Serving users with content in a web-based environment – a new challenge for information providers

Dr.Timo Borst

Information Systems and Publishing Technologies

ZBW – Leibniz Informationszentrum Wirtschaft

5th International Symposium on
Information Management in a Changing World

November 24-26, 2014, Antalya, Turkey

HACETTEPE UNIVERSITY

IMCW2014

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

ZBW is member of the Leibniz Association

Content provision – a brief history from a library‘s perspective

- For centuries, libraries provided information „in a bookshelf“ (in the sense of a particular, at some ages even hidden location)

- Inter Library Loan (ILL) has extended this model to physically remote places by posting physical items or copies based on explicit agreement

- With the emergence and popularity of the World Wide Web (WWW), the distribution of library‘s content has been tremendously flexibilized in terms of dissemination, access and circulation of items…

…but:

Is this situation so much different from the traditional ‚outside-in library' (Dempsey), while following the paradigm of information pull?

Three examples of externalizing library content…

…so it can be accessed from and integrated into existing user environments…

Example 1: Optimizing content for better indexing through search engines

- Popular search engines (Google, Yahoo, Bing,…) have become the most common user environment for retrieving and acccessing library content – thus not replacing, but further *popularizing* this content

- Without being indexed through search engines, library‘s content would be much less visible

- Thus, any library‘s content should be indexed resp. optimized for indexing

- Some library software already supports this (e.g., OPUS 3.x)

Optimizing content for better indexing through search engines

- Project background: EconStor = DSpace based repository for full texts in Economics, maintained since 2007, 83.000 documents (as of November 2014)
- HTML Jump-off pages from EconStor have been optimized for better indexing through Google Scholar
- OAI-Harvesting does not work, but instructions from GS
- Search Engine Optimization (SEO) has led to a significant increase in being accessed from Google Scholar‘s search results

Optimizing content for better indexing through search engines

Results at a glance:

- Repository items indexed by GS have more than doubled at the time of introducing SEO optimization (although content has not changed)

- As of November 2014, more than 41.000 documents (=**49 %**) are indexed by GS (without citations)

- **35%** of all document downloads from econstor.eu are triggered from GS hits (on fulltext or on jump-off page), whereas **3%** are triggered from searches within econstor.eu

Example 2: Integrating authority data for indexing and retrieval

- *Remember?* Libraries do not only provide books, but also metadata (THEIR content) on items / persons / organizations / concepts which is regarded as ‚normative', meaning a reference for others

- 10 years ago, authority data have already been identified as a good candidate for providing a backbone or ‚linking hub' for distributed content, see B. Tillett (2003): Authority Control: State of the Art and New Perspectives

- Providing this metadata in an user's environment can help to identify persons by linking to information in terms of authority data

Integrating authority data for indexing and retrieval

- Project context „SowiDataNet“
  - Funded by the Leibniz Association
  - Partners: German Institute for Economic Research (DIW), Berlin Social Science Center (WZB), GESIS, ZBW
  - Main goal: to establish an infrastructure for web-based archiving and management of research data from the empirical social sciences (basically, statistics…)

- How to overcome string based repository systems? Approach: Linking names (=strings) to persistent identifiers resolving to the authority record (which may contain links to other researcher identifier systems like VIAF, ORCID,…)

Identifying persons by names…

- Reference dataset: General authority file regularly published by the German National Library
- Filtering is crucial to exclude persons from other domains, therefore narrowing down to persons from Economics and the related domain of Social Sciences
- Subset *econ_pers*: 128.000 persons with 290.000 namings (average 2.26 labels per person)
- Pre-Sorting according to occurrences in EconBiz (subject portal for Economics)
- Numerous name matches on preferred label (15 times "Peter Schmidt" in Economics or Social Sciences)
- Further information on names for the purpose of identifying persons will become necessary (life data, profession, affiliation, etc.)

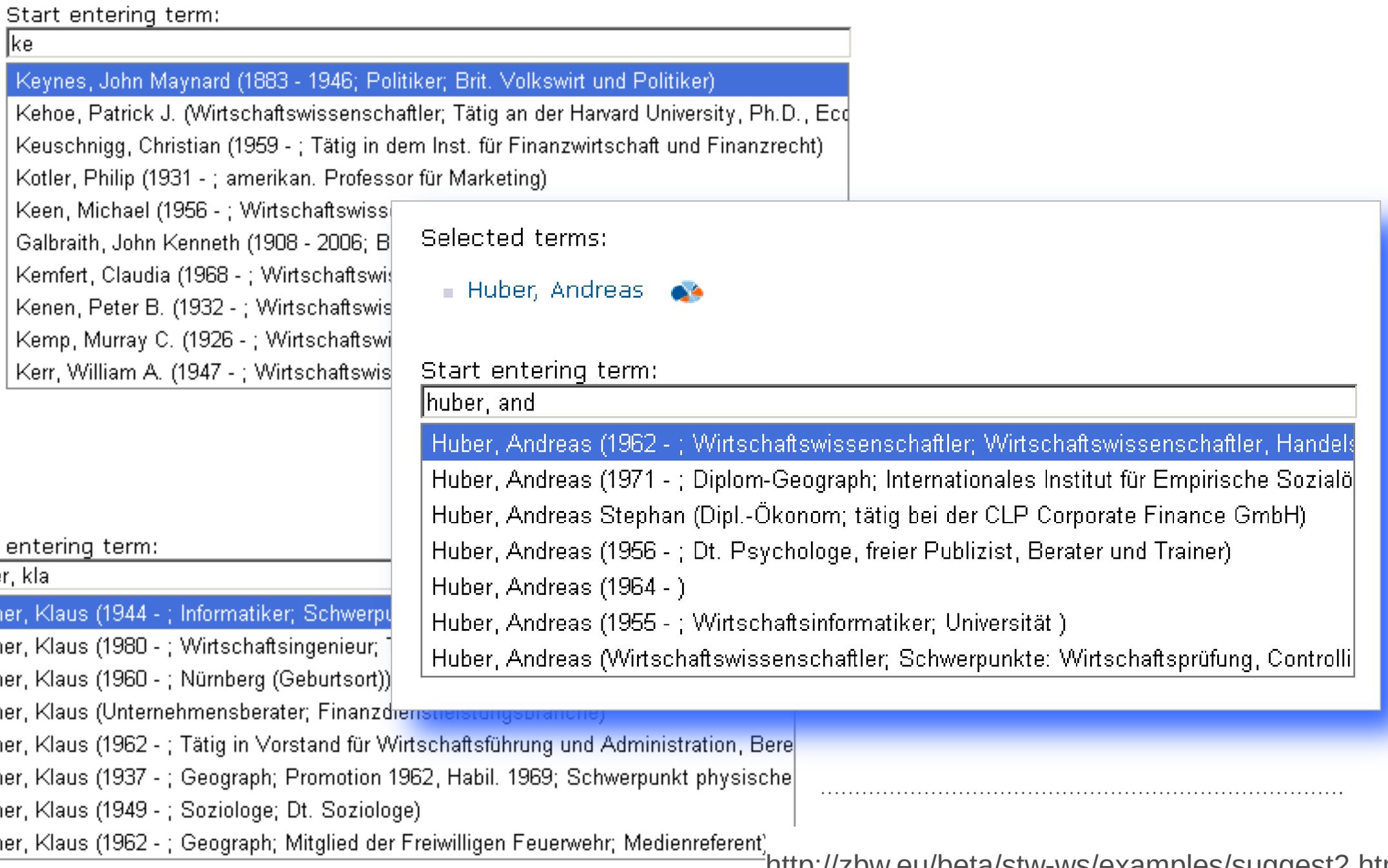
Technical solution for name disambiguation: Autosuggest for Economists and Social Scientists

- ZBW Lookupservice (beta) for Economists and Social Scientists available at http://zbw.eu/beta/econ-ws

- Querying the webservcie delivers the results in machine readable formats (RDF, JSON-LD) for better integration into existing applications

- Javascript code for lightweight integration is available at http://zbw.eu/beta/econ-ws/examples/suggest2.html

Examples of auto-suggest of persons



http://zbw.eu/beta/stw-ws/examples/suggest2.html

Example 3: Recommending literature within an authoring environment

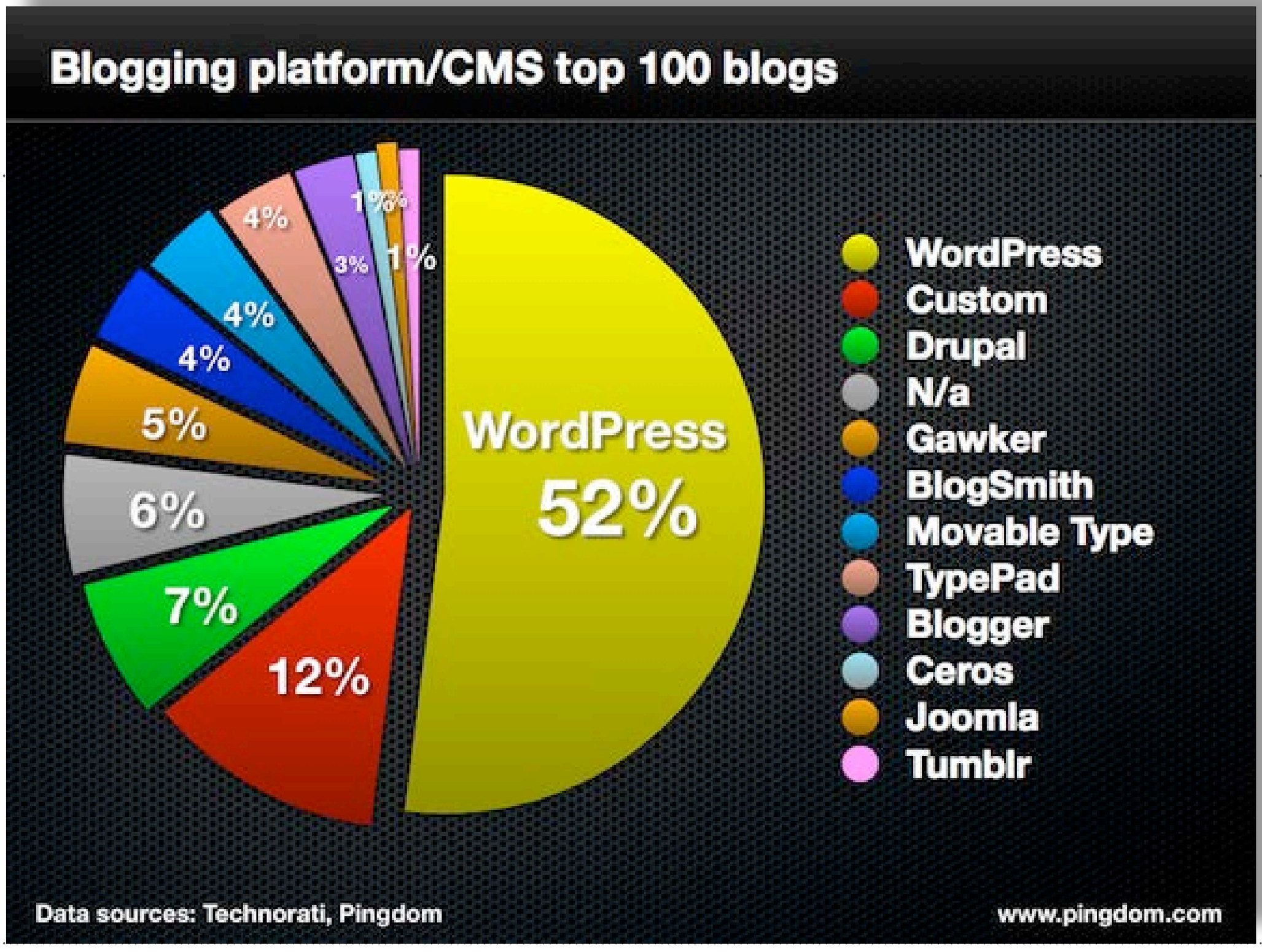
- Content management platforms (e.g. for blogs) have become very popular to generate, publish and deliver content to the web

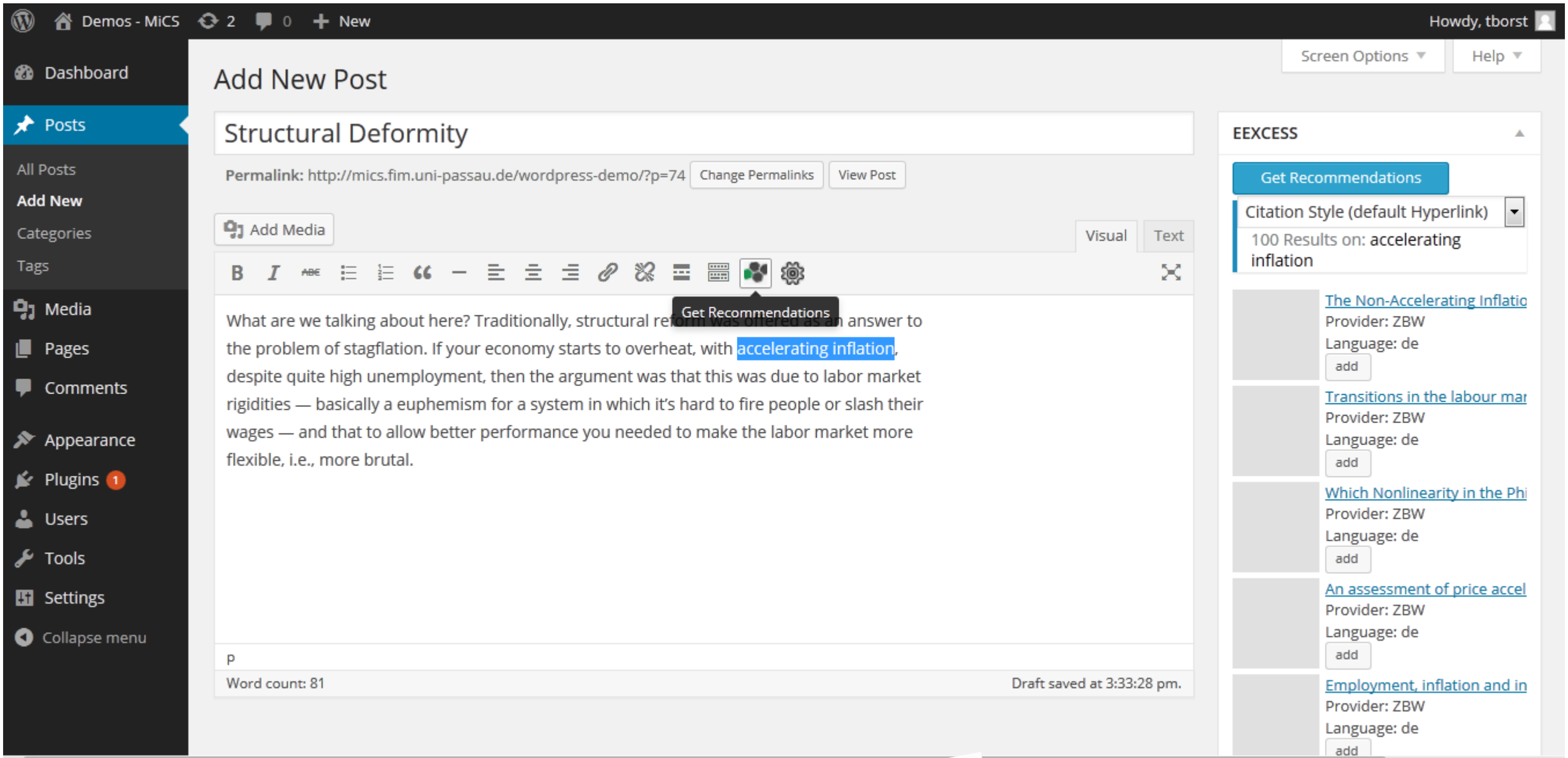
- Citing others is quite easy just by linking to the external content and/or by embedding citation strings for identifying/resolving the resource

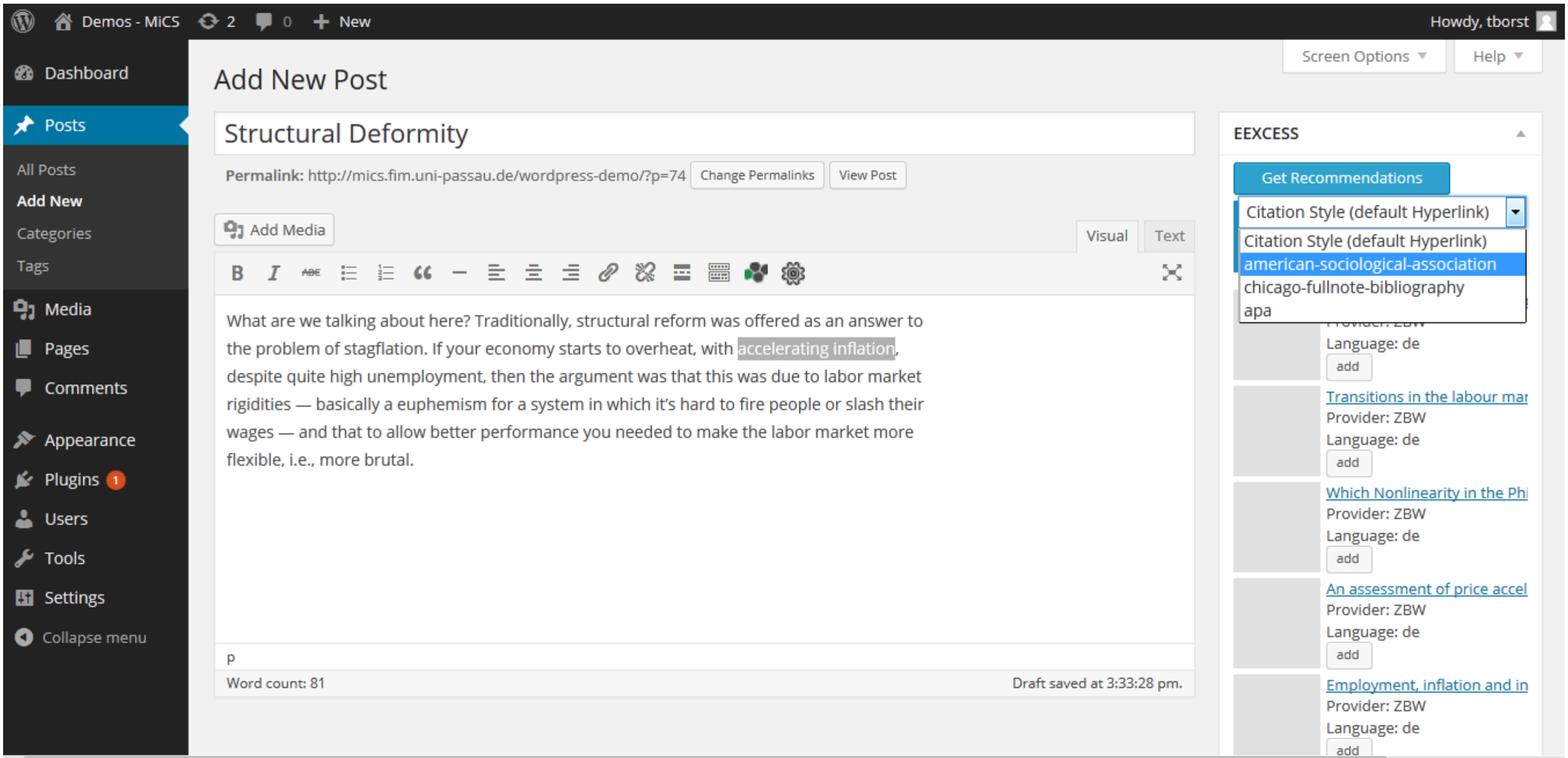
- However, this approach means that the resource is
  a. already known to the author
  b. accessible (also for your readers!)

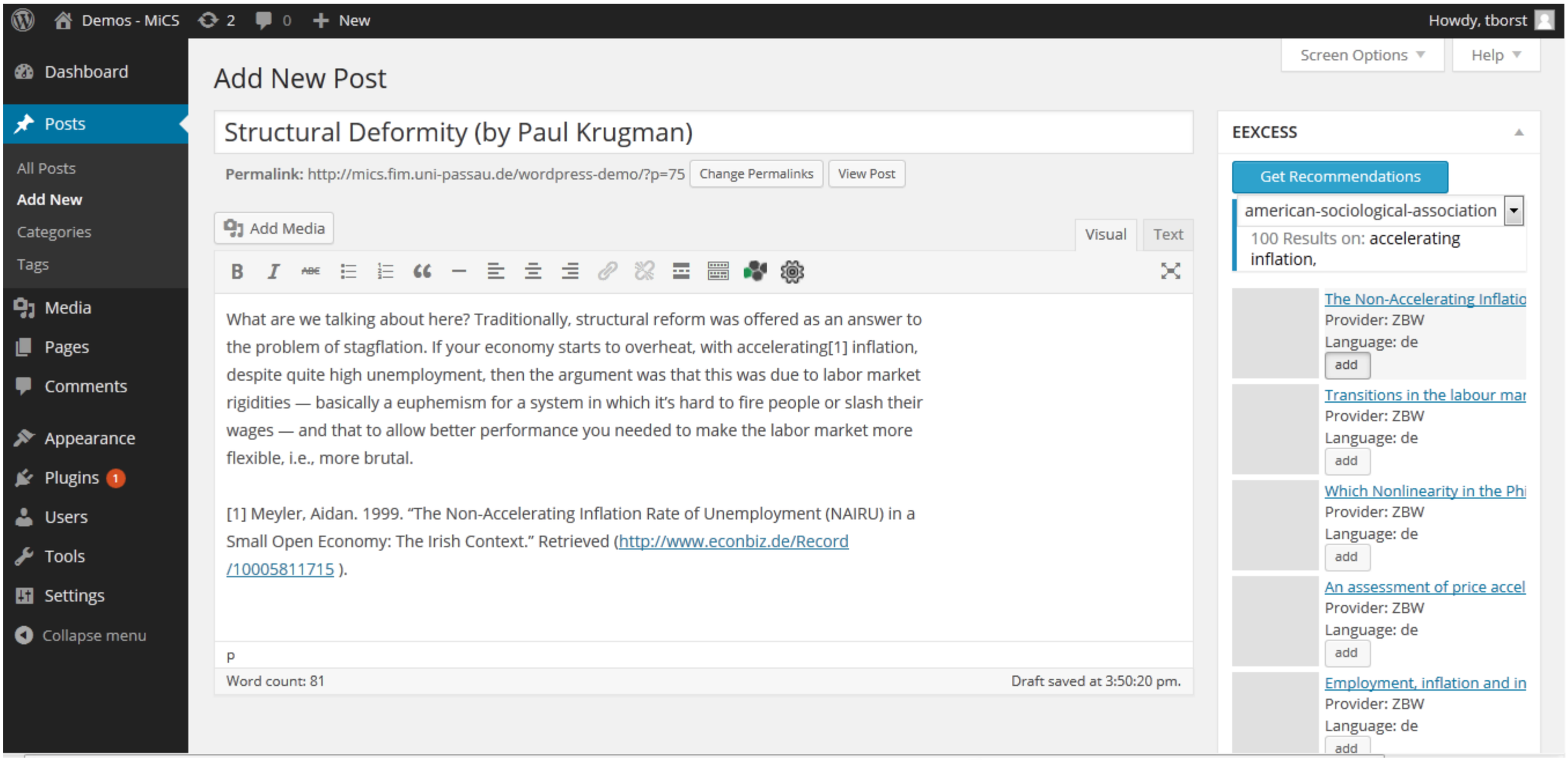
Recommending literature within an authoring environment

- WordPress-Plugin for recommendations
- EEXCESS = EU funded project / duration: 3.5 years / budget: / (content) partners (amongst others) BitMedia (Austria), Mendeley (UK), ZBW (Germany)
- WordPress = Most popular software package for blogging, also widely used by academics and online journalists


Blogging platform/CMS top 100 blogs
WordPress
52%
12%
7%
6%
5%
4%
4%
4%
3%
1%
1%
WordPress
Custom
Drupal
N/a
Gawker
BlogSmith
Movable Type
TypePad
Blogger
Ceros
Joomla
Tumblr
Data sources: Technorati, Pingdom
www.pingdom.com







Recommending literature within an authoring environment

User feedback (so far…):

- Not too many blogging scientists responded – maybe just because a blogging platform like WordPress actually is not used as a native authoring environment (at least in Economics), but as a platform for publishing/posting content which is copied and pasted from other native authoring environments (e.g., Word)
- Several suggestions on
  - editing references
  - (re-)sorting the recommendations (relevance ranking?)
  - embedding images (nice feature, but what about copyright issues?!)

Summary

- The World Wide Web (WWW) has become a place not only to expose and access a library‘s content, but also to inject this content into web-based user environments

- This is achieved by by technical operations like Search Engine Optimization, or with technical infrastructures like APIs or WebServices

- Requirements for setting up an API or a Web Service for accessing and integrating a library‘s content? This is another story, but put this on your agenda…

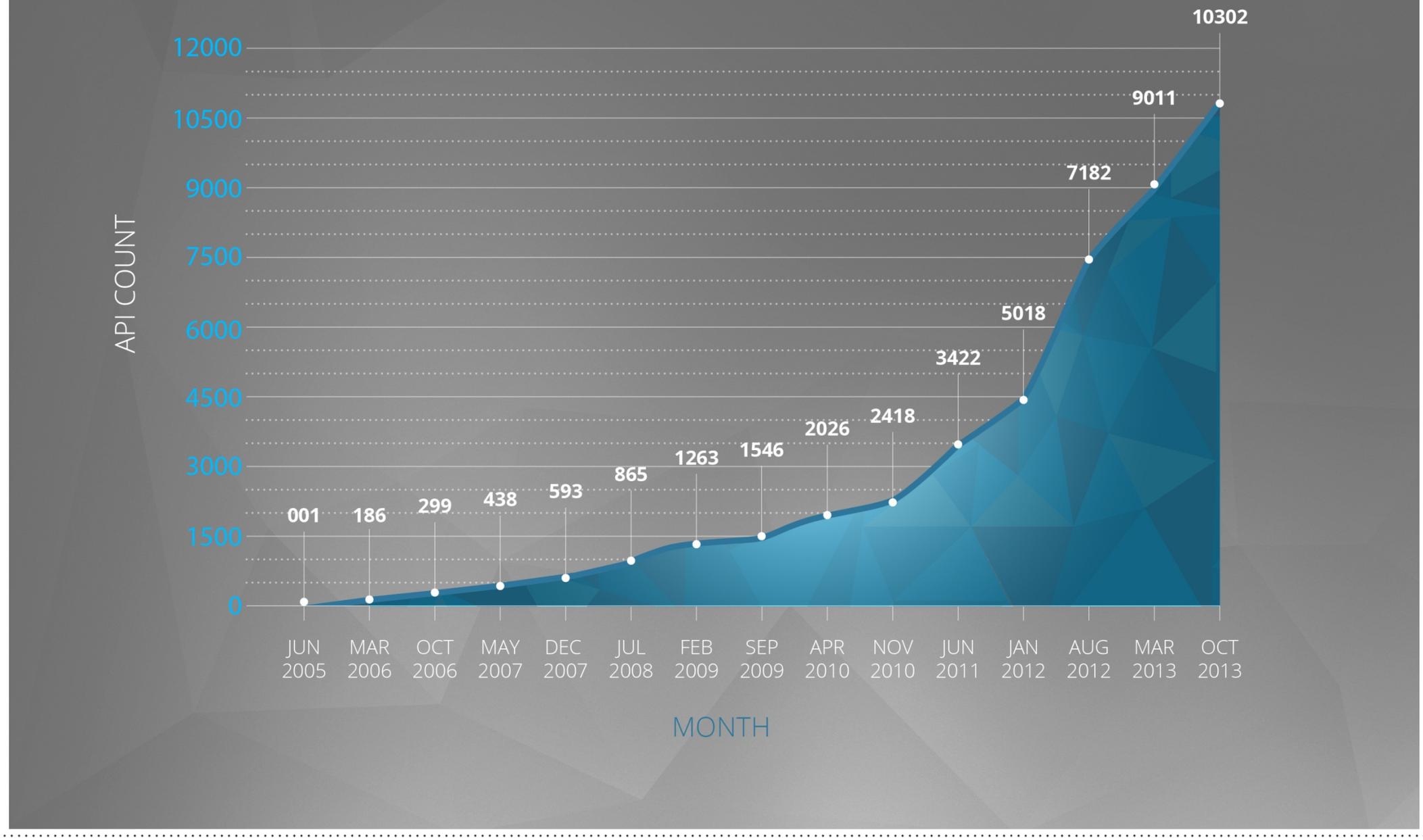
…to become a real ‚inside-out library‘!

Timo Borst

t.borst@zbw.eu

ProgrammableWeb

Growth In Web APIs Since 2005



Content

- Three examples for providing library‘s content in heterogenous digital user environments
  - WP-Plugin for literature recommendations (API?) (Demo)
  - Integrating authority data for indexing and retrieval (API?)
  - Central search index for distributing metadata in different applications (API?)
  - Boosting for better indexing through search engines
- Technical solutions: APIs, Widgets, Formats
- Technical Workflow: Encoding / Exposing / Accessing Information
- Content/Metadata APIs from Libraries and Information Providers
  - „Programmable Web“, Developer Contests, Lightweight Integration

Technical approaches for injecting library content into user environments

- (Meta-)Data integration (Google Scholar optimization) (heavy integration)
  - Use case: Harvesting and indexing distributed metadata and content
- Application Programming Interfaces (APIs) (Authority data) (middle-weight integration)
  - Use case: Providing data and functionality through Web services
- Widgets, JavaScript (lightweight integration)
  - Use case: Injecting external web content