

TEI and Scholarly publishing — experience from the PEER project

Laurent Romary
INRIA & HUB-ISDL

Overview

- The Text Encoding Initiative
 - Customization
- Scholarly publishing in the TEI world
 - Hot topic
- Some results from the PEER project
 - The publishing mess
- Further action
 - We're here to devise this...

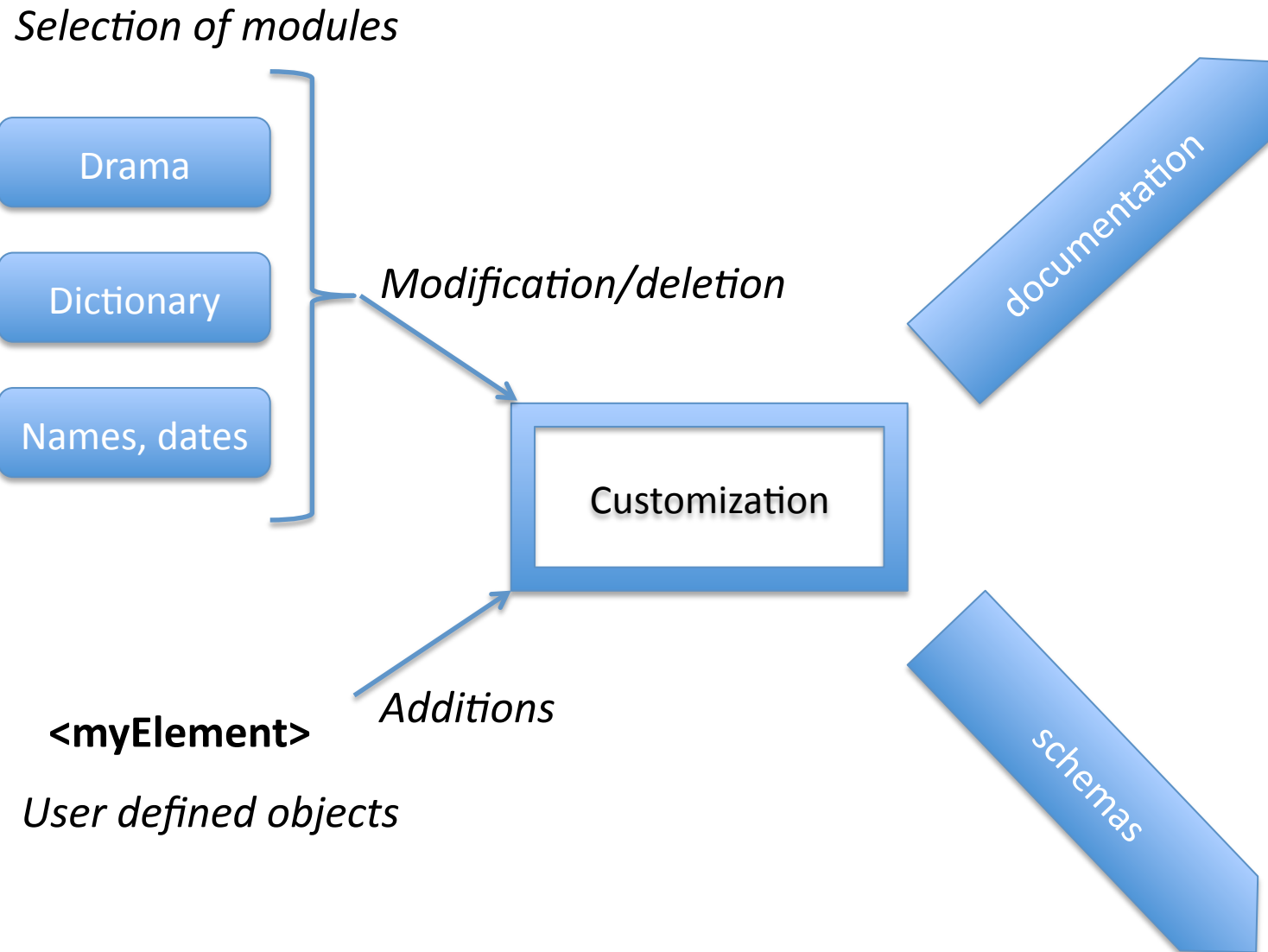
The Text Encoding Initiative

- Initiated in 1987 by major international text centers
 - Adoption of SGML than XML
 - Important contributions to the development of XML
- Organized as a membership consortium since 2000
 - 5 hosts (Virginia, Brown, Oxford, Nancy, Leithbridge)
 - Board (management) and council (technical content)
- Five editions of TEI guidelines (current P5)
 - Large community of users, continuous maintenance of content, evolution towards additional domains (e.g. manuscript transcription)

Main technical features of the TEI

- Modularity
 - Core modules: header text descriptions; bibliography
 - Thematic modules: drama; dictionaries; manuscript description
 - Additional components: time, names and dates; annotations;
- Customizability
 - ODD (one document does it all): specification language of the TEI

TEI customization



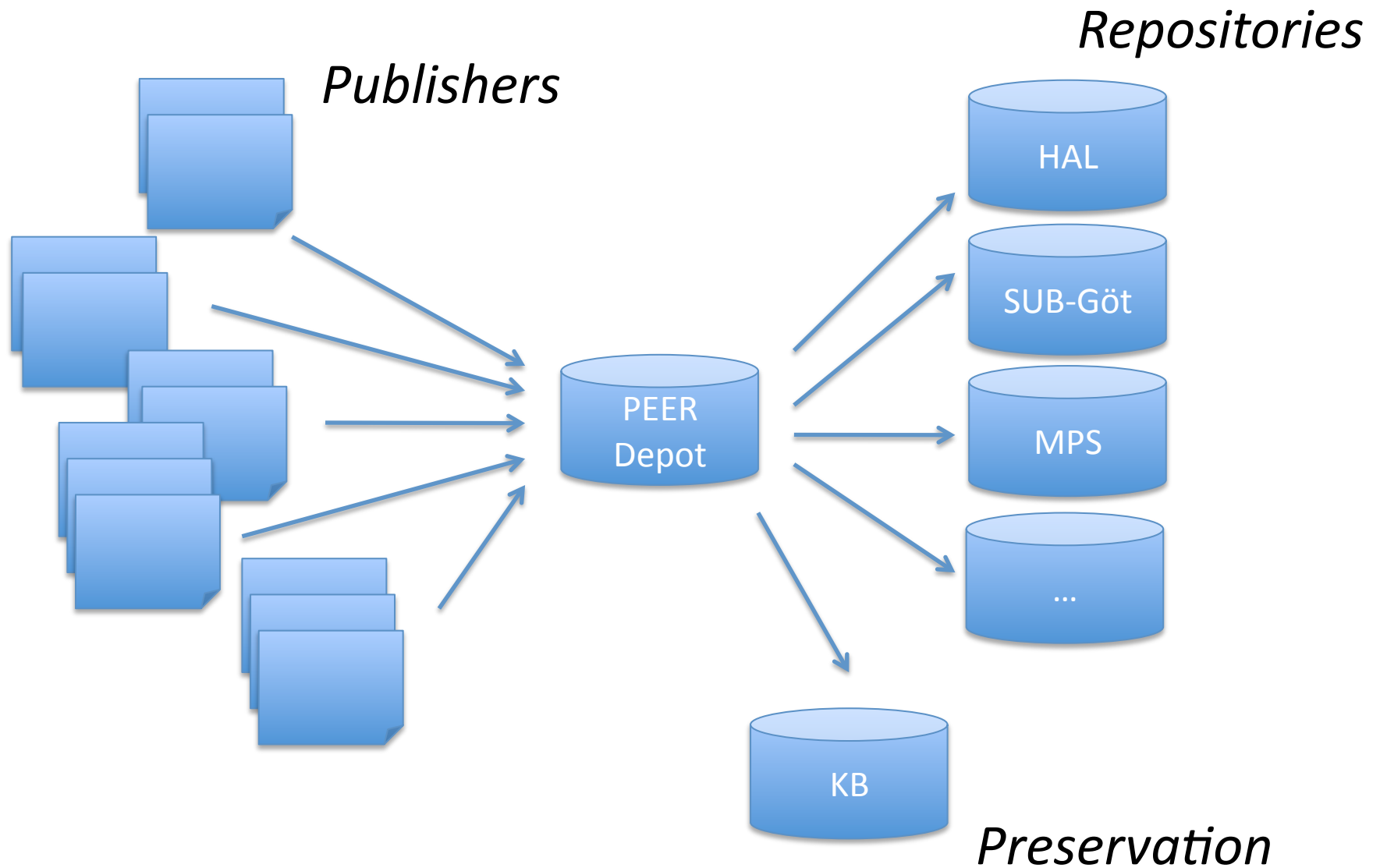
Scholarly publishing in the TEI world

- Several existing publishing initiatives
 - TEIJournal, Revues.org, DHQ (Digital Humanities Quaterly), Digital Humanists, PUC, OAPEN...
- A strategic issue for the TEI
 - Common framework for Primary sources and publications
 - Publication as primary sources in the humanities
 - Continuation of the TEI endeavor to provide reference customizations (cf. TEI Tite for digitization projects)
 - Creation of a TEI publishing SIG
- Strong demand from research libraries and academic publishers (e.g. AAUP)
 - Maintenance and customization issues related to NLM

The PEER project

- Initiated by the EU commission (DG INFSO)
- Objective: study the impact of systematically archiving stage-two outputs
 - on journals and business models
 - on wider ecology of scientific research
- Consortium
 - STM, European Science Foundation (ESF),
Goettingen State and University Library (UGOE),
Max Planck Gesellschaft (MPG), INRIA

The PEER deposit workflow



Mapping formats

- Great heterogeneity of format within publishers
 - Meta data (and full-text)
 - Proprietary, ScholarOne, NLM 2.0, NLM 3.0
- Various issues
 - Affiliations
 - Publication date information
 - ISO 639 codes (countries)
 - Bibliographical references
 - Proprietary metadata fields

Examples

- Article title
 - article-title/title | ArticleTitle | article-title | ce:title | art_title | article_title | nihms-submit/title | ArticleTitle/Title | ChapterTitle
- Journal title
 - j-title | JournalTitle | full_journal_title | jrn_title | journal-title
- ISSN (print)
 - JournalPrintISSN | issn[@issn_type='print'] | issn[@pub-type='ppub'] | PrintISSN | issn-paper
- First page of a paper
 - spn | FirstPage | ArticleFirstPage | fpage | first-page

A TEI customization for scholarly publishing

- A family of formats based on the TEI customization facilities
 - Core editing customization (to be further extended – minimal tool support)
 - Reference customization family for archiving
 - Can be extended to specific domains: Maths, physics, SVG graphics, etc.
 - Precise representation of bibliographic information
 - Specific support through associated tool:
 - XSLT stylesheets (html, pdf TEI2NLM)
 - PDF 2 TEI facility (Grobid)
 - Open Office 2 TEI facilities (maintained at Oxford)
 - MSWord 2 TEI facilities (TEI project with ISO)

Coordination with the NLM endeavor

- Context
 - Origin of NLM
 - Quick uptake (BMC)
 - No standardization strategy
 - Divergence (cf. PEER)
- TEI as a possible background for the maintenance of the NLM “DTD”
 - ODD specification, maintenance of a family of schemas
 - Schemas + Documentation available as a TEI customization
 - Roundtrip transformation with TEI

Invitation

- TEI council Symposium
 - Dublin, 28 April 2010 – Academy of sciences
 - Defining a strategy for the TEI and priorities for our workplan
 - Would be great to have input from a Publisher and/or STM