

PubMan Days - Entwicklerworkshop

Datenmigration in PubMan



PubMan Days
München, 19. September 2012

Matthias Walter
walter@mpdl.mpg.de

 Except where otherwise noted, this work is licensed under
<http://creativecommons.org/licenses/by/3.0/de/>



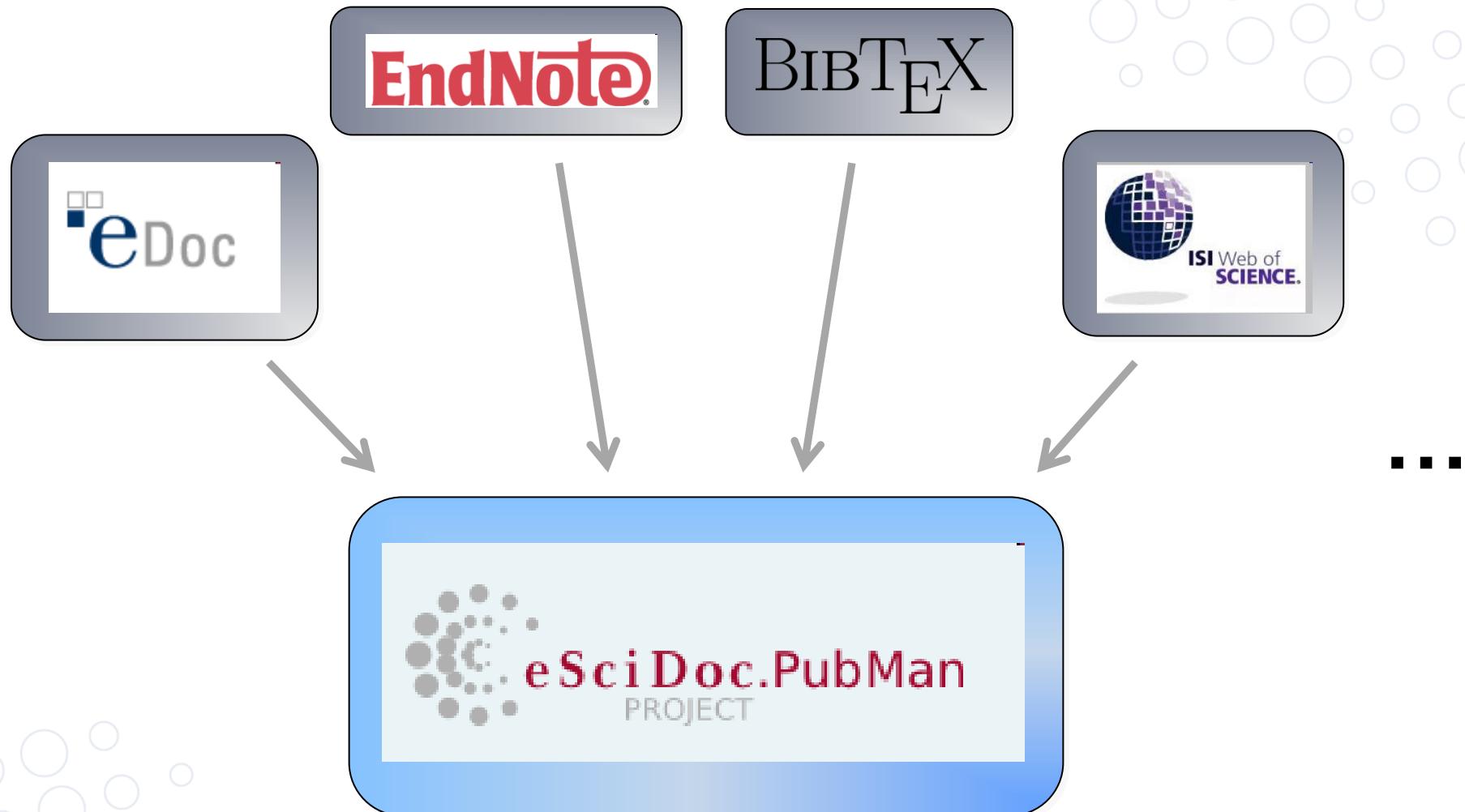
MAX PLANCK
digital library

<http://bit.ly/UiEn8v>

Agenda

- **Einführung**
- Ziele
 - Autoren in CoNE importieren
 - Transformation konfigurieren
 - CoNE-Autorenabfrage in Transformation einbinden

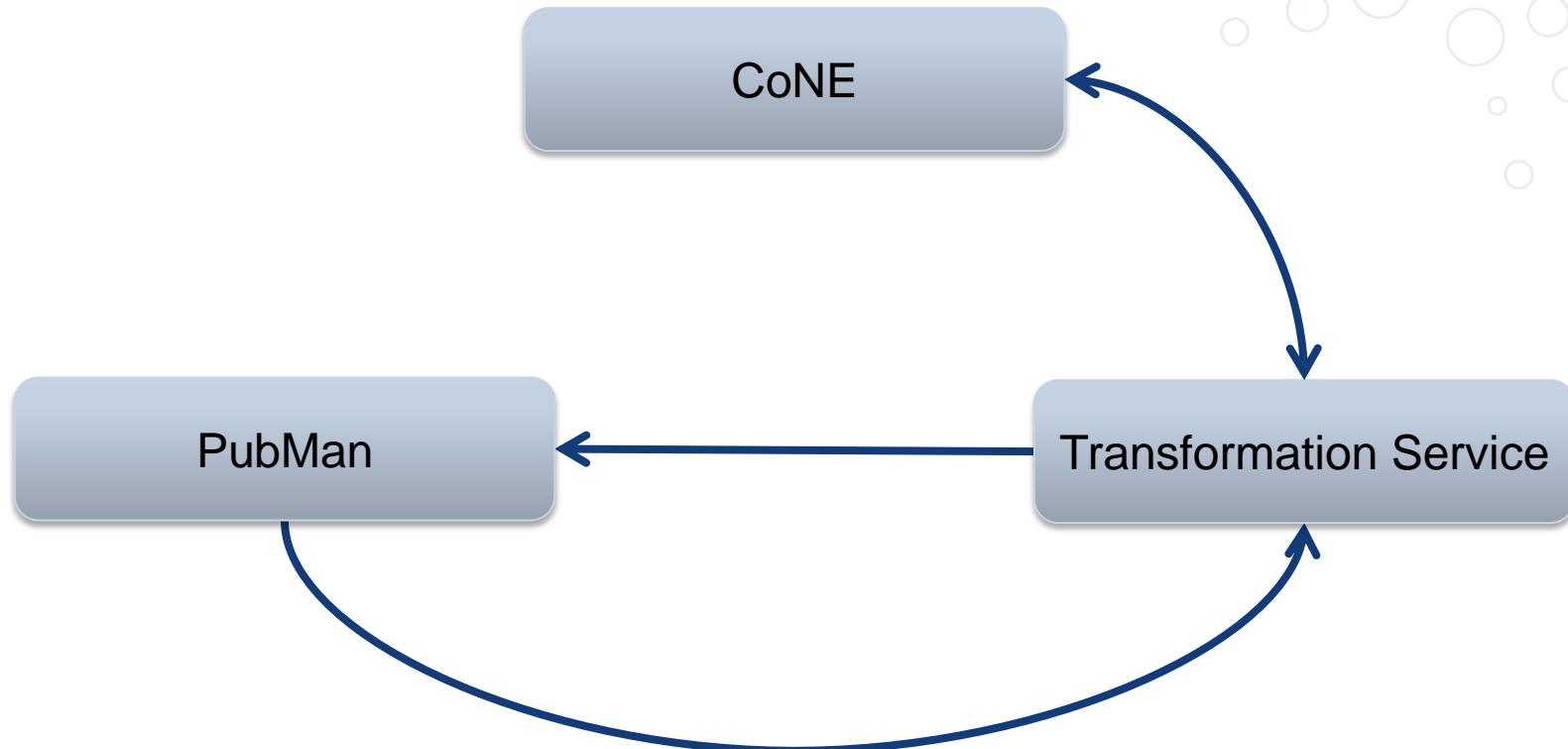
Formate



Voraussetzungen in der MPG

- Größe/Alter der Institute:
 - Zahl der Mitarbeiter
 - Zahl der Publikationen (80 – 20.000)
- „Kulturelle“ Unterschiede der Fachrichtungen:
 - Workflows
 - Publikationsverhalten
- Datenquelle:
 - eDoc, EndNote, BibTex, WoS, Lokale Systeme...
 - Qualität/Konsistenz der Daten (Mapping-Sonderwünsche)
- Wer ist Ansprechpartner?
 - Bibliothek
 - EDV
 - Öffentlichkeitsarbeit

Zusammenspiel



Ziele

!

Autoren in CoNE importieren

Transformation konfigurieren

CoNE-Autoren in Transformation einbinden

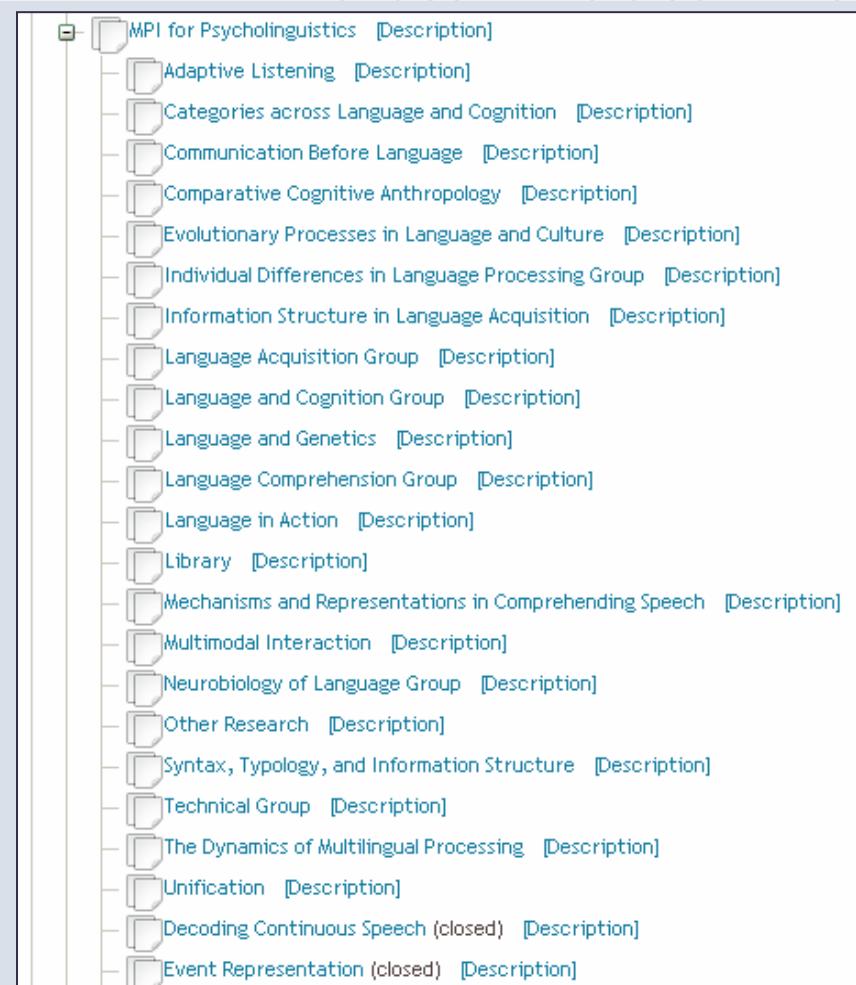
Agenda

- Einführung
- Ziele
 - Autoren in CoNE importieren
 - Transformation konfigurieren
 - CoNE-Autorenabfrage in Transformation einbinden

Voraussetzung: Organisationseinheiten strukturieren



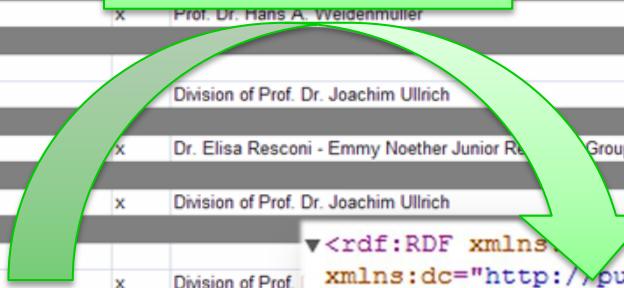
hierarchische Struktur



flache Struktur

Vorgehen

excel2cone.xls



The diagram illustrates the import process from an Excel file to an RDF XML representation. A green arrow points from the Excel file on the left to the RDF XML on the right, indicating the transformation.

	A	B	C	D	E
1	A3			<i>f_x</i>	
2	Nachname	Vorname	Eintr.	Abteilung	Abschluss/Titel
3	Aberle	Christoph	x	Division of Prof. Dr. Manfred Lindner	Dipl. Phys.
4	Abt	Iris			
5	Abul-Magd	A. Y.	x	Prof. Dr. Hans A. Weidenmüller	
6	Achtelik	A.			
7	Achtelik	Andreas		Division of Prof. Dr. Joachim Ullrich	Dipl. Phys.
8	Achterberg	A.	x	Dr. Elisa Resconi - Emmy Noether Junior Research Group	
9	Afaneh	F.	x	Division of Prof. Dr. Joachim Ullrich	
10	Agari	M.			
11	Agari	Michaela	x	Division of Prof.	
12	Agarwal	Jessica	x	Dr. Ralf Srama -	
13	Aharonian	F.			
14	Aharonian	F.A.			
15	Aharonian	F.A.			
16	Aharonian	Felic			
17	Aharonian	Felix A.	x	Division of Prof.	
18	Akemann	Gernot	x	Prof. Dr. Hans A.	
19	Akhmedov	Evgeny	x	Division of Prof.	
20	Akherjanian	A.			
21	Akherjanian	A. G.	x	Division of Prof.	
22	Akram	Uzma	x	Division of Prof.	

```

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
           xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description xmlns:foaf="http://xmlns.com/foaf/0.1/"
                    xmlns:dc="http://purl.org/dc/elements/1.1/"
                    xmlns:escidoc="http://purl.org/escidoc/metadata/terms/0.1/"
                    rdf:about="http://dev-
pubman.mpdl.mpg.de/cone/persons/resource/persons46823">
    <foaf:family_name>Mustermann</foaf:family_name>
    <dc:title>Mustermann, Max</dc:title>
    <foaf:givenname>Max</foaf:givenname>
  <escidoc:position>
    <rdf:Description xmlns:dc="http://purl.org/dc/elements/1.1/"
                      xmlns:eprints="http://purl.org/eprint/terms/">
      <dc:identifier>escidoc:282001</dc:identifier>
      <eprints:affiliatedInstitution>Workshop</eprints:affiliatedInstitution>
    </rdf:Description>
  </escidoc:position>
</rdf:Description>
</rdf:RDF>

```

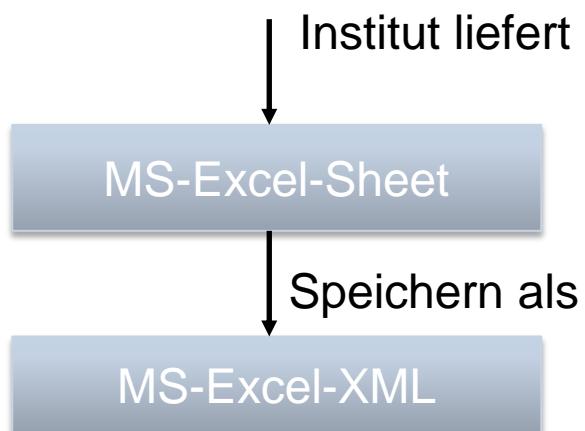
1 Fußnote: Das Format und die Platzierung der Fußnote ist so gewählt, dass man reichlich Platz zur Eingabe hat

Personen in CoNE importieren

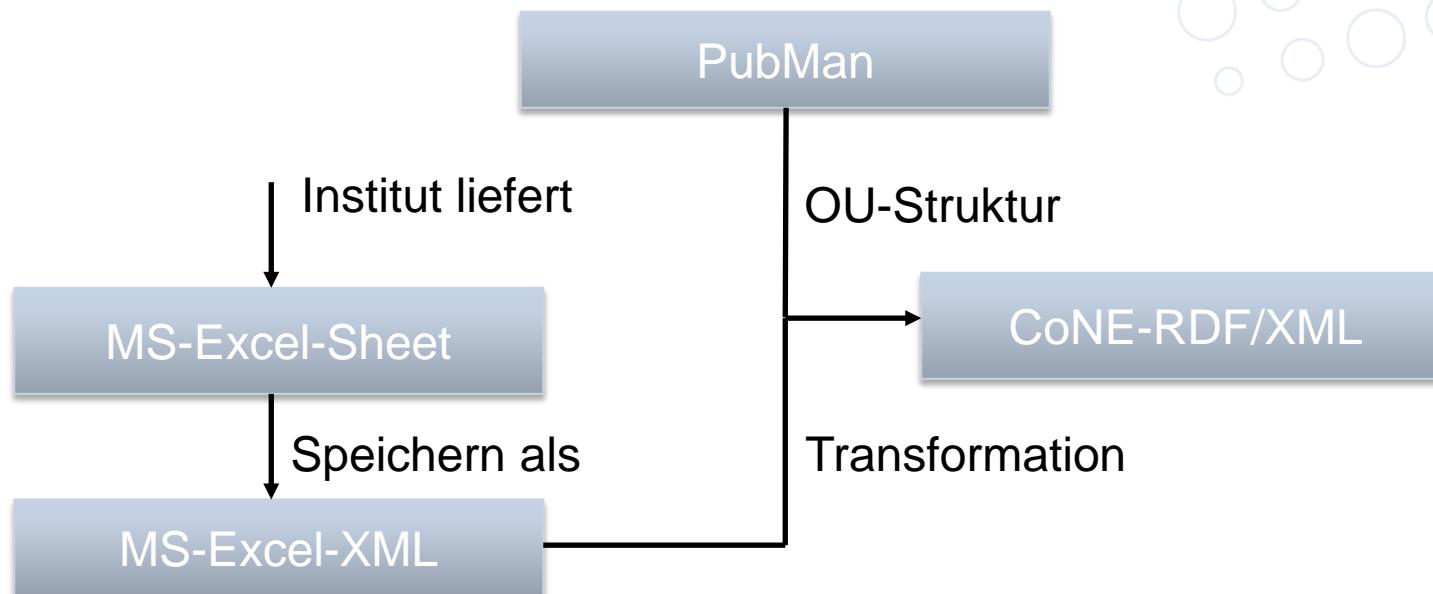
Institut liefert
↓

MS-Excel-Sheet

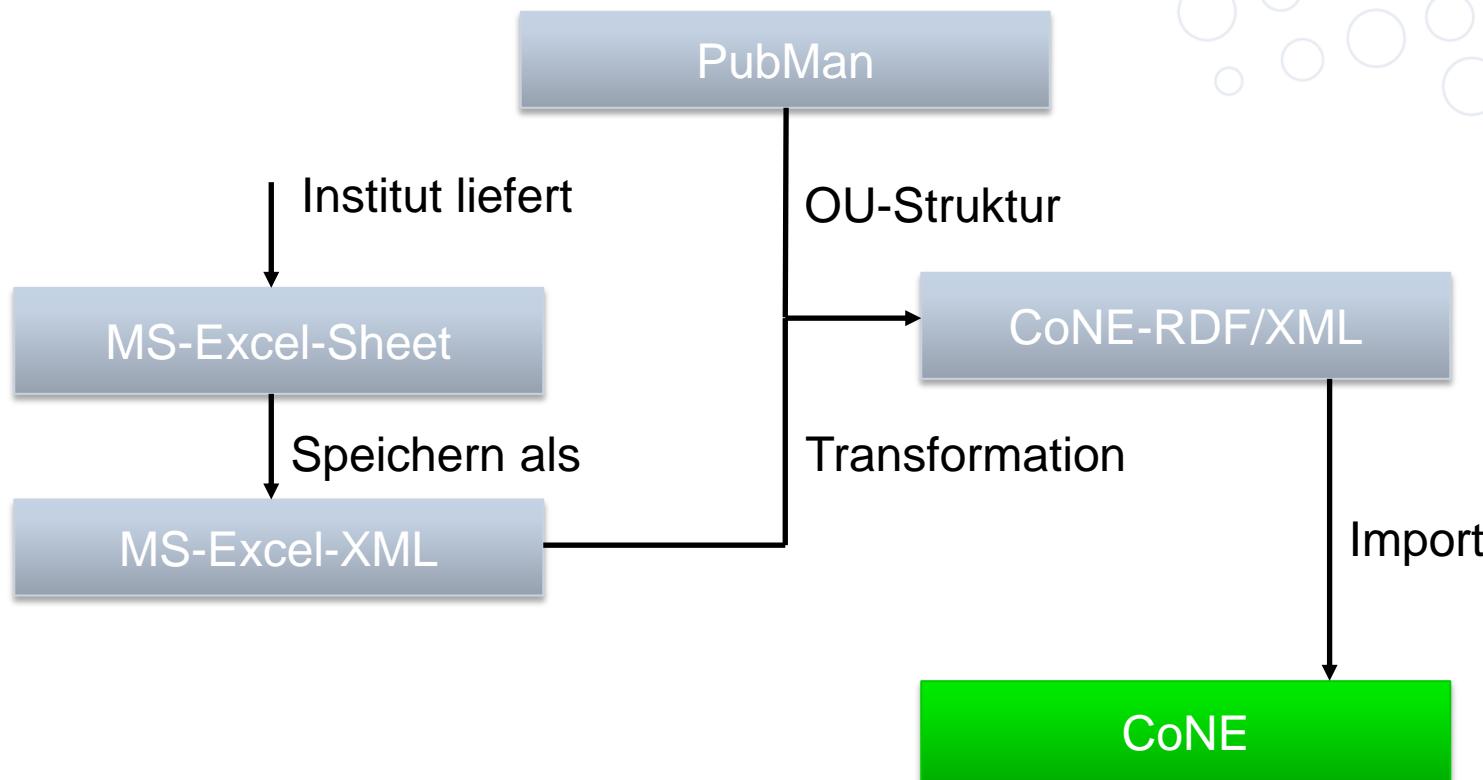
Personen in CoNE importieren



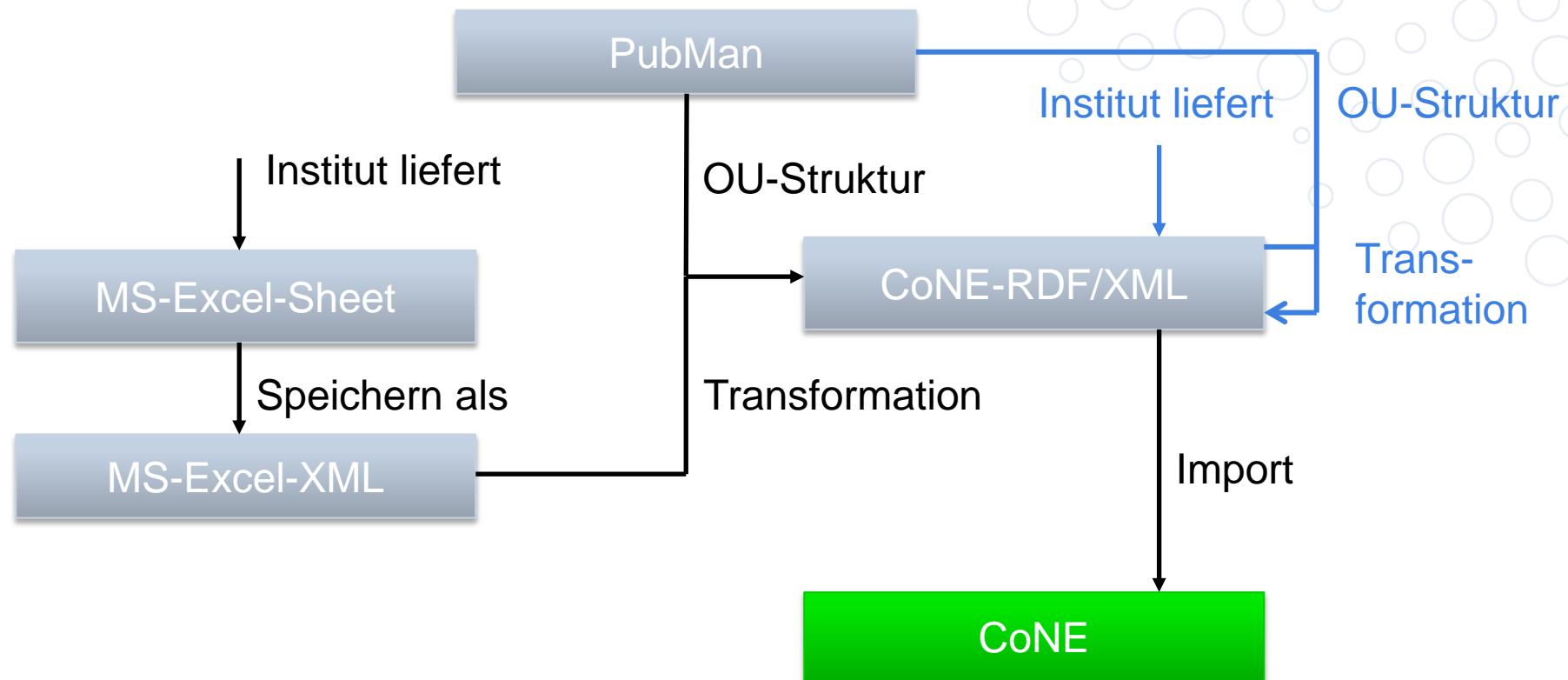
Personen in CoNE importieren



Personen in CoNE importieren



Personen in CoNE importieren



Ergebnis

```
▼<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"  
    xmlns:dc="http://purl.org/dc/elements/1.1/">  
    ▼<rdf:Description xmlns:foaf="http://xmlns.com/foaf/0.1/"  
        xmlns:dc="http://purl.org/dc/elements/1.1/"  
        xmlns:escidoc="http://purl.org/escidoc/metadata/terms/0.1/"  
        rdf:about="http://dev-  
pubman.mpdl.mpg.de/cone/persons/resource/persons46823">  
        <foaf:family_name>Mustermann</foaf:family_name>  
        <dc:title>Mustermann, Max</dc:title>  
        <foaf:givenname>Max</foaf:givenname>  
    ▼<escidoc:position>  
        ▼<rdf:Description xmlns:dc="http://purl.org/dc/elements/1.1/"  
            xmlns:eprints="http://purl.org/eprint/terms/">  
                <dc:identifier>escidoc:282001</dc:identifier>  
                <eprints:affiliatedInstitution>Workshop</eprints:affiliatedInstitution>  
            </rdf:Description>  
        </escidoc:position>  
    </rdf:Description>  
</rdf:RDF>
```

CoNE-ID

OU-ID

Beispiel

1 <http://www.jedit.org/>

Agenda

- Einführung
- Ziele
 - Autoren in CoNE importieren
 - **Transformation konfigurieren**
 - CoNE-Autorenabfrage in Transformation einbinden

Konfigurationsmöglichkeiten für Transformationen



Dieser Datensatz wird ein Teil der PubMan Opened.

Import | endnote_test_item.txt, 92B

Bitte wählen Sie Ihre Importparameter

bei Fehlern zurücksetzen

nicht nach Dubletten prüfen

Gefundene Dubletten nicht importieren

beim Finden von Dubletten nichts importieren

Name des Imports

Import Konfiguration | Flavor: CoNE OTHER: true

Abbrechen Importieren

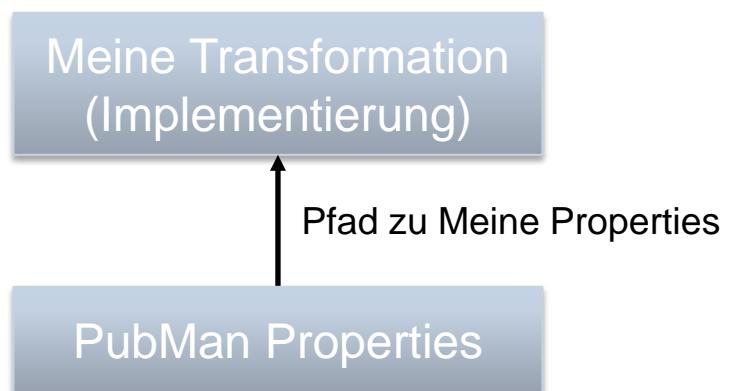
Allgemeine Importeinstellungen

Importspezifische Konfigurationsmöglichkeiten (z.B. Parameter für die Transformationen)

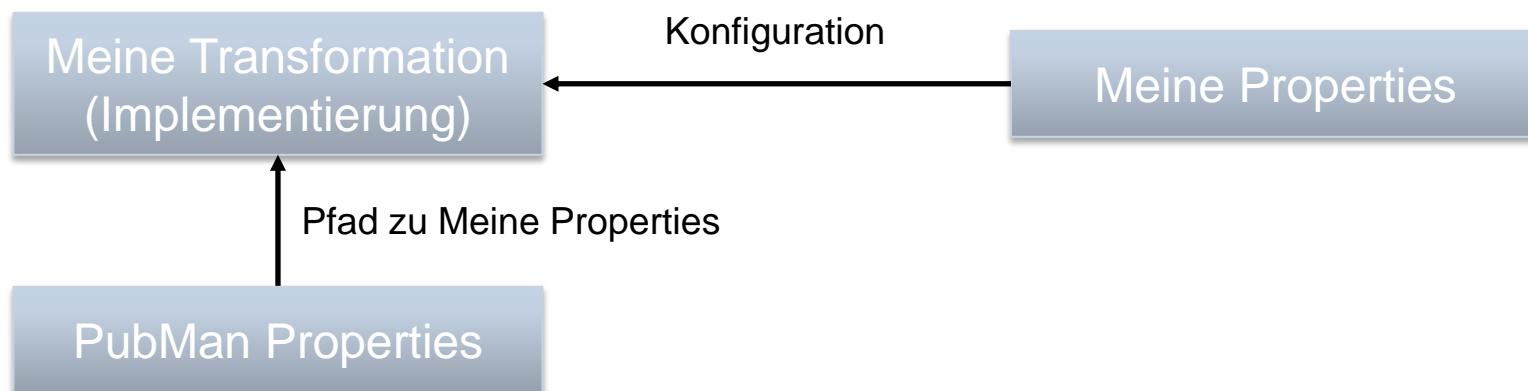
Struktur der Transformationsspezifischen Properties

Meine Transformation
(Implementierung)

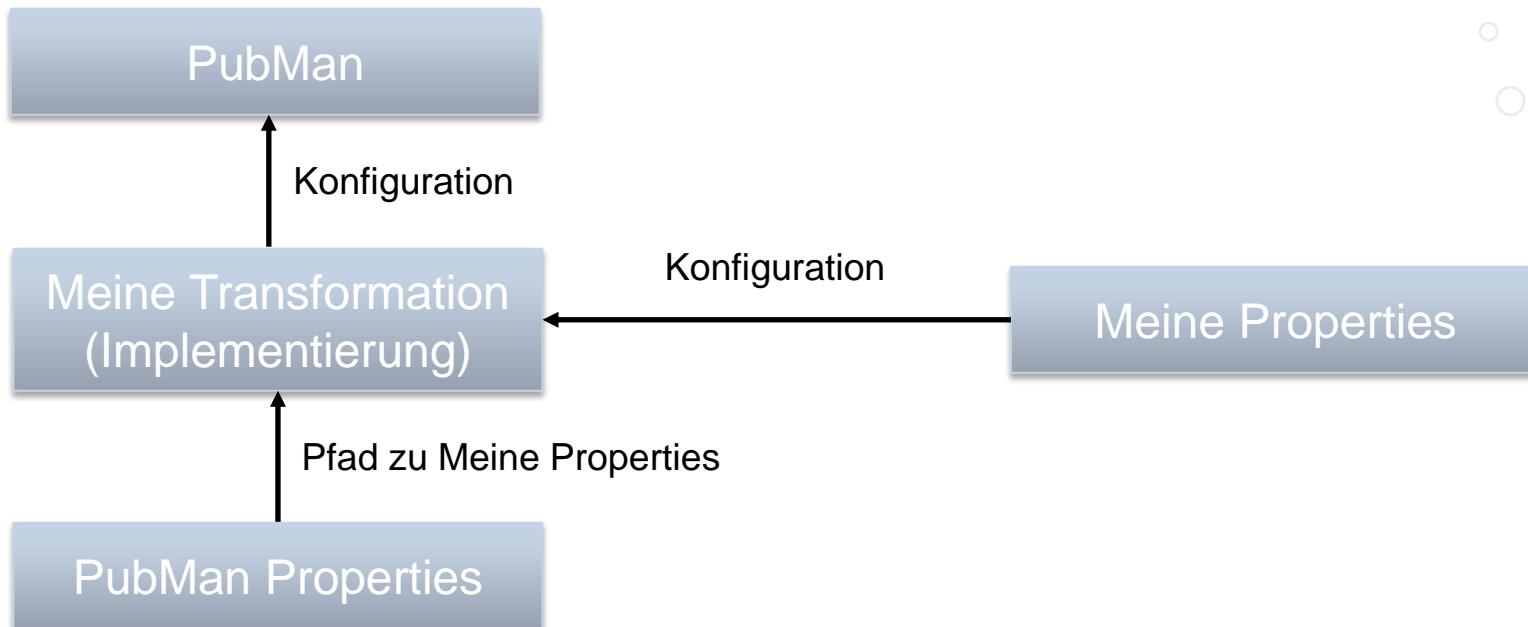
Struktur der Transformationsspezifischen Properties



Struktur der Transformationsspezifischen Properties



Struktur der Transformationsspezifischen Properties

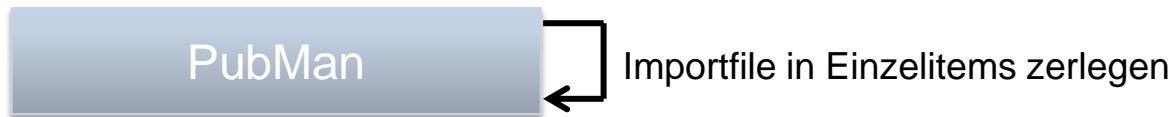


Beispiel

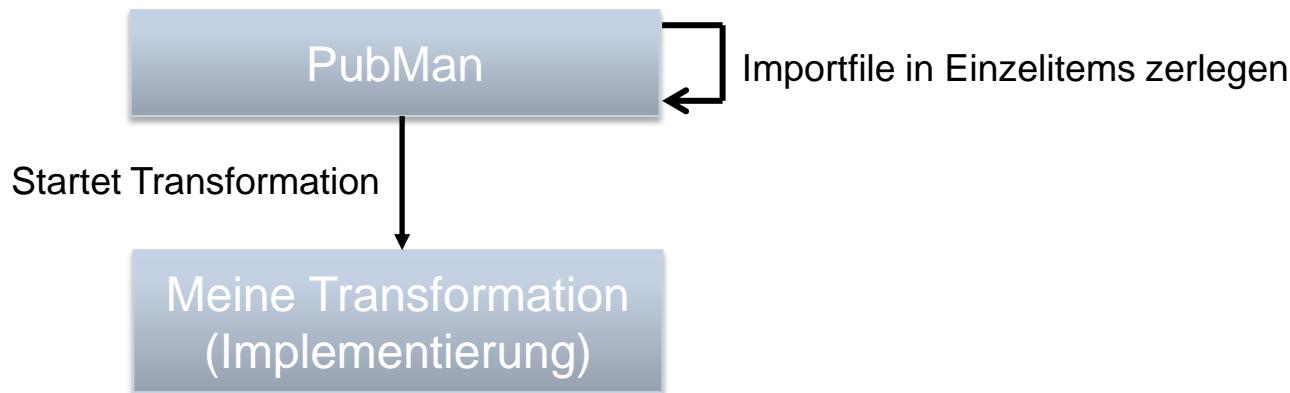
Agenda

- Einführung
- Ziele
 - Autoren in CoNE importieren
 - Transformation konfigurieren
 - CoNE-Autorenabfrage in Transformation einbinden**

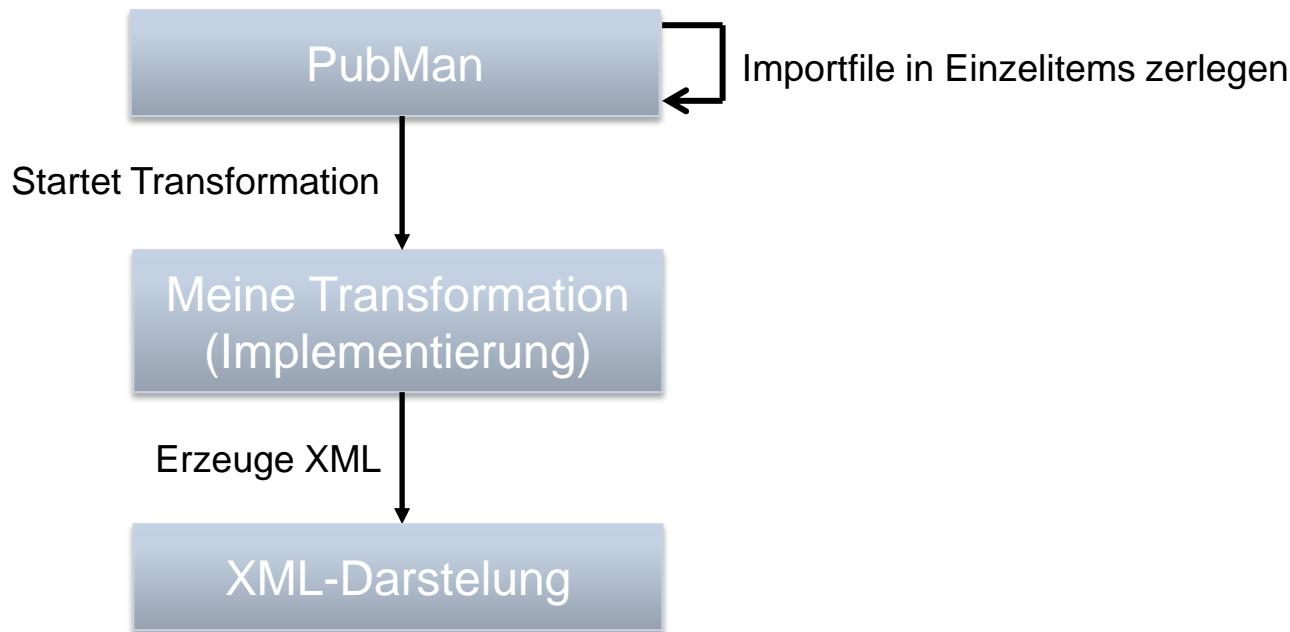
Struktur der CoNE-Autorenabfrage



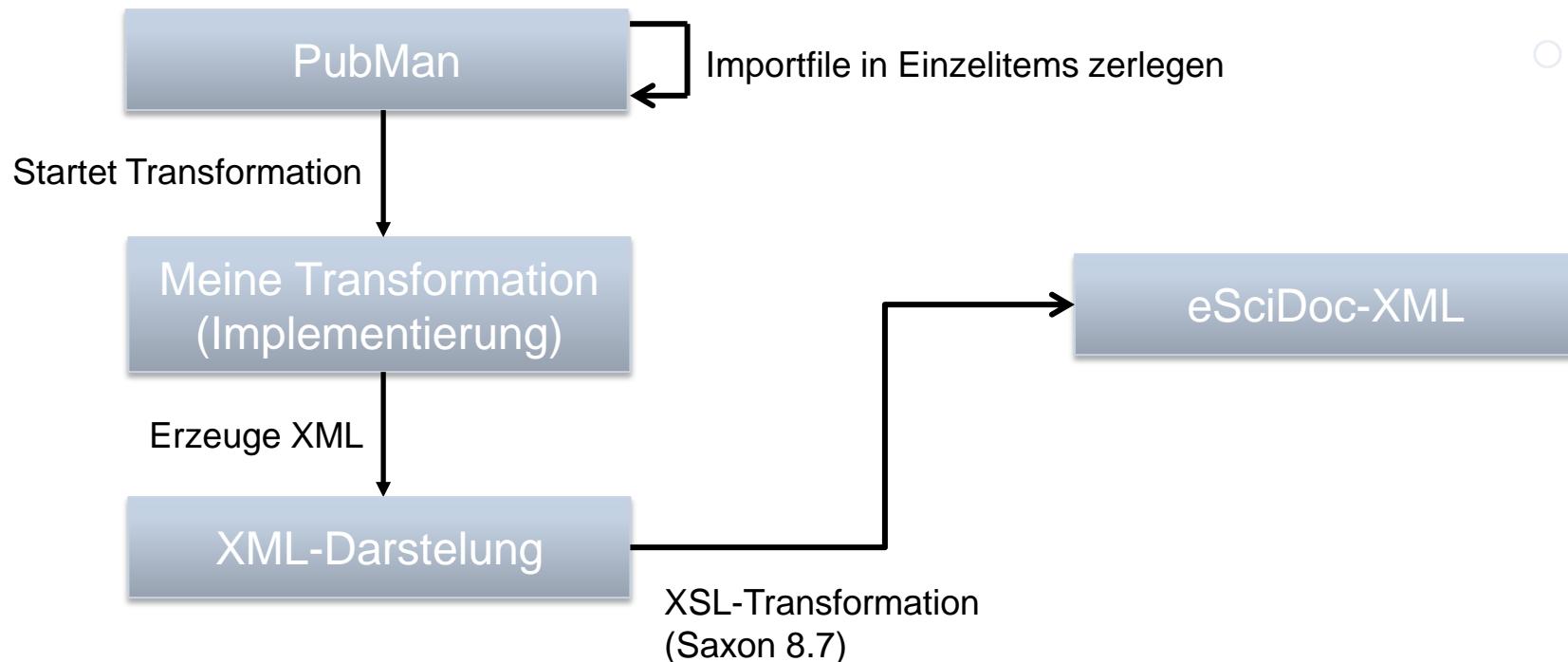
Struktur der CoNE-Autorenabfrage



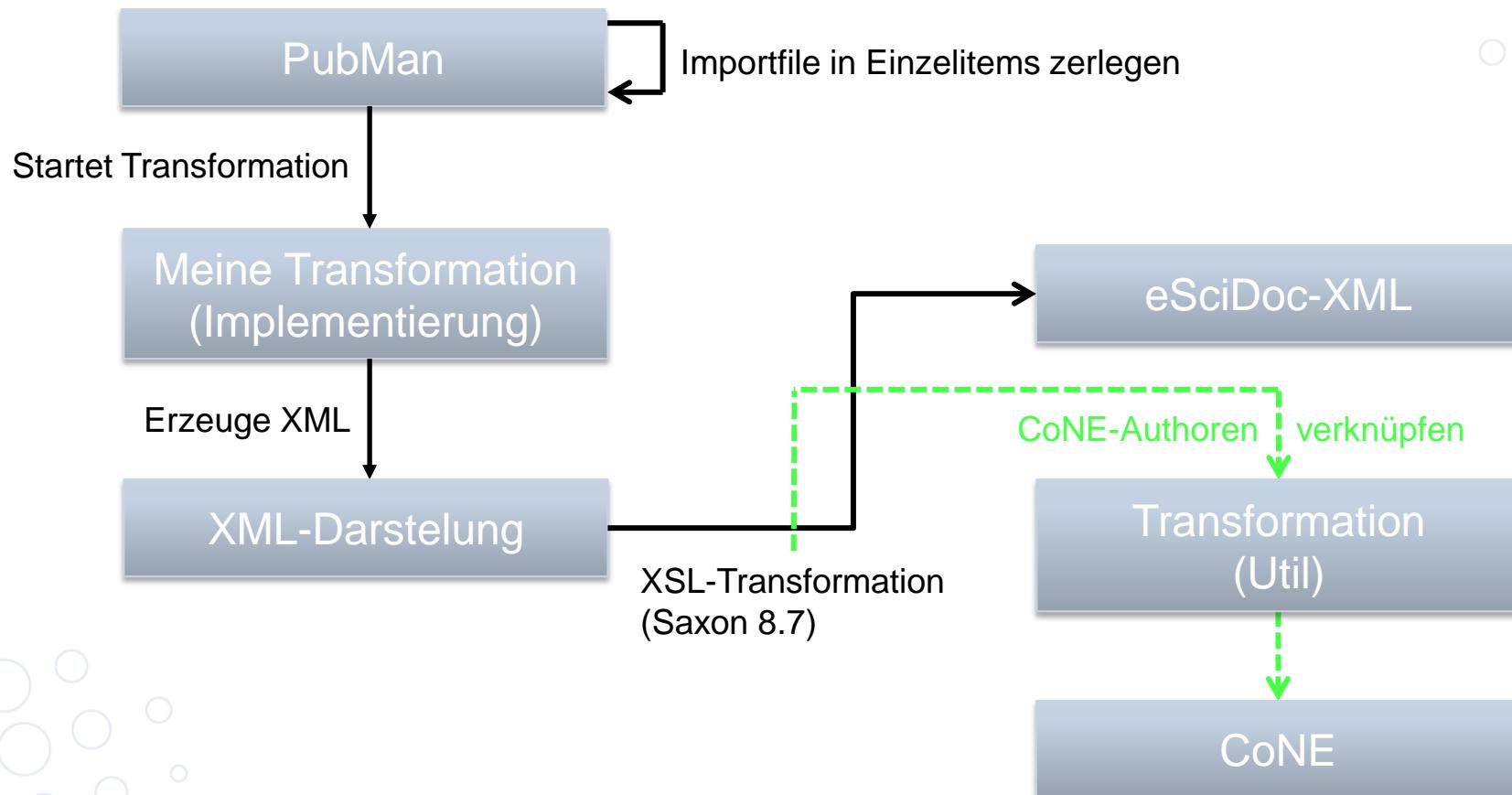
Struktur der CoNE-Autorenabfrage



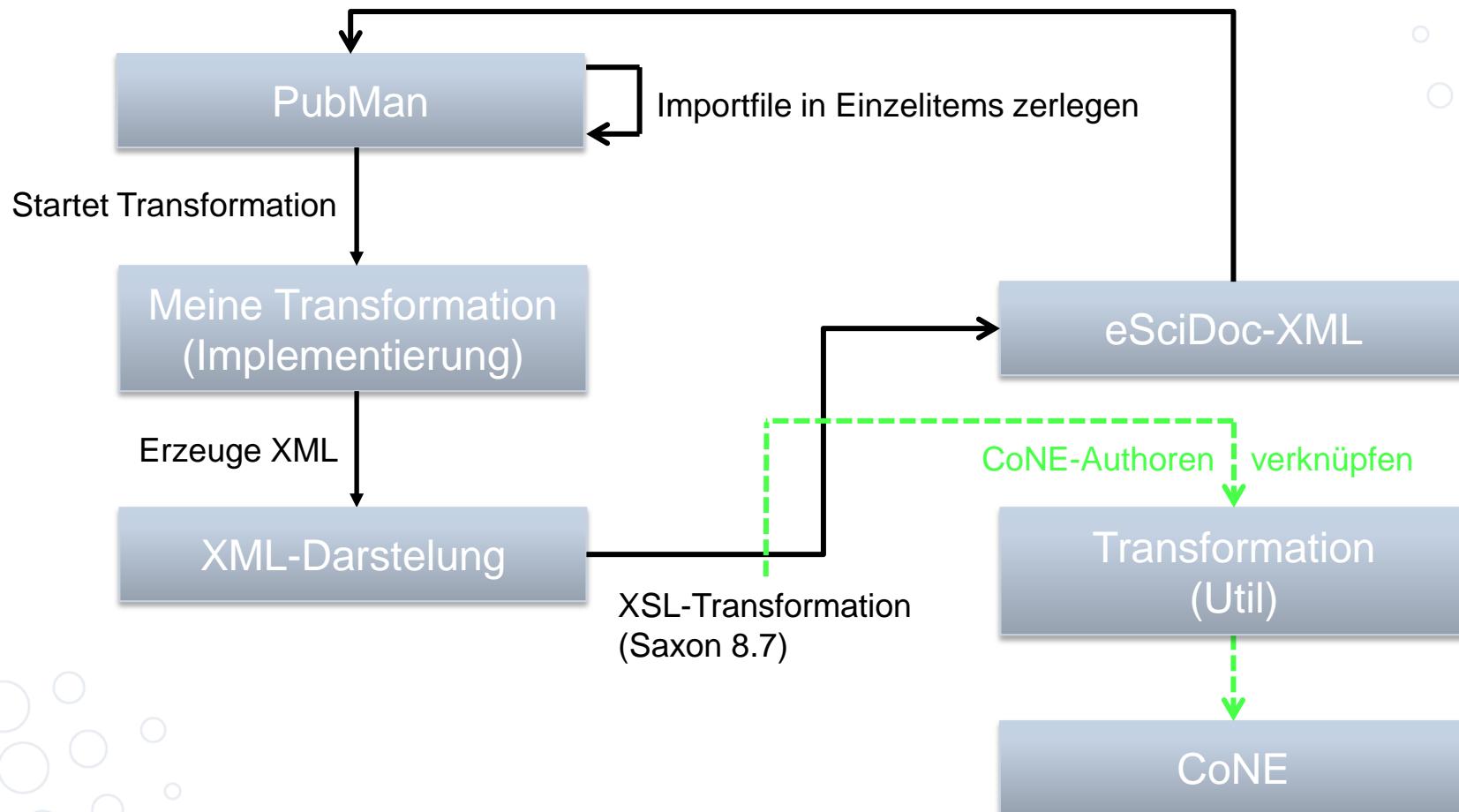
Struktur der CoNE-Autorenabfrage



Struktur der CoNE-Autorenabfrage



Struktur der CoNE-Autorenabfrage



Beispiel

Vielen Dank für Ihre
Aufmerksamkeit