CMS or Digital Object Store?

Managing (storing) Digital Information Objects in Hamburg UP, German Academic Publishers (GAP) and FIGARO

Dr. Stefan Gradmann

Regionales Rechenzentrum der Universität Hamburg

stefan.gradmann@rrz.uni-hamburg.de
Overview

- GAP, Figaro, Hamburg UP
- Objectives and expected results
- Partners and their roles
- Hamburg University Press as part of the GAP publication network
- Main technical issues in GAP/FIGARO
- Storing Objects in Hamburg UP and GAP/FIGARO
Why GAP and FIGARO?

- The critical situation in scholarly publication and communication ("journal crisis", "monograph crisis") forces universities to act in their role as content generators and users.

- The internet is evolving into the primary publication and communication platform in an increasing number of disciplines.

- Digital publication still is heavily modeled on the print-analogy: the innovative potential of electronic platforms is almost not used at all.

- Individual university presses are too weak (economically and technically speaking) to change these basic contextual parameters.

"German Academic Publishers" (GAP, funded by DFG, kicked off 01.12.2001) and FIGARO (funded by EC, kicked off 01.05.2002) to create a technical and organizational co-operation model for academic e-publishers.
Objectives

Overall: **stimulate and support scientific communication** and **return science to scientists** by

Building an **open, Germany/Europe-wide co-operation framework** for **federating academic e-publishing institutions** including

- **Shared/distributed technical facilities**, e. g.
  - Shared WWW-based workflow
  - Supporting tools for open, standard based object modeling
  - **Generic authentication layer** pluggable in SSO architectures

- **Common organisational/exploitation components**, e. g.
  - Business model
  - Legal framework

**Make this framework sustainable**

Investigate **new models of article publishing** (‘post-journals’) and **quality assurance** (‘public peer reviewing’)
The Federation Model

- Workflow
- Document modelling
- Authoring support
- Portal functions
Partners and roles in GAP

Regionales Rechenzentrum der Universität Hamburg
- Organization and marketing
- Project steering and managing

Staats- und Universitätsbibliothek Hamburg
- Content Provision

Bibliotheks- und Informationszentrum der Universität Oldenburg
- Development and implementation of workflow engine
- Authoring tools

Universitätsbibliothek Karlsruhe
- Presentation and portal functions
The FIGARO Consortium

- **Full Partners (Development and usability evaluation)**
  - Utrecht University (Consortium Leader) and Delft University (NL)
  - Hamburg University (Technical Coordination) and Oldenburg University (D)
  - Daidalos bv IT in Publishing (NL)
  - Firenze University (I)

- **Associate Partners (Content Provision)**
  - Academic content providers: Stichting Delft Cluster (NL), Leuven University (B), Lund University (S)
  - SME publishers: Uitgeverij LEMMA B.V. (NL) and Wydawnictwo DiG sc. (PL)
  - Association of Research Libraries/SPARC (US)

- **Subcontractor (XML based document modelling)**
  - SUN Microsystems/StarOffice (D)
Hamburg UP as part of the GAP-Network

- Hamburg University Press (Hamburg UP)
  - Founded 01.01.2002
  - Central service unit of Hamburg University located at the university’s computing center (RRZ)
  - Minimally staffed (1 director + 1 FTE)
  - Mission: E-Publish high quality content generated within Hamburg University including an option for printing on demand

- Systematically using GAP/FIGARO modules wherever possible within Hamburg UP serves 3 goals
  - Concentrate on content and authoring support
  - Make the UP effective with only minimal personal resources
  - Evaluate usability of GAP/FIGARO components
Standard Based Innovation

- Achieve functional innovation via integration and adaption of standard based (and wherever possible open source) building blocks and do not start own developments we cannot sustain

- Examples of such standards:
  - Dublin Core (DC) – plus extensions
  - OAI-protocol – but maybe not in a harvesting based pull-approach but using push-methods
  - Open, generic document models expressed in XML (Schema) and derived from operational modeling proposals such as DocBook and OO-XML
  - Open, generic authentication methods based on LDAP
  - ...

Use standard based, open models for digital information objects in authoring support and to support new and innovative publication objects

, electrified publishing

, true e-publishing

DOC
DVI
conversion
PDF

DOC
DVI
XML-Schema
XSLT
PDF
HTML
SHTML
???
Authentication & Authorization

- **WHO** - e.g. authors, customers, editors, reviewers, annotators ...

- May apply **WHAT** kind of operation - e.g. read, write (think of collaborative authoring!), annotate, stabilize (“freeze”), apply different status-levels such as ‘rejected’ ‘ready for public reviewing’, copy/attempt pirating

- On **WHICH** object (or which **specific part** of such an object) - e.g. overall document ID but also micro-structures to be referenced as part of compound MM-documents as well as of uniform complex objects ('books' and the like)

- In which **CONTEXT** - e.g. “scientific use” (teaching/studying) vs. commercial use, pre-publishing, public reviewing, publishing etc.

- In other words: identify **Actors, Entities, Operations, Context** and organize these in a **4-dimensional matrix** in a **secure, reliable** way using available building blocks and **standards** wherever possible
Workflow: functional building blocks

Back Office Processing

- Document Modeling and input processing (Doc/dvi to XML)
- Workflow components
- XML based Document Management and output (XML to pdf / html)

Pre-Publishing

- Reader(s)
- Author(s)
- Editor(s)
- Peer Reviewer(s)

Authentication Layer

- FO
- Peer-reviewed publication
- Public/open Peer-reviewing

Presentation/Portal Functions

- HTML/PDF
- Annotation and evaluation Functions
- Output

User(s)
Storing Objects in Hamburg

UP: Life is (relatively) simple!

Hamburg UP

Hamburg UP Portal Funcs

MILESS/MyCoRe

IBM-CM/EIP

DB2 (Objects)
Storing Objects in GAP: Life might still be kept simple ...
Storing Objects in FIGARO: Life gets complicated!

- Hamburg UP FO
  - MILESS/MyCoRe
  - IBM-CM/EIP
  - DB2

- Some Dutch FO
  - Portal
  - Oracle 9i
  - Oracle

- Some Polish FO
  - Portal
  - Zope
  - MySQL

- Some very small FO
  - Portal

- Function Layer
  - CMS (which one?)
  - Data Store
Lessons learned up to now

- We must remain extremely restrictive regarding our assumptions about what happens inside the components of the distributed and heterogeneous Figaro object store: the orange pointers are the essential glue of the overall architecture, and the structure of these pointers is an essential cornerstone of our projects.

- Full grown CMSs are degraded to simple digital object stores in such an approach.

- Details regarding pointers
  - URL will not do the job (mind the orange links, mind persistency aspects ... !)
  - XLink & cie. are yet to be observed, not yet a sure bet.
  - We may well go for URN – but then have to determine a syntax, find resolving partners etc.

- MyCoRe is well prepared for such complex scenarios: we are convinced to have made a valid choice in Hamburg, even if not all of our partners will make the same choice – or maybe just because we do not have to impose choices upon our partners this way.
C'est fini!

- Thanks for your patience and attention
- Questions?

stefan.gradmann@rrz.uni-hamburg.de