

# Introduction to Structured Authoring

What's the BIG DEAL??



## Valuable Content Transformed

- Document Digitization
- XML and HTML Conversion
- eBook Production
- Hosted Solutions
- Big Data Automation
- Conversion Management
- Editorial Services
- Harmonizer



## Experience the DCL Difference

DCL blends years of conversion experience with cutting-edge technology and the infrastructure to make the process easy and efficient.

- World-Class Services
- Leading-Edge Technology
- Unparalleled Infrastructure
- US-Based Management
- Complex-Content Expertise
- 24/7 Online Project Tracking
- Automated Quality Control
- Global Capabilities



# We Serve a Very Broad Client Base . . .



## ... Spanning All Industries

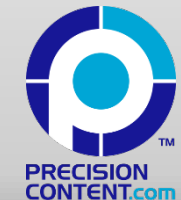
- Aerospace
- Associations
- Defense
- Distribution
- Education
- Financial
- Government
- Libraries
- Life Sciences
- Manufacturing
- Medical
- Museums
- Periodicals
- Professional
- Publishing
- Reference
- Research
- Societies
- Software
- STM
- Technology
- Telecommunications
- Universities
- Utilities



# About your presenter

- **Rob Hanna, ECMs**

- President of Precision Content Authoring Solutions Inc. and a director of AIIM First Canadian Chapter
- Expert in structured authoring and content management practices and technology
- Instructor at the University of Toronto School of Continuing Studies – Metadata and Controlled Vocabularies



# Who is Precision Content Authoring Solutions Inc.?

- We help organizations make their information easier to use
- Our solutions consist of
  - Content strategy
  - Detailed information architecture
  - Content lifecycle design and development
  - Turn-key content transformation
  - Tools selection and development
  - Multi-channel publishing
- [www.precisioncontent.com](http://www.precisioncontent.com)

**CHKISTIE**<sup>®</sup>



**Westoba**  
CREDIT UNION LTD.



Insurance Corporation  
of British Columbia

**xerox**



**KPMG**



**CRITERION**  
Accessibility Solutions, Inc.™

**Bell**

**BCAA** 



**AJCC**

American Joint Committee on Cancer

Validating science. Improving patient care.



*Lo rem ipsum dolor sit amet, consectetur adipiscing elit*

*In hendrerit vulputate condimentum. Curabitur eu*

*convallis ligula. Aenean eget cursus quam, non*

*porttitor arcu. Pellentesque facilisis sem*

*non augue faucibus, molestie*

*consequat leo iaculis. Duis nibh*

*odio, tincidunt ac elit eu.*

*rhoncus pulvinar dui.*

*Nullam cursus laoreet*

*voluptate*

*praesent*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

*malesuada*

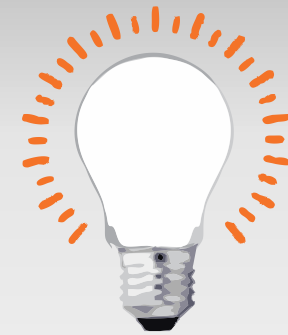
*malesuada*

*malesuada*

*malesuada*

*malesuada*

# WE ARE YOUR CONTENT TRANSFORMATION SPECIALISTS





# STOP THE INSANITY!

Stop throwing  
more technology  
at your content  
problems!

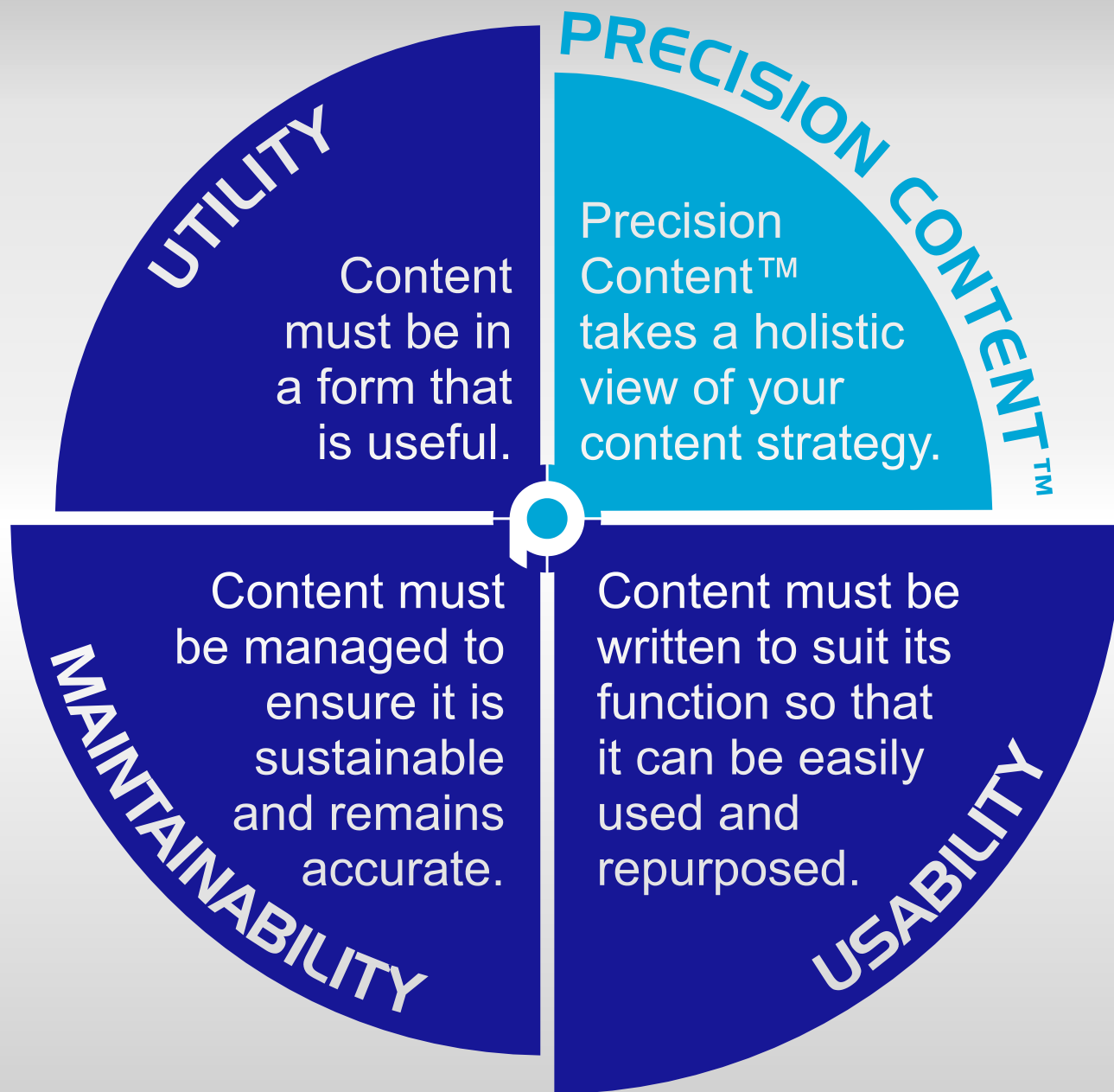


**Content problems  
require  
content solutions!**

## Importance of Content Strategy

Before any technology is considered, organizations must first consider their content strategy to fully understand what they need and how they are going to get there.

Several factors must be examined.



# Definitions

What is Structured Authoring?

# Do we have a problem that needs solving?

- 5,900,000,000 google searches per day (who did we ask these questions to before Google?)
- 540,000 word in the English language – 5 times more than in Shakespeare’s time
- 90% of the world’s data has been generated in the last 2 years

In the year ...	Human knowledge doubled ...
1900	100 years
1945	25 years
2014	13 months
2020	12 hours

# Structured authoring defined

- structured authoring means a standardised, methodological approach to content creation incorporating
  - systematic labelling
  - modular, topic-based architecture
  - constrained writing environments, and
  - the separation of content and form.

# Benefits of structured authoring

- Improved
  - Collaboration
  - Usability
  - Consistency, and
  - Publishing capabilities.

# Meeting 2 sets of fundamental needs

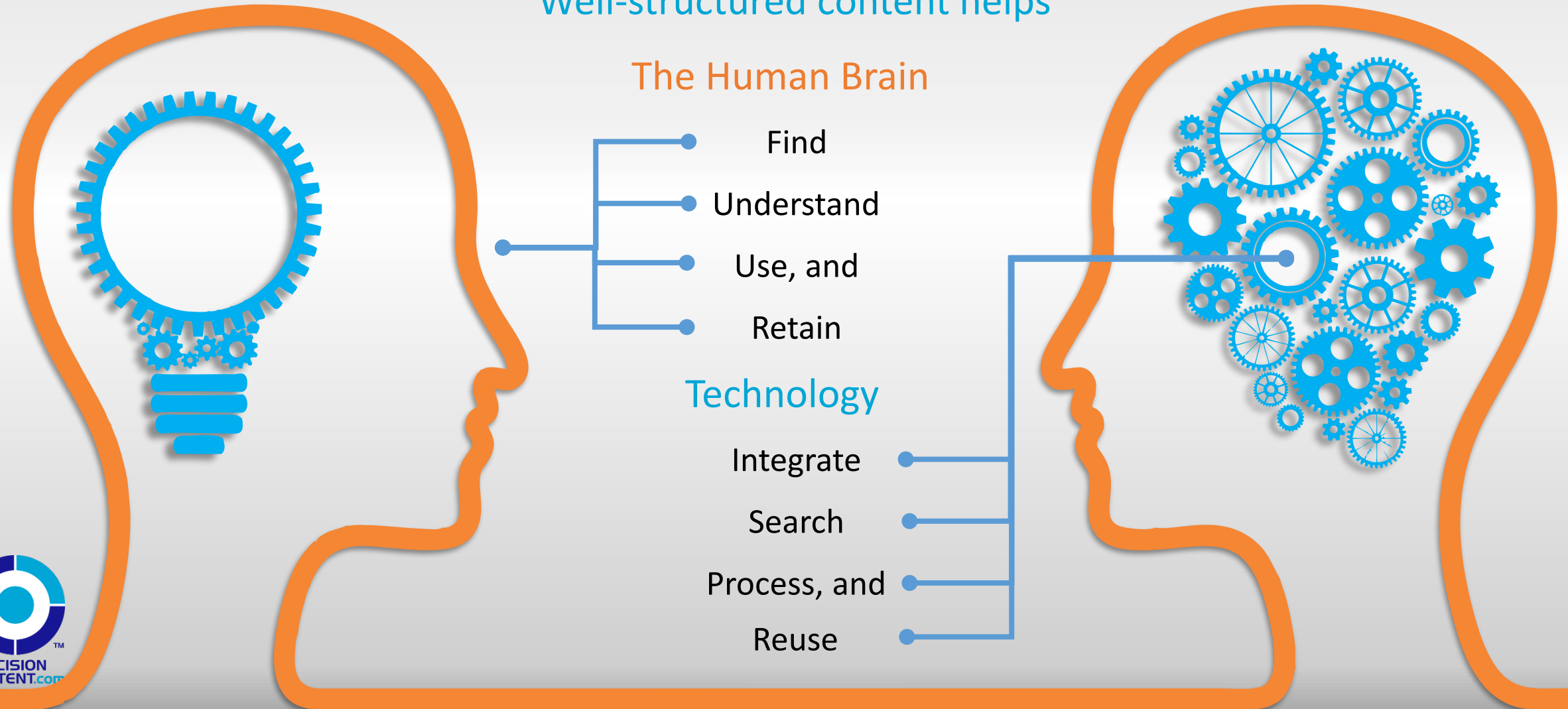
Well-structured content helps

## The Human Brain

- Find
- Understand
- Use, and
- Retain

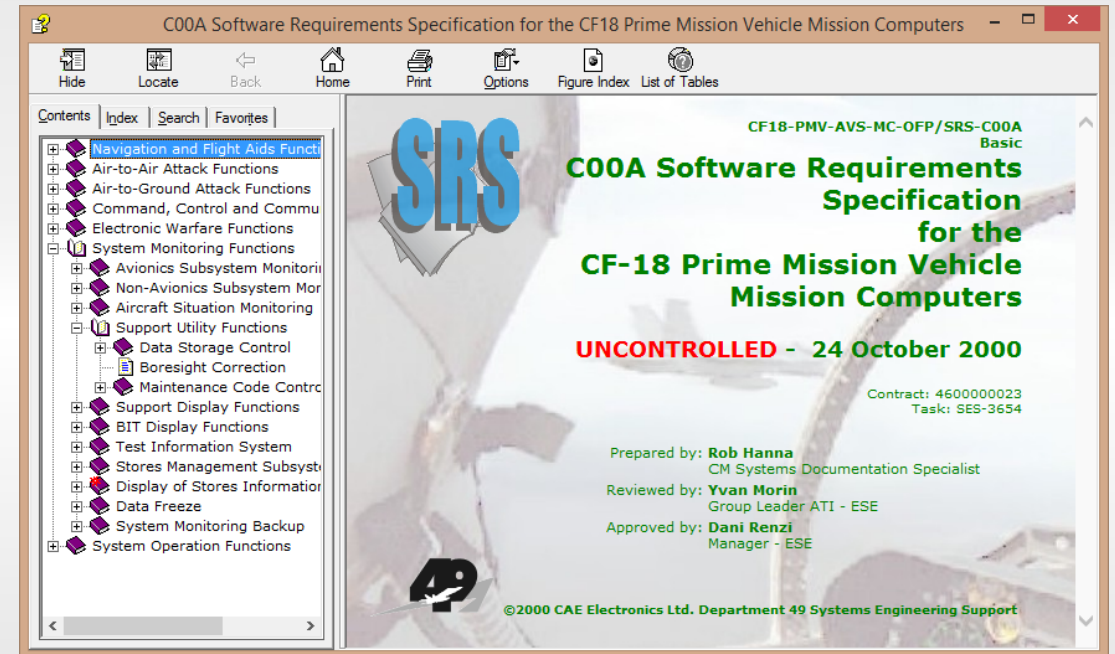
## Technology

- Integrate
- Search
- Process, and
- Reuse



# Early days in structured authoring

- 1997 GML
- Documentation process long and fraught with opportunities for human error
- Parsed GML into Access
- Created taxonomies
- Integrated structured data
- Output to HTML (.chm)

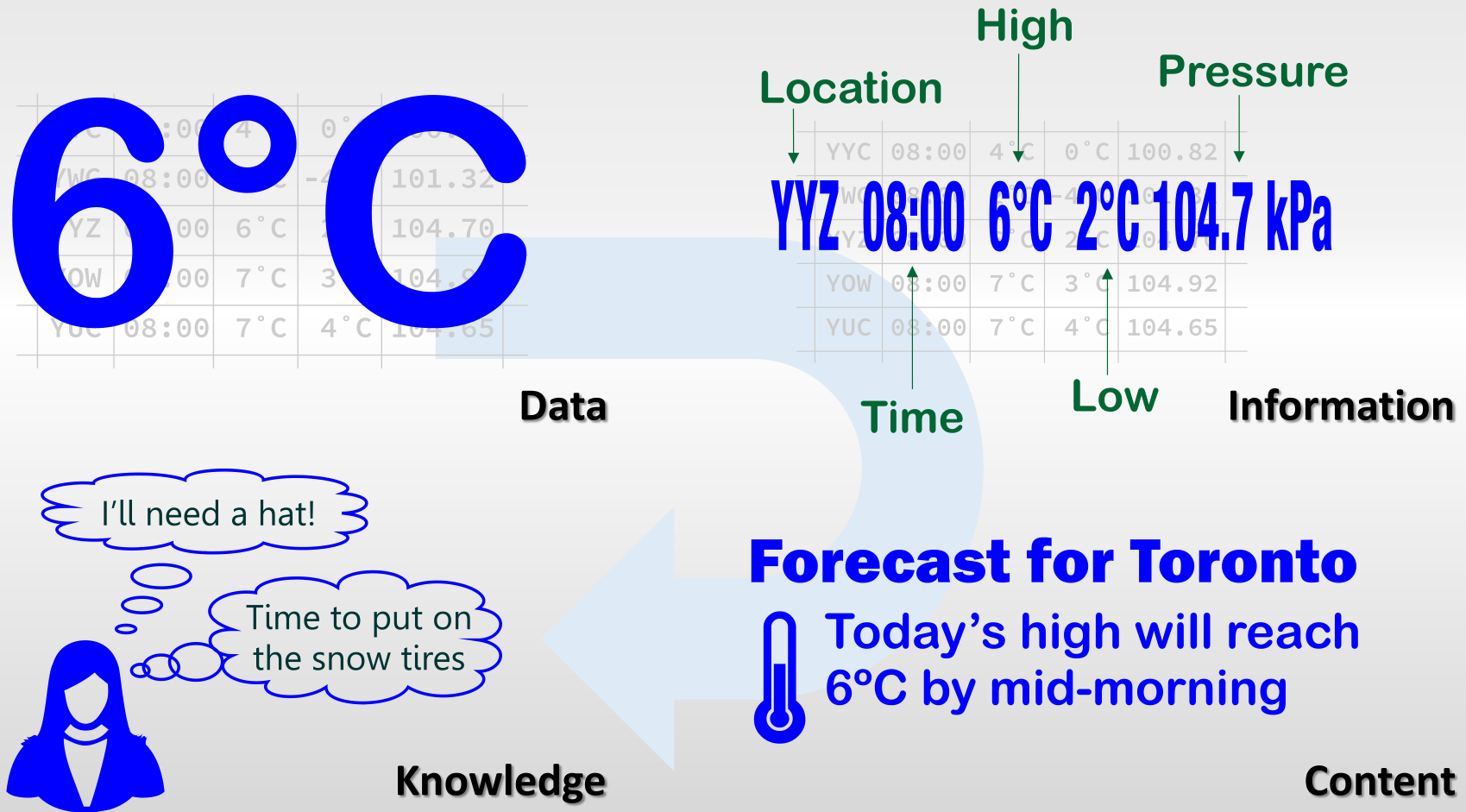




# Structured Authoring

The whys and whatfors

# What is Content?



# Language Arts

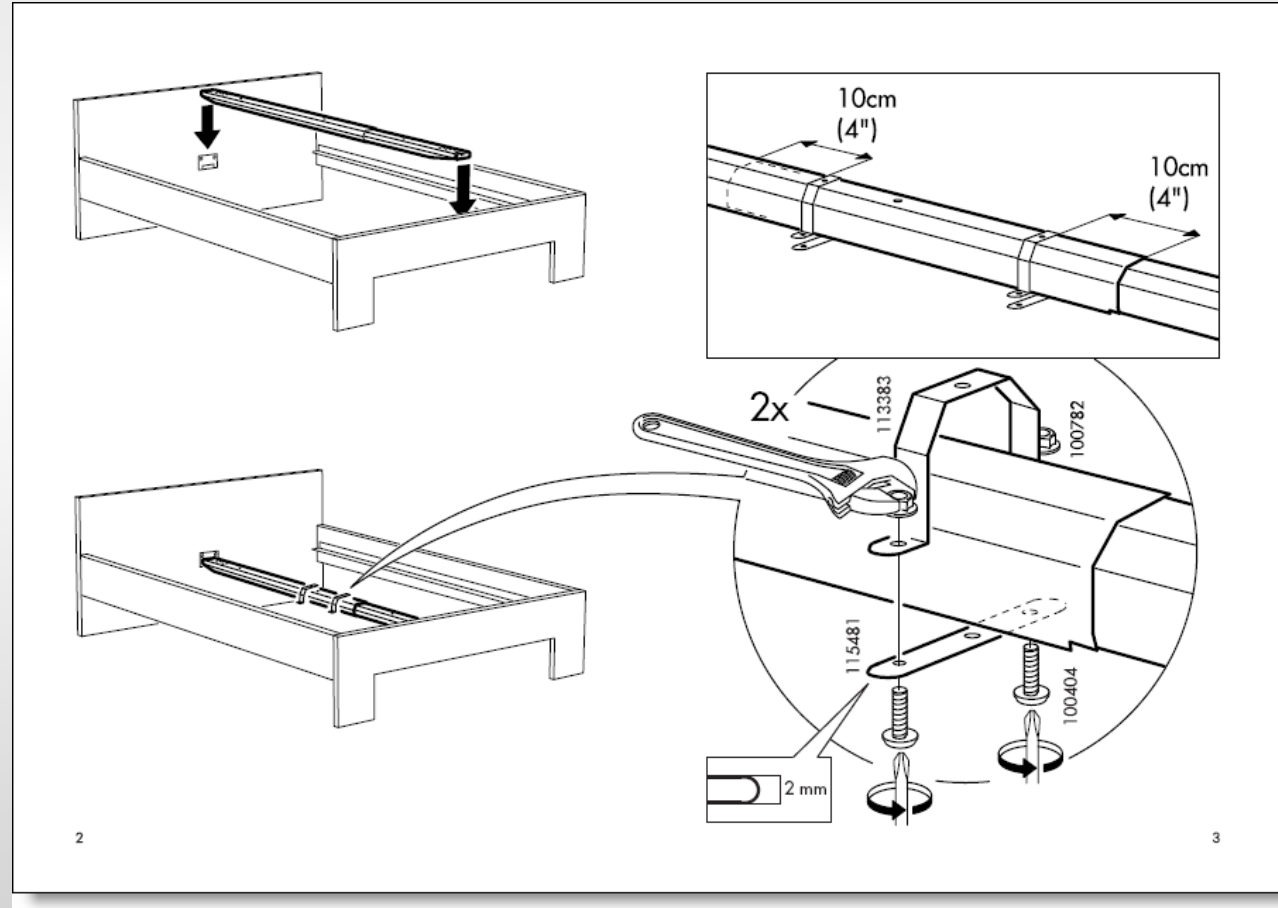
## Language Arts for Personal Response (LAFPR)

- To emotionally engage the reader
- Techniques:
  - narrative style
  - varied vocabulary & sentence structure
  - withholding information
- Writer driven
- Meant to be READ

## Language Arts for Information (LAFI)

- To convey information that readers need to use
- Techniques:
  - consistent modular structure
  - concise, direct vocabulary
  - use of graphics
- Reader driven
- Meant to be USED

# Ikea Instructions: LAFI



# Ikea Instructions: LAFPR

- If novelist Michael Ondaatje wrote Ikea instructions ....

“The eel-shaped talisman squirms inside the raspy recycled box.  
A series of quarter turns – clock hands marking time – bonds  
back to base. An alphabet of connections in English and French.  
A into groove B. C slots into D. Chipboard credenza  
communicating Swedish hegemony.”

- Author/parodist: Geoff Thomas  
Globe & Mail, August 27, 2009

# What is Topic-Based Authoring?



“Topic-based authoring is a modular content creation approach...”

“A topic is a discrete piece of content that is about a specific subject, has an identifiable purpose, and can stand alone...”

[http://en.wikipedia.org/wiki/Topic-based\\_authoring](http://en.wikipedia.org/wiki/Topic-based_authoring)

# Topic-Based Architecture

- Topics are standardized units of information based upon information type
- Topics require only navigational reference for context and can be read in any order
- Topics must all contain a descriptive title and normally include a body and metadata section
- Topics represent a single unit of work for authors

# Opportunities for Topic-Based Authoring

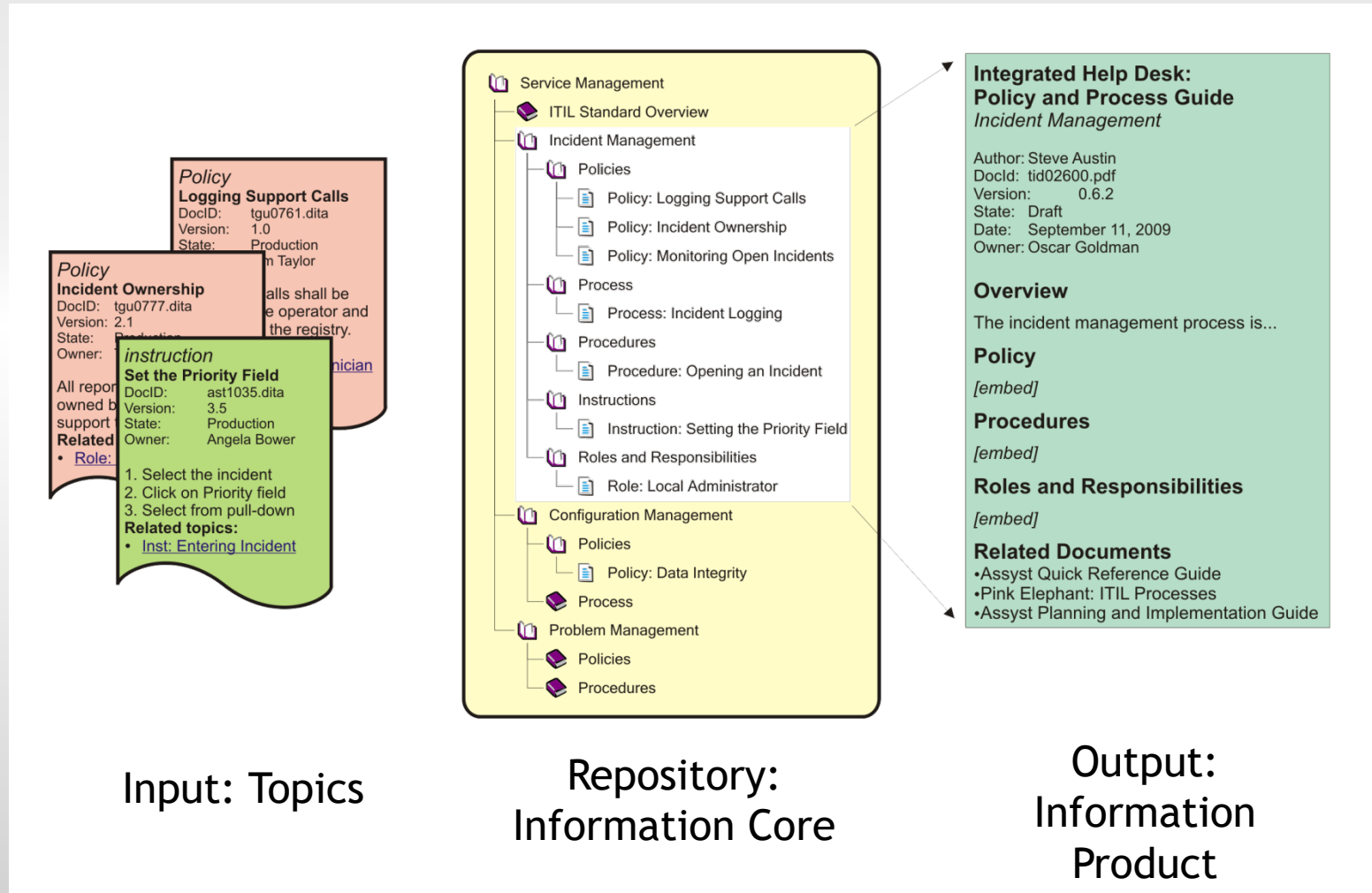
- Speed to market
- Reduced maintenance effort
- Better opportunities for reuse
- Balancing workload
- Finer control over project management
- Opportunities for collaboration
- Clearer, more concise minimalist content



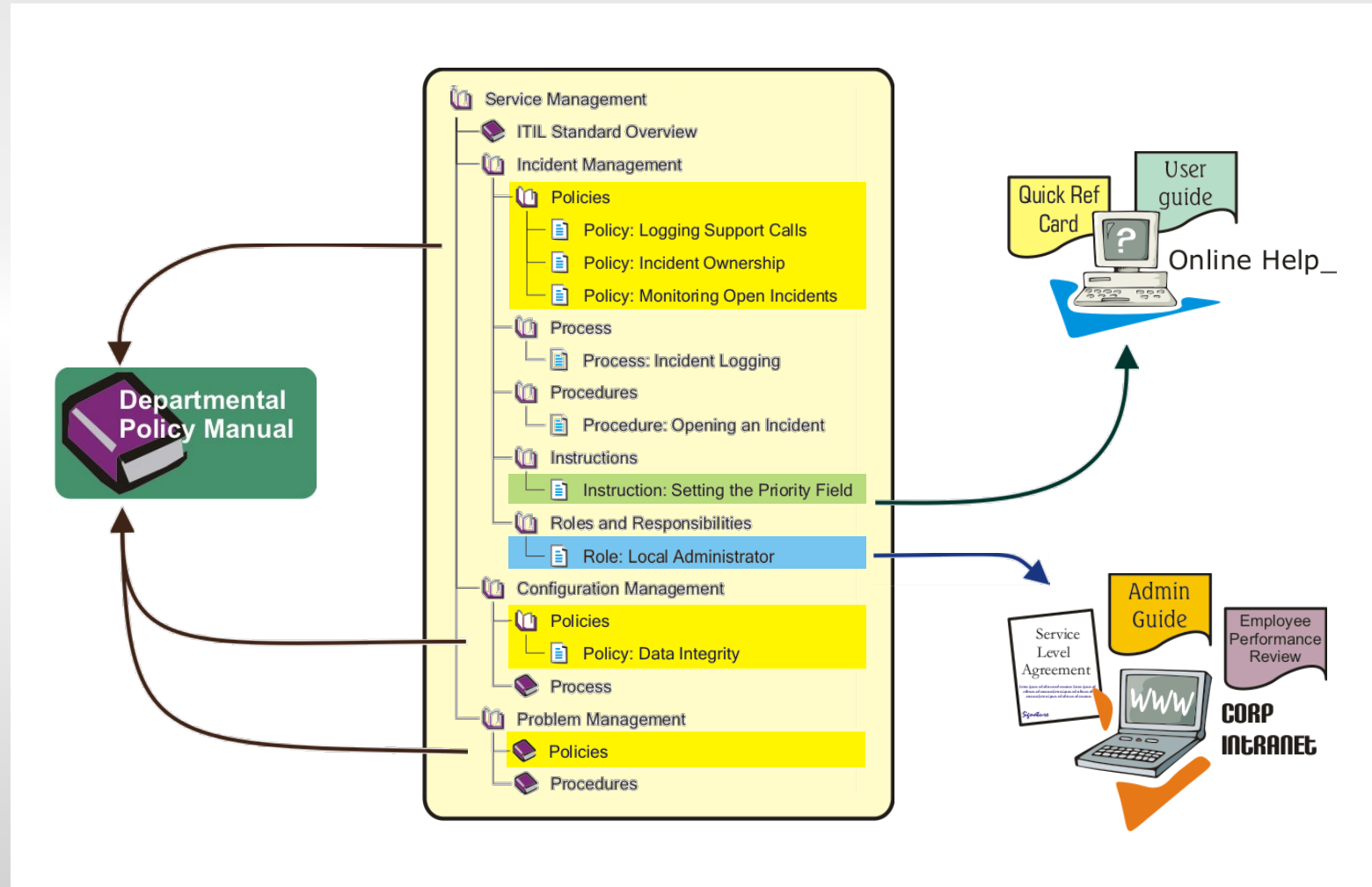
# Challenges for Topic-Based Authoring

- Writing process requires greater discipline
- Loss of content ownership for authors
- Less control over look and feel
- Responsibilities redefined
- Loss of context for SMEs and authors
- More to manage

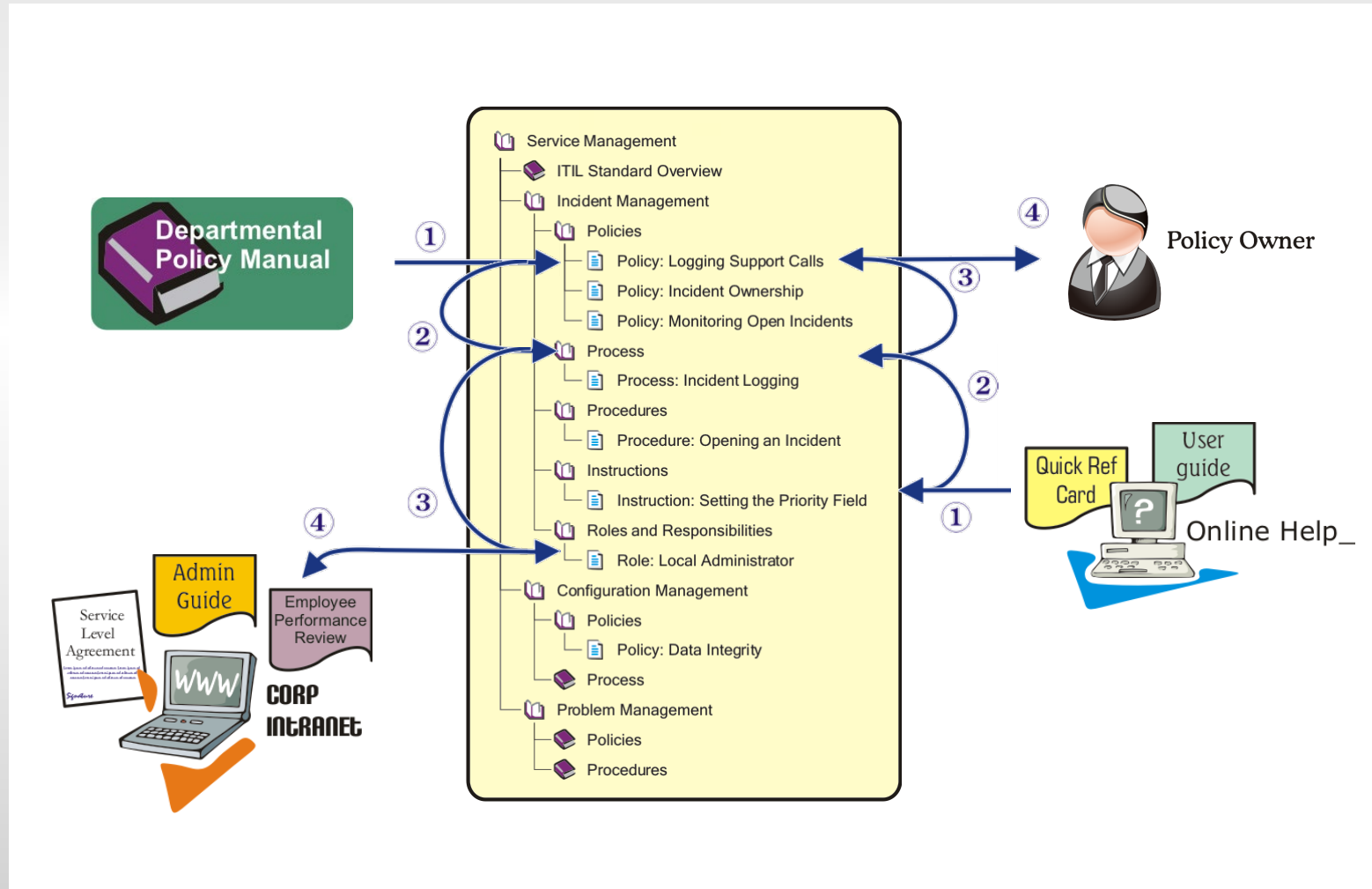
# Topic-based Content Lifecycle



# Reusability



# Traceability



# Working definitions

**Single-sourcing** is any process used to systematically create information products from a single defined source of information.

*or*

*“Writing information once  
and using it many times”*

- Ann Rockley, 2001

# Single-Sourcing

*Single-sourcing is not about content reuse  
- it is about reusable content.*

# Single-sourcing technology

**Single-sourcing** is a methodology, not a technology. Although the software tools associated with single-sourcing are complex, it is the modular writing, not technology, that ultimately determines the success of your single-sourcing project.

- *Kurt Ament – Single Sourcing: Building Modular Documentation, 2003*

# Can single-sourcing help?

- Single-sourcing will...
  - Improve the consistency of information
  - Save on maintenance and customization efforts
  - Improve the quality of the content
  - *Require significant upfront planning and investment*



# Return on investment (ROI)

- Expect return on investment if...
  - The document is deliverable to clients or is tied directly to a product or service
  - The document has a long life expectancy
  - Many updates can be expected over time
  - Several variants may exist at any one time
  - Parts of the document are reused elsewhere
  - I can expect to recover my *actual* costs

# Planning and process

- Defining your objectives
- Analysing your audience
- Defining your source and structure

# Rob Hanna, CIP IMCC

29-7340 Copenhagen Rd Mississauga, Canada L5N 2S5

rob@infoarchitects.ca

+1 (416) 723-4183

http://www.infoarchitects.ca

## Summary

Rob Hanna is a professional technical communicator and leading expert in structured XML authoring, DITA and content management. Over the course of his career, Rob has consulted to or worked for several Fortune 500 companies including:

- Cisco Systems Inc (NASDAQ: CSCO)
- Sun Microsystems (NASDAQ: ORCL)
- Xerox (NASDAQ: XRX)
- Research In Motion (TSE: RIM)
- CAE Ltd (TSE: CAE)
- Canadian Tire (TSX: CTC)
- Infineon (FWB: IFX)
- Wolters Kluwer (Euronext Amsterdam: WKL)

In 2014, Rob was awarded the rank of Fellow of the Society for Technical Communication (STC). Rob earned his certification from Information Mapping International as an Information Mapping® Certified Consultant™ (IMCC) in 2013 and holds a certification from AIIM as a Certified Information Professional (CIP).

Rob is the Chief Information Architect at Ascan Information Architects Limited in Mississauga, Canada.

Rob is a member of the OASIS DITA Technical Committee working on the Technical Documentation Subcommittee. He is an executive member of AIIM's First Canadian Chapter. He has served as a national director for the STC and as the vice-chairman for the STC Certification Commission.

## Certifications

Certified Information Professional *12/2011–12/2014*  
AIIM • License: CIP-92770

Information Mapping® Certified Consultant™ *05/2013–05/2015*  
Information Mapping International, Belgium

## Experience

Ascan Information Architects Limited *1991–Present*

Chief Information Architect

Technical communication consultancy specializing in content management, structured authoring, and single-source publishing. Services include strategic development, systems analysis and design, information architecture, training, DTD and stylesheet development, automation and scripting, database design, new media consulting and publishing.

Innovatia *11/2010–01/2013*

# A basic structured document

# What are our objectives?

- Key objectives include:
  - Produce a résumé that is easy to update and tailor for each new job application.
  - Maintain various versions: Short/Long; Technical Writer/Team Lead; Contractor/Employee.
  - Maintain various formats including: Word, HTML, PDF, and ASCII text.
  - Showcase help authoring skills by producing a Windows help version.

# Who is my audience

- HR Department
- Placement Agency
- Hiring Manager
- Automated systems

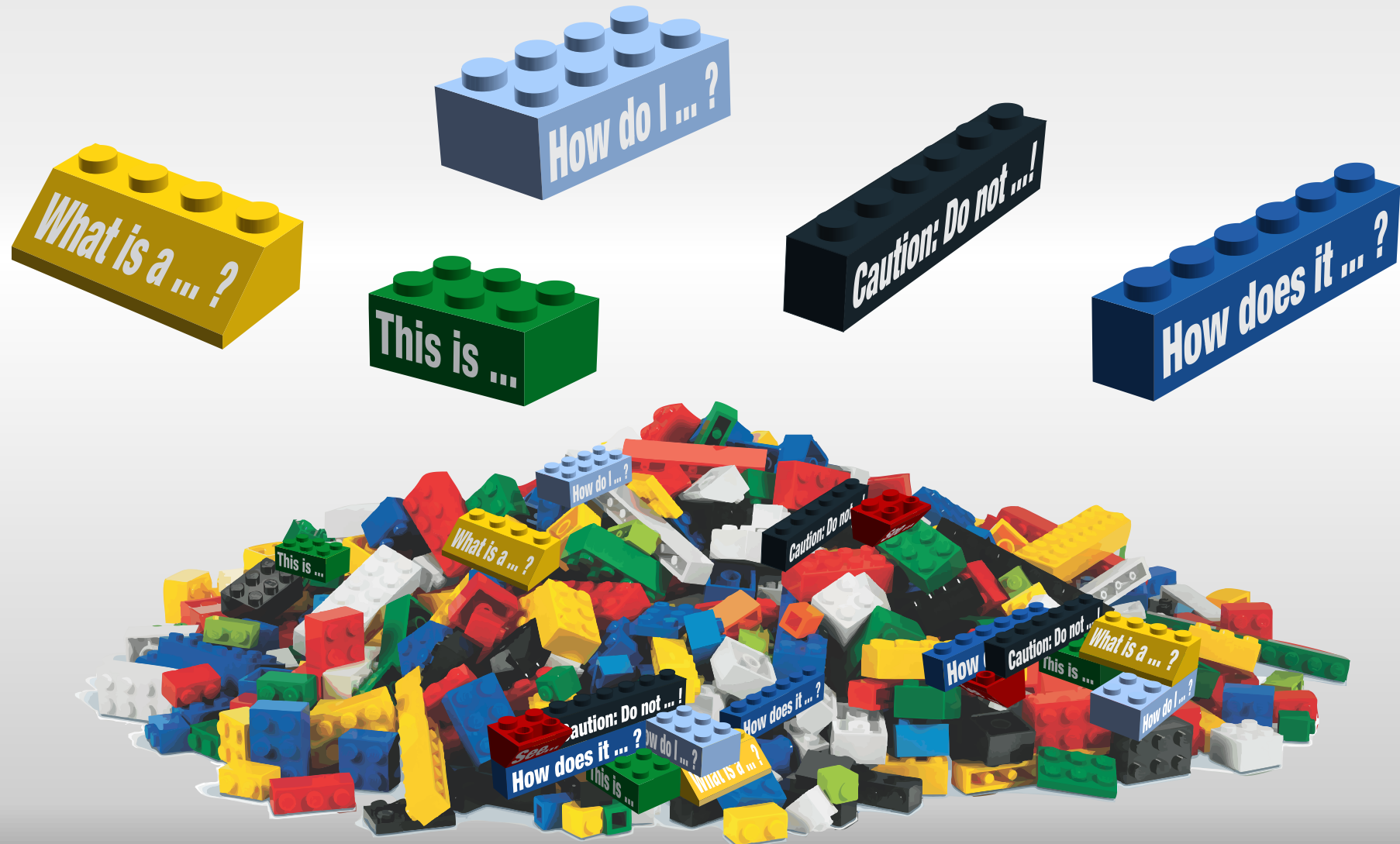
# Selecting a source

- Does a source already exist?
- Do multiple sources exist?
- How will you resolve discrepancies between various sources?
- Create your definitive source of information

# Resume structure

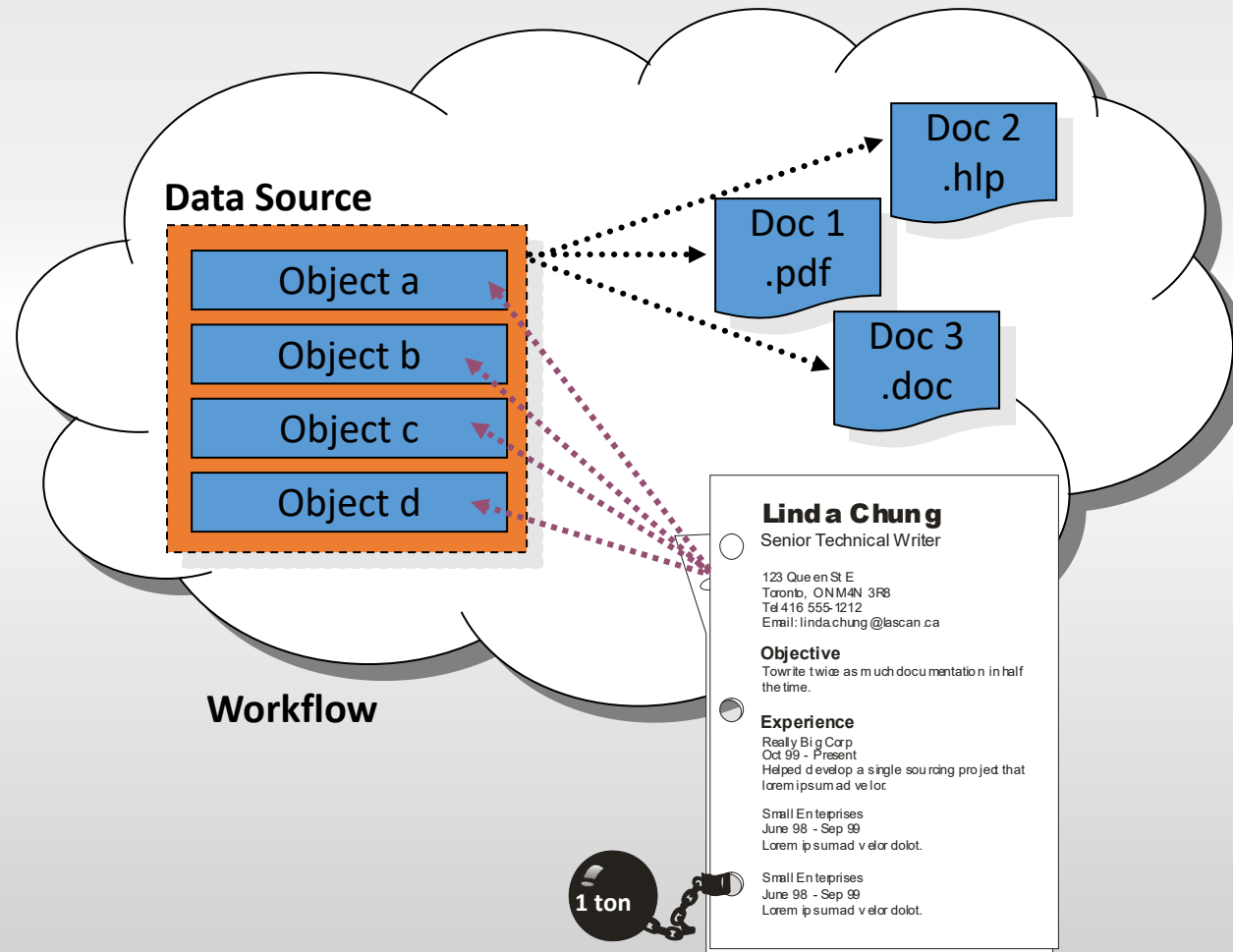
- Resume title
  - Objective
    - Target 1
    - Target 2
  - Personal Profile
  - Education
    - School
    - Location
    - Date
    - Degree
  - Professional Experience
    - Company
    - Location
    - Position
    - Tasks
    - Accomplishments
- Skills
  - Skill name
  - Experience level
  - Last used
  - Years used
- References
  - Name
  - Position
  - Company
  - Contact
    - E-mail
    - Telephone
  - Description

# Structured building blocks of information





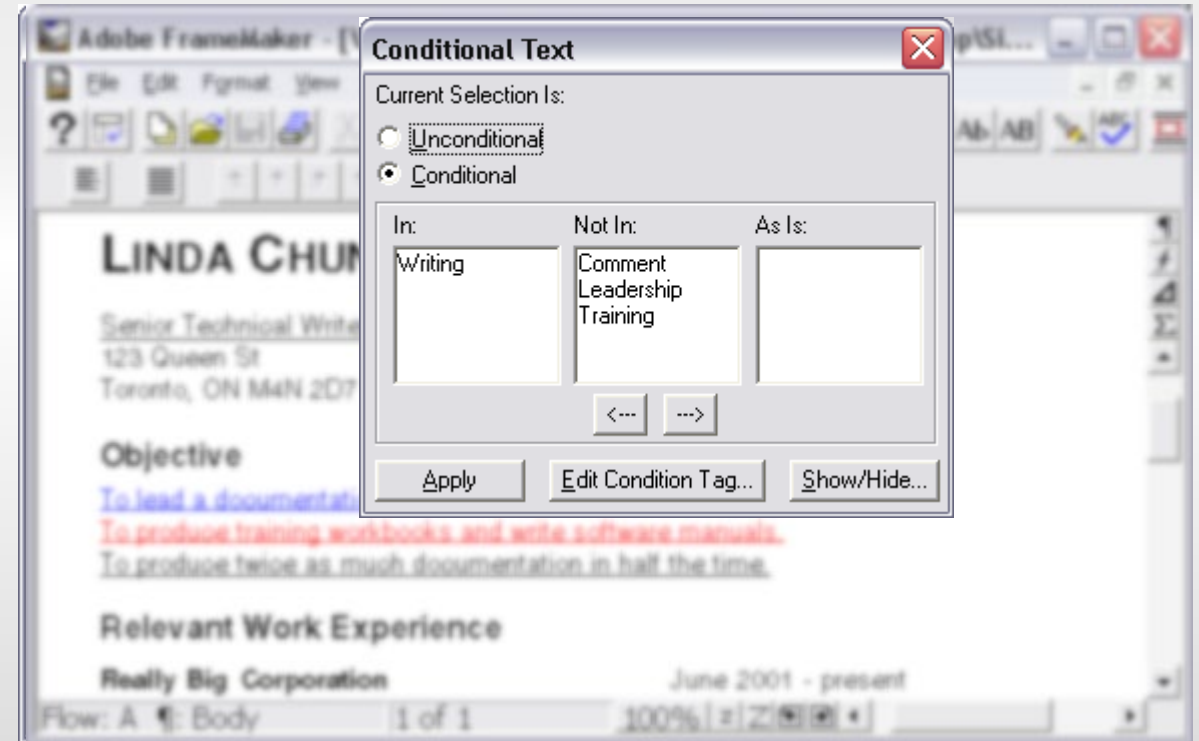
# Single-sourcing with MS Word



- Ball & Chain vs. single-sourcing

# Single-sourcing with FrameMaker

- Consistency is required for good output
  - Consistent style of writing
  - Consistent mark-up of conditions
- FrameMaker limitations
  - Conditions for multiple purposes
  - Text tagged with multiple conditions
  - Cannot use conditional text to rearrange your content



# XML and Markup Languages

Now we get technical...

# What is XML?

- (eXtensible Markup Language) is an open standard for the exchange of information
  - first published in 1998 by W3C
  - to encode electronic documents readable by
    - human, and
    - machine
  - for a multitude of applications ranging from
    - corporate financial reporting applications, to
    - Microsoft Word





**What's with all this talk about XML, Madge?**

**Relax, dear. You're soaking in it.**

## XML is Everywhere

XML defines meaningful data structures for documents and data. It is a human-readable file format used to power

- manufacturing assembly lines
- medical devices
- military applications, and
- many other things.

XML is the language of the Web. It enables smart phones and web browsers.

# What are markup languages?

- pre-date desktop publishing and the Internet
- tell computers how to handle data
  - such as how to render electronic content on a page
- categorized as either
  - presentation, or
  - semantic markup

# Document Content As We Know It

## Set up the writer's name as a variable in Oxygen Author

Comments in documents should always be attributed to the reviewer. Oxygen will insert the writer's name into draft comments automatically once it is configured properly.

### Before you begin

Ensure that the Precision Content framework is installed on your version of Oxygen Author.

### Procedure

The following steps will guide you through the setup of a custom user variable in your Oxygen environment.

1. Open Oxygen Author
2. Select Options > Preferences > Custom Editor Variables
3. Click **New**
4. Enter the appropriate values to create a new custom variable
  - Name: prec\_d\_mapper
  - Value: *{Your Name}*
  - Description: Writer name
5. Click **OK**

*The contents of a document are opaque to the computer...*

*So we have to label them!*

# Presentation markup

- With electronic presentation markup, we markup the paragraph and italicize the citation for publication
- This is typical of web pages using hypertext markup (HTML)

```
<p><i>The Cancer Journal: The Journal of Principles & Practice of Oncology</i> provides an integrated view of modern oncology across <i>all</i> disciplines.</p>
```

*The Cancer Journal: The Journal of Principles & Practice of Oncology* provides an integrated view of modern oncology across *all* disciplines.



# Enter Structured XML

```
<topic><title>Set up the writer's name as a variable in Oxygen Author</title>
<shortdesc>Comments in documents should always be attributed to the reviewer. Oxygen will insert
the writer's name into draft comments automatically once it is configured properly.</shortdesc>
<body><section><title>Before you begin</title>
<p>Ensure that the Precision Content framework is installed on your version of Oxygen
Author.</p></section>
<section><title>Procedure</title>
<p>The following steps will guide you through the setup of a custom user variable in your Oxygen
environment.<ol>
  <li>Open Oxygen Author</li>
  <li>Select Options > Preferences > Custom Editor Variables</li>
  <li>Click New</li>
  <li>Enter the appropriate values to create a new custom variable
    <lines>Name: prec_d_mapper
    Value: {Your Name}
    Description: Writer name</lines></li>
  <li>Click OK</li></ol></p></body></topic>
```

**Getting  
better...**

# Semantic markup

- With semantic markup, we markup the content to describe the *meaning* of the text
- Publishing stylesheets interpret the meaning from the markup and apply appropriate styles specific to the publishing context

```
<intro><cite>The Cancer Journal: The Journal of Principles &  
Practice of Oncology</cite> provides an integrated view of  
modern oncology across <em>all</em> disciplines.</intro>
```

The Cancer Journal: The Journal of Principles & Practice of Oncology provides an integrated view of modern oncology across **all** disciplines.

# Enter Structured Semantic XML

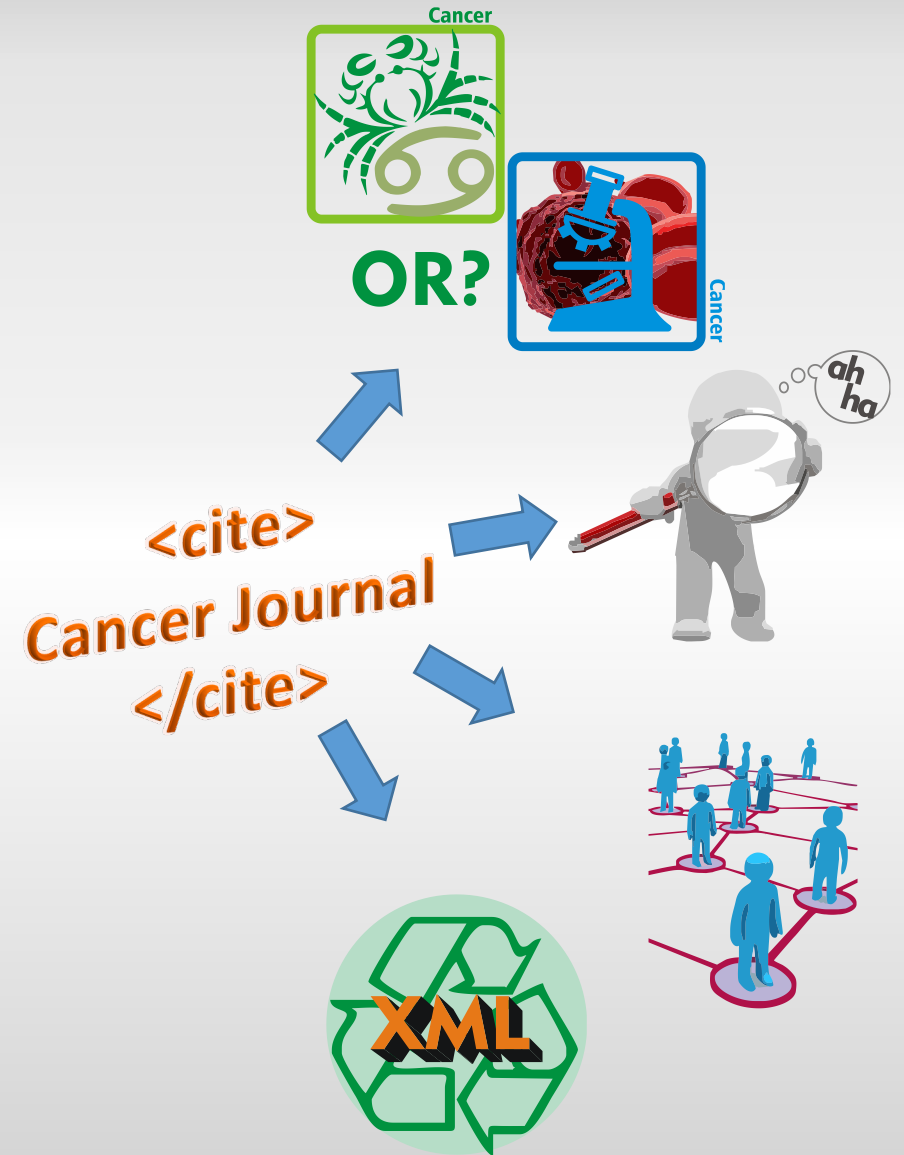
```
<task><title>Set up the writer's name as a variable in Oxygen Author</title>
<shortdesc>Comments in documents should always be attributed to the reviewer. Oxygen will
insert the writer's name into draft comments automatically once it is configured properly.</shortdesc>
<taskbody><prereq><p>Ensure that the Precision Content framework is installed on your
version of Oxygen Author.</p></prereq>
<steps><stepsection>The following steps will guide you through the setup of a custom user
variable in your Oxygen environment.</stepsection>
<step><cmd>Open Oxygen Author</cmd></step>
<step><cmd>Select
<menucascade><uicontrol>Options</uicontrol><uicontrol>Preferences</uicontrol><uicontrol>
Custom Editor Variables</uicontrol></menucascade></cmd></step>
<step><cmd>Click <uicontrol>New</uicontrol></cmd></step>
<step><cmd>Enter the appropriate values to create a new custom variable</cmd>
  <info><lines>Name: prec_d_mapper
  Value: <varname>{Your Name}</varname>
  Description: Writer name</lines></info></step>
<step><cmd>Click <uicontrol>OK</uicontrol></cmd></step></steps></taskbody></task>
```

**Structure  
and  
Meaning!**

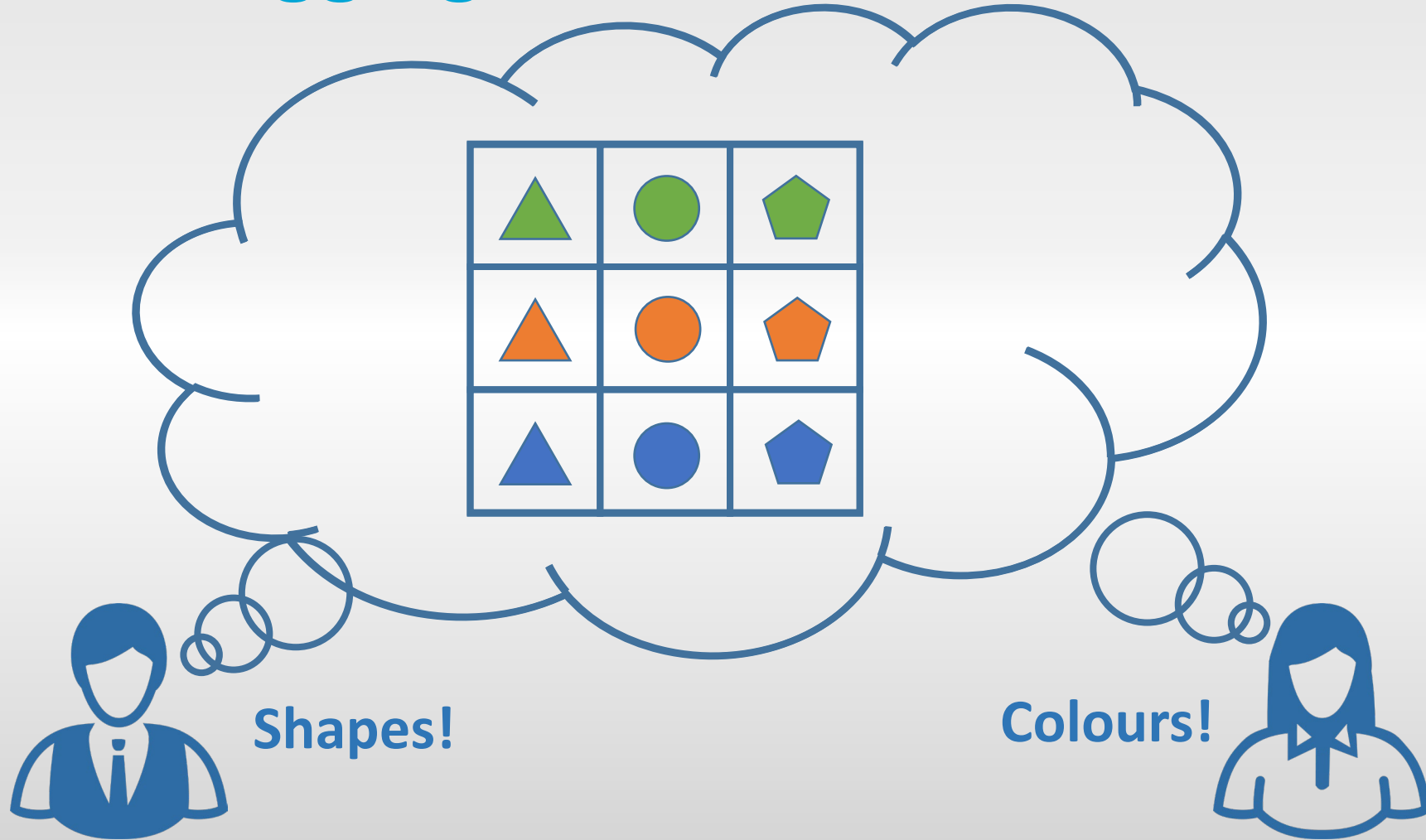
**Enable access!**

# Semantic markup

- Using semantic markup, we can
  - disambiguate content
  - search based on meaning
  - connect to other content, and
  - reuse or substitute new text.



# Semantic Tagging Adds Context to Content



# Intelligent Content

- Content that is
  - not limited to one
    - purpose
    - technology, or
    - output
  - structurally rich and semantically aware, making it
    - discoverable
    - reusable
    - reconfigurable, and
    - adaptable.

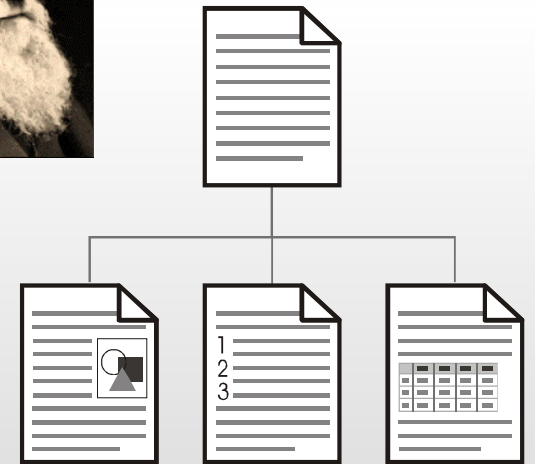
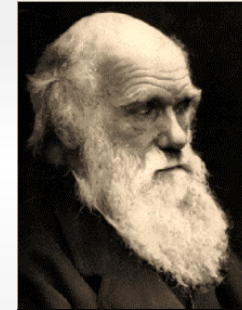
# Introduction to DITA/XML

Head-swimmingly technical...

# What is DITA? *(n. dit-uh)*

- (Darwin Information Typing Architecture) is an XML standard
  - developed in late 90's at IBM, and
  - given to the open source community in 2004
- used for topic-based, structured authoring
- designed for scalability using mechanisms for specialization and inheritance
- defining an extendable set of information types

<dit a>





# Information types

What are my battery specifications?

How do I change the oil?

What does an airbag do?

What does "ABS" mean?



Concept



Task

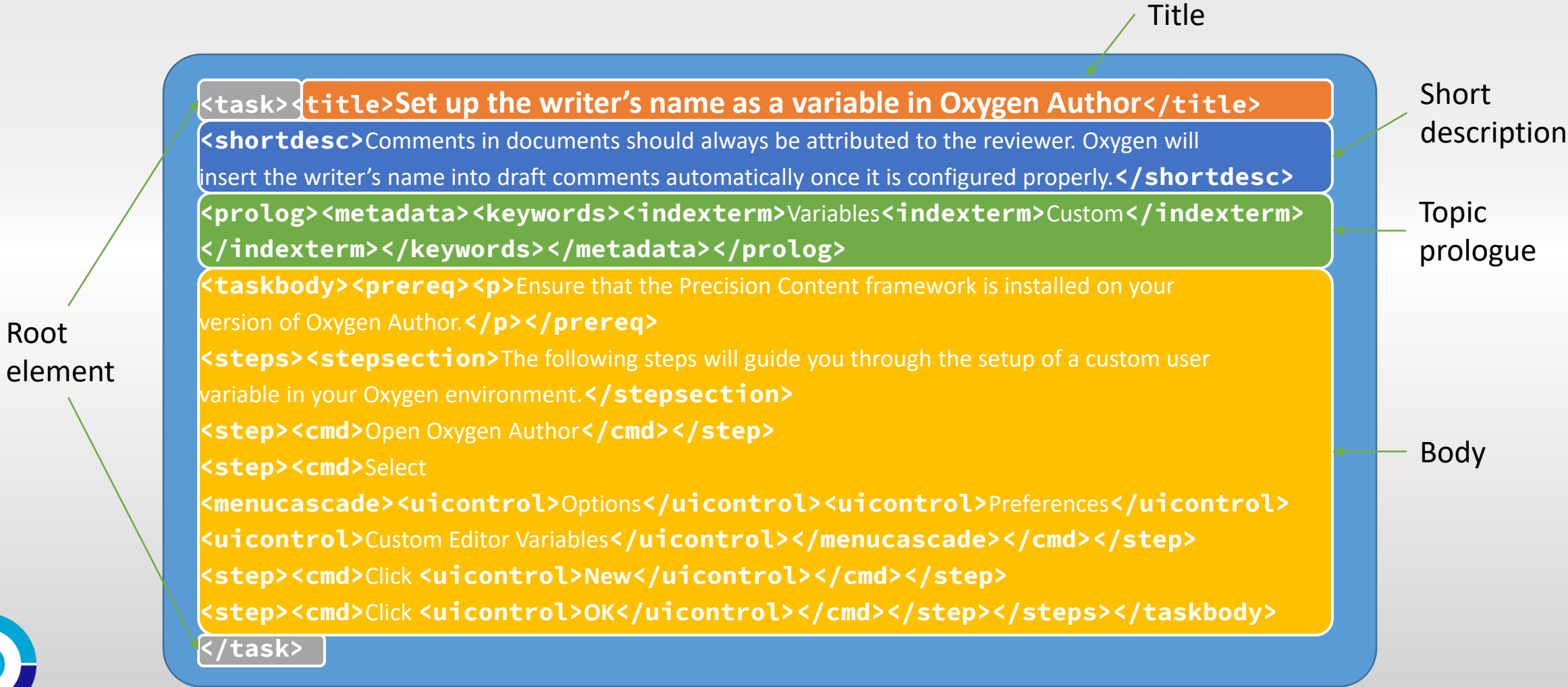


Reference



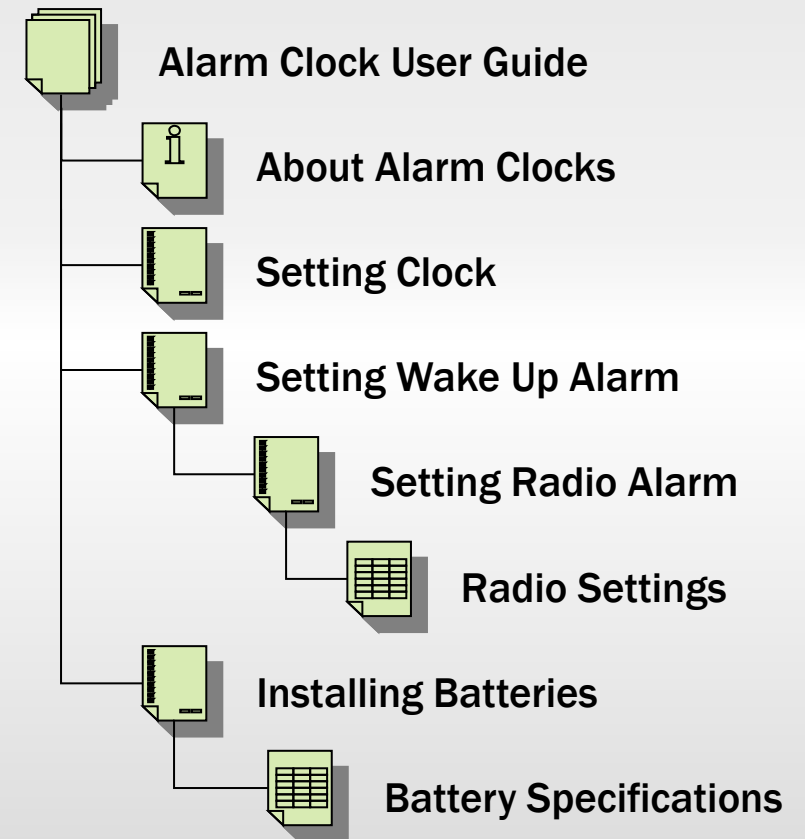
Glossary

# DITA Topic Structure

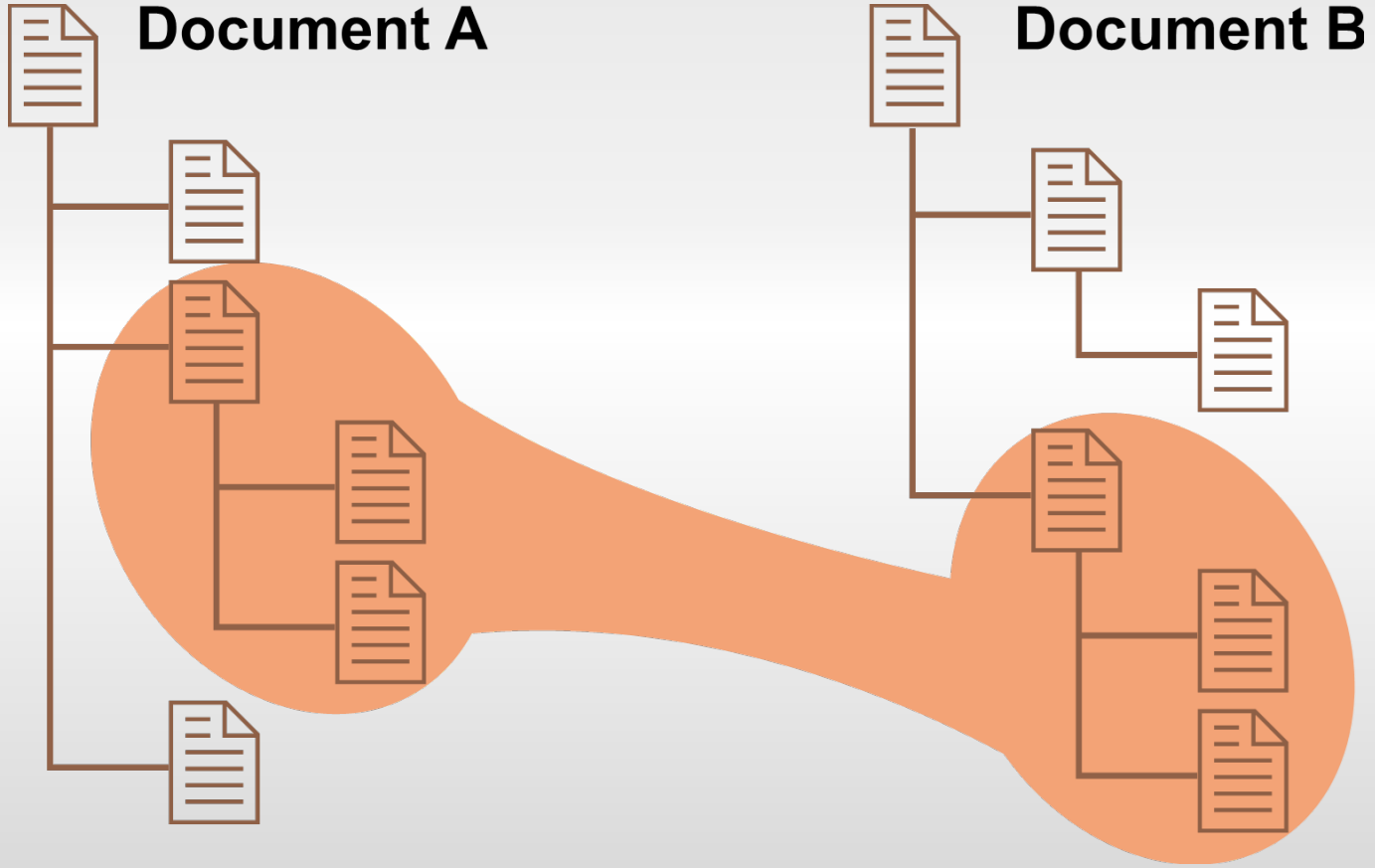


# DITA maps

- Standard DITA information model
  - identifies and presents user tasks
  - maps topics to task flow
  - supplements with
    - conceptual and
    - reference material.
- Maps organize topics into context for publication
- They manage relationships between all topics



# Topic Reuse



# Conditional Reuse (Profiling)

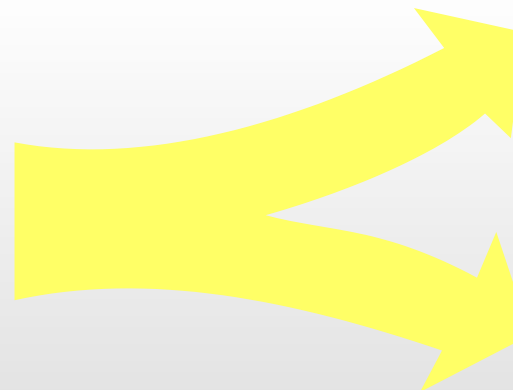
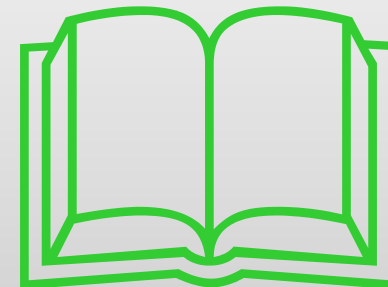
**Topic**



**Variant A**



**Variant B**

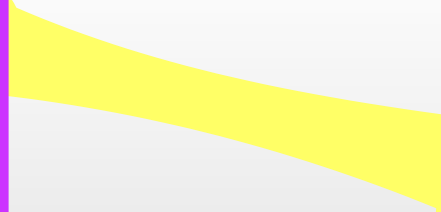
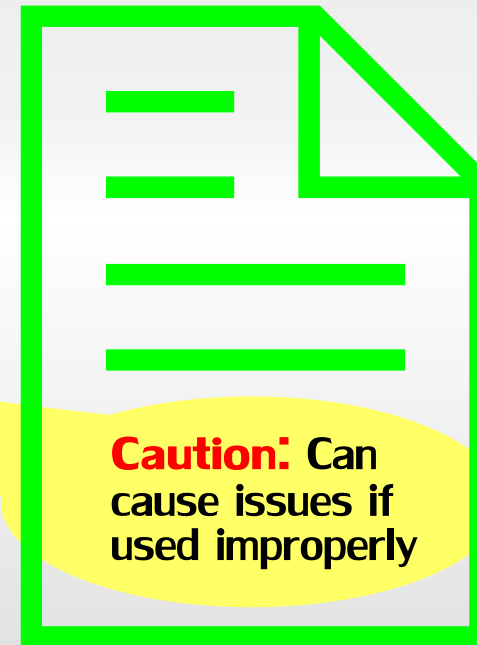


# Fragment-Level Reuse

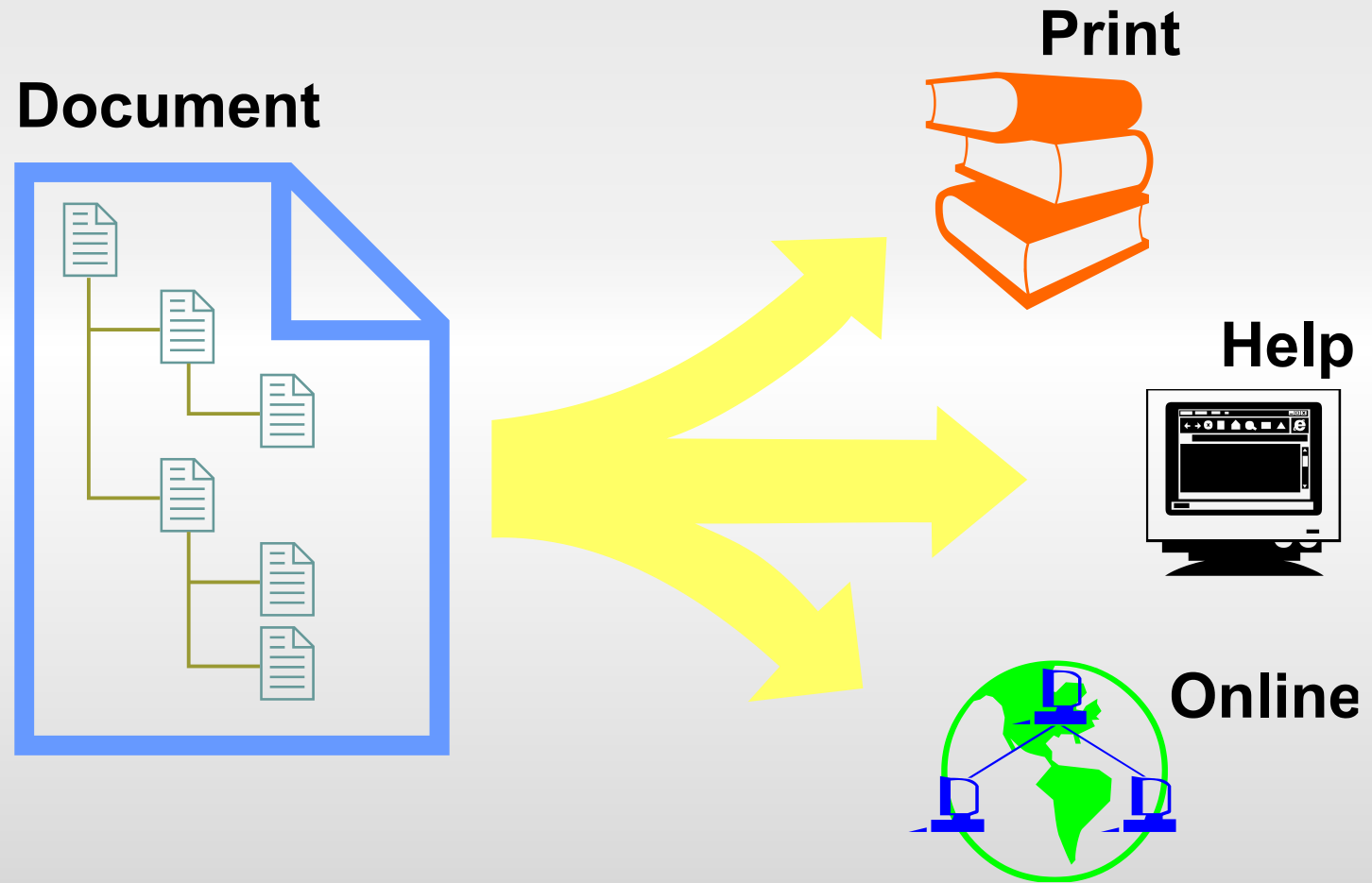
## Topic A



## Topic B



# Repurposing (Multi-Channel Publishing)



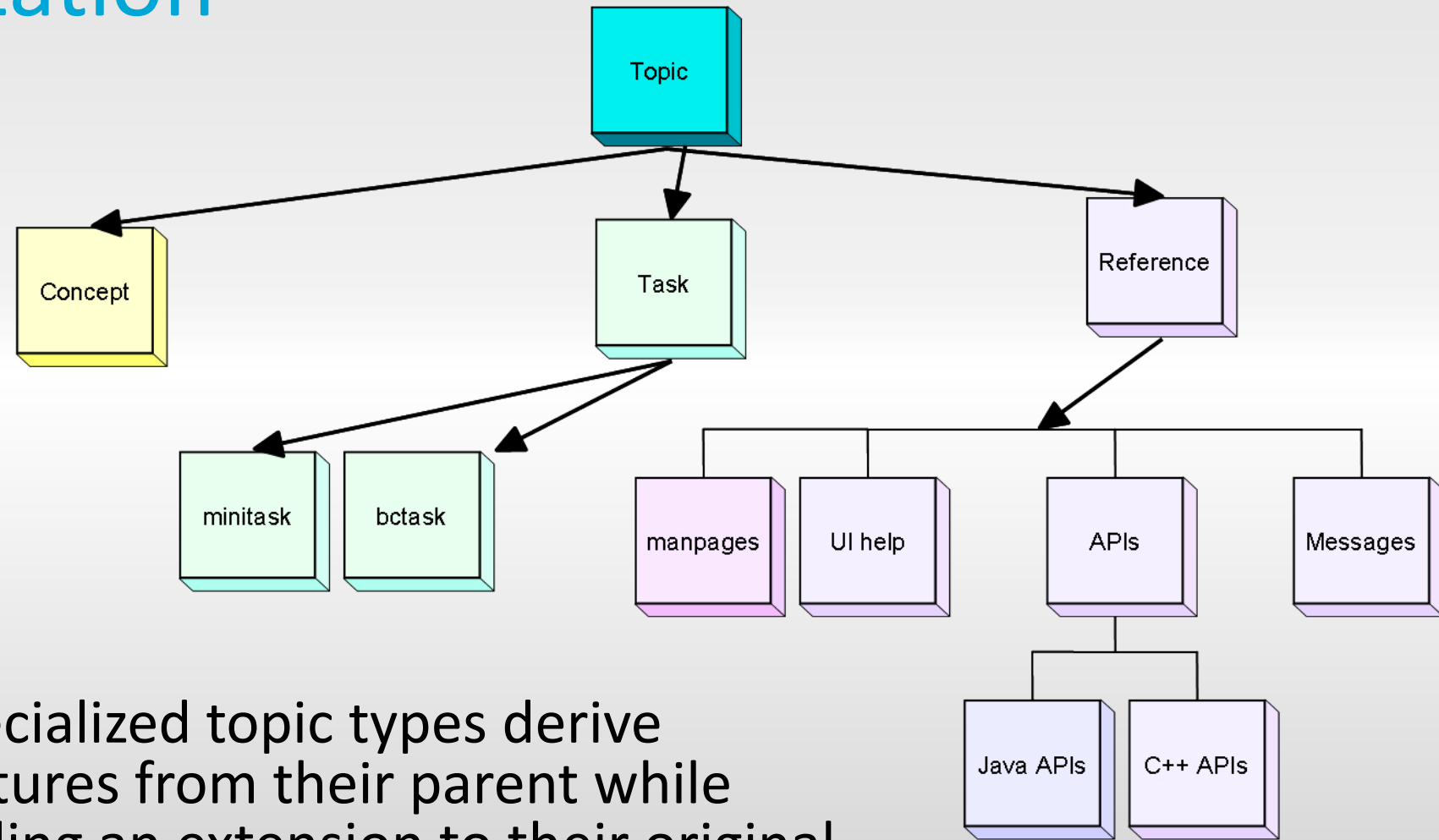
# Multi-Channel publishing

- Supports complex, multi-channel publishing to many common output formats
- Add new formats or styles easily





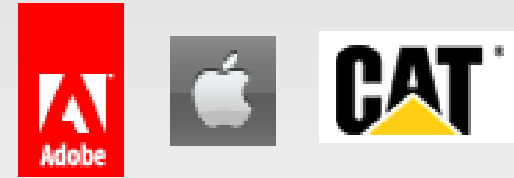
# Specialization



- Specialized topic types derive features from their parent while adding an extension to their original function

# Who Uses DITA?

- Hundreds of companies worldwide, including:
  - Adobe, Apple, Caterpillar, Dell, Elekta, IBM, Juniper Networks, McAfee, Nokia, PayPal, Philips, RIM, SAP, SDL, Xerox and many more
- IBM publishes 60M pages of content in 40 languages using DITA



# Questions?

Rob Hanna, ECMs

+1 (289) 290-4337

[www.linkedin.com/in/singlesourceror](http://www.linkedin.com/in/singlesourceror)

rob@precisioncontent.com

