

## Messen Zitate, Downloads und Altmetrics (am Beispiel von “Leserhäufigkeiten” bei Mendeley) dasselbe?

Christian Schlögl  
Institut für Informationswissenschaft  
und Wirtschaftsinformatik  
Karl-Franzens-Universität Graz

### Projektteam



Juan **Gorraiz**  
Team Bibliometrie, Universität Wien



Christian **Gumpenberger**  
Team Bibliometrie, Universität Wien



Kris **Jack**  
Mendeley, London (UK)



Peter **Kraker**  
Know-Center, Graz



Christian **Schlögl**  
Institut für Informationswissenschaft und Wirtschaftsinformatik  
Karl-Franzens-Universität Graz

## Danksagung

This paper is partly based on anonymous ScienceDirect usage data and Scopus citation data kindly provided by Elsevier within the framework of the Elsevier Bibliometric Research Program (EBRP).

## Inhalt

1. Einführung
2. Forschungsfragen und Datenquellen
3. Vorgehensweise
4. Ergebnisse
  - Downloads
  - Zitate
  - Mendeley „Leserhäufigkeiten“
  - Gemeinsamkeiten/Unterschiede
5. Schlussfolgerungen

## EINFÜHRUNG

- Es gibt schon einige Studien, die Downloads mit Zitaten verglichen haben
- Quellen für Download-Analysen:
  - Repositories/Preprint-Archive: z. B. Chu and Krichel (2007) - RepEc, Brody et al. (2006) - arxiv
  - einzelne e-Journals: Moed (2005), Coats (2005)
  - Volltextdatenbanken (z. B. ScienceDirect): z. B. Schlögl & Gorraiz (2010), Schloegl & Gorraiz (2011)
- Kollaborative Literaturverwaltungssysteme (z. B. Mendeley) sind eine weitere vielversprechende Quelle
- Erste Studien, die “Lesehäufigkeiten” mit Zitationshäufigkeiten verglichen haben (Bar-Ilan 2012, Li and Thelwall 2012, Kraker et al. 2012, Schlögl et al. 2013, Gorraiz et al. 2013, Haustein et al. 2014)
- Gegenstand der vorliegenden Studie: Zitate vs. Downloads vs. “Leser”

## FORSCHUNGSFRAGEN

1. Werden die am öftesten zitierten Artikel auch am häufigsten heruntergeladen bzw. am häufigsten von Mendeley-Nutzern “gelesen”?  
 “**Leserhäufigkeit**” (“readership”) eines Artikels: Wie viele Nutzer haben einen bestimmten Artikel in ihre **persönliche Nutzerbibliothek hinzugefügt!**
2. Entwickeln sich Zitate, Downloads und Leserhäufigkeiten (von Dokumenten) im Zeitablauf ähnlich?
3. Gibt es andere Aspekte, bei denen sich Zitate, Downloads und Leserhäufigkeiten unterscheiden?
4. Gibt es möglicherweise disziplinspezifische Unterschiede (am Beispiel einer Linguistik- und einer Wirtschaftsinformatikzeitschrift) in Hinblick auf FF 1 - FF 3?

## ANALYSE-OBJEKT

### Journal of Phonetics :

- “covers phonetic aspects of language and linguistic communication processes.” (Elsevier 2014)
- Inhalte:
  - Lautbildung
  - Lautwahrnehmung
  - Sprachsynthese
  - Automatische Sprach- und Sprechererkennung
  - Spracherwerb
  - usw.
- 4 Hefte pro Jahr
- “Peer reviewed”
- Starke anglo-amerikanische Dominanz: 75% der AutorInnen, USA: 50 %

## DATENQUELLEN

- Datenquellen:
  - ScienceDirect (SD): monatliche Artikel-Downloads (PDF & HTML)
  - Scopus: monatliche Artikelzitate
  - Mendeley: monatliche „Leserhäufigkeiten“ – „von wie vielen Nutzern wurde ein Artikel (nur FLA) im jeweiligen Monat in die entsprechenden Nutzerbibliotheken hinzugefügt“
- Analysezeitraum: 2002 – 2011
- Analysierte Dokumente: 395 (ScienceDirect)

## DATENQUELLEN: Mendeley

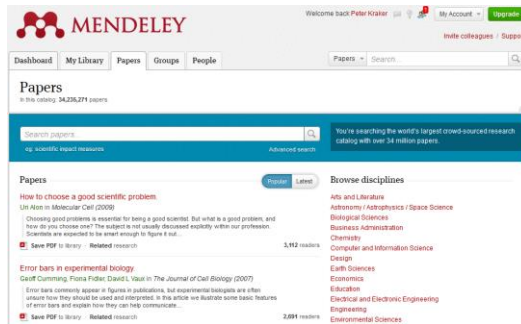
### Social reference management system: kollaboratives Literaturverwaltungssystem

- jeder Nutzer kann seine Literatur in einer persönlichen Nutzerbibliothek verwalten
- Nutzer kann auch ein persönliches Nutzerprofil anlegen

### Crowdsourced Mendeley research catalog:

- aus den einzelnen Nutzerbibliotheken generiert
- > 2.5 Mio. Nutzer
- > 110 Mio. (eindeutige) Dokumente
- "Leserhäufigkeiten" eines Dokuments: wie viele Nutzer haben ein bestimmtes Dokument ihrer Nutzerbibliothek hinzugefügt

<http://www.mendeley.com/research-papers/>



## VORGEHENSWEISE

- Vorverarbeitung:
  - Vergleich des gleichen Dokuments in ScienceDirect (SD) und Scopus
    - Kein durchgängiger Schlüssel für Dokumente in for SD **und** Scopus!
    - Verschiedene Dokumenttypen in SD und Scopus
    - Matching: Name der Zeitschrift, Bandnummer, erste Seite des Beitrags
  - Vergleich des gleichen Dokuments (nur FLA) in Scopus und Mendeley : Artikeltitlel
- Auswertung:
  - Deskriptive Statistik: meist Häufigkeitsverteilungen in Bezug auf Dokumenttyp, „Nutzungs“-Jahr für Zitate, Downloads und Leserhäufigkeiten
  - Korrelationen: Download- vs. Zitations- vs. Leserhäufigkeiten

## ERGEBNISSE Downloads: Downloads je Dokumenttyp

	n	% docs	% downloads (DL)	DLs per doc - relations
Announcement	2	0.5%	0.1%	1.8
Book review	1	0.3%	0.1%	1.7
Contents list	2	0.5%	0.1%	1.9
<b>Discussion</b>	9	2.3%	2.7%	<b>8.7</b>
Editorial Board	30	7.6%	1.1%	1.1
<b>Editorial</b>	5	1.3%	1.5%	<b>8.7</b>
Erratum	3	0.8%	0.5%	4.4
<b>Full length article (FLA)</b>	324	82.0%	92.3%	<b>8.2</b>
Index	1	0.3%	0.1%	1.8
Miscellaneous	9	2.3%	0.4%	1.3
Other contents	1	0.3%	0.1%	2.1
Personal report	2	0.5%	0.2%	3.4
Publishers note	3	0.8%	0.1%	1.0
Short communication	2	0.5%	0.3%	4.8
<b>Short survey</b>	1	0.3%	0.3%	<b>9.9</b>
	<b>395</b>	<b>100%</b>	<b>100%</b>	

- FLAs (82%) werden am öftesten heruntergeladen (92% aller Downloads)
- Downloads per Dokument für "discussions", "editorials", "FLAs" und "short surveys" am höchsten

## ERGEBNISSE Downloads je Dokumenttyp „Journal of Strategic Information“ Systems (JoSIS)

Document type	n	% docs	% downloads	Downloads per doc – relations
Announcement	5	1.6%	0.4%	5.9
Book review	4	1.2%	0.3%	5.5
Contents list	29	9.0%	0.4%	1.0
Editorial Board	29	9.0%	0.6%	1.5
Editorial	49	15.3%	3.3%	4.6
Erratum	1	0.3%	0.1%	5.7
<b>Full length article</b>	<b>181</b>	<b>56.4%</b>	<b>94.1%</b>	<b>35.4</b>
Index	12	3.7%	0.2%	1.3
Miscellaneous	9	2.8%	0.2%	1.8
Publishers note	2	0.6%	0.2%	7.0
	321	100%	100%	

Source: ScienceDirect; n=321

- FLAs (56%) wurden ebenfalls mit Abstand am öftesten heruntergeladen (94.1% aller Downloads)
- Downloads je Dokumenttyp von "Editorial Policy" der Zeitschrift und von Disziplin (z. B. Linguistik vs. Wirtschaftsinformatik/Information Systems) abhängig

## ERGEBNISSE Downloads

Entwicklung der Downloads im Zeitablauf (Verhältniszahlen!)

PY	n	Download year										all
		2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	
2002	28	0.2	1.6	1.5	1.3	1.3	1.5	1.5	1.4	1.3	1.0	12.6
2003	29		2.2	<b>3.4</b>	2.4	2.2	2.1	3.1	2.8	2.5	1.9	22.4
2004	21		0.3	2.7	2.6	2.0	2.3	2.8	2.9	2.5	2.1	20.3
2005	20			0.0	<b>3.1</b>	2.5	1.9	2.0	2.2	1.7	1.4	14.9
2006	22				0.6	<b>4.4</b>	4.7	4.1	4.1	3.5	2.9	24.4
2007	29					0.9	<b>5.4</b>	5.1	4.2	3.1	2.7	21.3
2008	35						0.2	<b>6.6</b>	6.3	4.3	3.3	20.7
2009	32							0.3	<b>6.7</b>	5.3	3.0	15.3
2010	51							0.0	0.7	<b>7.8</b>	6.8	15.3
2011	57									0.3	<b>10.4</b>	10.7
all	324	0.2	4.1	7.6	10.1	13.3	18.2	25.5	31.2	32.2	35.6	178.0

- Am meisten Downloads überlicherweise im Veröffentlichungsjahr
- Download half-life<sub>2011</sub> = 2.2 Jahre

## ERGEBNISSE Downloads - JoSIS

Entwicklung der Downloads im Zeitablauf (Verhältniszahlen!)

PY	n	DL-year										all
		2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	
2002	13	1.0	2.3	1.7	1.3	1.2	1.4	2.4	2.8	2.8	2.7	19.6
2003	21	0.0	1.3	2.2	1.0	1.0	0.9	1.5	1.3	1.5	1.1	11.9
2004	17			1.7	2.6	2.1	2.2	2.4	2.7	2.9	2.3	18.9
2005	18				1.7	2.3	1.8	2.0	2.4	2.6	2.2	15.0
2006	14				0.2	2.4	2.1	1.8	2.1	2.0	2.0	12.5
2007	18					0.0	2.7	3.6	3.4	3.5	2.9	16.1
2008	16						0.0	2.9	3.5	3.0	2.4	11.8
2009	14								3.1	4.0	3.1	10.2
2010	21									3.9	4.4	8.3
2011	29									0.3	5.6	5.9
all	181	1.0	3.7	5.6	6.8	8.9	11.1	16.6	21.4	26.4	29.0	130.4

Source: ScienceDirect; FLA only (n=181)

- Downloads in den meisten Fällen **ein Jahr nach** dem Veröffentlichungsdatum
- Download half-life<sub>2011</sub> = 3.5 Jahre (I&M: 5 Jahre)

## ERGEBNISSE Zitate: Zitate je Dokumenttyp

Doc type	n	Uncited	% uncited	Cites	% cites	Cites per doc type
Article	316	74	23%	2331	84%	<b>7.4</b>
Review	17	0	0%	432	16%	<b>25.3</b>
Editorial	5	3	60%	6	0%	1.2
Letter	3	0	0%	15	1%	5.0
Notes	1	1	100%	0	0%	0.0
Erratum	3	3	0%	0	0%	0.0
	<b>345</b>	<b>81</b>	<b>23%</b>	<b>2784</b>	<b>100%</b>	<b>8.1</b>

- verschiedene Dokumenttypen in Scopus und ScienceDirect (FLA ≈ articles + conference papers + reviews)
- “reviews” werden mit Abstand am häufigsten zitiert
- ca. 25% aller Dokumente wurden gar nicht zitiert (primär “editorials”, “notes” und “erratum”)

## ERGEBNISSE Zitate - JoSIS: Zitate je Dokumenttyp

Doc type	no. docs	% uncited	Cites	Cites per doc type
Article	151	15%	2563	<b>14.8</b>
Conference paper	13	69%	8	0.4
Editorial	33	79%	13	0.2
Review	18	6%	383	<b>20.2</b>
All	<b>215</b>	<b>27%</b>	<b>2967</b>	<b>10.9</b>

Source: Scopus; n=215



## ERGEBNISSE Zitate: Entwicklung der Zitate im Zeitablauf

PY	n	Citation year										all
		2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	
2002	28	5	12	<b>44</b>	45	46	53	<b>73</b>	73	72	80	503
2003	37		55	23	<b>60</b>	<b>78</b>	<b>91</b>	64	<b>93</b>	<b>120</b>	<b>107</b>	691
2004	23		4	6	42	48	53	61	51	81	57	403
2005	18			1	7	17	27	28	35	48	32	195
2006	23					12	34	40	57	86	97	326
2007	29						5	41	59	71	58	234
2008	35						1	11	52	67	66	197
2009	32								7	44	74	125
2010	51								1	16	49	66
2011	57										23	23
all	333	5	71	74	154	201	264	318	428	605	643	2763

- nur wenige Publikationen werden bereits im Publikationsjahr zitiert - Zitationsmaximum wird erst mehrere Jahre danach erreicht
- deutliche Unterschiede im Vergleich zu Downloads

## ERGEBNISSE Zitate - JoSIS: Entwicklung der Zitate im Zeitablauf

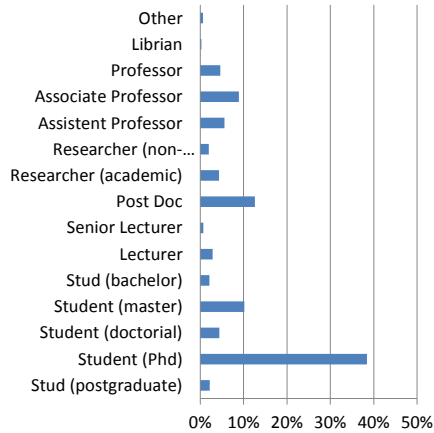
Special Issue on "Trust in the Digital Economy"

Pub year	n	Citation year										all	cites per doc
		2002	2003	2004	2005	2006	2007	2008	2009	2010	2011		
2002	13	2	19	<b>38</b>	<b>69</b>	<b>88</b>	<b>105</b>	<b>158</b>	<b>165</b>	<b>194</b>	<b>199</b>	1037	79.8
2003	14		1	6	21	27	39	35	41	40	39	249	17.8
2004	17			0	15	40	56	74	78	88	<b>107</b>	458	26.9
2005	19				0	16	46	78	76	<b>93</b>	99	408	21.5
2006	14				1	2	14	<b>31</b>	31	53	49	181	12.9
2007	18						1	31	74	92	85	283	15.7
2008	15							3	30	69	83	185	12.3
2009	14								3	34	57	94	6.7
2010	18									5	40	45	2.5
2011	8										14	14	1.8
all	150	2	20	44	106	173	261	410	498	668	772	2954	

Special Issue with conference papers

Source: Scopus; Document types: articles, reviews, conference papers; only cited documents (n=150)

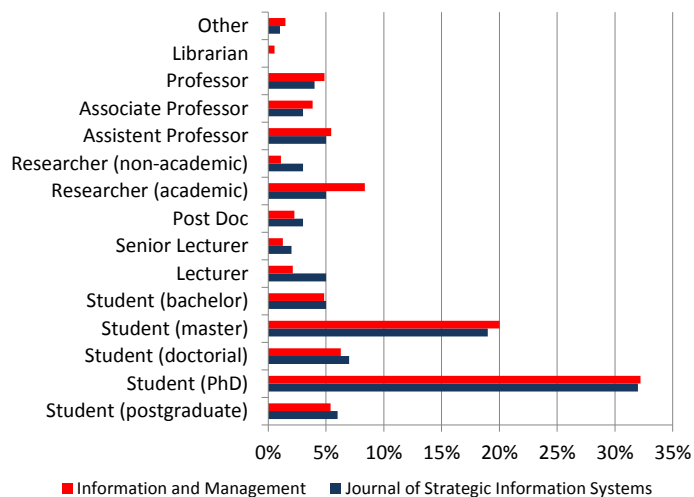
## ERGEBNISSE Mendeley: Nutzerstruktur



Source: Mendeley; doc type: FLA; n=4741

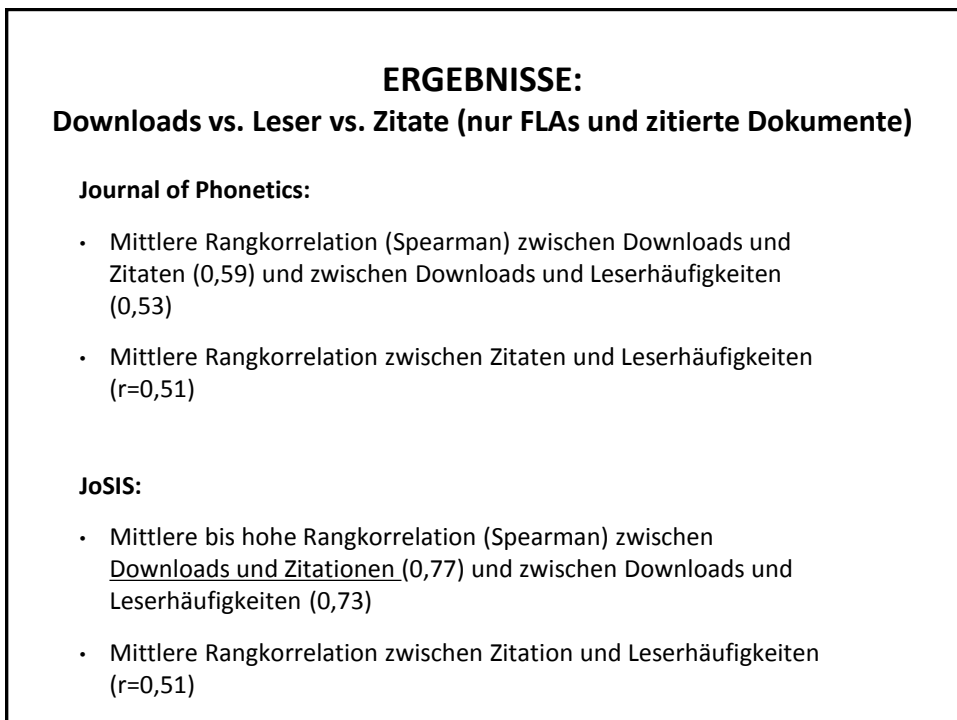
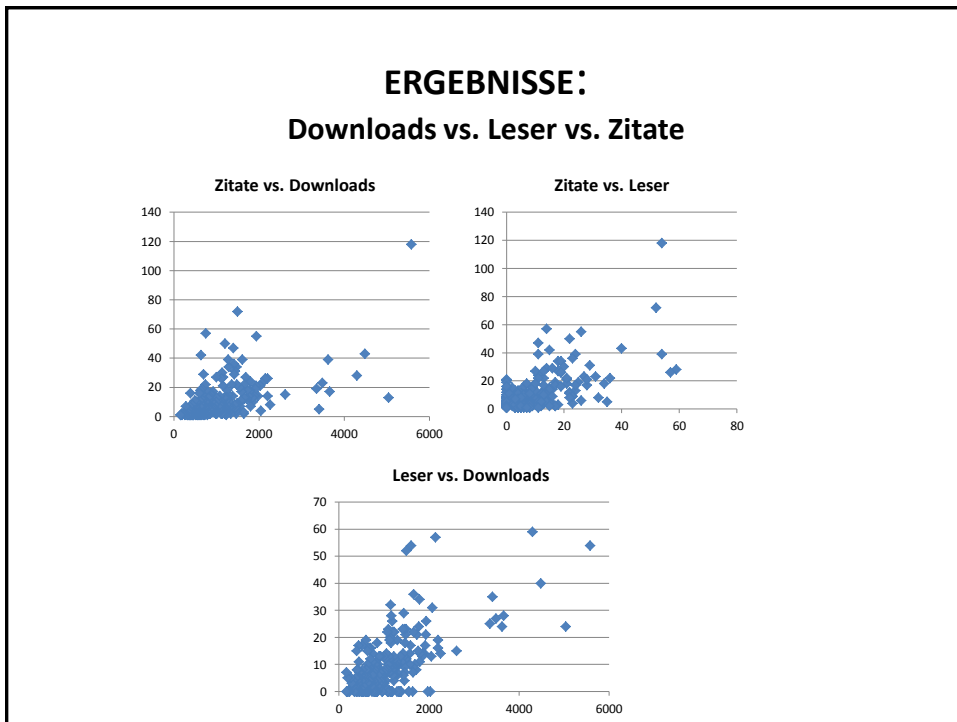
- **75%** aller FLA (Scopus) in Mendeley-Nutzerbibliotheken enthalten
- **57%** der Leserhäufigkeiten von **“Studenten”** generiert
- **13%** durch **PostDocs**, **20%** durch **Professoren**

## ERGEBNISSE Mendeley – JoSIS/I&M: Nutzerstruktur



■ Information and Management ■ Journal of Strategic Information Systems

- **97%/88%** aller FLAs in Mendeley-Nutzerbibliotheken enthalten
- **2/3** der Leserhäufigkeiten durch **Studenten**
- **3%/2%** durch **PostDocs**, **12%/14%** durch **Professoren**



## SCHLUSSFOLGERUNGEN

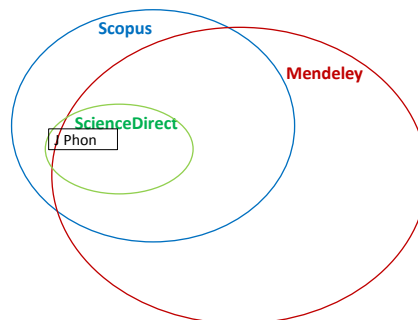
- Vergleiche bei der Verwendung von verschiedenen Datenquellen oft nicht ganz „einfach“ (ScienceDirect vs. Scopus!)

### Hypothesen:

- mittlere Korrelation zwischen Downloads und Zitaten sowie zwischen Leserhäufigkeiten und Downloads bzw. Zitaten (FF 1)
- Downloads, Leserhäufigkeiten und Zitate greifen wissenschaftliche Publikationen unterschiedlich schnell auf (Veralterung) (FF 2)
- disziplinspezifische Unterschiede (FF 3) hinsichtlich
  - Höhe der Zitate, Downloads, Leserhäufigkeiten
  - „Aufnahmegeschwindigkeit“ (Obsolescence) bei Zitaten, Downloads und Leserhäufigkeiten
- Weitere Unterschiede (FF 4) bestehen hinsichtlich
  - Nutzer- und Dokumentpopulationen
  - Höhe der Zitate, Downloads und Leserhäufigkeiten je Dokumenttyp

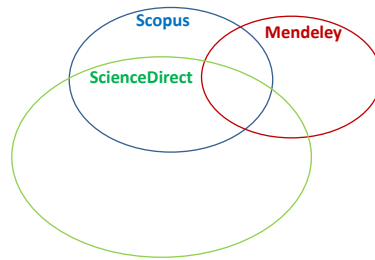
## SCHLUSSFOLGERUNGEN

Unterschiedliche Größen (in Bezug auf enthaltenen Dokumente)



## SCHLUSSFOLGERUNGEN

Unterschiedliche Nutzerpopulationen:



## SCHLUSSFOLGERUNGEN

→ Zitate, Downloads und Leserhäufigkeiten messen unterschiedliche Aspekte der „Zeitschriftennutzung“