

PubMan – Anforderungen und Ideen



Marion Schmidt

Max-Planck-Institut
für Kognitions- und Neurowissenschaften Leipzig

Dublettencheck #1

Drei Konstellationen:

- eine WissenschaftlerIn möchte einen Artikel eingeben, den ihre Ko-AutorIn aus dem Institut schon eingegeben hat
- eine WissenschaftlerIn/Sekretärin aus Institut *a* möchte einen Artikel eingeben, den die Ko-AutorIn aus Institut *b* schon eingegeben hat
- wir möchten mit Importen aus dem Web of Science den MitarbeiterInnen und Sekretärinnen Schreibaufwand ersparen, gleichzeitig sollen keine Dubletten von bereits eingegebenen Arbeiten entstehen

Dublettencheck #2

Situation zur Zeit

- WissenschaftlerInnen geben ihre Publikationen in ein Intranet-System selbst ein, Sekretärinnen erledigen dies meist für höherrangige (Senior Researchers, DirektorInnen)
- vielen WissenschaftlerInnen ist die Pflege Ihrer Publikationen auf den Staff Pages sehr wichtig und sie geben schon Arbeiten im *Submitted*- und *In Press*-Stadium ein, die auch als solche angezeigt werden
- andere kümmern sich überhaupt nicht darum
- vor Fachbeiratsevaluationen gibt es Deadlines zur Eingabe von Publikationen für den Forschungsbericht, Sekretärinnen geben dann einige Wochen lang im Akkord Publikationen ein
- ein gemischtes Team aus Sekretärinnen und Bibliotheksmitarbeiterinnen überprüft und korrigiert die Metadaten anhand des WoS und von Bibliotheksmeta- und Buchhandelskatalogen; und überprüft die Formatierung nach APA

Dublettencheck #3

Ideal wäre eine Prozedur, mit der man entweder

- vorhandene Arbeiten durch ein automatisches Check-Up auf Titel-/Doctype-/Personenebene mit dem WoS abgleichen kann und die PubMan-Daten mit WoS-Metadaten überschrieben werden incl. der WoS-ID
- für den Rest recherchiert und importiert man WoS-Daten, Dublettencheck geschieht auf der Basis der WoS-IDs

... oder

- es werden WoS-Daten importiert, es findet eine Dublettenkontrolle auf Titel-/Doctype-/Personenebene statt und Metadaten werden überschrieben

(oder beides)

Dublettencheck #4

Für beide Lösungen gilt:

- entweder jeweils mit Dialog (ist das der gleiche Datensatz? Soll alles/sollen einzelne Felder überschrieben werden?)
- oder mit einem Log-Protokoll, anhand dessen man im Nachhinein checken kann, was überschrieben wurde (WoS-Daten enthalten selten (Tipp-)Fehler o.ä., die eingegeben Daten der Wissenschaftler sind dagegen vor allem relativ häufig unvollständig)
- es sollte voreingestellt oder konfigurierbar sein, dass bestimmte Felder nicht überschrieben werden (d.h. Volltexte sollten durch einen Import **niemals** verschwinden; WoS-Daten enthalten keine DOIs, diese sind aber praktisch zum Verlinken auf den Staff Pages)
- die im WoS durchgehend in Großbuchstaben geschriebenen Journaltitel sollten angepasst werden
- Dubletten aus verschiedenen MPIs können aufgrund der Schreibrechte notwendig sein

Migration aus *Web of Science*

Intracerebral hemorrhage related to systemic gas embolism during hysteroscopy

Less

Creators	Guillard, Emilie; Nancy, Benedicte; Floch, Herve; Henckes, Anne ...
Dates	Created: 2010
Genre	Journal Article
Files	-
Locators	-
Publishing info	-
Event Title	-
Source Genre	Journal
Source Title	UNDERSEA & HYPERBARIC MEDICINE
Volume / Issue	37 (2)
Start / End Page	89 - 93
Sequence Number	-
Source Publ. Info	-

Citation Style: APA

Guillard, E., Nancy, B., Floch, H., Henckes, A., Cochard, G., Arvieux, J., et al. (in preparation).
Intracerebral hemorrhage related to systemic gas embolism during hysteroscopy.

Dublettencheck #5

Was braucht man zum Dublettencheck?

Auf der Ebene PubMan (Fälle 1 und 2):

- Test: eDoc-weite Suche nach Titelwörtern aus etwa zwei Dutzend Artikeln unseres letzten Forschungsberichtes: i.d.R. reichen drei Wörter aus dem Titel aus, in einigen Fällen vier; ich habe keinen Fall gefunden, in dem eine eindeutige Identifizierung auf der Basis des Titels nicht möglich gewesen wäre

Auf der Ebene PubMan – WoS:

- Test: WoS-Suche nach etwa einem Dutzend Artikeln: die Beschränkung auf die Titelebene scheint auch hier oft auszureichend zu sein (eindeutige Identifizierung hat immer geklappt)
- allerdings können Artikel genauso heißen wie (in eDoc nicht vorhandene) Meeting Abstracts (oder Talks), deshalb sollte der Dokumenttyp wohl auf jeden Fall auch abgefragt werden

Metadatenextraktion aus PDFs

- muss nicht 100%ig korrekt funktionieren; es wäre sehr positiv, wenn ein Großteil der Schreibarbeit abgenommen würde und der Eingebende nur ein paar Details korrigieren muss
- funktioniert ganz gut in Mendeley

Management der eigenen Daten

- es sollten im Workspace eines Wissenschaftlers tatsächlich alle seine Arbeiten auftauchen, nicht nur diejenigen, die er selbst eingegeben hat
- diese Funktion wäre zentral für das fehler- und unfallfreie Managen der eigenen Publikationen auf den Staff Pages
- die Definition von einzelnen Normdaten für *Local Tags* wäre sinnvoll, insofern diese für das *Feeding* der Staff Pages relevant sind, da man so sicherstellen könnte, dass tatsächlich alle Mitarbeiter das gleiche Tag z.B. für *Selected Publications* auswählen

Usability #1

Eingabe

mehr Genrespezifität in der *Full*, z.T. auch in der *Easy Submission* wäre sinnvoll

- Konfigurierbarkeit von Dokumenttypen?
- zu viele Datumsangaben: wozu braucht man *Date Modified*? Statt *Date Submitted* und *Date Accepted* wären Klickboxen ausreichend (der Status und nicht das Datum ist die relevante Information); *Date Created* allenfalls für unpublizierte *Working Papers* relevant
- z.B.: Poster und Teachings haben eigentlich keine Publikationsdaten, ein Book Chapter kann oder sollte eigentlich keine Quelle „Journal“ haben
- z.B.: Für Artikel werden meines Wissens in Zitationsstilen nie Herausgeber, Verlag, Verlagsort und Gesamtseitenanzahl des Journals angegeben

Usability #2

Genrespezifität

The screenshot shows a web form for entering source details. The form is titled 'Source1' and has a 'hide details' button. The fields are organized into sections:

- Genre:** A dropdown menu with 'Journal' selected.
- Title:** A text input field containing 'Journal of cognitive Neuroscience'.
- Creator role:** A dropdown menu with '-' selected.
- Type:** Radio buttons for 'Person' (selected) and 'Organization'.
- Family name:** A text input field.
- Given name:** A text input field.
- Organization Name and Address:** A text input field with a 'select' button below it.
- Details of Source:**
 - Volume:** A text input field containing '18'.
 - Total no. of pages:** A text input field.
 - Publisher:** A text input field.
 - Place:** A text input field.
 - Identifier:** A dropdown menu with '-' selected and a text input field.
 - Issue:** A text input field containing '11'.
 - Start Page:** A text input field containing '1924'.
 - Endpage:** A text input field containing '1936'.
 - Sequence Number:** A text input field.

At the bottom, there is a 'hide Source1' button.

... für Zitation (nach APA): *Journal of Cognitive Neuroscience*, 18(11), 1924-1936

Usability #3

Eingabe

- rote Markierung von unvollständig ausgefüllten Pflichtfeldern wäre nicht schlecht
- Fehlermeldungen des Systems sind manchmal nicht ganz stimmig
- ein nicht-belegtes Autorfeld sollte einfach mitabgespeichert werden können und nicht minimiert werden müssen

Suche

- Fuzzy Search oder *Did you mean?*

Dokumenttyp Meeting Abstracts

- Abstracts von Vorträgen werden in Sonder-Issues von Zeitschriften oder in Proceedingsbänden veröffentlicht; Conference Papers sind dagegen längere Arbeiten von ca. 6-12 Seiten, in der Regel mit Peer Review. Im Unterschied zu Talks sind die Abstracts „publiziert“ und können nachgelesen werden
- Metadaten Zeitschriftensuppl.: Autor, Titel, Titel Zeitschrift, Jahr, Volume, Issue, Seiten, URL, DOI
- Metadaten Proceedingsband: Autor, Titel, Titel Proceedings, Jahr, Verlag, Verlagsort
- denkbar wäre auch eine Klick-Option, mit der ein Artikel zu einem Meeting Abstract gemacht werden kann (allerdings hat ein Meeting Abstract in einem Proc. wenig mit einem Artikel gemein und könnte entsprechend unintuitiv sein; sinnvoll für Artikel-Subkategorien wie Letter, Review etc.)

Ansetzung Personen

- keine Felder für Präfixe (von, van, de usw.), Mittelnamen und zweite Vornamen
- für die Suche nicht essentiell, aber kein Garant für hochwertige Metadaten insb. von MPG-externen Personen außerhalb des CoNE-Services

Cramon, D.Yves von
Cramon, D. Y. von
von Cramon, Detlef Yves
von Cramon, Detlev Y.
von Cramon, D. Y.
von Cramon, D.Yves
von Cramon, D. Yves
Von Cramon, D. Yves
von Cramon, D. Y.ves
von Cramon, Yves D.

Normdaten-Mapping

Begrenzung des Mappings auf Personennamen wäre sinnvoll, am besten auf den Wortanfang

*Creator role

*Type Person Organization

*Family name Given name

Antonchick, Andrey (Research Group Biosynthesis / NMR, MPI for Chemical Ecology)
Antonchick, Andrey (Research Group Biosynthesis / NMR, MPI for Chemical Ecology)
Baranovsky, Alexander (Research Group Biosynthesis / NMR, MPI for Chemical Ecology)
Baranovsky, Alexander (Research Group Biosynthesis / NMR, MPI for Chemical Ecology)
Brand, Silke (Research Group Biosynthesis / NMR, MPI for Chemical Ecology)
Brand, Silke (Research Group Biosynthesis / NMR, MPI for Chemical Ecology)
Daolio, Cristina (Research Group Biosynthesis / NMR, MPI for Chemical Ecology)

Kontakt

mschmidt@cbs.mpg.de