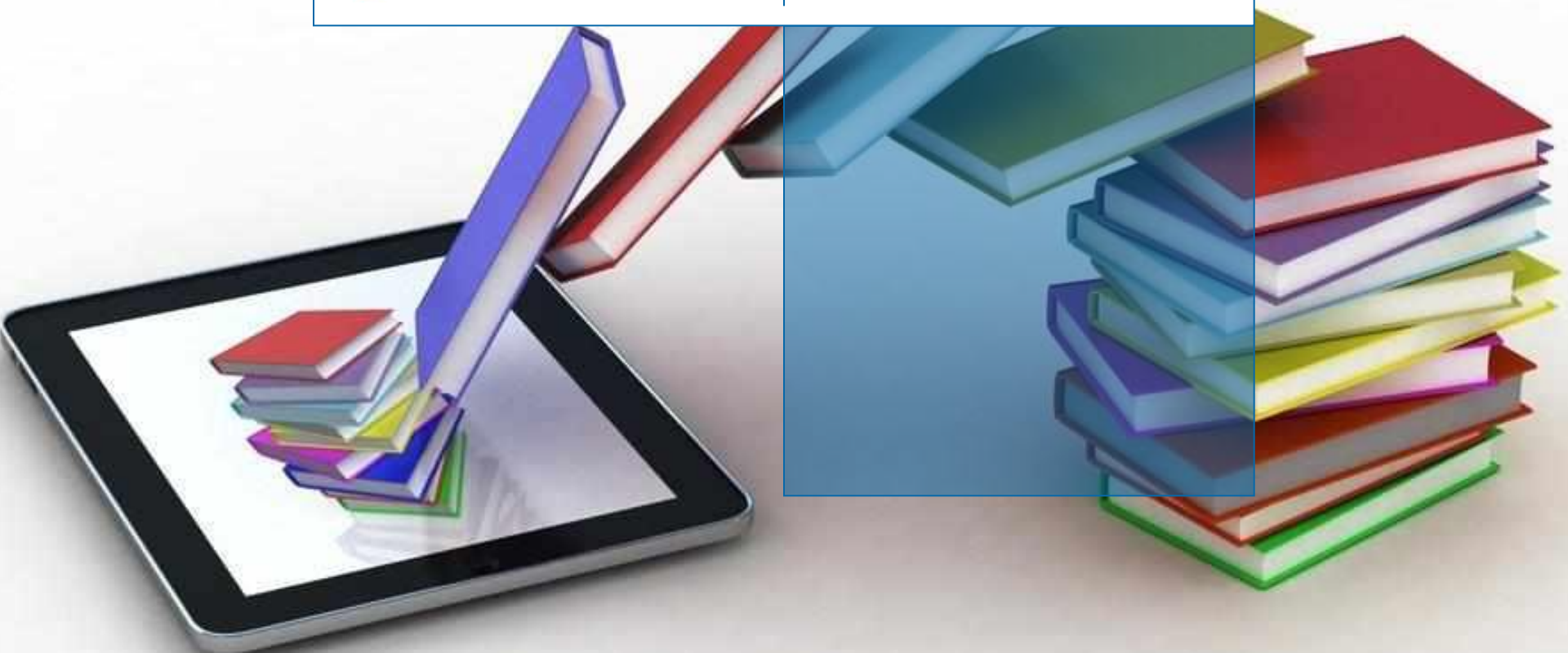
XML last UND XML first - Erfahrungswerte aus dem Verlagsalltag von Wolters Kluwer Deutschland

Sven Krantz-Knutzen

22.11.2013

Wolters Kluwer



Über Wolters Kluwer

Die Wolters Kluwer Deutschland GmbH ist ein Wissens- und Informationsdienstleister mit Kernkompetenzen in den Bereichen Recht, Wirtschaft und Steuern. Für den professionellen Anwender bietet Wolters Kluwer Deutschland fundierte Fachinformationen in Form von Literatur, Software und Services. Das Unternehmen mit Hauptsitz in Köln beschäftigt an über 20 Standorten rund 1.200 Mitarbeiter und agiert seit über 25 Jahren auf dem deutschen Markt.

Wolters Kluwer Deutschland ist Teil des führenden internationalen Informationsdienstleisters Wolters Kluwer n.v. mit Sitz in Alphen aan den Rijn (Niederlande). Die Kernmärkte des Unternehmens sind Recht, Wirtschaft, Steuern, Rechnungswesen, Unternehmens- und Finanzdienstleistungen und Gesundheitswesen; Zielgruppe sind professionelle Anwender. Die Aktien sind an der Euronext Amsterdam (WKL) gelistet, außerdem in der AEX und im Euronext 100 Index. Wolters Kluwer hat einen Jahresumsatz (2012) von 3,6 Milliarden Euro, beschäftigt weltweit rund 19.000 Mitarbeiter und ist in über 40 Ländern in Europa, Nordamerika und im Raum Asien/Pazifik und Lateinamerika vertreten.

Weitere Informationen finden Sie unter www.wolterskluwer.de

Zur Person

Sven Krantz-Knutzen (Jurist) hat von 2002 bis 2006 als selbstständiger Web Solution Developer gearbeitet. 2003 übernahm er die Leitung der juristischen Dokumentation Normen bei der LexisNexis Deutschland GmbH. Ab dem Jahr 2006 war er bei LexisNexis Deutschland für die Bereiche Copy Editing, Datenaufbereitung und elektronische Produktion verantwortlich, ab 2009 auch für LexisNexis Österreich.

Seit 2011 leitet Sven Krantz-Knutzen die Editorial Services bei der Wolters Kluwer Deutschland GmbH und ist dort für Copy Editing, XMLfirst-Prozesse, Contentaufbereitung, Verlinkung und Qualitätsmanagement verantwortlich.

Agenda

1. **XML first vs./und XML last**
2. Anforderungen an XML bei Wolters Kluwer
3. Strategie
4. QS-Tools
5. Ausblick

XML first vs. XML last

© http://www.preussenweb.de/kriege6.htm

XML first und XML last

Voraussetzungen

Damit eine Ergänzung zwischen beiden Wegen stattfinden kann, müssen bestimmte Voraussetzungen gegeben sein:

- Klare Zielvorstellungen
- Konstante Kostenüberprüfung
- Gemeinsame QS-Tools und QS-Regeln
- Gleiche dtd(s) bzw. Schema(ta)
- Daten müssen den Anforderungen des benutzten Systems entsprechen.

Ein „gutes XML“ erfüllt die Anforderungen, wenn es

- Angemessene inhaltliche und semantische Tiefe,
- Kostengünstige Erstellung und
- Angemessene Erstellungszeiten kombiniert.

Kann es Unterschiede geben?

Zu entscheiden ist die Frage, ob Unterschiede im XML zugelassen werden sollen/müssen?

- Tiefe der Strukturierung
- Kohärenz
- Zeitpunkt der XML-Verfügbarkeit im Prozess
- Enhancement
  - Semantische Auszeichnung
  - Anreicherung mit Metadaten
  - QS der Metadaten durch den Autor/die Autorin.

Agenda

1. XML first vs./und XML last
2. **Anforderungen an XML bei Wolters Kluwer**
3. Strategie
4. QS-Tools
5. Ausblick

Anforderungen an WKD-XML 1

- Medienneutralität:
  - Print (Buch, Zeitschrift, Loseblatt, Broschüre)
  - Online (Plattform, Stand-alone-Produkt, Intranet-Einbindung; noch:DVD)
  - Epub
  - Roundtrip nach MS Word für Folgeauflage
- Metadaten:
  - Grundlage für Autoklassifikation
  - Grundlage für Search
  - Produktspezifische Informationen
  - Beziehungen (Anker für RDF-Triple + open data)

Anforderungen an WKD-XML 2

- Harmonisierungen:
  - Zitierweisen (Anker für automatische Verlinkung)
  - Abkürzungen (opt)
  - Vergleichbare Inhalte gleich benutzbar (UX, intuitiv)

- Erweiterbar:
  - Mehr Metadaten
  - Mehr medienspezifische Inhalte
    - Farb-/s/w-Abbildungen
    - Unterschiedliche Auflösungen
    - Interaktive Formulare/Muster
  - Integration in Kundenworkflow
    - Muster/Formulare
    - Verlinkungen auf lokale Inhalte

Enhancement

- Zitierfähigkeit über alle Medienformen hinweg
  - Randnummern
- Aktualisierungen mit Zwischenauflagen in bestimmten Formaten
- „Familienbezüge“ herstellen
  - Z. B. zwischen zusammengehörigem Kommentar, Handbuch und Formularsammlung

Agenda

1. XML first vs./und XML last
2. Anforderungen an XML bei Wolters Kluwer
3. **Strategie**
4. QS-Tools
5. Ausblick

Strategie

- XML first wenn
  - Online-zur-Verfügung-Stellung vor Print erforderlich
    - Sequentielles Veröffentlichen
    - Zwischenauflagen nur online
  - Enhancement mit inhaltlicher QS erforderlich (Autorenbeteiligung)
  - Enhancement im editoriellen Prozess erforderlich (Kostengründe)
  - XMLfirst-Produktion am billigsten
  - Medienspezifische Ausformungen vorhanden
- XML last wenn
  - 1:1-Darstellung Print/Online
  - Mangelnde Standardisierung
  - Kein Enhancement erforderlich oder möglich

Agenda

1. XML first vs./und XML last
2. Anforderungen an XML bei Wolters Kluwer
3. Strategie
4. **QS-Tools**
5. Ausblick

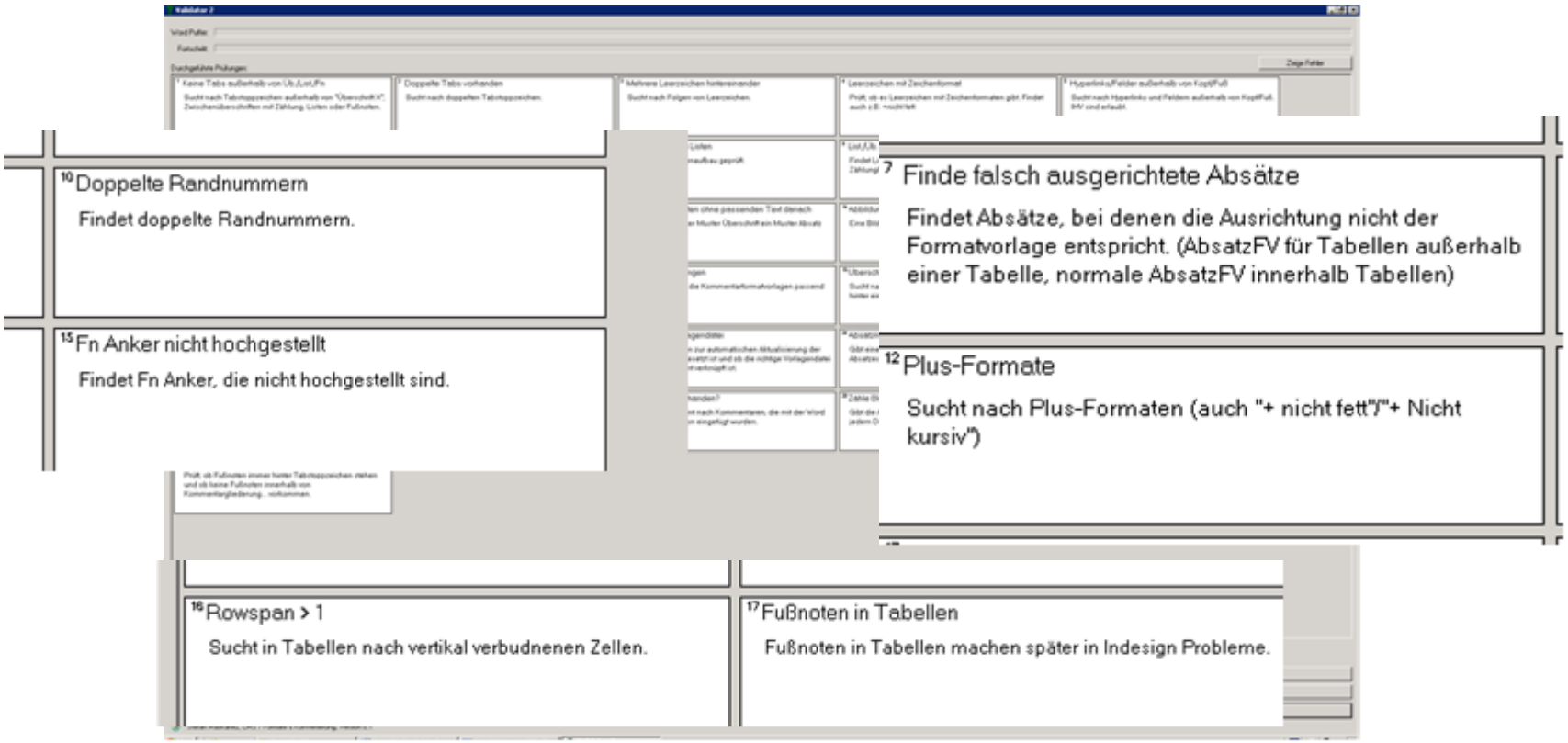
QS-Tools

- Word-Validator (nur XML first)
- Contempo (beide Varianten)
- Überprüfungen beim workflowgesteuerten Einchecken in das CMS (beide Varianten)
  - Validität
  - Encoding und Entities
  - Linkauflösung (sprechende Links werden mit Beziehungen zu echten Dokumenten versehen)
  - Verknüpfungen mit Externen Dokumenten werden mit IDs versehen, obwohl die Dokumente noch nicht vorliegen (müssen)

Word-Validator

„Validitäts- und Kohärenz-Prüfung“ innerhalb des Editors MS Word

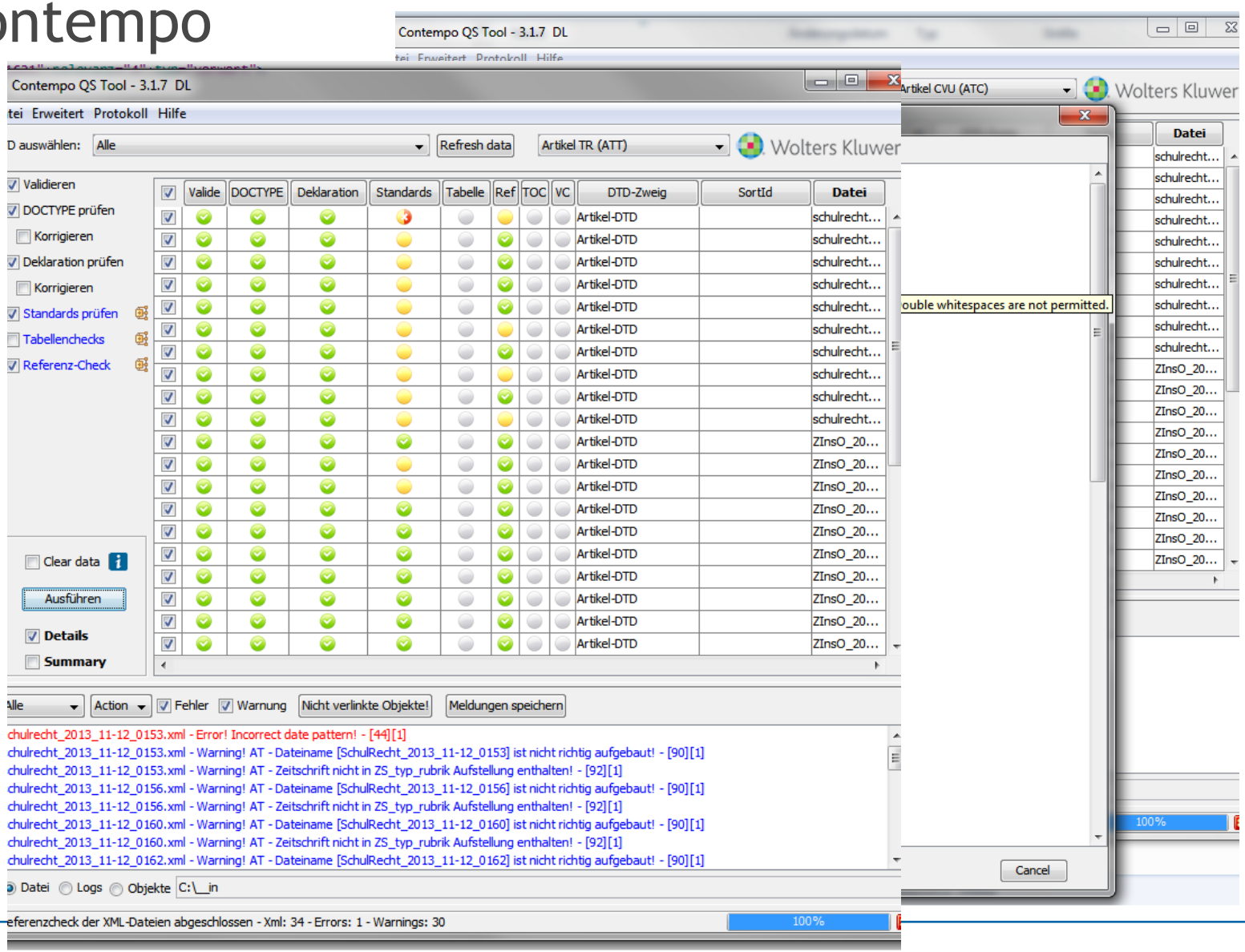
Aufruf über die Symbolleiste



Contempo - Überblick

- Contempo ist ein Tool zur Qualitätssicherung von Wolters Kluwer Deutschland. Es stellt sicher, dass die xml-Erstellungsprozesse vorgabengemäß durchlaufen worden sind und die erwarteten Produkte erzeugt werden können.
- Contempo ist ein batch xml parser mit dem ein oder mehrere Ordner parallel verarbeitet werden können.
- Ziele:
  - Erkennen und identifizieren von Fehlern
  - Fehlerbehebung
  - Fehlerreporting
  - Visualisierte Datenanalyse
  - Troubleshooting und Hilfe

Contempo


Contempo QS Tool - 3.1.7 DL
tei Erweitert Protokoll Hilfe
D auswählen: Alle
Refresh data
Artikel TR (ATT)
Wolters Kluwer
Validieren
DOCTYPE prüfen
Korrigieren
Deklaration prüfen
Korrigieren
Standards prüfen
Tabellenchecks
Referenz-Check
Valide
DOCTYPE
Deklaration
Standards
Tabelle
Ref
TOC
VC
DTD-Zweig
SortId
Datei
Artikel-DTD
schulrecht...
ZInsO_20...
Clear data
Ausführen
Details
Summary
Alle
Action
Fehler
Warnung
Nicht verlinkte Objekte!
Meldungen speichern
chulrecht_2013_11-12_0153.xml - Error! Incorrect date pattern! - [44][1]
chulrecht_2013_11-12_0153.xml - Warning! AT - Dateiname [SchulRecht_2013_11-12_0153] ist nicht richtig aufgebaut! - [90][1]
chulrecht_2013_11-12_0153.xml - Warning! AT - Zeitschrift nicht in ZS_typ_rubrik Aufstellung enthalten! - [92][1]
chulrecht_2013_11-12_0156.xml - Warning! AT - Dateiname [SchulRecht_2013_11-12_0156] ist nicht richtig aufgebaut! - [90][1]
chulrecht_2013_11-12_0156.xml - Warning! AT - Zeitschrift nicht in ZS_typ_rubrik Aufstellung enthalten! - [92][1]
chulrecht_2013_11-12_0160.xml - Warning! AT - Dateiname [SchulRecht_2013_11-12_0160] ist nicht richtig aufgebaut! - [90][1]
chulrecht_2013_11-12_0160.xml - Warning! AT - Zeitschrift nicht in ZS_typ_rubrik Aufstellung enthalten! - [92][1]
chulrecht_2013_11-12_0162.xml - Warning! AT - Dateiname [SchulRecht_2013_11-12_0162] ist nicht richtig aufgebaut! - [90][1]
Datei
Logs
Objekte
C:__in
eferenzcheck der XML-Dateien abgeschlossen - Xml: 34 - Errors: 1 - Warnings: 30
100%
Artikel CVU (ATC)
ouble whitespaces are not permitted.
Cancel

Contempo - erkennen und identifizieren von Fehlern

Contempo verwendet zwei Herangehensweisen um die Datenqualität sicherzustellen:

1. Datenvalidierung (xml-validation, declaration und doctype-check) aufgrund der WKD-DTDs
2. Fehlererkennung und -korrektur.

Um die Fehler erkennen zu können, verwendet Contempo unterschiedliche Regeln für die verschiedenen Publikationstypen: jurkommentar, handbuch, artikel, esa-eintrag. Zur Vereinfachung können in Contempo für jede dtd eigene Suchprofile hinterlegt und werkspezifisch angepasst werden.

Die Prüfung erfolgt in drei Schritten:

1. Standards
2. Tabellen
3. Referenz-Check

Contempo - Fehlerbehebung

Während der Erkennungsphase checkt Contempo die eingeladenen Daten und wirft entweder Fehlermeldungen oder Warnungen.

Fehlermeldungen: der Benutzer muss jede dieser Meldungen kontrollieren und die referenzierten Fehler gemäß der Publikationsregeln korrigieren.

Warnungen: Diese Meldung zeigt an, dass evtl. mit den Daten etwas nicht stimmen könnte. Das xml muss kontrolliert aber nicht notwendigerweise verändert werden. Über diese Meldung können auch generell fehleranfällige Stellen erkannt und angezeigt werden.

Die Fehlerbehebung erfolgt durch manuelles Eingreifen. Dabei kann entweder der eingebaute XML-Editor oder jeder beliebige vorhandene Editor benutzt werden.

Contempo - Fehlerreporting

Contempo kann auch als Reporting-Tool für die Datenqualität benutzt werden:

1. Zusammenfassender Bericht über Anzahl und Typ der gefundenen Fehler/Warnungen aus den überprüften XML-Daten
2. Zusammenfassender Bericht über die durchgeführten und abgewählten Check-Routinen

Agenda

1. XML first vs./und XML last
2. Anforderungen an XML bei Wolters Kluwer
3. Strategie
4. QS-Tools
5. **Ausblick**

Ausblick

- Schemagesteuerte Autorenumgebung
- Metadatenintegration
- RDF Integration - Poolparty
- Online first
  - Sequentielles Veröffentlichen
  - Ausprintung nicht mehr von jeder Auflage
- Bessere Integration in Kundenenvironment
  - Kundenworkflow
  - Open data-Integration
  - Konfigurierbarkeit

WKD und Semantic Web - LOD2

Projektstruktur

Wolters Kluwer Deutschland

OPENLINK SOFTWARE

NUI Galway OÉ Gaillimh

TenForce

UNIVERSITÄT LEIPZIG

Zemanta

exalead connect the dots

INSTYTUT INFORMATYKI GOSPODARCZEJ

VŠE V PRAZE

Freie Universität Berlin

institute MIHAILO PUPIN

CWI

SEMANTIC WEB COMPANY school • consulting • projects • events • media

KAIST

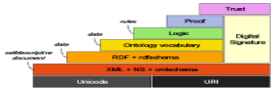
15 Partner aus 12 Ländern

Projektdimension

- 95 Personenjahre
- 10.2 M€ Gesamtbudget
- 7.2 M€ Funding der EU
- Laufzeit 4 Jahre ab 09/2010

Projekthighlights

Open Government Initiative wie in den USA:
Gemeinsame europäische Plattform **PublicData.eu**

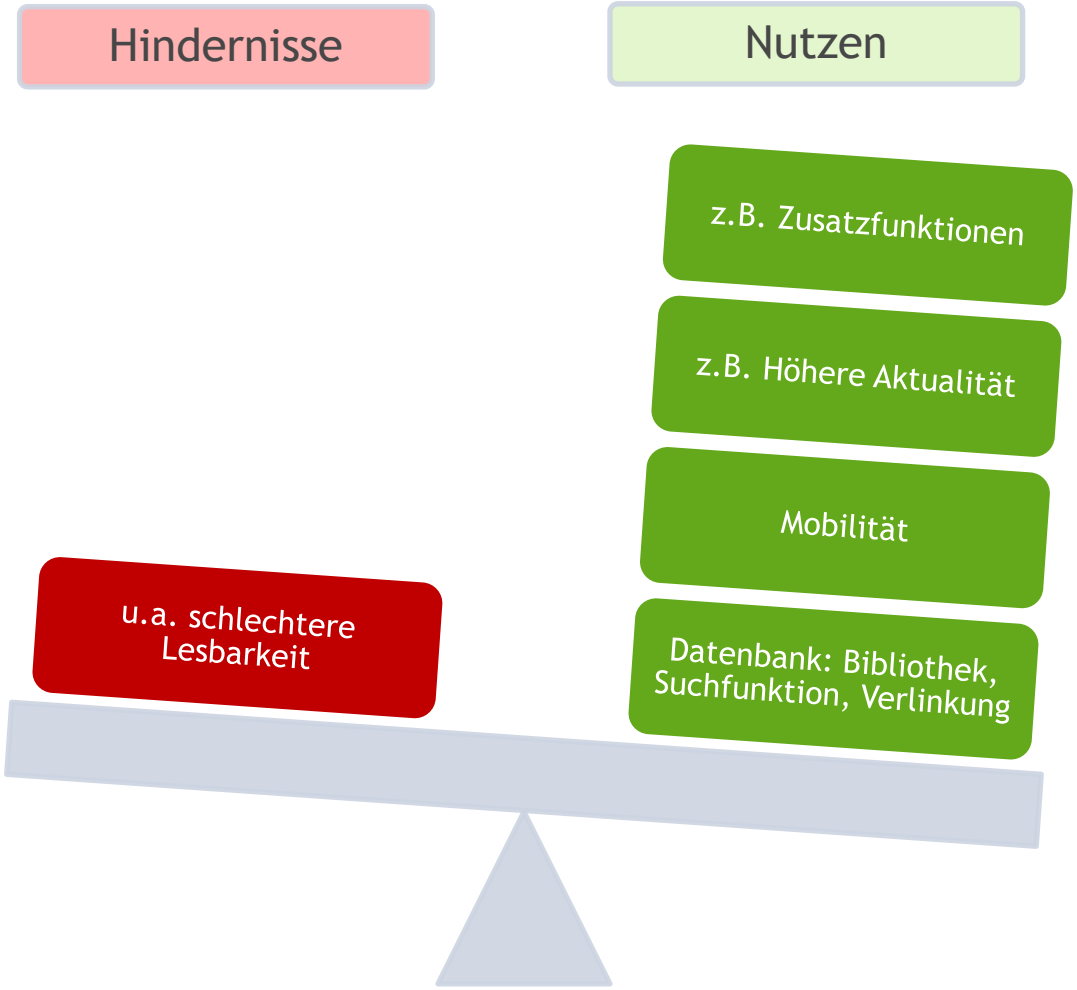


Führende **Web 3.0-Technologien** ziehen im Konsortium an einem Strang, z.B. DBpedia, Virtuoso, Sindice, Silk

Zugang zu üblichen Datenbanklösungen

Hindernisse	Nutzen
Wenig Arbeitserleichterung	Mobilität
u.a. schlechtere Lesbarkeit	Datenbank: Bibliothek, Suchfunktion, Verlinkung

Zugang zu zukünftigen Informationslösungen



Beispiele für erweiterten Kundennutzen

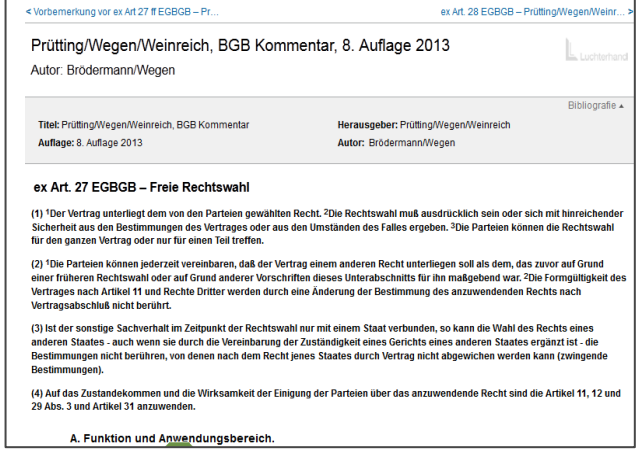
Zusätzliche Metadaten

<autor>...</autor>
<verwandte_werke>...</verwandte_werke>
<grundsatzentscheidung>...</grundsatze...dung>
<rechtsgebiet>...</rechtsgebiet>
<stichwort>...</stichwort>

Zusatzfunktionen, z.B.

...

Zusatzcontent



Aktualisierter Content

Latest News

Kontaktdaten

Sven Krantz-Knutzen

Leiter Editorial Services

Wolters Kluwer Deutschland GmbH

Luxemburger Str. 449

50939 Köln

skrantz-knutzen@wolterskluwer.de

+49 (0) 221 94373 7099