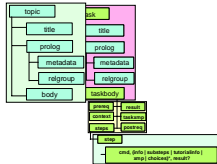


# Taking advantage of XML to improve information: Darwin Information Typing Architecture

DA Schell, M Priestley, G Hargis



© IBM Corporation Copyright, 2001. All rights reserved. Page 2

## Introduction to the problem and the solution

- I. Background
  - ▶ Our background on markup languages
  - ▶ Promise of XML
  - ▶ Reality gap
- II. DITA overview
- III. DITA advantages and challenges

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

© IBM Corporation Copyright, 2001. All rights reserved. Page 3

## Background: Markup languages at IBM®

Language	1970s: ISIL	1980s: BookMaster, IPF	1990s: SGML, HTML
Output	Printed books	Printed and online books, online help	Online information, webs, printable & printed books
	Limited reuse Single purpose	Reuse: embedding Multipurpose: conditional code Multiformats: style files Document types	One book-centered DTD Information architecture

---

---

---

---

---

---

---

---

---

---

## Growing needs in the 1990s

- Upheaval in the ranks: away from proprietary tools, toward WYSIWYG
- Shorter cycles, fewer people
- Componentization of products, multiplatform, open systems
- Need for faster, cheaper, more usable alternative to printed book
- e-business: WWW for display and distribution of products and information

---

---

---

---

---

---

---

---

---

---

---

---

## Clash in the late 1990s between SGML and HTML

- SGML
  - ▶ Authoring with DTD oriented to a linear document, and processing for delivery
- HTML
  - ▶ Authoring with an information architecture, and processing for delivery. Homegrown tools and lower barrier to entry.

---

---

---

---

---

---

---

---

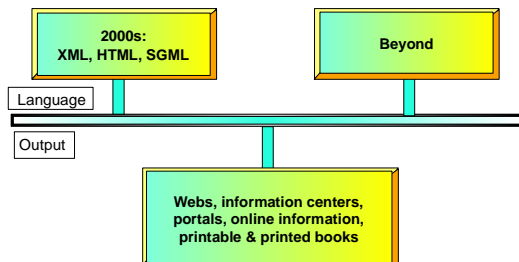
---

---

---

---

## Looking at 2000s for IBM User Technology



Reuse, architected information (Darwin Information Typing Architecture), specialized DTDs

---

---

---

---

---

---

---

---

---

---

---

---

## Promise of XML

- Separate content from form to allow reuse of content in different presentation media
- Use specific markup to describe content
  - ▶ Apply presentation styles intelligently
  - ▶ Mine information for reuse
  - ▶ Enable intelligent online search
- Use standard solution to enable easy exchange of information

---

---

---

---

---

---

---

---

---

---

---

---

## Reality gap

- Form applies to structure, not just fonts and page breaks
- Current model of DTD development: heavy upfront investment and long-term payoffs
- Standard solutions: not specific to your needs

**Tradeoff: The more useful your markup is to you, the more it will cost you, and the fewer people share the costs.**

---

---

---

---

---

---

---

---

---

---

---

---

## Information interchange, tools management, and extensibility

- Knowledge representation is strongly related to corporate culture
- Most companies rely on many delivery systems, or process their information in ways that differ widely from company to company.
- Most attempts to formalize a document description vocabulary have been done as information modeling exercises to capture the current business practices of data owners.

---

---

---

---

---

---

---

---

---

---

---

---

## Part II: DITA overview

- Fulfilling the promise of XML
- Development of DITA
- Principles of DITA
- Validation

---

---

---

---

---

---

---

---

---

---

## Fulfilling the promise of XML: DITA

- Separate content from context for reuse
- Maintain compatibility with standards by using a base and then using specialization
  - ▶ Minimize time for DTD development
  - ▶ Create specific markup quickly and cheaply
  - ▶ Support corporate innovation
- Create an alternative for interchange by larger non-IBM community

---

---

---

---

---

---

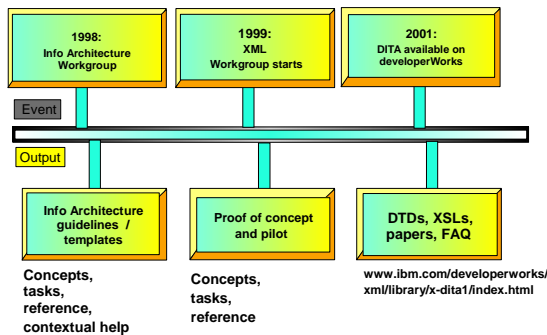
---

---

---

---

## Development of DITA



---

---

---

---

---

---

---

---

---

---

## XML workgroup

- **Phase I: Analyze the landscape**
  - ▶ Preserve the huge investment in legacy documents of both IBM and our customers
  - ▶ Focus XML activity on the new information architecture
  - ▶ Support diverse authoring and delivery needs
  - ▶ Support coexistence of article-based and linear content
  - ▶ Support active outreach across IBM to others developing expertise in XML

---

---

---

---

---

---

---

---

---

---

---

---

## XML workgroup (*continued*)

- **Phase II: Prototype**
  - ▶ Develop prototype DTDs
  - ▶ Develop prototype XSL and CSS stylesheets; utilize XSLT
  - ▶ Make widely available for non-IBMers to use
  - ▶ Deliver DITA to developerWorks
  - ▶ Work to establish as standard across technical writing professionals
  - ▶ Continue to develop the XML WG team members so that they have the skills to lead independent work using these technologies
  - ▶ Conduct site specific technical vitality meetings
  - ▶ Move authoring skills into rank and file
- **Phase III: Integrate**
  - ▶ Implement in tools

---

---

---

---

---

---

---

---

---

---

---

---

## Principles of DITA

- **Topic orientation**
  - ▶ The topic is the smallest independently maintainable unit of reuse.
- **Information types**
  - ▶ Information types are a delivery-independent way to describe content.
- **Inheritable design**
  - ▶ Specialized information types can be created from a more general type of topic.
- **Inheritable process**
  - ▶ DITA-aware processes, such as publishing and translation, can also be specialized themselves.

---

---

---

---

---

---

---

---

---

---

---

---

## Topics and information types

- Consider information types as delivery-neutral descriptions of topic content
  - ▶ A task is a task, regardless of where you read it
  - ▶ Elements like <chapter> and <appendix> are factored out of the topic and into the context
- Consider topics as the building blocks for documents. Topics can be:
  - ▶ Assembled into book structures
  - ▶ Or linked into webs
  - ▶ Or served up from portals

---

---

---

---

---

---

---

---

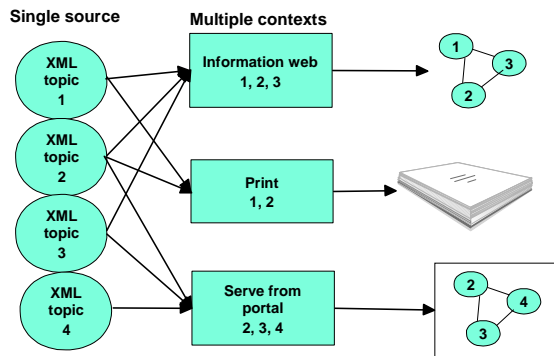
---

---

---

---

## Single source and multiple contexts




---

---

---

---

---

---

---

---

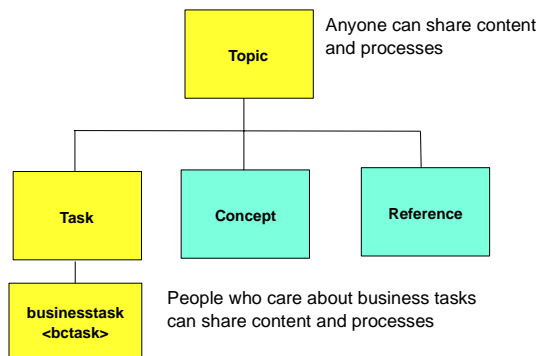
---

---

---

---

## Specialization of task information type




---

---

---

---

---

---

---

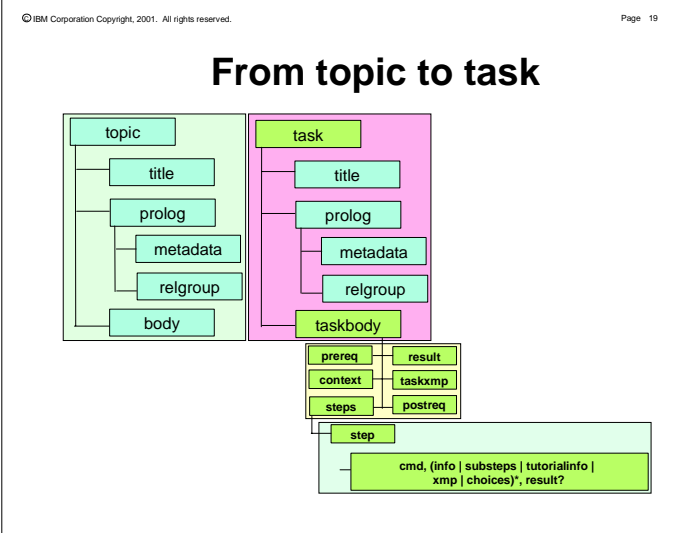
---

---

---

---

---




---

---

---

---

---

---

---

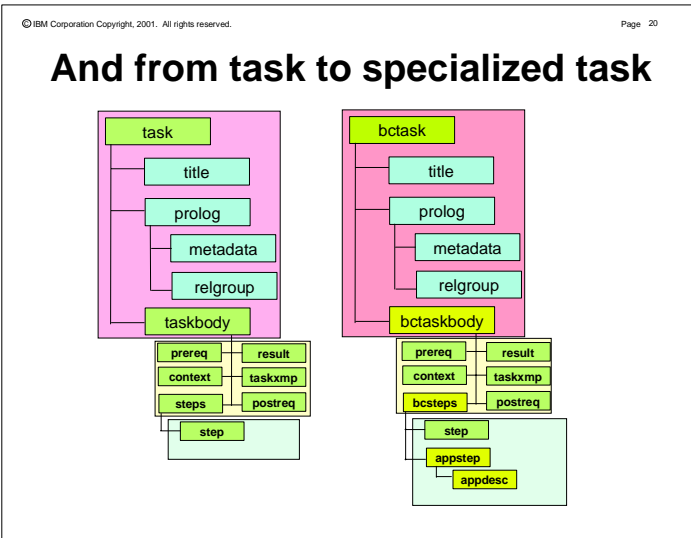
---

---

---

---

---




---

---

---

---

---

---

---

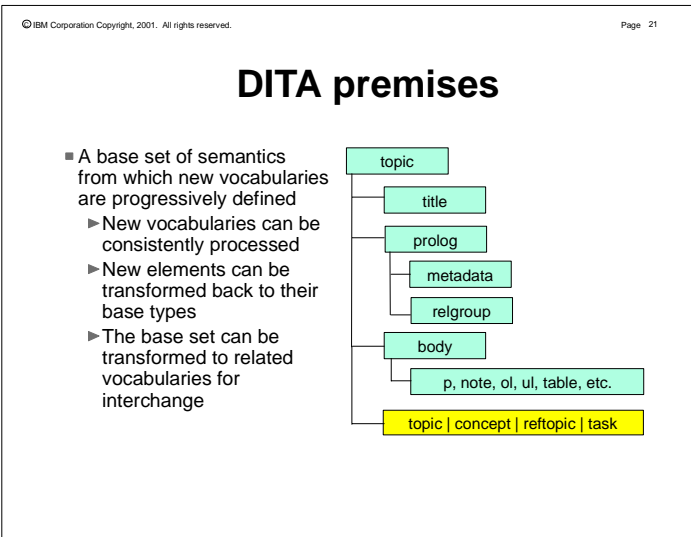
---

---

---

---

---




---

---

---

---

---

---

---

---

---

---

---

---

### A point of validation: Lotus Notes "client" help into DITA XML

- Processing assumption: All topics are either tasks or concepts
  - ▶ Task topics - Any topic with a numbered list that's not in a table
  - ▶ Concept topics - All other topics
  - ▶ Loose ends -Troubleshooting, FAQs, Glossary

---

---

---

---

---

---

---

---

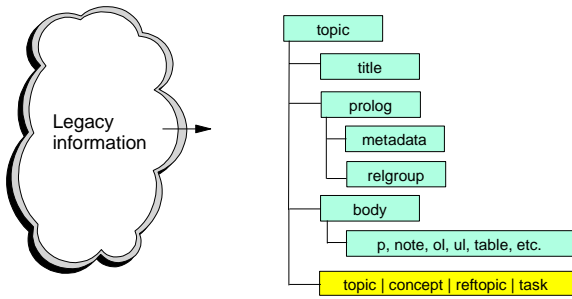
---

---

---

---

### Process: First migrate legacy information to "generic" topics




---

---

---

---

---

---

---

---

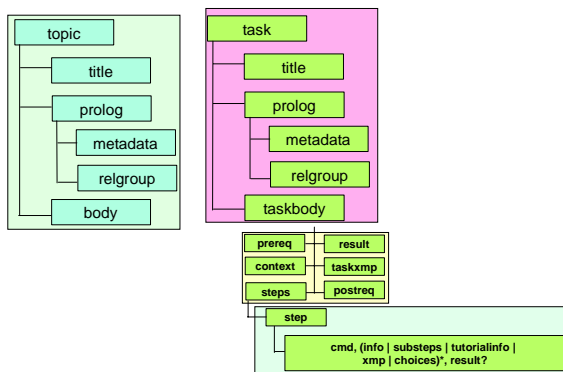
---

---

---

---

### Then "infotype" topic to task




---

---

---

---

---

---

---

---

---

---

---

---



## Validation results

- Examined 7,500 topics in Lotus Notes and Domino help systems
- Converted the 768 topics in the Notes 5.0.3 end-user help system to DITA
  - ▶ Available for download from Lotus Notes.net
- Successful because information was developed by good authors who wrote to a clear set of guidelines
  - ▶ DITA goal is to move guidelines to tools

---

---

---

---

---

---

---

---

---

---

---

---

## Part III: DITA advantages and challenges

- Advantages
  - ▶ Customers
  - ▶ Technical processes
  - ▶ Employees
- Challenges
  - ▶ Managing the change and legacy issues
  - ▶ Employees and processes
  - ▶ Implementation
- Conclusion

---

---

---

---

---

---

---

---

---

---

---

---

## Customer value: Deliver valuable information

- Customer-specific
  - ▶ Content and markup suited for customer and user needs
- Portable
  - ▶ Content can be interchanged with other companies
  - ▶ Markup can be processed by other companies
- Flexible
  - ▶ Content and markup can be easily migrated to new types when customer requirements change
- Pervasive
  - ▶ Content used across product families so that it can be integrated - with fallback support

---

---

---

---

---

---

---

---

---

---

---

---

## Technical: General DITA advantages

- An XML architecture proposal aligned directly to an architecture of information
- Topic-oriented focus
- Rigorous semantic specification of topic types based on audience and author requirements
- Designed to be repurposed and delivered to multiple contexts

---

---

---

---

---

---

---

---

---

---

---

---

## Technical: Design phase reduced

- With DITA you still need to understand your domain well enough to define its rules
- And you still need to understand DITA well enough to choose where to specialize, and what to reuse
  - ▶ More efficient, because you can reuse markup and rules from more general types, only declaring differences in the new type
- And you still need to create a new DTD (but it's only ten lines or so, instead of a few hundred)
  - ▶ Design is faster, because topics rather than documents

---

---

---

---

---

---

---

---

---

---

---

---

## Technical: Reuse and maintenance

- Because specialized types reuse general markup by reference, you can manage changes to multiple types at once by editing a common ancestor
- Specialized elements won't pick up changes to ancestor equivalents
- The type hierarchy becomes a way to pool common elements, reuse common design across multiple types, and reduce maintenance costs
- Specialized types still work with other people's processes; do not have to recreate infrastructure

---

---

---

---

---

---

---

---

---

---

---

---

## Employees: Training and skill development

- Reduced set of tags to be learned
  - ▶ Not a monolithic DTD
- Benefits of specialization
  - ▶ Reinforce what was learned in school related to writing task-oriented material
  - ▶ Reinforce content authoring
  - ▶ More robust processes, less reliance on training
  - ▶ Since you can match processes more closely to your data, catch errors before they happen with tighter DTD rules
- Challenge: Training required
  - ▶ Performing task analysis and understanding its importance
  - ▶ Having authors understand topic writing

---

---

---

---

---

---

---

---

---

---

---

---

## Getting to benefits, managing for change

- Challenges
  - ▶ Legacy data formats
  - ▶ Level of maturity of tools
  - ▶ People, processes, and organizations
  - ▶ Acceptance and implementation

---

---

---

---

---

---

---

---

---

---

---

---

## Managing the change

- Identify strategy and purpose
  - ▶ Align with corporate strategic directions
  - ▶ Identify value (customer, corporate, community)
- Preserve investments
  - ▶ Prototype, proof of concept, acceptance test
    - Millions of pages in BookMaster
    - Millions of pages in IBMIDDoc
- Develop intellectual capital
- Identify barriers and challenges
- Develop implementation strategy

---

---

---

---

---

---

---

---

---

---

---

---

© IBM Corporation Copyright, 2001. All rights reserved. Page 34

## Preserve investments: Legacy data format considerations

M - migrate  
E - export  
N - author new

From \ To	SGML	HTML	XHTML	XML-DITA
SGML	N	M or E	M or E	M or E
HTML	M	N	M or E	M
XHTML	M	M or E	N	M
XML-DITA	M or E	M or E	M or E	N

- Support conditional processing and transforms
- Develop a set of strategic guidelines to determine when and if

---

---

---

---

---

---

---

---

---

---

---

---

© IBM Corporation Copyright, 2001. All rights reserved. Page 35

## Challenge: Implementation maturity

- Existing, workable solutions
  - ▶ Content management systems
  - ▶ Not deployed widely
- Still shaky
  - ▶ Editors
  - ▶ Formatters
  - ▶ Browsers
  - ▶ End-to-end customizable toolkits or workbenches
- Innovation and invention needed
  - ▶ Context maps
  - ▶ Linking guidelines

---

---

---

---

---

---

---

---

---

---

---

---

© IBM Corporation Copyright, 2001. All rights reserved. Page 36

## Challenge: Employees

- Education
  - ▶ Education for processes
  - ▶ Education for tools
- Dealing with change
  - ▶ Loss of familiar book model
  - ▶ Training for topic orientation
  - ▶ What will cleanup after migration to XML be like?
- Managing the core team
  - ▶ Managing the hype and the delivery

---

---

---

---

---

---

---

---

---

---

---

---

## Challenge: Processes and organization

- Processes
  - ▶ Identify overlap and organizational boundaries
- Identify required changes in other organizations
  - ▶ Publication hosting
  - ▶ Publication printing
  - ▶ Translation processes

---

---

---

---

---

---

---

---

---

---

---

---

## Challenge: Acceptance and implementation

- Gain acceptance
  - ▶ External organizations (standards bodies)
  - ▶ Customers
  - ▶ Business partners
  - ▶ Tool makers
- Implement
  - ▶ Develop tooling
  - ▶ Develop education and strategy roll-out
  - ▶ Monitor continuously
- Continuing evolution

---

---

---

---

---

---

---

---

---

---

---

---

## Conclusion: A walk in the park<sub>(which park?)</sub>

- Challenges in markup and practices
  - ▶ Reduce complexity
  - ▶ Separate style, content, delivery context
  - ▶ Support product integration and solution development
- Use DITA to deliver on the promise of XML
  - ▶ Support minimal set of base DTDs
  - ▶ Support specialization
- Managing the change to get benefits
  - ▶ Align with strategy
  - ▶ Align with customer value
  - ▶ Assist people, process, organizational change, and change agents

---

---

---

---

---

---

---

---

---

---

---

---

© IBM Corporation Copyright, 2001. All rights reserved. Page 40

## DITA resources

- <http://www.ibm.com/developerworks/xml/library/x-dita1/index.html>
  - DeveloperWorks site (intro, specialization paper, FAQs, and the actual DTDs themselves, as well as a discussion forum), link to Notes.net. The source for all updated information.
- [http://www.ibm.com/ibm/easy/eou\\_ext.nsf/Publish/1819](http://www.ibm.com/ibm/easy/eou_ext.nsf/Publish/1819)
  - Make IT Easy paper which details the history, introduction, advantages, and validation of DITA.
- Michael Priestley, Gretchen Hargis, Susan Carpenter. 2001. "DITA: An XML-based technical documentation authoring and publishing architecture." *Technical Communication, Journal of the Society for Technical Communication* 48:352-367.
  - Details about DITA from end to end for the technically inclined.

---

---

---

---

---

---

---

---

---

---

---

---

---

---

© IBM Corporation Copyright, 2001. All rights reserved. Page 41

## Relevant books on articles and online presentation

- ▶ Marlana Coe. *Human Factors for Technical Communicators*, (New York: Wiley, 1996). Particularly Chapter 9 "Choosing a Medium".
- ▶ JoAnn Hackos and Dawn Stevens. *Standards for Online Communication* (New York: Wiley, 1997), which is filled with case studies and research. The pertinent chapter here is Chapter 6 "Structuring Your Online System"
- ▶ Peter Morville & Louis Rosenfeld. *Information Architecture for the World Wide Web*. (Sebastopol, CA: O'Reilly, 1998). Best on online navigation structures.
- ▶ Schriver, Karen. *Dynamics in Document Design*. (New York: Wiley, 1997). All of it since it's a great mix of research review and case studies.

---

---

---

---

---

---

---

---

---

---

---

---

---

---

© IBM Corporation Copyright, 2001. All rights reserved. Page 42

## Additional contact information

- Dave A Schell  
dschell@us.ibm.com
- Michael Priestley  
mpriestl@ca.ibm.com
- Gretchen Hargis  
ghargis@us.ibm.com
- Don Day  
dond@us.ibm.com

---

---

---

---

---

---

---

---

---

---

---

---

---

---