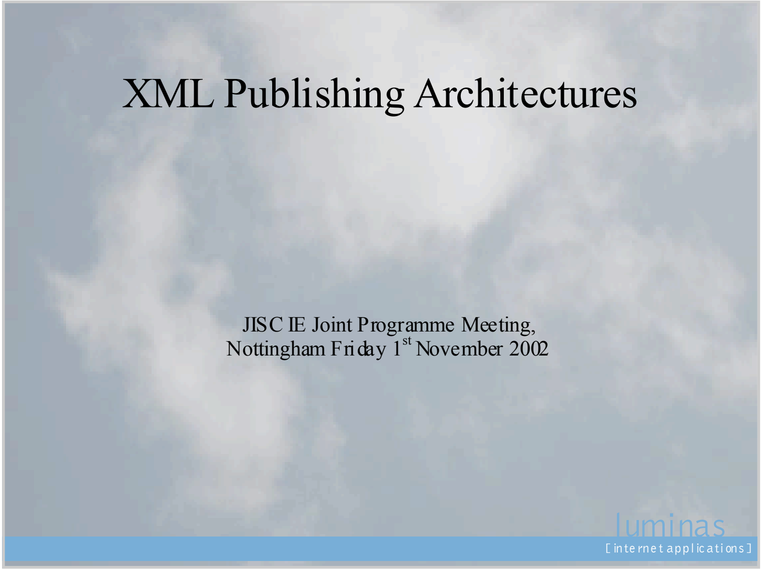
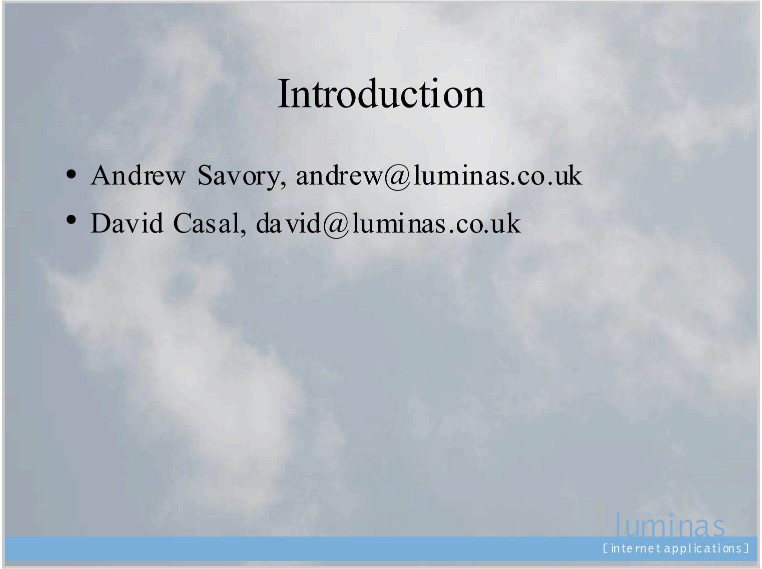


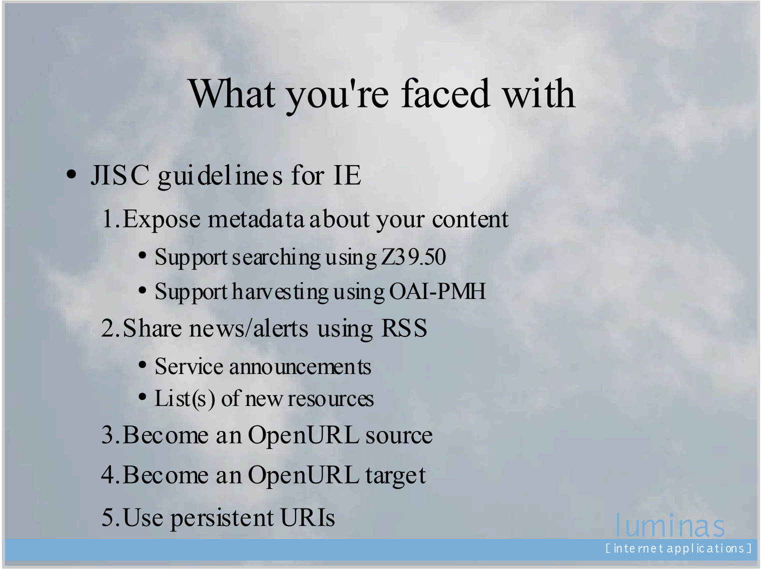
XML Publishing Architectures

JISC IE Joint Programme Meeting,
Nottingham Friday 1st November 2002



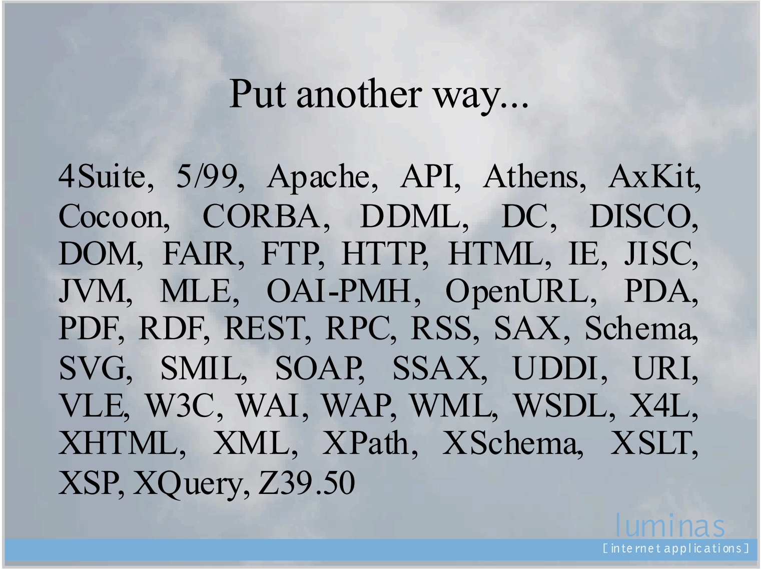
Introduction

- Andrew Savory, andrew@luminas.co.uk
- David Casal, david@luminas.co.uk



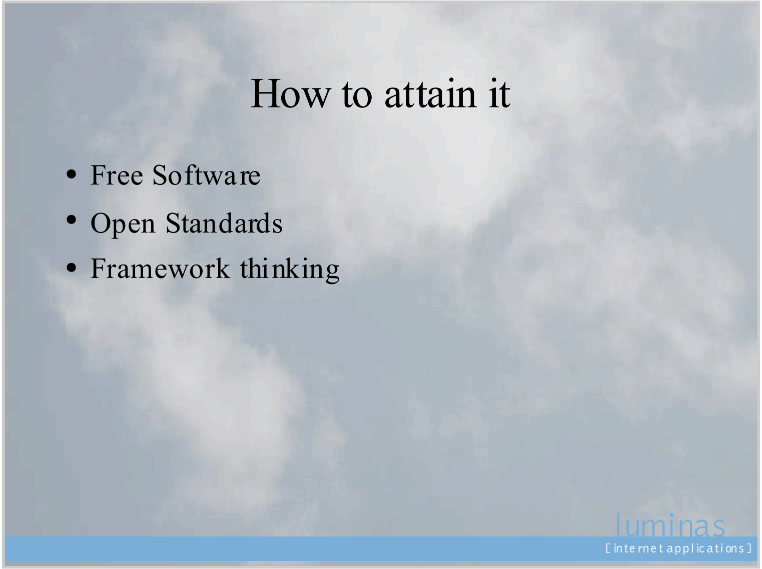
What you're faced with

- JISC guidelines for IE
  1. Expose metadata about your content
     - Support searching using Z39.50
     - Support harvesting using OAI-PMH
  2. Share news/alerts using RSS
     - Service announcements
     - List(s) of new resources
  3. Become an OpenURL source
  4. Become an OpenURL target
  5. Use persistent URIs



Put another way...

4Suite, 5/99, Apache, API, Athens, AxKit, Cocoon, CORBA, DDML, DC, DISCO, DOM, FAIR, FTP, HTTP, HTML, IE, JISC, JVM, MLE, OAI-PMH, OpenURL, PDA, PDF, RDF, REST, RPC, RSS, SAX, Schema, SVG, SMIL, SOAP, SSAX, UDDI, URI, VLE, W3C, WAI, WAP, WML, WSDL, X4L, XHTML, XML, XPath, XSchema, XSLT, XSP, XQuery, Z39.50

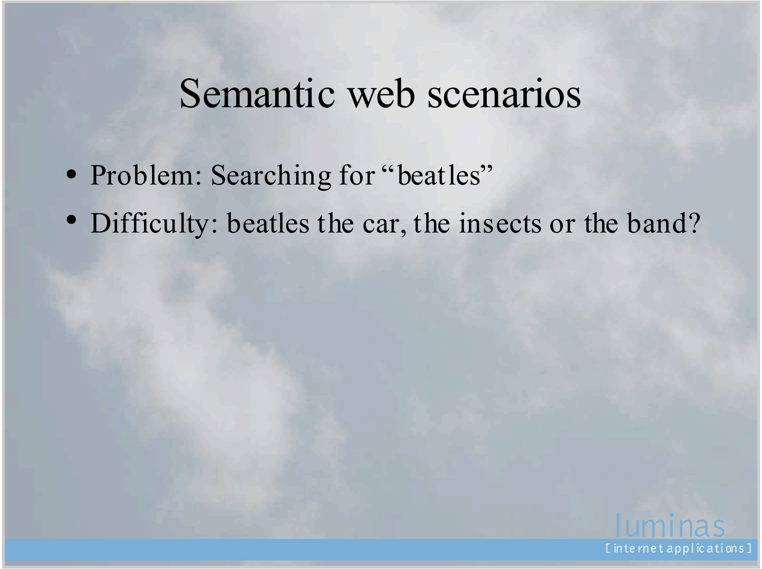


How to attain it

- Free Software
- Open Standards
- Framework thinking

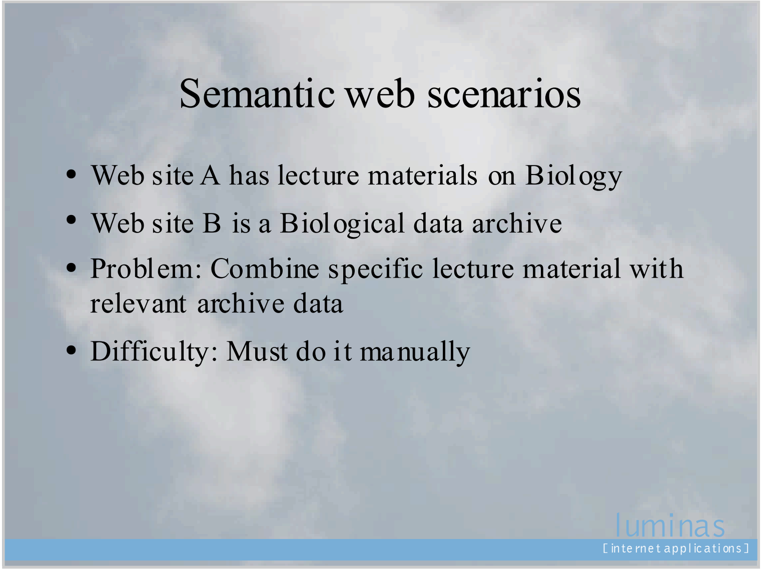
Semantic Approach

(Semantawha?)



Semantic web scenarios

- Problem: Searching for “beatles”
- Difficulty: beatles the car, the insects or the band?



Semantic web scenarios

- Web site A has lecture materials on Biology
- Web site B is a Biological data archive
- Problem: Combine specific lecture material with relevant archive data
- Difficulty: Must do it manually

The Semantic Web

- What is the Semantic Web?
  - Evolution of the current Web, in which meaning is machine-processable
- Purpose of the Semantic Web
  - To allow us (and machines) to find, share, combine and re-purpose information more easily.

Semantic Approach

- Metadata (RDF, Schemas, Dublin Core)
- Repurposable content
- Exposing content across domains (OAI-PMH, Z39.50, RDF)

You were going to
tell me how to attain it!

What?

XML

(XML is meaningless!)

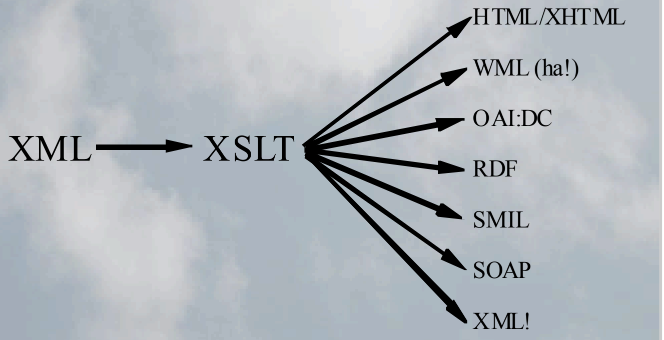
Why XML?

- HTML is dark data
- XML is structured content (visible but meaningless)
- Let others figure out what to do with it!

Reasons for using XML

- Gives meaning to content
- Serve different sites based on user agent
  - Text-based (text readers, slow links, low-tech)
  - Netscape / Internet Explorer specific, TV, PDA
- Serve different formats based on need
  - Text, HTML, PDF, SVG, SMIL, VoiceXML...
- Interoperate with other web applications
  - - Business, semantic web, future-proofing?

Repurpose this!



XML → XSLT

HTML/XHTML

WML (ha!)

OAI:DC

RDF

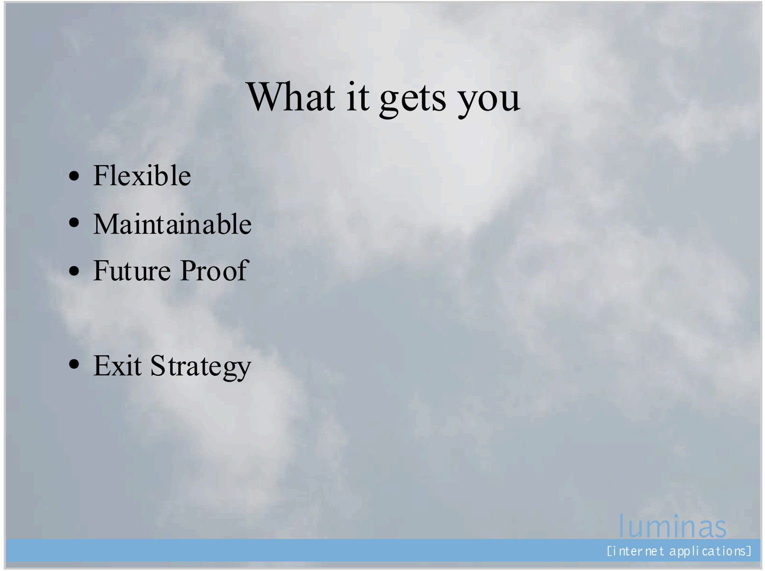
SMIL

SOAP

XML!

So... how?

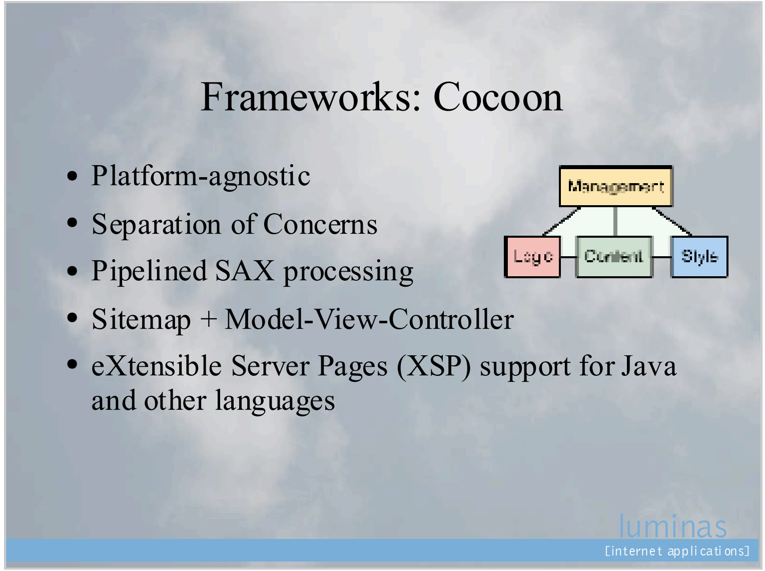
- Apache AxKit
  - C and Perl
- Apache Cocoon
  - Java
- 4Suite
  - Python and C
- Scheme Simple API for XML (SSAX)
  - Scheme



What it gets you

- Flexible
- Maintainable
- Future Proof

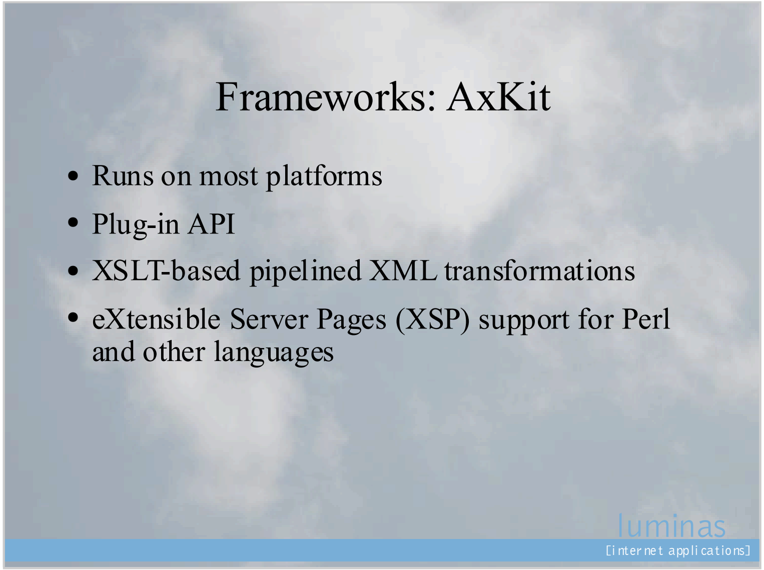
- Exit Strategy



Frameworks: Cocoon

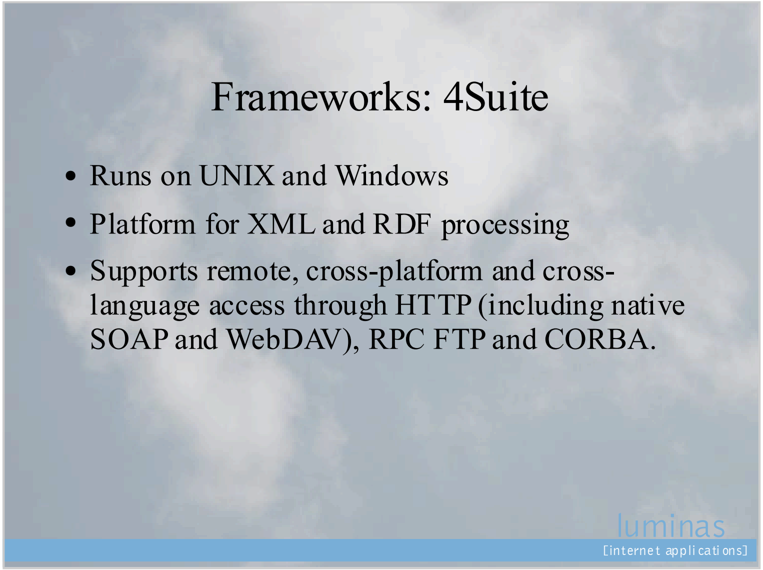
Management
Logic
Content
Style

- Platform-agnostic
- Separation of Concerns
- Pipelined SAX processing
- Sitemap + Model-View-Controller
- eXtensible Server Pages (XSP) support for Java and other languages



Frameworks: AxKit

- Runs on most platforms
- Plug-in API
- XSLT-based pipelined XML transformations
- eXtensible Server Pages (XSP) support for Perl and other languages



Frameworks: 4Suite

- Runs on UNIX and Windows
- Platform for XML and RDF processing
- Supports remote, cross-platform and cross-language access through HTTP (including native SOAP and WebDAV), RPC FTP and CORBA.

Frameworks: SSAX

- Platform-agnostic (including running inside JVM)
- Functional programming for the web
- Language-native tree processing using transformation by example
- Servlets + REST
- Ability to do “robust” parsing if required

Apache Cocoon

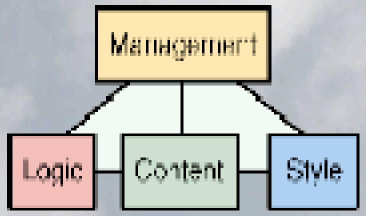
- A publishing and web application framework
- A document generator
- A small revolution

The Cocoon project aims to change the way web information is created, rendered and served.

This new paradigm is based on fact that document content, style and logic are often created by different individuals or working groups.

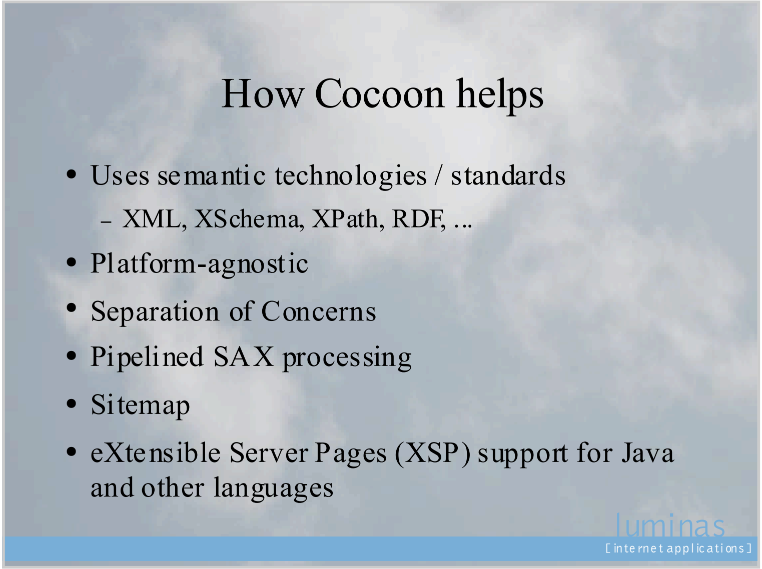
Cocoon aims to a complete separation of the three layers, allowing the three layers to be independently designed, created and managed, reducing management overhead, increasing work reuse and reducing time to market

Cocoon: Separation of Concerns



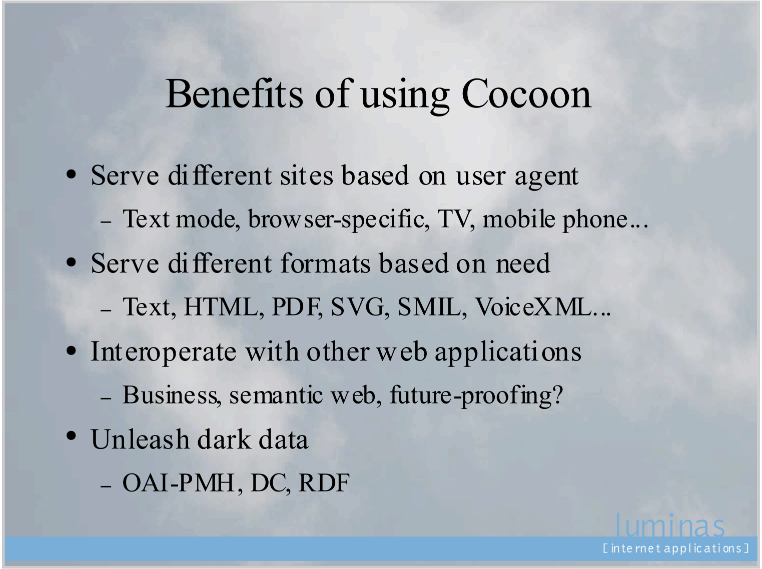
Benefits of SoC

- Management: easier!
- Logic: maintainable and discrete
- Style: designers get freedom from code
- Content: easier to repurpose, available for semantic applications



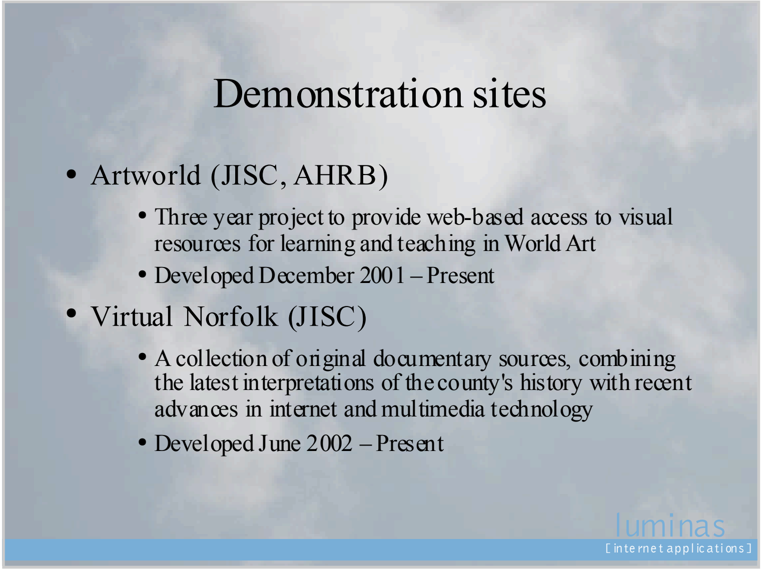
How Cocoon helps

- Uses semantic technologies / standards
  - XML, XSchema, XPath, RDF, ...
- Platform-agnostic
- Separation of Concerns
- Pipelined SAX processing
- Sitemap
- eXtensible Server Pages (XSP) support for Java and other languages



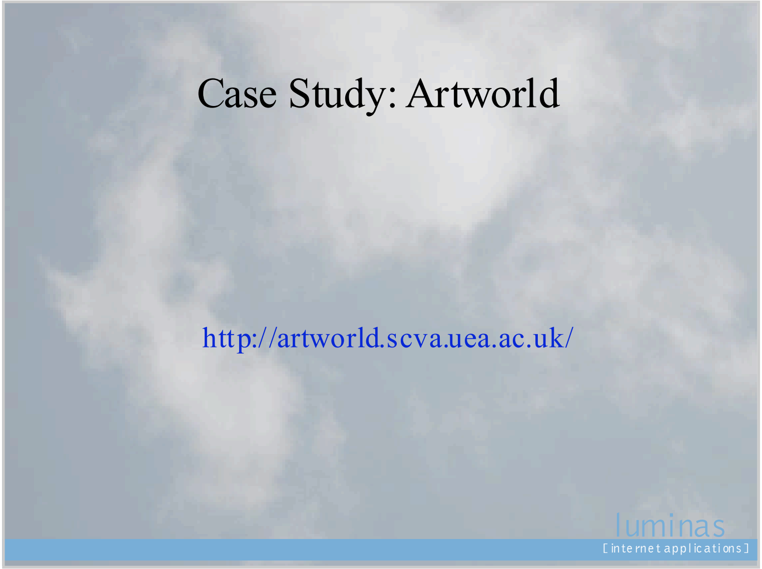
Benefits of using Cocoon

- Serve different sites based on user agent
  - Text mode, browser-specific, TV, mobile phone...
- Serve different formats based on need
  - Text, HTML, PDF, SVG, SMIL, VoiceXML...
- Interoperate with other web applications
  - Business, semantic web, future-proofing?
- Unleash dark data
  - OAI-PMH, DC, RDF



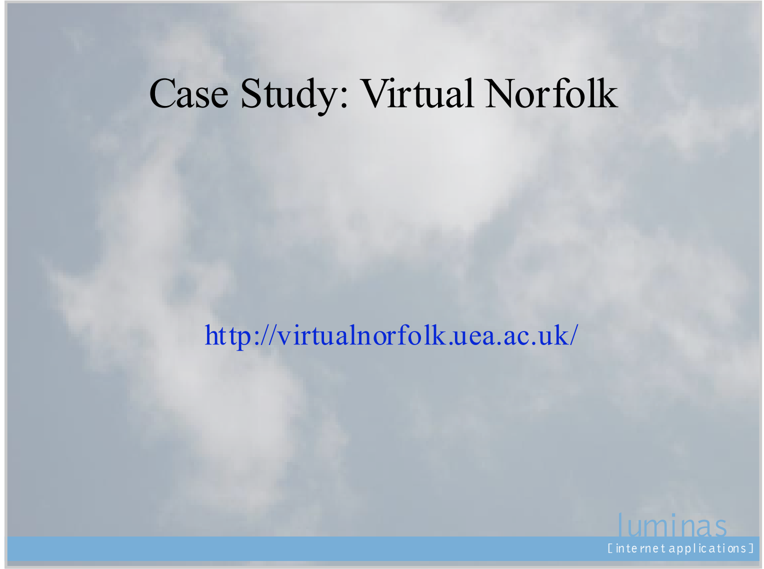
Demonstration sites

- Artworld (JISC, AHRB)
  - Three year project to provide web-based access to visual resources for learning and teaching in World Art
  - Developed December 2001 – Present
- Virtual Norfolk (JISC)
  - A collection of original documentary sources, combining the latest interpretations of the county's history with recent advances in internet and multimedia technology
  - Developed June 2002 – Present



Case Study: Artworld

http://artworld.scva.uea.ac.uk/



Case Study: Virtual Norfolk

http://virtualnorfolk.uea.ac.uk/

Useful Links

- http://www.luminas.co.uk/
- http://xml.apache.org/cocoon/
- http://www.w3c.org