

University of Sydney
Workshop on Web-based Enterprise Information Systems

Content Model for Reuse:
Systems for Enterprise Content Reuse

Edmund Balnaves, PhD

ejb@prosentient.com.au

<http://www.prosentient.com.au>



The University of Sydney
Australia

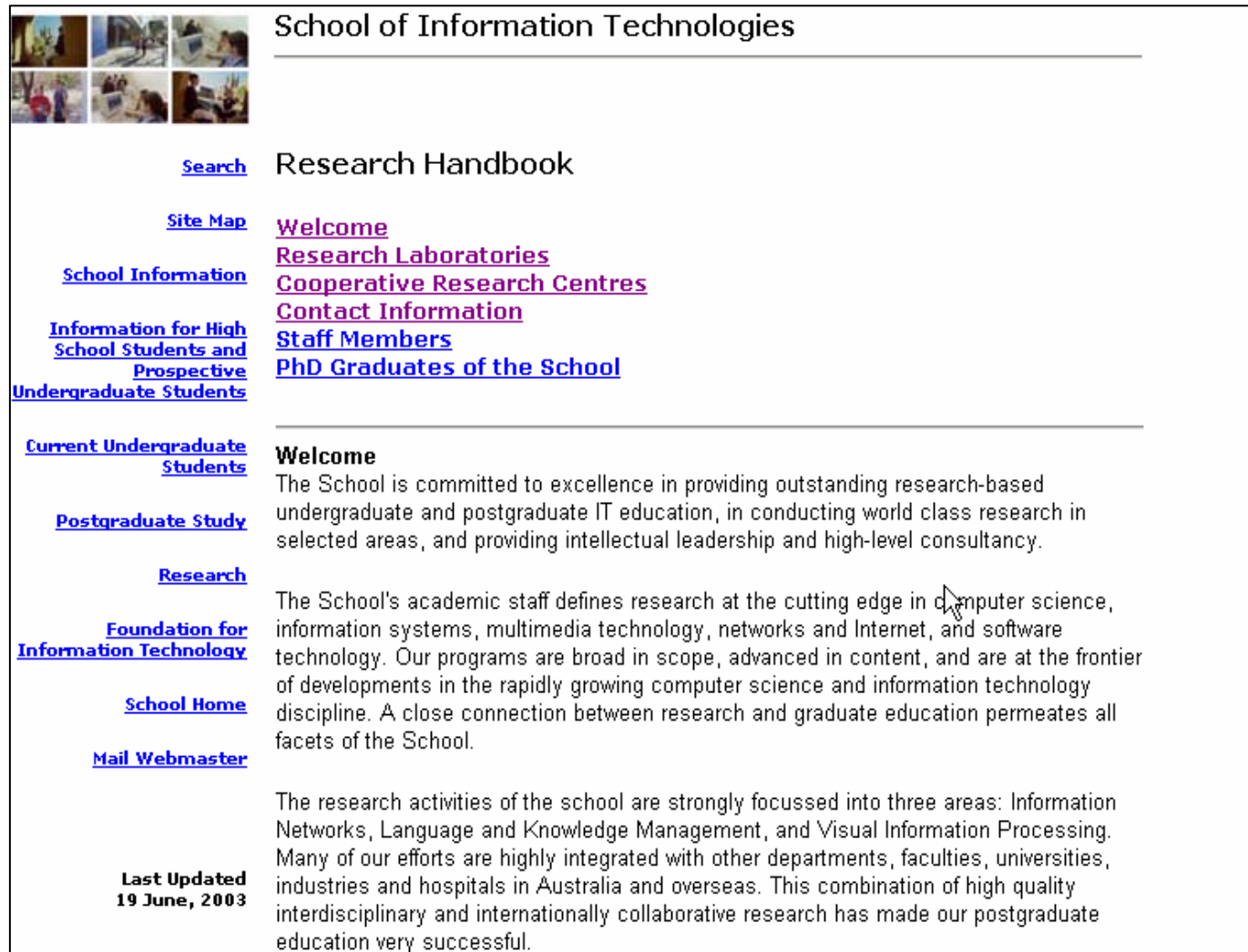



Enterprise CMS - background

- Enterprise CMS
- Extended website editing
- Reuse Situation
 - RSS is popular for simple syndication
 - XML is not the total solution
 - There are many functionally rich web publishing systems, including Open Source but often with little content reuse



Illustration of simple content reuse



 School of Information Technologies

[Search](#) [Research Handbook](#)

[Site Map](#) [Welcome](#)

[School Information](#) [Research Laboratories](#)

[Information for High School Students and Prospective Undergraduate Students](#) [Cooperative Research Centres](#)

[Current Undergraduate Students](#) [Contact Information](#)

[Postgraduate Study](#) [Staff Members](#)

[Research](#) [PhD Graduates of the School](#)

[Foundation for Information Technology](#)

[School Home](#)

[Mail Webmaster](#)

Last Updated
19 June, 2003

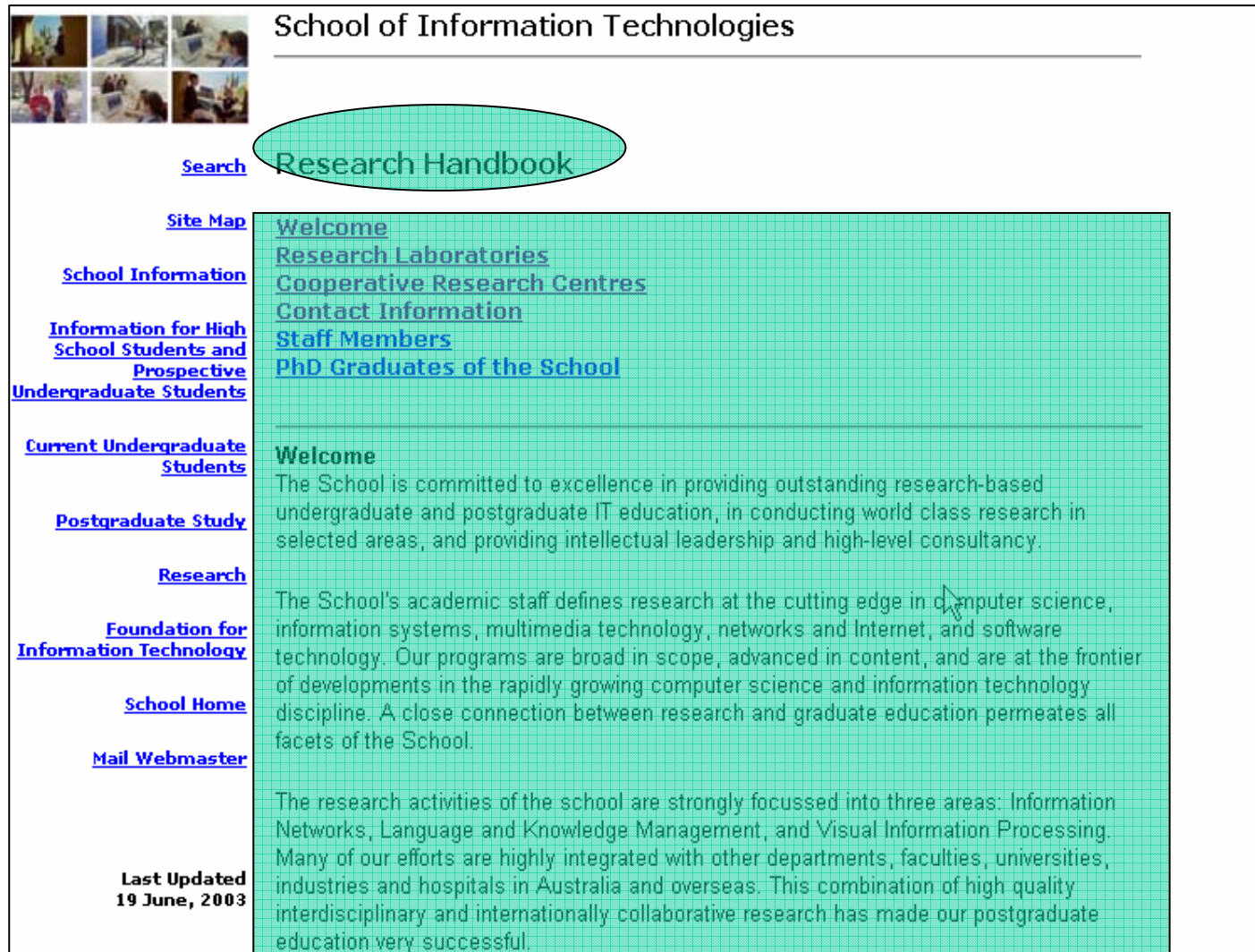
Welcome
The School is committed to excellence in providing outstanding research-based undergraduate and postgraduate IT education, in conducting world class research in selected areas, and providing intellectual leadership and high-level consultancy.

The School's academic staff defines research at the cutting edge in computer science, information systems, multimedia technology, networks and Internet, and software technology. Our programs are broad in scope, advanced in content, and are at the frontier of developments in the rapidly growing computer science and information technology discipline. A close connection between research and graduate education permeates all facets of the School.

The research activities of the school are strongly focussed into three areas: Information Networks, Language and Knowledge Management, and Visual Information Processing. Many of our efforts are highly integrated with other departments, faculties, universities, industries and hospitals in Australia and overseas. This combination of high quality interdisciplinary and internationally collaborative research has made our postgraduate education very successful.



Trivial content reuse



School of Information Technologies

[Search](#) [Research Handbook](#)

[Site Map](#)

[School Information](#)

[Information for High School Students and Prospective Undergraduate Students](#)

[Current Undergraduate Students](#)

[Postgraduate Study](#)

[Research](#)

[Foundation for Information Technology](#)

[School Home](#)

[Mail Webmaster](#)

Last Updated
19 June, 2003

Welcome
[Research Laboratories](#)
[Cooperative Research Centres](#)
[Contact Information](#)
[Staff Members](#)
[PhD Graduates of the School](#)

Welcome
The School is committed to excellence in providing outstanding research-based undergraduate and postgraduate IT education, in conducting world class research in selected areas, and providing intellectual leadership and high-level consultancy.

The School's academic staff defines research at the cutting edge in computer science, information systems, multimedia technology, networks and Internet, and software technology. Our programs are broad in scope, advanced in content, and are at the frontier of developments in the rapidly growing computer science and information technology discipline. A close connection between research and graduate education permeates all facets of the School.

The research activities of the school are strongly focussed into three areas: Information Networks, Language and Knowledge Management, and Visual Information Processing. Many of our efforts are highly integrated with other departments, faculties, universities, industries and hospitals in Australia and overseas. This combination of high quality interdisciplinary and internationally collaborative research has made our postgraduate education very successful.



More complex content separation

A collection of staff profiles generates this list

Associate Professor Hossam ElGindy, University of New South Wales, Australia
Dr Vladimir Estivill-Castro, University of Newcastle, Australia
Dr Antonios Symvonis, University of Ioannina, Greece

Funding

A collection of publications generates this list

- "Visual Information Retrieval", ARC-APAI (Feng, Jin), \$101,060, 1997-1999.
- "Human Vision Systems", Academy of Sciences (Feng, Sun), \$4,500, 1998-1999.
- "Non-invasive Measurement", NH&MRC (Feng, Fulham, Meikle), \$138,108, 1998-2000.
- "Novel FI Data Storage & Retrieval Techniques", ARC (Feng, Fulton), \$154,941, 1999-2001.
- "Functional Imaging for SPECT", ARC, (Feng, Eberl), \$200,000, 2000-2002.
- "Quantification of functional images" ARC Large (Feng) \$174,762, 2001-2003.
- "A web-based multimedia tele-medicine system" ARC-SPIRT (Feng, Fulham, Eberl) \$335,860, 2001-2003.
- "Knowledge acquisition in content-based image retrieval" ARC-APAI (Jin) \$66,876, 2001-2003.
- "Modelling and cataloguing video data" ARC-APAI (Jin) \$66,876, 2001-2003.
- "Modelling, cataloguing, processing and retrieving video data" ARC-IREX (Jin with Tsinghua University, China) \$8,300, 2001, 2002-2003 tba.
- "Design and analysis of simple greedy dynamic on-line routing algorithms", ARC (Symvonis), \$210,000, 1999-2001.
- "Clustering Methods for Graphs", ARC Large Grant, (Eades) \$168,000 2000-2002.

All resources hyperlinked to further details and cross-linked back to the research page



Digital Collection Risks

SHORT TERM (0-5 years)

System continuity.

Duplication of subscriptions with different aggregate database providers.

Heterogeneity of architectures to support

Hardware, Networking and Operating System interruption.

MEDIUM TERM (6-15 years)

Changes to media storage and content encoding.
The cost of retrospective conversion of content into current delivery methods

Business failure or ownership changes in key service suppliers.

Licensing changes in content supply.

LONG TERM (> 15 years)

Business failure of digital subscription suppliers

Inability to sustain subscription charges for retrospective material.

National boycotts

Proliferation of online resources

Long-term content location and identification

Long-term persistence of access

Media longevity for archival digital content.



Principles for reuse

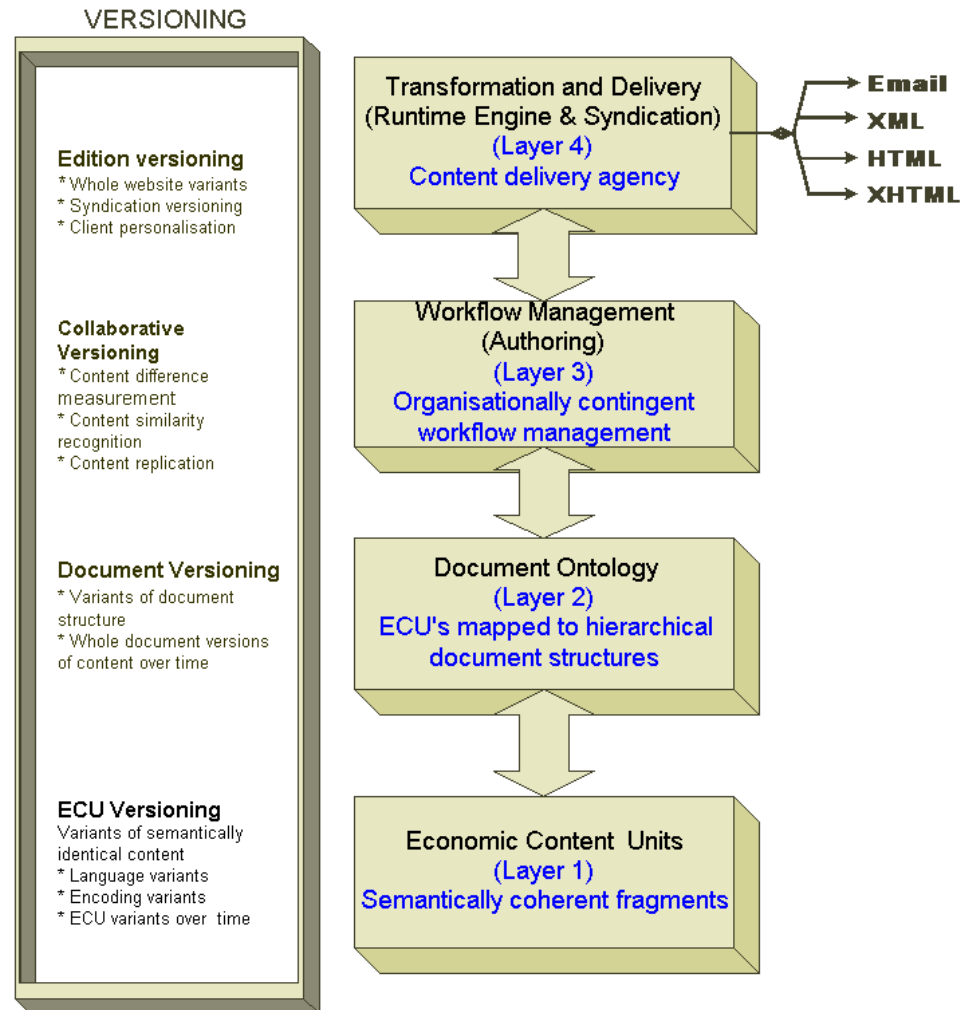
Content Fragmentation	<ul style="list-style-type: none"> • Content reuse requires the organisation of content in a manner that allows the separation of content from its final presentational form. • Content discovery, and semantic information that facilitates this discovery, is an essential element of content reuse. • Fragmentation of content is essential to achieve content reuse, although the granularity of this fragmentation is contingent on the reuse objectives.
Document Organisation	<ul style="list-style-type: none"> • Content syndication and reuse rely on agreed ontologies for the description of semantic content. Industry implementations remain domain specific, but there is a convergence toward RDF. The ability to address multiple semantic ontologies remains important. • A consistent document exchange ontology with strong community support is an essential prerequisite to content exchange. The CMS should accommodate the flexible mapping of content and metadata to accepted document ontologies. • Stability of the document ontology over time will improve content discovery and reuse. • Content discovery is facilitated by: <ul style="list-style-type: none"> ◦ Semantic descriptors (using controlled vocabularies) supplementing the original content. ◦ Support for complex query structures. ◦ The ability to persist search content queries and their resulting document collections. • Multimedia content reuse characteristically requires flexibility in the support of a heterogeneity of document models for authoring, publishing and syndication.
Content authoring	<ul style="list-style-type: none"> • Facilities for content discovery and structured capture of semantic information to facilitate content discovery are essential to content reuse. • Content discovery is facilitated by Semantic descriptors (using controlled vocabularies) supplementing the original content, support for complex query structures and the ability to persist search content queries and their resulting document collections. • The end user should be masked as far as possible from the technical aspects of the publishing process to encourage currency of content. • The issue of content reuse goes beyond specification for a document exchange ontology and requires a systems approach that is currently characterised by the Content Management System. The core functions of content management that support content reuse are: modelling, semantic metadata management, authoring workflow management, publishing workflow management, temporal content management and versioning.
Content publishing	<ul style="list-style-type: none"> • The delivery of digital content in its final presentational form will involve some form of transformational process. • Content syndication relies on strong ontologies, which achieve industry acceptance. • The transformation engine should support a flexible, possibly multi-stage, sequence of transformation processing - that is, a flexible generative process. This can be regarded as the “publishing workflow”. • Content reuse over long periods requires a framework that facilitates transformation of content into new media formats.



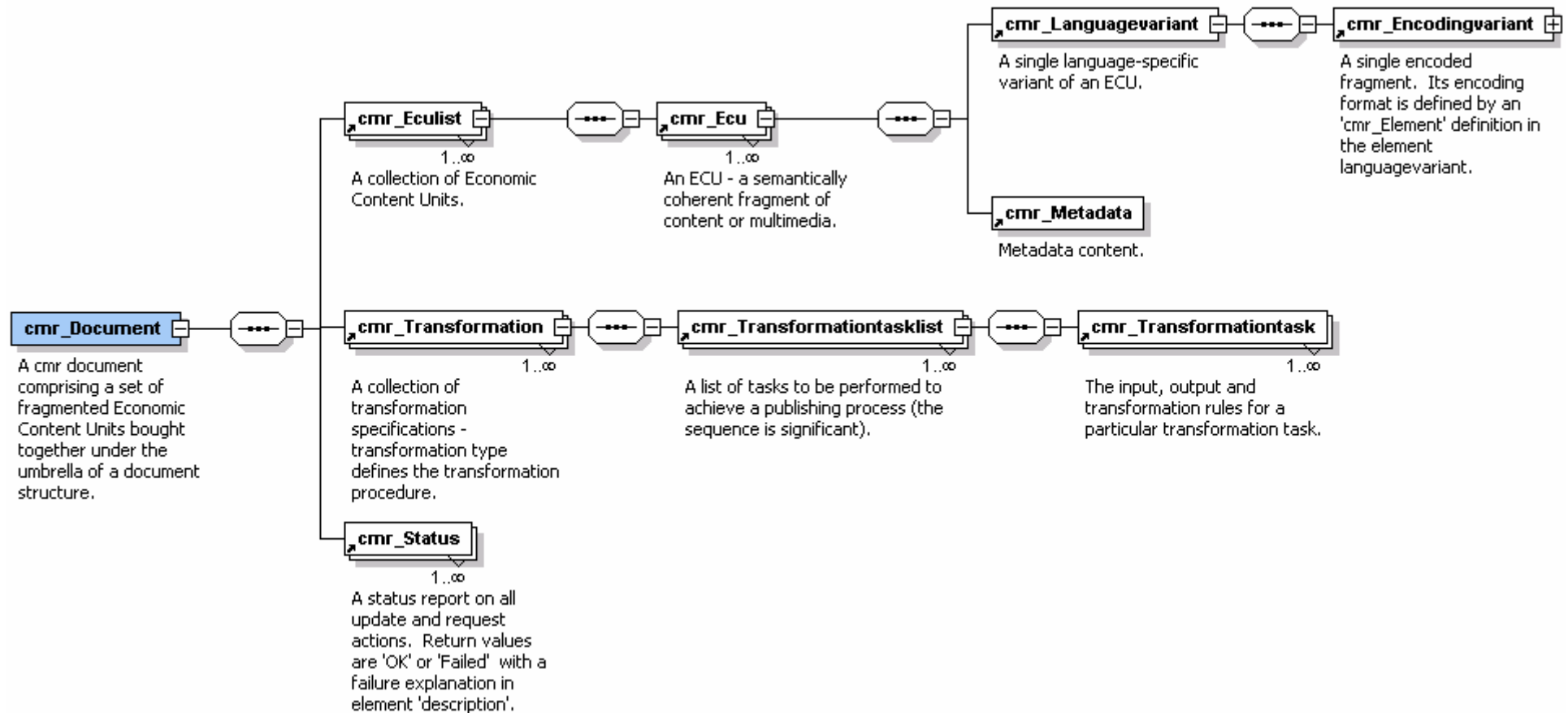
Research Completed 2004

- To formulate a theoretical systems model for the management of digital multimedia resources directed to systematic reuse and regeneration of content.
- To evaluate the effectiveness of the model through a non-trivial prototype and evaluation against current industry practice.

The Content Model for Reuse



A supporting Document Ontology

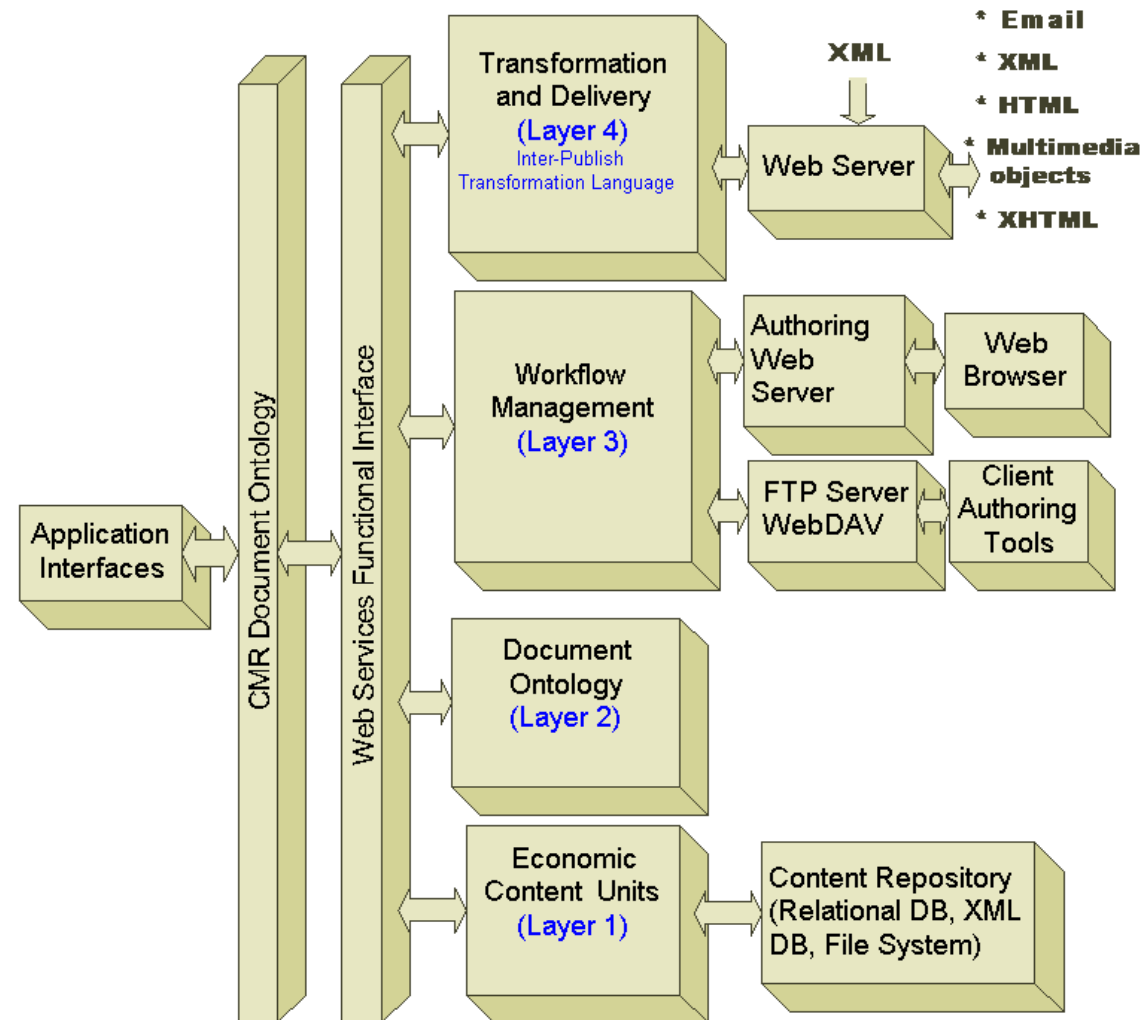


Application design

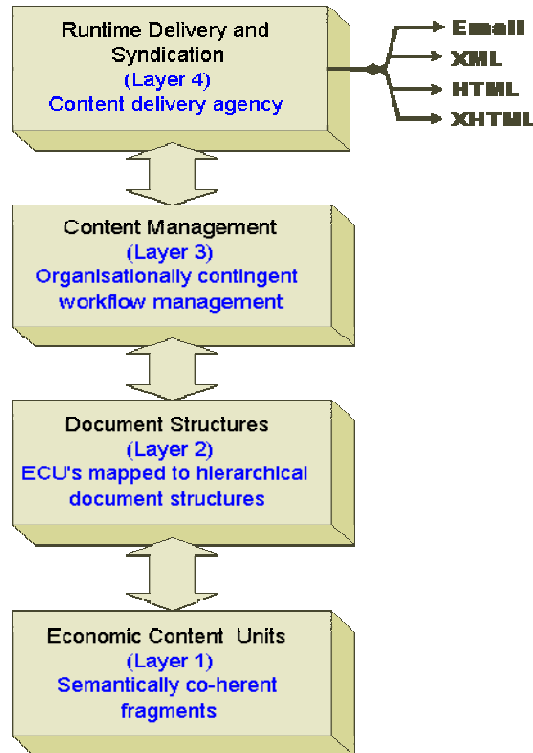
- Linear dissection of ECU content tied together by the Document Ontology and stored in an SQL database
- Acyclic Document Ontology
- Workflow supported through
 - A custom-designed web server providing WebDAV extensions
 - A custom-designed ftp server giving a “virtual” view of content units
 - Web Services at several levels (trusted RPC, un-trusted Web Services with authentication)
- Publishing supported through
 - Inter-Publish Workflow Language
 - Custom-build indexing engine supporting transient and static search lists



Architecture



Field Trial vs. Lab Test



- To test the implementability of the model
- To test the effectiveness of the model in Content Regeneration and Reuse
- ECU in a Web Services document exchange framework
- Strong external validity but weak internal validity



Metrics for Content Reuse

Information Turnover. “Freshness of content”

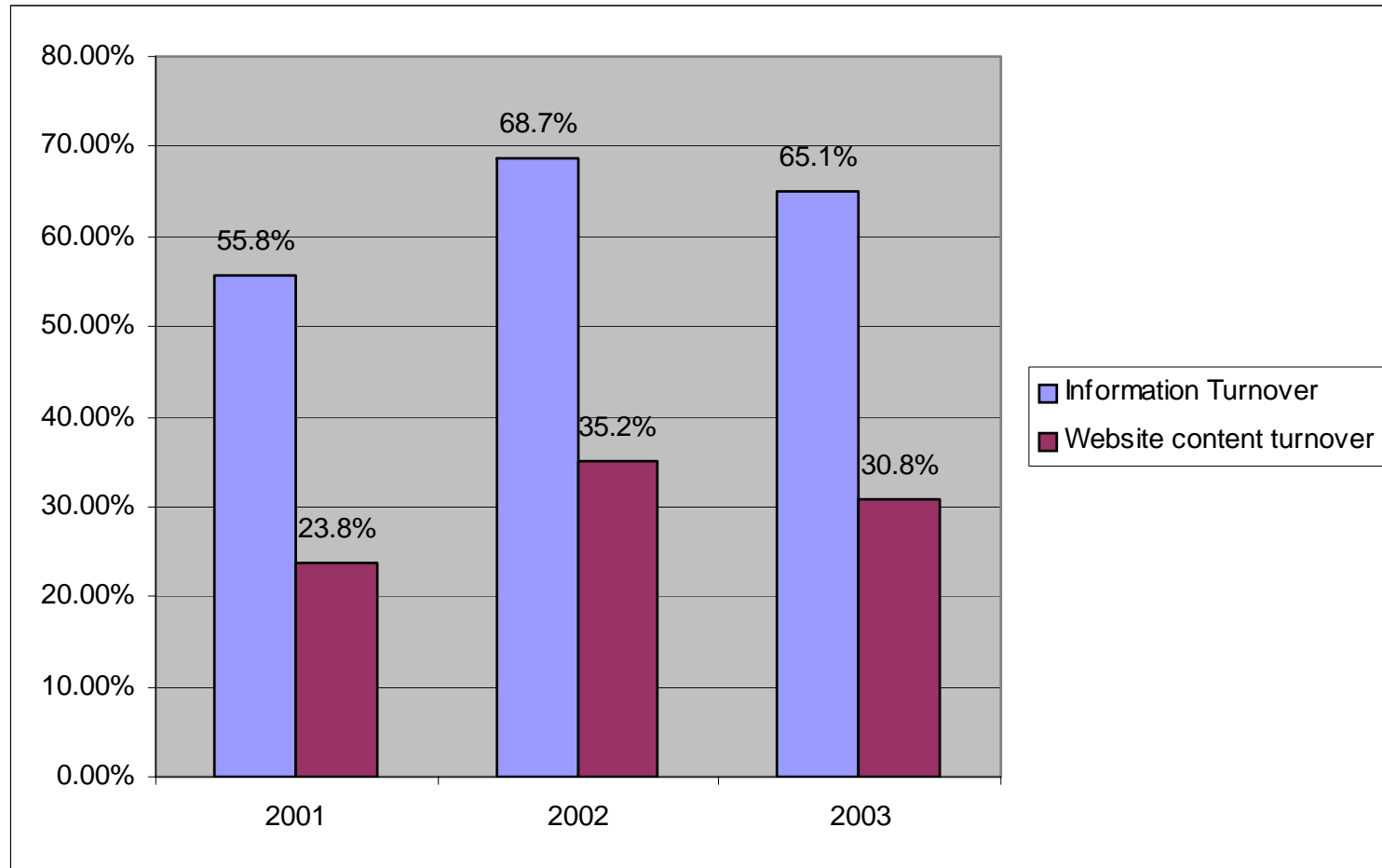
ECU elements / Total Original Pages

Website Turnover. “Visibility of content”

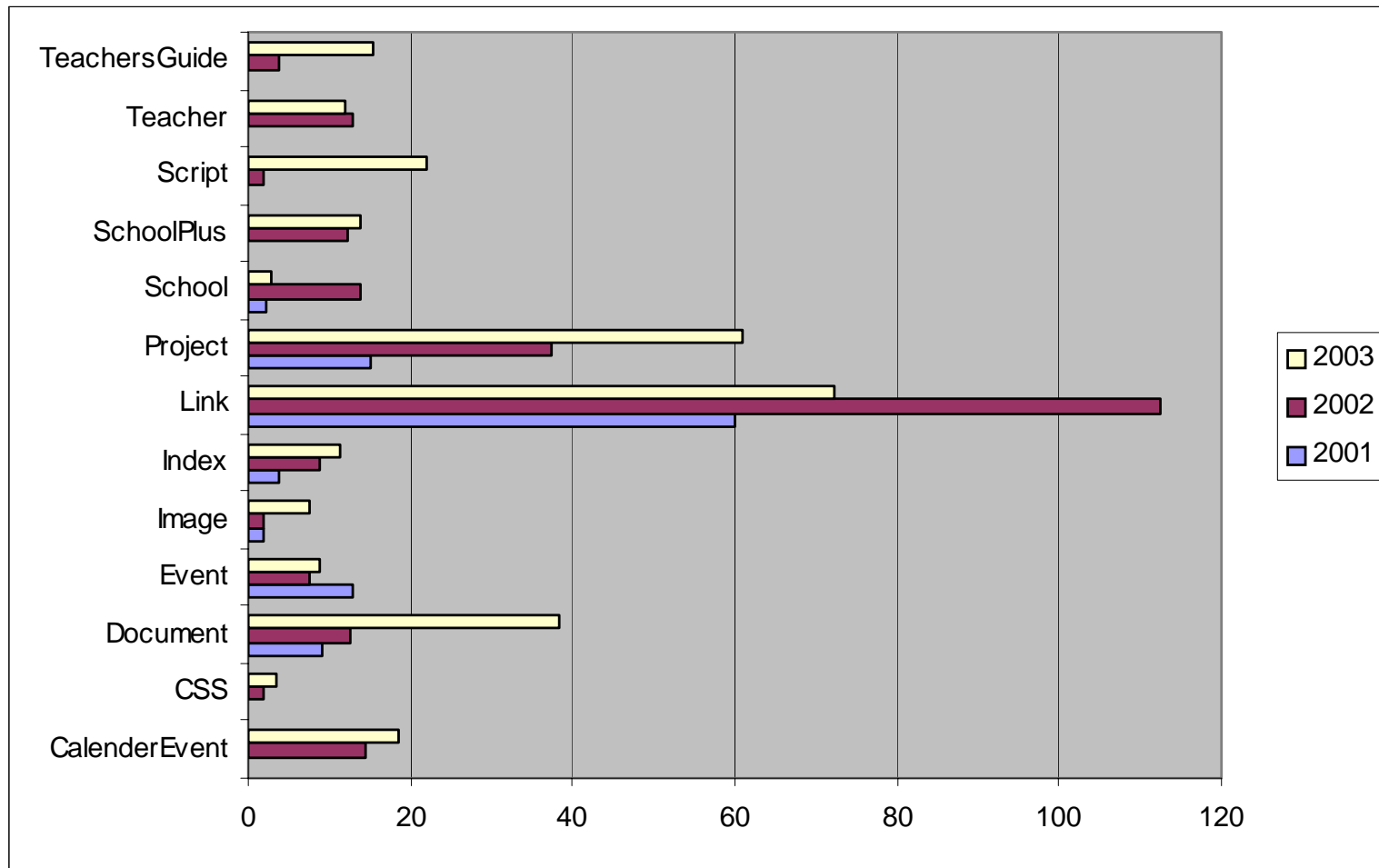
Pages changed / Total Website Pages



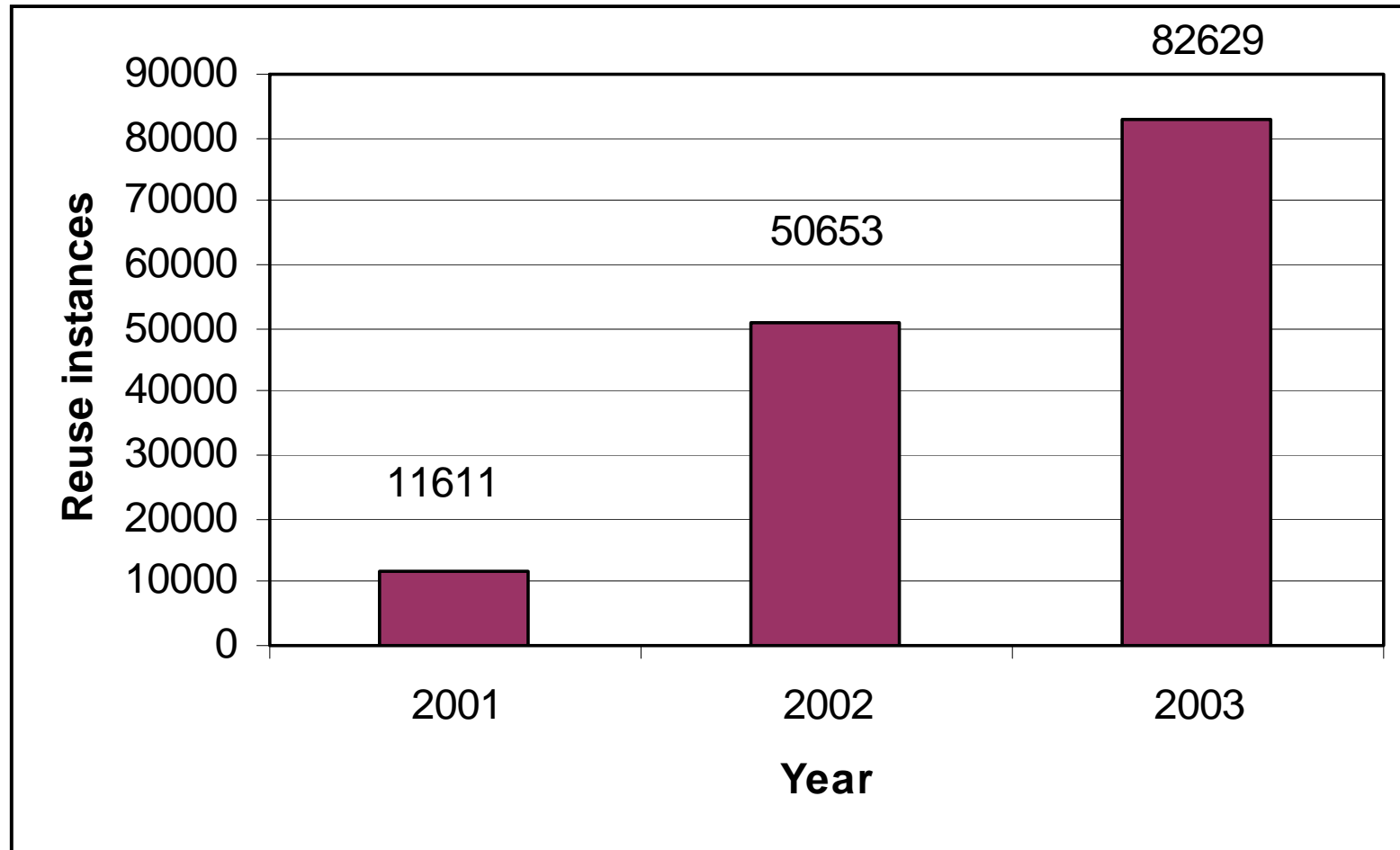
Turnover Metrics



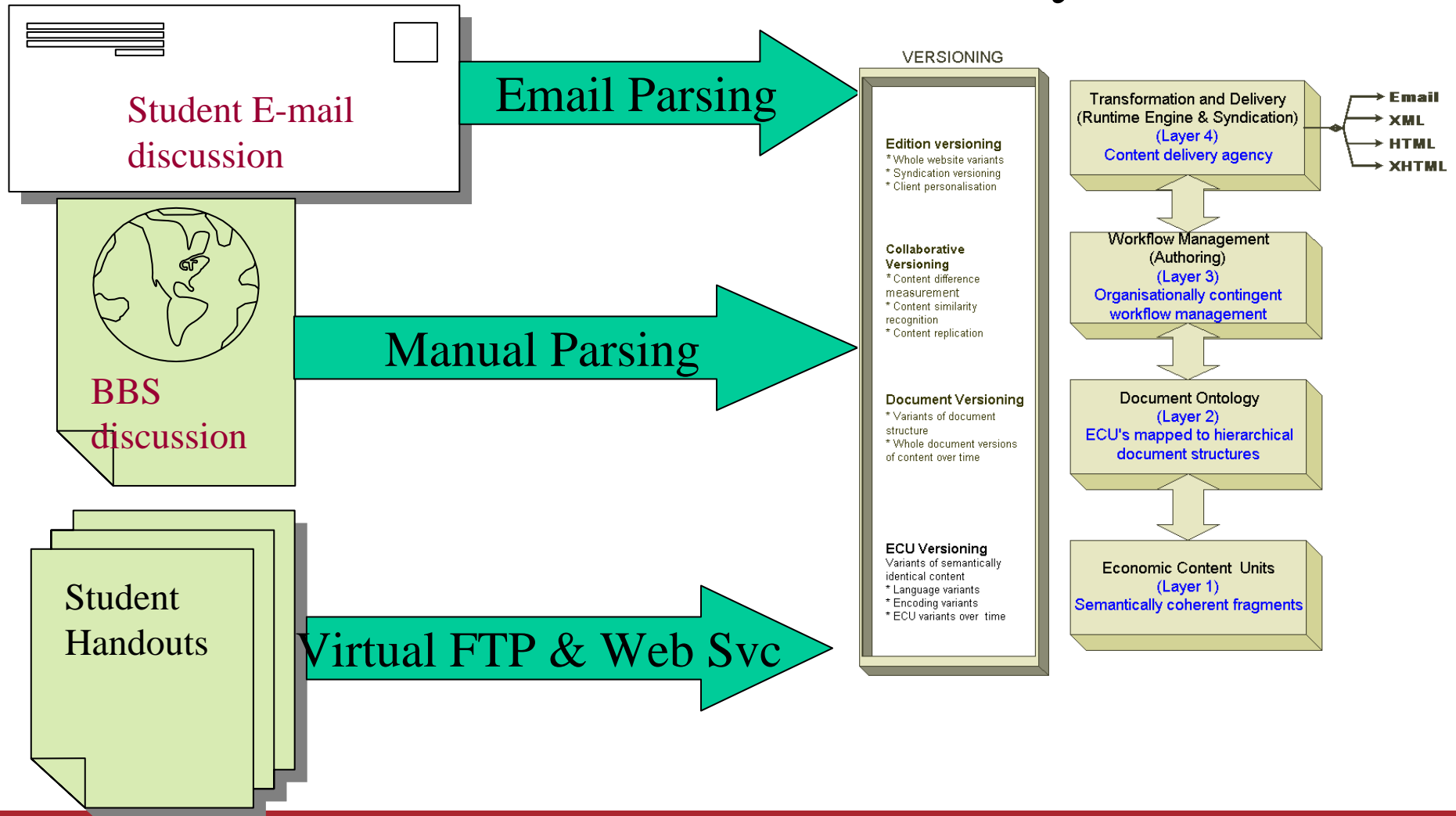
Measurement of Content Reuse



Measurement of Content reuse



Other applications: Reuse in course delivery



Digital Resource Management in Community Based Networks



GratisNet: <http://www.gratisnet.org.au> (Gratisnet is a Health-focussed network of 350 libraries)



GLASS - <http://www.glassdd.org> (a network supporting the GLINN and LOTS consortia of NSW government and semi-governmental libraries)



QShare: <http://www.qshare.org.au> (a network of 15 Queensland government libraries)



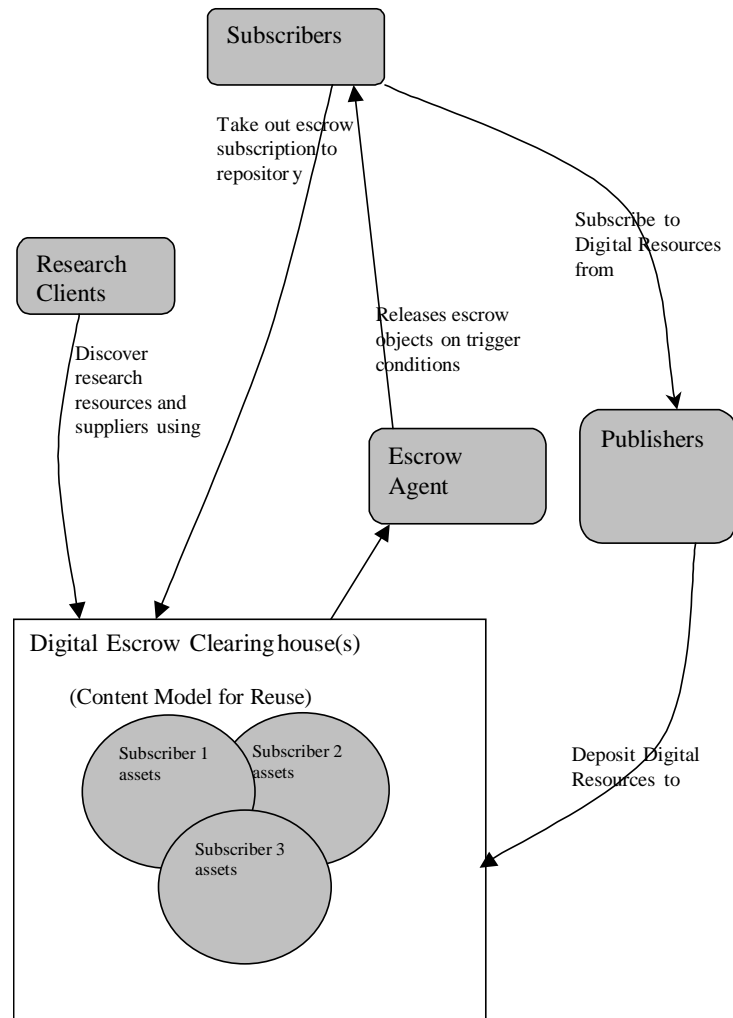
Australasian Libraries in the Emergency Sector: <http://www.alies.org.au> a network of 28 libraries in Emergency Service.



The University of Sydney
Australia



Digital Resources Clearinghouse



Future Research

- Integrating Data Mining with Authoring to improve metadata gathering
- Identifying early key content reuse items
- Evaluating the model in other Industry contexts
- Exploring XML database systems
- The role of the “reuse champion

