

Einbindung von Linked Data in existierende Bibliotheksanwendungen

Dr. Timo Borst

Joachim Neubert

IT-Entwicklung

Deutsche Zentralbibliothek für Wirtschaftswissenschaften /

Leibniz-Informationszentrum Wirtschaft

Kiel/Hamburg

SWIB09 – Semantic Web in Bibliotheken

Köln, 24./25. November 2009





- 1) Linked Data in der ZBW: Kurze Historie
- 2) Design des STW-Webservice und API
- 3) Integration von Linked Data in Bibliotheksanwendungen
 - a. Motivation (am Beispiel von Repositorien)
 - b. Anwendungsfall_1: Verschlagwortung
 - c. Anwendungsfall_2: Suche
 - d. Anwendungsfall_3: Erfassung von Autoren
 - e. Mehrschichten-Architektur
- 4) Zusammenfassung und Fazit: Die Bedeutung von Linked Data für Bibliotheken

Linked Data in der ZBW: Kurze Historie



- Die Deutsche Zentralbibliothek für Wirtschaftswissenschaften / Leibniz-Informationszentrum Wirtschaft sammelt und erfasst wirtschaftswissenschaftliche Literatur seit über 90 Jahren
- Hoher Grad an eigener systematischer Indexierung
- Der “Standard-Thesaurus Wirtschaft” (STW) wurde in den 90er Jahren gemeinsam mit anderen ökonomischen Instituten im Rahmen eines Projekts gefördert
- Ziel des Projekts war die Vereinheitlichung wirtschaftswissenschaftlicher Schlagwörter
- Mittlerweile wird der STW von der ZBW weiter gepflegt, aktuell umfasst er 5.800 Schlagwörter in Deutsch und Englisch
- März 2009 wurde der STW als Linked Data im Netz publiziert. Merkmale: Aufbereitung als SKOS, Webpräsentation mit RDFa-Tags, Download als RDF/XML-Datei, SPARQL-Endpoint
- Veröffentlicht unter einer Creative Commons-Lizenz zum nicht-kommerziellen Gebrauch (by-nc-sa)

- Home
- Alphabetical descriptor list

- ▶ A General descriptors
- ▶ B Business economics
- ▶ G Geographic names
- ▶ N Related subject areas
- ▶ P Commodities
- ▼ V Economics
 - V.00 Economics
 - ▶ V.01 Economic Theory
 - ▶ V.02 Microeconomics
 - ▶ V.03 Macroeconomics
 - ▶ V.04 Business Cycle, Growth, and Economic Structure
 - ▼ V.05 Money and Financial Markets
 - V.05.01 Monetary Theory
 - V.05.02 Monetary Policy
 - V.05.03 Inflation
 - V.05.04 Interest Rate
 - V.05.05 Credit
 - ▼ V.05.06 Financial Markets

Financial crisis

Finanzmarktkrise (german)

used for: Financial instability

Narrower Terms

- Banking crisis 
- Currency crisis 
- Stock market crash 

Broader Terms

- Economic crisis 

Related Terms

- International financial market 

Subject Categories

- V.05.06 Financial Markets
- V.07.07 International Financial Markets

Persistent Identifier (for bookmarking and linking)

- <http://zbw.eu/stw/descriptor/19664-4>



(Längerfristige) Ziele

- Den STW als Linked Data in ZBW-eigene Applikationen einbinden
- Auch für Semantic Web-Unkundige (*“RDF(a)? OWL?? Turtles?? Graphen?? Ontologien???”*) anbieten
- Möglichkeit zur “leichtgewichtigen” Integration in andere Bibliotheksanwendungen schaffen
- Damit auch Dritten die Möglichkeit zu verhältnismäßig unaufwendiger Nachnutzung bieten
- Externe Linked Data selber wiederum möglichst “leichtgewichtig” integrieren



Entwurfsentscheidungen

- Aktuelle Testversion öffentlich verfügbar unter <http://zbw.eu/beta/stw-ws>
- „Kommandozeilen“-Mimik zur Vermeidung komplexer SPARQL-Anfragen:

<http://zbw.eu/beta/stw-ws/suggest?query=finanz>

ist äquivalent mit:

```
SELECT DISTINCT ?term ?concept ?prefLabel
WHERE {
  {
    ?concept skos:prefLabel ?term .
  } UNION {
    ?concept skos:altLabel ?term .
  } UNION {
    ?concept skos:hiddenLabel ?term .
  } .
  FILTER (regex(str(?term), "^finanz", "i")) .
  ?concept skos:prefLabel ?prefLabel .
}
ORDER BY ?term

LIMIT 10
```



Entwurfsentscheidungen

- An typischen Anwendungsfällen im Retrieval (Konzept-/Labelsuche, Synonymsuche, Eingrenzung/Erweiterung von Suchanfragen) orientiert
- Liefert alle notwendigen Informationen in einem Round-Trip
- Pragmatischer Ansatz: Kein Vollständigkeitsanspruch im Gegensatz zu SKOS API (2004) und anderen SOAP-basierten Ansätzen
- Anwender (Programmierer) können selbstgebaute Queries gegen den öffentlichen SPARQL-Endpoint ausführen und die Ergebnisse mit demselben Interface ihren eigenen Anwendungen anbieten



Beispiel: Synonymsuche

/synonyms

Erhält einen Term (Ein- oder Mehrwort) und gibt die dazugehörigen Synonyme aus.

<http://zbw.eu/beta/stw-ws/stw-ws-wrapper.php?service=synonyms&query=bank>

STW Web Service Query Results

synonym
"Bank" @en
"Bank" @de
"Kreditbank" @de
"Geschäftsbank" @de
"Kreditwesen" @de
"Kreditwirtschaft" @de
"Geldinstitut" @de
"Bankwesen" @de
"Kreditinstitut" @de
"Bankgewerbe" @de



Übersicht zu den bisher angebotenen Services

Servicename	Beschreibung
/concepts	Liefert den Term als Linked Data (d.h. als URI)
/narrower	Liefert zu einem Term alle Unterbegriffe
/labels	Liefert alle (mehrsprachigen) Namen für ein Konzept
/synonyms	Liefert alle Vorzugsbenennungen eines Terms
/suggest	Liefert Vorschläge für einen Suchterm (auf Basis der eingegebenen Anfangszeichen)
/combined1	Vereinigung der obigen Operationen, zur Integration in Suchanwendungen



Parameter des STW-Webservice

dataset (stw / jel / ???)

query (Suchstring)

concept (URI des Konzepts)

lang (en / de)

output (html / sparql-xml / sparql-json)

callback (Javascript-Funktion)



Motivation (am Beispiel von Repositorien)

- Warum Repositorien oder Repository-SW?
- Repositorien sind allgemein genutzte Erfassungs-/Retrieval-/Verbreitungssysteme („Data Provider“), zumeist OSS und Community-getrieben
- Zahlreiche Installationen an Hochschulen, Bibliotheken, Rechenzentren
- Einschlägige OSS-Repositoriensoftware (DSpace, EPrints, OPUS) unterstützt derzeit noch nicht die Integration extern kontrollierter Normdaten

„The major difficulty we have found is with DSpace’s handling of metadata. While we feel that the number of fields in Dublin Core is adequate for most if not all uses (DCMI Usage Board 2006), we are troubled by the lack of authority control when completing its fields. Without some control over uniform titles, authors and subjects accessing the items in the future will very problematic.“ [4]

„Neither the standards nor the software underlying institutional repositories anticipated performing naming authority control on widely disparate metadata from highly unreliable sources.“ [2]



Typische Mängel

- Daten (Autorennamen, beschreibende Schlagwörter, Herausgeber, Dokumenttypen etc.) werden ohne semantische und syntaktische Restriktionen teilweise im Freitext erfasst
- (Syntaktische) Suchen liefern nur eingeschränkte oder uneindeutige Ergebnisse, erfordern ggf. aus Anwendersicht subtilere Suchkenntnisse (Trunkierung etc.) (Suche nach „Abbott, Philip“ liefert nur Treffer für Nr. 2/3)
- Namensänderungen sind praktisch nicht nachvollziehbar
- Selbst wenn syntaktische und semantische Eindeutigkeit in einem Repository lokal erzielt werden können, erneuert und verschärft sich das Problem auf Seiten der Aggregatoren

Abbott, P.C.

Abbott, Philip

Abbott, Philip C.

Anwendungsfall_1: Verschlagwortung



- DER Anwendungsfall für Bibliothekare
- Nutzergruppen: Bibliothekare + Wissenschaftler (?) + Bibliotheksnutzer (?)
- Vorgang: Eingabe von beschreibenden Metadaten, die häufig aus einem kontrollierten Vokabular stammen
- Bestandteil von Open Access relevanten Vorgängen wie das „self-publishing“, das „self-archiving“ oder das „social tagging“
- Frage: Können wir den Vorgang der Verschlagwortung mit Hilfe des STW Webservice unterstützen, so dass wir bei der Eingabe zu „besseren“ normierten Metadaten gelangen?

Anwendungsfall_1: Verschlagwortung



Demo

Anwendungsfall_1: Verschlagwortung



Erfassungsmaske unter <http://econstor.eu>

Geben Sie eine Zusammenfassung des Inhalts der Publikation ein.

Abstract

Impact of the financial crisis on the computer industry.

Geben Sie hier die STW-Begriffe ein.

STW-Begriffe

Computerindustrie

financia

Weitere hinzufügen

Klassifizierung der Publikation entsprechende Kategorie

JEL

Bitte wählen Sie die Sprache aus, in der

Sprache

N/A

STW Begriffe:

- Financial Engineering
- Financial Futures
- Financial administration → Finanzverwaltung
- Financial aid → Kapitalhilfe
- Financial arbitrage → Arbitragegeschäft
- Financial asset market → Finanzmarkt
- Financial assets → Betriebliches Finanzanlagevermögen
- Financial audit → Wirtschaftsprüfung
- Financial centre → Finanzplatz
- Financial crisis → Finanzmarktkrise

terature Classification System"

fügen

< Zurück

Weiter >

Abbi

trieben von der [Deutschen Zentralbibliothek für Wirtschaftswissenschaften](#) [Finanzmarktkrise](#) [Akt - Über Eco](#)

Anwendungsfall_1: Verschlagwortung



Erfassungsmaske unter <http://econstor.eu>

Geben Sie eine Zusammenfassung des Inhalts der Publikation ein.

Abstract

Impact of the financial crisis on the computer industry.

Geben Sie hier die STW-Begriffe ein.

STW-Begriffe

Computerindustrie

Finanzmarktkrise

Weitere hinzufügen

Klassifizierung der Publikation entsprechende Kategorien aus dem "Journal of Economic Literature Classification System"

JEL

L6

Weitere hinzufügen

Bitte wählen
Sprache

JEL Klassifikation:

- L6 - Industry Studies: Manufacturing
- L60 - Industry Studies: Manufacturing: General
- L61 - Metals and Metal Products; Cement; Glass; Ceramics
- L62 - Automobiles; Other Transportation Equipment
- L63 - Microelectronics; Computers; Communications Equipment**
- L64 - Other Machinery; Business Equipment; Armaments
- L65 - Chemicals; Rubber; Drugs; Biotechnology

Bitte Publikation verfasst ist.

< Zurück

Weiter >

Abbi

trieben von der [De](#)

[Wirtschaftswissenschaften \(ZBW\)](#) - [Kontakt](#) - [Über Eco](#)

Anwendungsfall_2: Suche



- Der häufigste Anwendungsfall in Bibliotheksanwendungen
- Nutzergruppen: Bibliothekare + Wissenschaftler + Bibliotheksnutzer
- Vorgang: Zumeist einfache Volltextsuche über Metadatenfelder und/oder Volltexte (soweit letztere verfügbar sind und im Suchmaschinenindex liegen)
- Frage: Können wir den Vorgang der Suche mit Hilfe des STW Webservice unterstützen, so dass wir zu besseren Suchergebnissen gelangen?

Anwendungsfall_2: Suche



Demo

Anwendungsfall_2: Suche



Erweiterte Suche unter <http://econstor.eu>

Eingabe: „Telework“

Treffer: 9

EconStor.
DIGITAL ARCHIVE FOR ECONOMICS AND BUSINESS STUDIES

In EconStor suchen

[Erweiterte Suche](#)

- [Startseite](#)

Blättern in:

- [Institutionen & Reihen](#)
- [Titeln](#)
- [Autoren](#)
- [Stichwörtern](#)
- [Nach Datum](#)

Anmelden für:

- [Abonnements](#)
- [Mein EconStor](#)
Registrierte Benutzer
- [Profil bearbeiten](#)

[DSpace@ZBW](#) >

Suchergebnisse

Suchen:

nach

Telearbeit Telekommunikation Virtuelles Unternehmen

Ergebnisse 1-9 von 9.

Treffer in Publikationen:

Datum	Titel	Autoren
2001	Why Do Firms Recruit Internationally?	Winkelmann, Rainer
2007	Flexicurity : Reconciling Social Security with Flexibility ; Empirical Findings for Europe	Seifert, Hartmut / Tangian, Andranik
2007	Is flexible work precarious? : a study based on the 4th European survey of working conditions 2005	Tangian, Andranik S.

Anwendungsfall_2: Suche



Mouse-over „Telearbeit“

[DSpace@ZBW](#) >

Suchergebnisse

Suchen:
nach

Telearbeit Telekommunikation Virtuelles Unternehmen

-9 von 9.

Suchterme:

- Telework
- Telearbeit
- Teleworking
- Telecommuting
- Distance working
- Teleheimarbeit
- Elektronische Heimarbeit
- EDV-Heimarbeit

Relevanz: 6
Treffer (match) der Suchanfrage:
- Telework

Datum	Titel	Verfasser
20...	...nationally?	Wink Raine
20...	...ocial Security Findings for	Seife Tang

Anwendungsfall_2: Suche



Trefferzahl: 20 (!) durch die Erweiterung mit Synonymen

[DSpace@ZBW](#) >

Suchergebnisse

Suchen:
EconStor gesamt
nach

Telearbeit Telekommunikation Virtuelles Unternehmen

Ergebnisse 1-10 von 20.

Treffer in Publikationen:

Datum	Titel	Au
2001	Qualifikationsanforderungen und Arbeitsformen in der Neuen Ökonomie	Christenser
2007	Flexicurity : Reconciling Social Security with Flexibility ; Empirical Findings for Europe	Seifert, Har Tangian, Ar

Anwendungsfall_3: Erfassung von Autoren



- Der Normalfall in Katalogen - in anderen Erfassungssystemen bisher der Ausnahmefall
- Nutzergruppen: Bibliothekare + Wissenschaftler (?) + Bibliotheksnutzer (?)
- Vorgang: Eingabe von Autorennamen
- Frage: Können wir die Autorenerfassung mit Hilfe des STW-Webservice und vor allem mit extern kontrollierten Normdaten so unterstützen, dass wir bei der Eingabe zu „besseren“ Autordaten gelangen?

Anwendungsfall_3: Erfassung von Autoren



Demo

Anwendungsfall_3: Erfassung von Autoren



Erfassungsmaske unter <http://87.106.250.18/beta/econstor/>

Publikationstyp des eingereichten Dokuments aus. Für die Auswahl mehrerer Typen halten Sie die STRG

Publikationstyp

- Zeitschriftenaufsatz / Article
- Buchaufsatz / Book Part
- Konferenzbeitrag / Conference Paper
- Dissertation / Doctoral Thesis
- Buch / Book
- Konferenzband / Proceedings

Geben Sie den Titel der Publikation ein.

Titel Financial asset returns, direction-of-change forecasting

Geben Sie den Namen des Autors bzw. die Namen der Autoren ein. Reihenfolge: Nachname, Vorname(n)

Autor(en) sie

Weitere hinzufügen

Abstract

Personen Normdaten:

- Sieb, Peter (1642-1665)
- Sieb, Petrus → Sieb, Peter (1642-1665)
- Siebach, Konrad (1912-1995)
- Siebe, Hans (1919-2001)
- Siebe, Josefine → Siebe, Josephine (1870-1941)**
- Siebe, Josephine (1870-1941)
- Siebe, Józefa → Siebe, Josephine (1870-1941)
- Siebelis, Carl G. → Siebelis, Carl Gottfried (1769-1843)
- Siebelis, Carl Gottfried (1769-1843)
- Siebelis, Carolus G. → Siebelis, Carl Gottfried (1769-1843)

tion ein.

Weitere hinzufügen

STW-Begriffe

haltlichen Klassifizierung

JEL

of Economic Literature Classificati

Weitere hinzufügen

DEUTSCHE
NATIONAL
BIBLIOTHEK

Treffer gesamt: 232

NORMDATEN: *Personenname (107475073)*

Siebe, Josephine

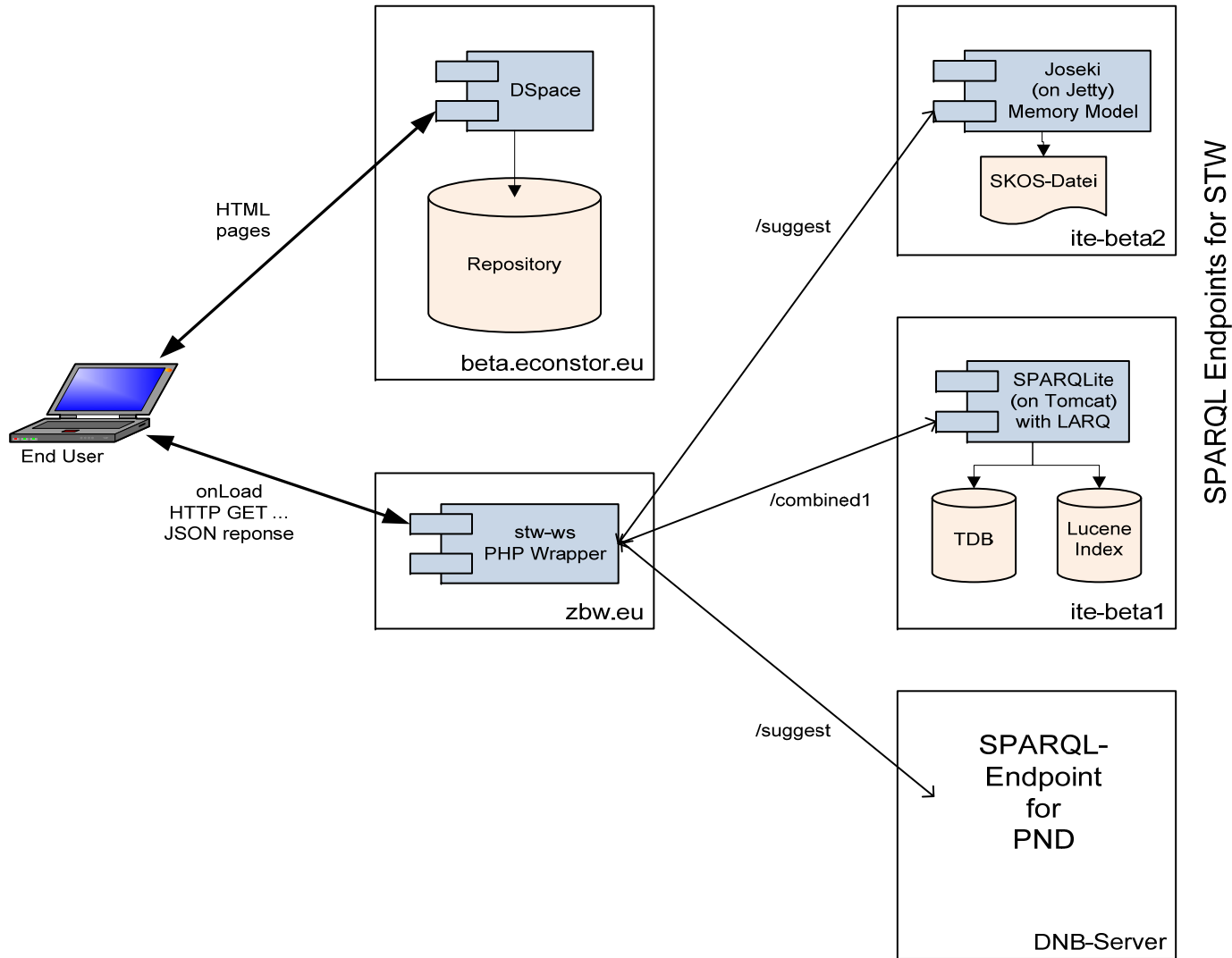
Lebensdaten: |1870-1941

1. Beruf/Funktion: 4053311-6 Schriftstellerin

Verweisungsformen: Siebe, Józefa

Siebe, Josefine

Mehrschichten-Architektur





Vorteile

- „Leichtgewichtige“ Erweiterungen bestehender (Alt-)Systeme
- Strategie des „minimalen Eingriffs“
- Keine Änderungen am Datenmodell oder am Systemkern nötig
- Typen von (Web-)Bibliotheksanwendungen, die sich auf diese Weise zwanglos erweitern ließen:
 - Repositorien (zumeist OSS, Plugin-Architektur)
 - Kataloge (zumeist proprietäre SW und monolithisch)
 - Portale (zumeist OSS, heterogene Inhalte hinsichtlich Art und Herkunft) -> Webservice zur Erzeugung von Meshups
 - ...sowie alle offenen Systeme mit kollaborativer verteilter Verschlagwortung (z.B. Academic Linkshare)



Komplementäre Ansätze

DSpace 1.6 (in Planung, siehe [1]):

- Umfassende Unterstützung bei der Konfiguration von Auswahlfeldern und ihrer Belegung mit Authority-Plugins („Choice Management“)
- Erweiterung des Datenmodells um eine Spalte für den „Authority Key“ (nach dem Linked Data-Modell eine URI), ferner um eine Spalte für die sprachspezifische Anzeige (im Fall von mehrsprachigen Thesauri -> AGROVOC)
- Anpassungen bei den Eingabe- bzw. Bearbeitungsmasken, bei der Suche und bei der Darstellung

Zusammenfassung und Fazit: Die Bedeutung von Linked Data für Bibliotheken



- Normdaten liegen häufig nicht bei der Einrichtung selbst, sollten gerade deshalb aber eingebunden werden (Bibliotheken als Abnehmer von Linked Data zur Veredelung der im Haus generierten Metadaten)
- Damit Normdatenbestände sowohl im Haus als auch von Dritten genutzt werden können, bedürfen sie der Exponierung als Linked Data
- Konsequente Weborientierung: Selbst die passive Nutzung von Linked Data erfordert tiefergehende Kenntnisse über Webtechnologien
- Webservices können die Schwelle zur Nutzung dieser Daten senken, indem sie ein simples API anbieten
- Wenn das verbreitet geschieht, haben die Betreiber zumal von Repositorien und anderen webbasierten Bibliothekssystemen die Wahl, welche Schlagwortlisten/Thesauri/Normdaten sie einbinden und welche Tools sie zur Unterstützung von Indexierung und Recherche nutzen.



Vielen Dank!

Links:

<http://87.106.250.18/beta/econstor/>

<http://zbw.eu/beta/stw-ws>

<http://zbw.eu/beta/sparql>

<http://zbw.eu/stw>

On Publishing STW as SKOS and Linked Data:

http://events.linkeddata.org/ldow2009/papers/ldow2009_paper7.pdf

Timo Borst

Joachim Neubert

German National Library of
Economics (ZBW) /

Leibniz Centre for Economics
{t.borst,j.neubert}@zbw.eu



- [1] http://wiki.dspace.org/index.php/Authority_Control_of_Metadata_Values
- [2] <http://minds.wisconsin.edu/handle/1793/31735>
- [3] <http://dsug09.ub.gu.se/index.php/dsug/dsug09/paper/view/22/3>
- [4] <http://subjectobject.net/2006/11/09/the-dspace-digital-repository-a-project-analysis/>
- [5] <http://code.google.com/p/dspace-agrisap/wiki/ThesaurusAddOn>
- [6] <http://edoc.hu-berlin.de/conferences/dc-2008/subirats-imma-199/PDF/subirats.pdf>