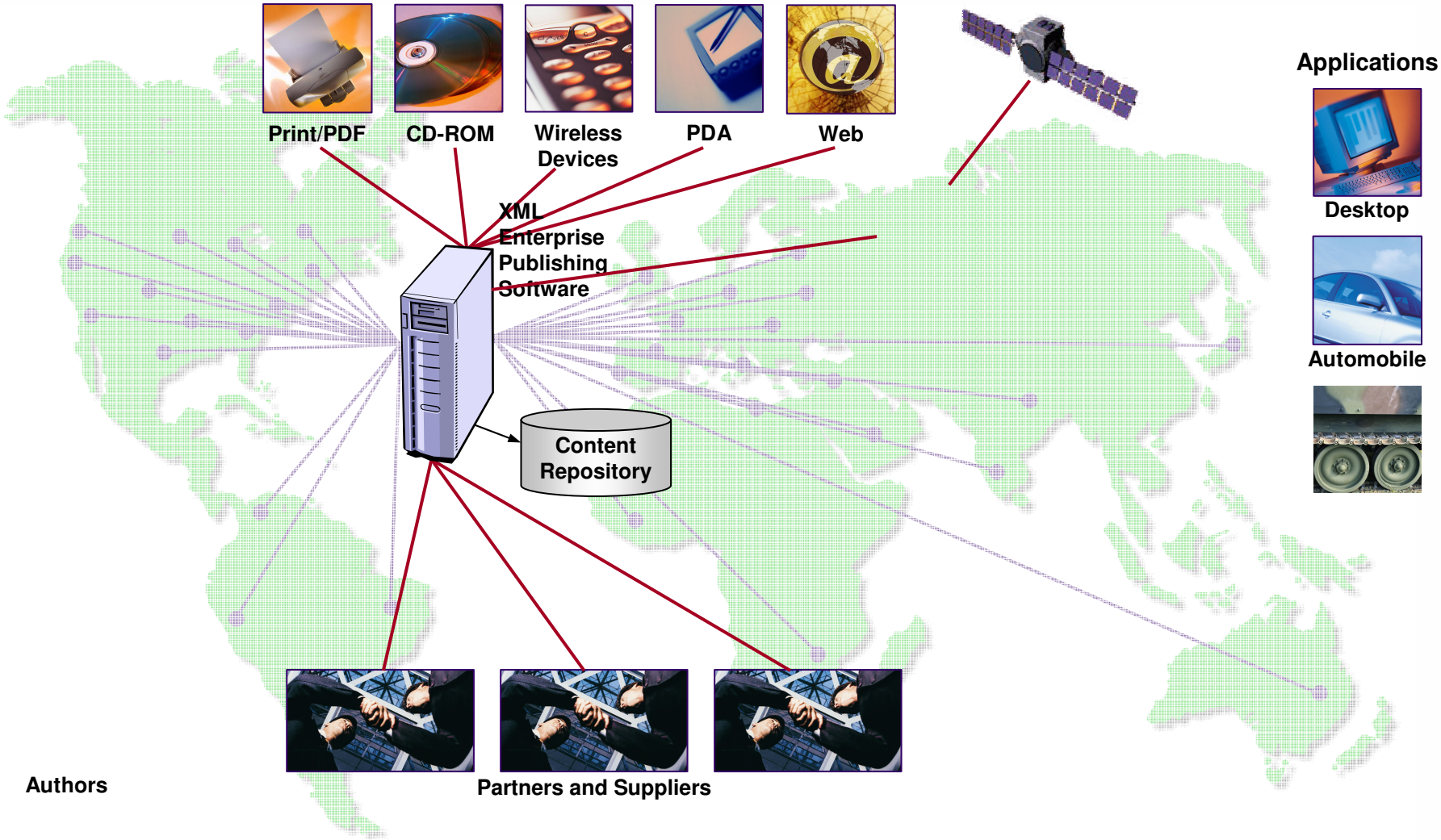
Demo

Publishing Better, Faster, Cheaper



Arbortext

Integrated Global Publishing Environment



Arbortext

```
ERROR: undefined
OFFENDING COMMAND: f`~

STACK:
```