

## Szientometrie

Terje Tüür-Fröhlich\*

# Eine „autoritative“ Datenbank auf dem Prüfstand: Der Social Sciences Citation Index (SSCI) und seine Datenqualität

<https://doi.org/10.1515/iwp-2018-0050>

**Zusammenfassung:** Zitatdatenbanken bilden die Datengrundlagen für zahlreiche szientometrische Untersuchungen, Evaluationen wissenschaftlicher Leistungen und Uni-Rankings. In der Literatur finden sich kaum Hinweise auf endogene Fehler (Original richtig, Datenbankeintrag falsch) in den kostenpflichtigen Datenbanken. Banale Fehler (z. B. Falschschreibung der Namen von Autorinnen oder Autoren) in Datenbanken hätten nur geringe Relevanz. Die Fehlersuche zu Pierre Bourdieu als „cited author“ im SSCI (Vergleich Original – SSCI-Record) ergab mehr als 85 Mutationen. Die Fallstudien zeigen eine hohe Anzahl endogener Datenbankfehler. In den Rechtswissenschaften übliche Referenzen in Fußnoten laufen große Gefahr, in Phantomreferenzen verwandelt zu werden (Fallstudie Harvard Law Review: 99 Prozent Fehler). Dem Anspruch des SSCI, die „relevanten“ globalen Sozialwissenschaften abzubilden – für alle im SSCI erfassten Disziplinen –, stehen offenbar Mängel in Datenerfassung und -verarbeitung im Wege.

**Deskriptoren:** Empirische Untersuchung, Fehler, Qualität, Datenbank, Impact Factor, Zitierhäufigkeit, Kritik, Sozialwissenschaften, Social Science Citation Index

**Abstract:** Citation indexes provide the data for many scientometric studies, evaluations of scientific advancements and university rankings. In the literature, there are hardly any indications of endogenous errors (i.e. original references are correct, data base entry contains errors or is faulty) in the licensed databases. Trivial mistakes (e.g. misspelling of author names) in databases are counted of little relevance. Debugging Pierre Bourdieu as a cited aut-

---

**Anmerkung:** Dieser Beitrag beruht auf der „österreichischen“ Kurzversion Tüür-Fröhlich 2017.

---

\*Kontaktperson: Dr.in phil. Terje Tüür-Fröhlich, MSc., Institut für Philosophie und Wissenschaftstheorie, Johannes Kepler Universität Linz, Altenberger Straße 69, 4040 Linz, Österreich, E-Mail: [terje.tuur@jku.at](mailto:terje.tuur@jku.at)

hor in the SSCI revealed more than 85 mutations. The case studies show a high number of endogenous database errors. Footnotes, common in jurisprudence, run the risk of being turned into phantom references as was the case for an article in Harvard Law Review case study: 99 % of error rate. The claim of SSCI to map the “relevant” global social sciences – for all disciplines covered in the SSCI – appears to be hampered by deficiencies in data collection and processing.

**Deskriptoren:** Empirische Untersuchung, Fehler, Qualität, Datenbank, Impact Factor, Zitierhäufigkeit, Kritik, Sozialwissenschaften, Social Science Citation Index

**Résumé:** Les bases de données de citations constituent la base des data pour de nombreuses analyses scientométriques, d'évaluations de travaux scientifiques et de classements d'universités. Dans la littérature, il n'y a pratiquement aucune indication d'erreurs endogènes (original correct, saisie dans la base de données incorrecte) dans les bases de données payantes. Les erreurs banales (telles que les fautes dans les noms d'auteur) dans les bases de données n'auraient que peu d'importance. La recherche d'erreurs pour Pierre Bourdieu en tant qu'auteur cité dans le SSCI (comparaison original – enregistrement SSCI) a révélé plus de 85 mutations. Les études de cas montrent un nombre élevé d'erreurs endogènes dans les bases de données. Les notes de bas de page, courantes dans les sciences juridiques, risquent d'être transformées en références fantômes (étude de cas de la *Harvard Law Review* : 99 % d'erreurs). Donc, l'affirmation de SSCI de représenter les sciences sociales globales « pertinentes » – pour toutes les disciplines couvertes dans le SSCI – semble être entravée par des lacunes dans la saisie et le traitement des données.

**Descriptors:** Empirical Study, Citation frequency, Quality, Database, Criticism, Impact factor, Mistakes, Social sciences, Social Science Citation Index

## 1 Zitatdatenbanken als Datenbasis für quantitative Evaluationen wissenschaftlicher Leistungen

Zitatdatenbanken bilden die Grundlage für viele szientometrische/bibliometrische<sup>1</sup> Untersuchungen. Sie liefern die Datengrundlagen für die Evaluation wissenschaftlicher Leistungen (von Zeitschriften, Fächern, Forschenden) und für die Berechnung von Universitätsrankings und anderen Kennzahlen, von denen Gelder, Posten und Karrieren abhängen können.

Der *Social Sciences Citation Index* gilt unter Fachleuten aus der Evaluations-/Szientometrie-Szene als „autoritative“ Datenbank (im Gegensatz etwa zur angeblich unseriösen Zitatsuchmaschine *Google Scholar*). Bis vor kurzem hatte der SSCI jahrzehntelang das Monopol auf dem Gebiet der Zitaindexierung.

Die SSCI-Records (= die Repräsentationen von Zeitschriftenartikeln in der Datenbank, die klassischen Dokumentationseinheiten) enthalten bekanntlich zusätzlich zu den üblichen bibliographischen Angaben die komplette Literaturliste der zitierten Zeitschriftenartikel, zumindest dem Anspruch nach. Daher kann die Suche umgedreht werden und es kann gefragt werden: Wer hat eine Autorin oder einen Autor oder eine Zeitschrift zitiert? Der SSCI wird täglich weltweit dazu verwendet, Zeitschriften, Institutionen, einzelne Forschende der Sozialwissenschaften im weitesten Sinne des Begriffs (inkl. Ökonomie, Rechtswissenschaften), die Zeitschriften, in denen sie publizieren und ihre Institutionen zu suchen und/oder zu bewerten.

Die akribische Prüfung der Datenqualität dieser SSCI-Records stand im Fokus der Dissertation der Verfasserin.<sup>2</sup> Ihre wichtigsten Ergebnisse werden in diesem Beitrag skizziert. Technischer Jargon soll dabei nach Möglichkeit vermieden werden, Vereinfachungen sind daher unvermeidlich. Die Beschränkung auf den SSCI erfolgt aus pragmatischen Gründen: Der AHCI (*Arts & Humanities Citation Index*, für Geistes- und Kulturwissenschaften) ist an der JKU Linz, der Heimatinstitution der Verfasserin, nicht verfügbar und eigene Forschungserfahrungen wie wissenschaftliche Kompetenzen der Verfasserin beziehen sich neben den Kulturwissenschaften auf die Sozialwissenschaften im weitesten Sinne.

<sup>1</sup> Szientometrie steht für Wissenschaftsvermessung und bezieht sich auf alle quantitativen Aspekte von Wissenschaft; Bibliometrie eher auf das Publikationswesen, wobei die quantitativen Aspekte von Publikationen im Fokus stehen.

<sup>2</sup> Siehe Tüür-Fröhlich 2016.

## 2 Defizite der Fehlerforschung/ Grundannahmen der Mainstream-Szientometrie/ Bibliometrie

Unter sogenannten banalen (trivialen) Fehler verstehen wir gewöhnlich z. B. Tipp-, Schreib- oder Druckfehler der Namen von Autorinnen und Autoren oder Initialen, Journaltitel, Namen wissenschaftlicher Einrichtungen; Falschklassifizierung von Dokumenten; fehlende Einträge (z. B. Universität statt Universitat, Salzburg statt Salzburg, Tur statt Tuur, Wiese statt Weise).

Nach allgemeiner Meinung unter Wissenschaftlerinnen und Wissenschaftlern sind banale Fehler von geringer Relevanz. Vor allem in der Kommunikationswissenschaft, der Szientometrie und bei Datenbank Anbietern ist die feste Ansicht verbreitet, dass Fehler in wissenschaftlichen Publikationen und Datenbanken irrelevant waren. Ja, es gabe viele Fehler, – aber sie wurden sich gegenseitig ausgleichen (Garfield 2005). Im Gegensatz zu dieser weit verbreiteten Meinung habe ich fur mein Dissertationsprojekt meine Forschungsthese T2 (Tuur-Frohlich 2016, 30) formuliert – kurz: Fehler sind nicht zufallig verteilt, sondern treten gehauft bei der Indexierung nicht-englischsprachiger Namen und Publikationen (LOTE= Languages Other Than English) auf, und eher bei weniger bekannten Autorinnen und Autoren und bei niedrigrangigen Journalen bzw. Verlagen. Die Fehler sind hartnackig, nicht zuletzt, weil Autorinnen und Autoren ihre Literaturlisten oft aus den (fehlerhaften) Angaben aus Artikeln oder Datenbanken ohne Sichtung des Originals ubernehmen (vgl. Simkin/Roychowdhury 2002).

Ein kurzer kritischer uberblick uber die Behandlung der Fehlerthematik in der Literatur fuhrt zu folgendem Ergebnis:

- 1) Die systematische Literatursuche nach Fehlererkenntnis- bzw. Fehlermanagementliteratur zeigt viele Publikationen in Psychologie und Management Science. Diese thematisieren den Umgang mit Katastrophen wie Tschernobyl – hervorgerufen aufgrund menschlichen Fehlverhaltens.
- 2) „Typo“- und „accuracy of references“-Studien wurden vor allem in Medizin, Pflege-, Bibliotheks- und Informationswissenschaftsjournalen gefunden. Aus dieser Literatur lassen sich folgende Verallgemeinerungen ableiten: Die Mehrzahl der Studien klassifiziert Fehler entweder als geringfugig (minor) oder schwerwiegend (major). Es gibt allerdings keine allgemeingultigen Definitionen. Oft wird ein Fehler als minderwertig betrachtet, wenn ein Artikel trotz der fehlerhaften biblio-

graphischen Angaben im Internet gefunden werden konnte. Ein Fehler wurde als wichtig erachtet, wenn der Artikel überhaupt nicht gefunden werden konnte. Für die Fehler werden die Autor(en) der Publikationen verantwortlich gemacht. Laut Unver et al. (2009) schleichen sich Fehler in Referenzlisten ein: aufgrund mangelnder Aufmerksamkeit für Details, „unvorsichtiger“ Transkription von bibliographischen Daten, oder der „Delegation von Verantwortung“, d. h. der Abwälzung der Überprüfung der Zitate auf unqualifizierte Assistenten. Die Verantwortung für die Korrektheit der Zitate liege bei den Autoren. Nur wenige Publikationen erwähnen beiläufig, dass Datenbanken nicht fehlerfrei sind.

- 3) Interessanterweise zeigt die Literatur zu Datenbankverzerrungen und / oder Datenbankfehlern in Finanzanalysen (z. B. Finanzinformationen zu öffentlichen Unternehmen)<sup>3</sup> eine kritischere Haltung. Sie kritisieren u. a. Auswahl, Auslassungen und Fehlklassifikationen sowie Kodierungsrichtlinien der inspizierten Datenbanken und schlagen Methoden der Qualitätskontrolle vor. Ich denke, die Informationswissenschaften könnten von diesem Forschungsgebiet lernen.

Zusammengefasst: In der Fachliteratur finden sich kaum Hinweise auf Fehler in Datenbanken, am seltensten auf Fehler in den kostenpflichtigen Datenbanken SCI/SSCI/AHCI.<sup>4</sup> Hingegen kritisieren mehrere Studien die schlechte Datenqualität (Doubletten, Phantomautoren) der (gratis nutzbaren) Zitatsuchmaschine Google Scholar.<sup>5</sup>

Wenn Datenbankfehler überhaupt erwähnt werden, macht die Fachliteratur für die Entstehung von Datenbankfehlern die *Autorinnen und Autoren* der Originalpublikation verantwortlich: Die Datenbanken würden deren Fehler übernehmen. Belege für diese die Datenbankproduzenten entlastenden These finden sich jedoch keine. Wenn Fehler angesprochen werden, dann handelt es sich um recht „harmlose“ banale Fehler (etwa falsche Seitenzahlen, leichte Tippfehler).

<sup>3</sup> Kellogg School of Management, *Database Biases and Errors*. [online]. 2011. Available at: <http://www.kellogg.northwestern.edu/rc/crsp-cstat-references.htm> [cit. 16.10.2013].

<sup>4</sup> Siehe vor allem die Studien von Franceschini et al. 2013. Dieselben Autoren (2016) führten auch eine kritische Studie zur Datenqualität der kostenpflichtigen Zitatdatenbank Scopus des Informationskonzerns Elsevier durch.

<sup>5</sup> Vgl. dazu die Studien von Péter Jacsó (z. B. 2009), der sich fast ausschließlich der Kritik an Google Scholar widmet.

Nur einige wenige kritische Publikationen (und Blogs) erwägen die Möglichkeit *endogener* (= durch die Datenbankproduktion selbst verursachte) Fehler in Zitatdatenbanken.

### 3 Eigene Fallstudien zur Datenqualität von SSCI-Records

Von diesem Mainstream abweichend, wurden im Dissertationsprojekt der Verfasserin folgende methodische Vorgangsweisen gewählt:

- 1) Es wurde die – für den normalen wissenschaftlichen Hausverstand – naheliegende Herangehensweise gewählt: Der Vergleich zwischen den Literaturlisten der *Originalpublikation* mit den Datenbank-Einträgen in den *SSCI-Records* (in der Folge: *Original-Record*), so wie sie bei einer Zitatrecherche in der Datenbank (lizenzierter Zugriff über die JKU Linz) sichtbar sind. Diese Methode wurde bislang zwar bei der Suche von bibliographischen Fehlern in Zeitschriften, nicht jedoch bei der Fehlersuche in Datenbanken eingesetzt.
- 2) Auf Grundlage des Vergleichs Original – Record kann nach *endogenen* Datenbankfehlern gesucht werden: Die Feststellung *Originalangaben richtig – Referenzen im SSCI-Record fehlend oder falsch* lässt nur einen Schluss zu: Die gefundenen Fehler müssen bei der Dateneingabe bzw. -verarbeitung zur Produktion des SSCI entstanden sein.

### 4 Ausgangspunkt: der Autorenname Pierre Bourdieu

Die Vergleiche Original-Record wurden nicht an willkürlich ausgewählten Zeitschriftenartikeln vorgenommen (es wurden also nicht einzelne SSCI-Records mit schwerwiegenden Fehlern herausgegriffen). Bei der Auswahl der Vergleichspaare wurde konsequent nach der *Schneeballmethode* vorgegangen.<sup>6</sup>

Ausgangspunkte der Suche waren die Fehlschreibungen, Mutationen, Verstümmelungen des Namens des französischen Philosophen und Soziologen *Pierre Bourdieu* als zitierter Autor im SSCI (d. h. im Datenfeld „cited aut-

<sup>6</sup> Details zur methodischen Vorgangsweise siehe Tüür-Fröhlich 2016, 50 f.

hor“). Die Auswahl dieses Autors erfolgte aus pragmatischen Gründen:

Um Fehlern in Zitatdatenbanken mit Hilfe der *Ping-Pong-Methode* (siehe Kapitel 5) auf die Spur zu kommen, ist die Kenntnis des Gesamtwerks eines Autors oder einer Autorin inklusive Übersetzungen, Neubearbeitungen etc. unverzichtbar. Bei Pierre Bourdieu konnte auf eine an der JKU Linz seit vielen Jahren aufwändig erstellte Hypertext-Rekonstruktion seines Gesamtwerks zurückgegriffen werden.<sup>7</sup>

Der Name Pierre Bourdieu enthält (sowohl im Vor- wie auch im Familiennamen) nur ASCII (= American Standard Code for Information Interchange)-Zeichen, daher sollte die Verarbeitung seines Namens durch einen nordamerikanischen Datenbankproduzenten keine Schwierigkeit bereiten.

## 5 Fallstudie PIERRE BOURDIEU: Mutationen, Mutilationen, Totalverluste

Bei der Fehlersuche zum Thema Pierre Bourdieu als zitierter Autor im SSCI wurde nach der selbst entwickelten *Ping-Pong-Methode* vorgegangen. Darunter ist das Hin- und Herspringen zwischen den Datenfeldern (vor allem: Cited Author, Cited Work/Title) zu verstehen. Schwere Fehler und fehlende Einträge im Datenfeld Cited Author können nicht direkt gesucht werden, aber es kann z. B. im Datenfeld „Cited Work“ nach Zitaten des Bourdieuschen Hauptwerkes „La distinction“ gesucht werden. Findet sich im so gefundenen SSCI-Eintrag im Datenfeld Cited Author ein Fehler oder fehlt dort ein Eintrag, ist damit eine neue Fehlervariante gefunden. Abgesehen von fehlenden Einträgen bzw. „anonymous“-Einträgen können diese fehlerhaften Autorennamen wieder ins Datenfeld Cited Author eingegeben werden und unter Umständen zu weiteren Fehlern im Datenfeld „Cited Work“ führen. Dieses Wechseln der Datenfelder lässt sich fast unendlich oft wiederholen.

Als Resultat zeigten sich mehr als 85 Mutationen, Verstümmelungen, Totalverluste. Folgende *Typologie* konnte erstellt werden, wobei die zahllosen falschen Initialen, die Bourdieu „angedichtet“ wurden, im Folgenden erst gar nicht erwähnt werden:

- *Fehlertyp 1*: Fehler eher banalerer Art („Bordieu“ ohne „u“, „Boudieu“ ohne „r“); diese sind aber keineswegs

irrelevant, weil die Datenbank-Software Zeichenketten (strings) vergleicht, und hier sind bereits kleine Abweichungen folgenschwer.

- *Fehlertyp 2*: Schwerwiegende Fehler, z. B. „\*BP“, „Bourdieu, (P)“, „ourdiu P“,
- *Fehlertyp 3*: Substitution durch andere Autorennamen (z. B. fand sich der Leistungspsychologe Rowland Atkinson als Autor von Bourdieus Hauptwerk „La distinction“),
- *Fehlertyp 4*: Missings (Totalverluste): leeres Feld beim Autorennamen oder Angabe „anonymous“ (obwohl im Original Bourdieu als Autor korrekt angegeben war),
- *Fehlertyp 5*: Verwechslung von Vor- und Nachnamen bei der Indexierung.<sup>8</sup>

Bourdieu-Zitierungen sind also im SSCI auch unter „Pierre B\*“, ja in einem Fall sogar unter „Pierri B\*“ auffindbar. SSCI-Records mit den Phantomautoren „Pierre B“ bzw. „Pierri B“ wurden in der Folge mit der Originalpublikation verglichen: War Bourdieu der einzige verstümmelte Autorennamen, oder geschah dies auch bei den übrigen Referenzen des jeweiligen Artikels?

## 6 Fallstudie SOFIO 2008a, Journal *Travail, genre et sociétés*<sup>9</sup>

Der erste als Volltext verfügbare Zeitschriftenartikel mit einem SSCI-Record, das „Pierre B“ enthielt, war der Artikel von Séverine Sofio 2008a:

*Sofio, Séverine (2008a): Les vertus de la reproduction. Travail, genre et sociétés 19, 23–39.*

Der SSCI-Record Sofio 2008a enthält den Eintrag: „PIERRE B, 1987, CAHIERS MUSEE NATL M, V19, P6“. Im Original lesen wir hingegen in der Bibliographie am Ende des Artikels: „BOURDIEU Pierre, 1987, « L'institutionnalisation de l'anomie », Cahiers du Musée national d'art moderne, 19–20, pp. 6–19.“

Es zeigte sich: Der Originalartikel enthält an seinem Ende eine korrekte Bibliographie, in der die zitierten Werke nach den Familiennamen alphabetisch gelistet werden. Die Nachnamen sind sogar in Kapitalia hervorgehoben. Die Verarbeitung dieser Bibliographie hätte also keine Herausforderung darstellen müssen.

<sup>8</sup> Zu einer Liste aller bisher gefundenen Mutanten und Verstümmelungen und Erläuterungen dazu siehe Tüür-Fröhlich 2016, 52ff.

<sup>9</sup> Zum methodischen Vorgehen und zu den Ergebnissen siehe ausführlicher Tüür-Fröhlich 2016, 57ff.

<sup>7</sup> Siehe HyperBourdieu.jku.at.

**Tabelle 1:** Phantomautorinnen und Phantomreferenzen im SSCI-Record für Sofio (2008a) nach Fehlerkategorie und Häufigkeit.  
Quelle: Tüür-Fröhlich 2016, S. 63.

Fehlerkategorie	Anzahl Referenzen (n=)
1 Fehlende Referenzen (SOFIO 2008b)	1
2 <i>Phantomautorinnen A:</i> Vornamen der Autorinnen wurden als angebliche Familiennamen indexiert; die Anfangsbuchstaben der Familiennamen als Initiale des Vornamens	20
3 <i>Phantomautor B:</i> Der zweite Vorname wurde als angeblicher Familienname indexiert, erste Buchstaben von Familiennamen und erstem Vornamen wurden als Initialen für den angeblichen Vornamen verwendet (“EHRlich WB”)	1
4 <i>Phantomautorin C:</i> Familienname zusammengeklumpt (fatale Mutation), Initiale korrekt (“CACOUAULTBITAUD M”),	1
5 Phantomreferenz A: Vorname als angeblicher Familienname; erster Buchstabe des Familiennamens als angebliches Initial. Vorgeblicher Buchtitel aus Einträge in verschiedenen Datenfeldern zusammengeklumpt, Dokumenttyp, Document PhD institution als Buchtitel (“CHARLOTTE Y”)	1
6 Phantomreferenz B: Vorname als angeblicher Familienname; vorgebliche Initialen der Vornamen aus den ersten Buchstaben des Doppelnamens gebildet, Buchtitel: Permutation von Buchstaben, nicht auffindbar (“ANDREE SS”)	1
Phantomeinträge/fehlende Einträge TOTAL (N=)	25

Trotzdem zeigte der Vergleich Original-Record (Tabelle 1) ein ernüchterndes Ergebnis: Nur *eine* Referenz von n=26 im Original korrekten Referenzen war fast richtig dargestellt (wenn auch unter Verletzung der Indexierungsregeln des SSCI). Alle anderen hatten schwere Mängel. Eine Literaturangabe fehlte ganz, n=23 im Original korrekt zitierte Autorinnen und Autoren mutierten zu Phantomautoren, bei denen der erste Vorname oder der mittlere Name als neuer Familienname verwendet wurden und der Nachname zur neuen Initiale schrumpfte. In einem Fall war zwar die Initiale richtig, aber der Familienname mit einem nirgendwo in der Literaturliste oder im übrigen Artikel auffindbaren Zeichenkette „bitaud“ zusammengeklumpt: aus Cacouault, M. wurde *Cacouaultbitaud, M.*

Fazit: Von einer einzigen halbwegs richtig wiedergegebenen Referenz abgesehen, gehen bei Sofio 2008a alle übrigen n=25 Referenzen verloren, sowohl für Zitate-recherchen als auch für Zitatezählungen, etwa für einen h-index.<sup>10</sup>

<sup>10</sup> Der h-Index, oder Hirsch Index, benannt von/nach seinem Erfinder, dem Physiker Jorge E Hirsch, soll den Publikationserfolg von Wissenschaftlerinnen und Wissenschaftlern „messen“. Statt der abstrakten Formel des Erfinders zur Erläuterung zwei fiktive Beispiele: Eine Autorin X hat einen h-Index von 3, wenn sie vier Artikel mit 150, 115, 3, 2 Zitaten veröffentlicht hat, denn sie hat n=3 Publikationen, die jeweils mindestens dreimal zitiert wurden. Autor Y mit drei Artikeln, die jeweils nur dreimal zitiert wurden, hat jedoch ebenfalls einen h-Index von 3. Der H-Index belohnt also die fließbandmäßige Produktion von Publikationen mit mittelmäßiger Resonanz und bestraft Autorinnen und Autoren mit wenigen Publikationen, aber mit durchschlagender Resonanz. Der H-Index kann mittels der Daten v. a. von SCI/SSCI, Scopus oder Google Scholar errechnet werden. Jede Daten-grundlage ergibt andere Werte. Der h-index von Google Scholar (GS)

Die Referenzenliste im SSCI-Record (und wie eine extensive Sichtung zeigte, die vieler anderer Artikel aus dem Journal *Travail, genre et sociétés*) kann wohl kaum einer Qualitätsüberprüfung unterzogen worden sein. Ein Vorname kann auch als Nachname fungieren, doch handelt es sich dabei eher um seltene Fälle. Ein solch massenhaftes Vorkommen von Autorinnen und Autoren mit Nachnamen wie *Albert, Andree, Anne, Bruno, Chantal, Charlotte, Eugene, Gerard, Isabelle, Margaret, Maria, Nathalie, Rene, Severine* wie in diesem SSCI-Record hätte jedoch auffallen müssen.<sup>11</sup>

## 7 Fallstudie BEEBE 2010, Journal *Harvard Law Review (HLR)*

Kommen wir zum Phantomautor „Pierri B“ im SSCI. Der zitierte Autoreneintrag „*Pierri B*“ enthält drei Fehler: Falscher Nachname, falscher Anfangsbuchstaben des Vornamens – beide Fehler können nur durch die Verwechslung von Nachname und Vorname zustande gekommen sein. Zusätzlich enthält der Vorname – der fälschlich als Nachname geführt wird – einen Schreibfehler (ein „i“ an-

ist meist am höchsten, da GS LOTE (LOTE = nicht-englischsprachige) Publikationen und vor allem Buchpublikationen und sonstige Literaturtypen weitaus stärker abdeckt, (siehe auch Tüür-Fröhlich 2016, 19 und dort zitierte Literatur).

<sup>11</sup> Zur kompletten Referenzliste des SSCI-Records siehe Tüür-Fröhlich 2016, 153. Auch ein automatischer Abgleich mit Vornamen-Datenbanken hätte ein Warnsignal ergeben müssen.

stelle eines „e“). Diese drei schwerwiegenden Fehler müssen beim Scannen bzw. der Umwandlung in maschinenlesbare Zeichen mittels OCR („Optical Character Recognition“) bzw. beim Parsen<sup>12</sup> entstanden sein.

Jener Artikel, der bei der Suche nach dem zitierten Autor „Pierri B“ vom SSCI nachgewiesen wurde, war Beebe 2010:

*Barton Beebe (2010): Intellectual property law and the sumptuary code. Harvard Law Review 123 (4), 809–889.*

Der SSCI-Record zu Beebe 2010 enthält den Eintrag: „PIERRI B, 1970, REPROD IN ED, P71“. Im Originalartikel von Beebe 2010 ist in Fußnote 23 hingegen angeführt:

„PIERRE BOURDIEU & JEAN-CLAUDE PASSERON, REPRODUCTION IN EDUCATION, SOCIETY AND CULTURE (Richard Nice trans., Sage Publ'ns 1977) (1970)“ (Beebe 2010, 814).

Der Vergleich Original – Record zu Beebe 2010 war außerordentlich komplex und zeitintensiv<sup>13</sup>:

a) Der Originalartikel hat 80 Seiten und enthält n=435 Fußnoten. Um das Untersuchungsergebnis vorwegzunehmen: Für n=277 (von insgesamt n=493) ursprünglichen Zitaten im Original gibt es *überhaupt keinen* Eintrag im SSCI-Record, d. h. 65 Prozent der Originalreferenzen gingen bei der Indexierung verloren. Dafür fanden sich für etliche Originaleinträge gleich mehrere Phantomreferenzen im Record (Beispiel Thorstein Veblen siehe Tabelle 2).

**Tabelle 2:** Fehlerhafte Triplets zu Thorstein VELEN 1912 im SSCI-Record zu Beebe 2010. Quelle: Tüür-Fröhlich 2016, 72; fett: zusätzlicher Eingabe- bzw. OCR-Fehler.

ORIGINAL	SSCI-RECORD
(Beebe 2010, 812, Fußnote 14)	THORSTEIN V, 1912, THORSTEIN VELEN, THE THEORY OF THE LEISURE CLASS 51 (2d ed. 1912)
	VELEN, INSISTENCE IT ELEMEN, P87
	VELEN S, SERVICEABILITY CONSU, V14, P87

b) Der ausgewertete Originalartikel enthält keine alphabetisch geordnete Bibliographie. Beebe 2010 verwendet die vom Harvard Law Review vorgeschriebene Zitierweise.

<sup>12</sup> Parsingprogramme sollen automatisch bestimmen, welche Wörter als Eigennamen oder Journaltitel aufzufassen sind etc.

<sup>13</sup> Zum Modus Operandi und zu den schwierigen Problemen, überhaupt die einander entsprechenden Paare (Original vs. Record) herauszufinden, siehe Tüür-Fröhlich 2016, 65 ff.

Demnach sind eher Fußnoten zu verwenden, die Fußnoten sollten den Bestimmungen der 19. Ausgabe des *Bluebook* folgen.<sup>14</sup> Das *Bluebook* kann als *das* juristische Zitierungssystem in den USA bezeichnet werden, seine Richtlinien müssten den nordamerikanischen Datenbankproduzenten bekannt sein.

Der Vergleich Original – Record erbrachte zahlreiche *Phantomautorinnen* und *-autoren* und *Phantomwerke* (zu einigen Beispielen siehe Tabelle 3):

Dabei unterlief SSCI oft eine „Todsünde“ der Indexierung, nämlich die Verwechslung von Datenfeldern. So mutiert im ersten Beispiel in Tabelle 3 der abgekürzte *Journaltitel* FAR E. ECON. REV. im Original, er steht für *Far Eastern Economic Review*, im Record zu einem *AutorInnen*-namen „Econ R“.

**Tabelle 3:** Phantomautoren, Phantomwerke im SSCI-Record zu Beebe 2010 (Beispiele). Quelle: Eigene Zusammenstellung, Tüür-Fröhlich 2016, 80ff; fett= „Übereinstimmungen“ Original-Record.

ORIGINAL	SSCI-RECORD
(Beebe 2010, 818, Fußnote 31) Velisarios Kattoulas, <i>Bags of Trouble</i> , FAR E. ECON. REV., Mar. 21, 2002, at 52, 52.	<b>ECON R</b> , 2002, BAGS TROUBLE, V52, P52
WALTER BENJAMIN, <i>The Work of Art in the Age of Mechanical Reproduction</i> , in ILLUMINATIONS 217, 224 (Hannah Arendt ed., Harry Zohn trans., Schocken Books 1968) (1955)	WALTFR B, <b>11LUMINKFIONS</b> , V217, P224
See H.P. Bovenkerk et al., <i>Preparation of Diamond</i> , 184 NATURE 1094 (1959); F.P. Bundy et al., <i>Man-Made Diamonds</i> , 176 NATURE 51 (1955).	BOVENKER HKP, 2008, <b>NXRURE FT BUNDY</b> , V51, P959

Im zweiten Fall wird der Vorname des Philosophen Walter Benjamin zum neuen Nachnamen, mit einem OCR-Fehler (F statt E), der Anfangsbuchstabe seines Nachnamens wird zum fiktiven Initial, und der Titel des Sammelbandes wird aufgrund weiterer schwerer OCR-Fehler fast völlig verstümmelt, aus „ILLUMINATIONS“ wird „**11LUMINKFIONS**“.

Im dritten Fall wird der Autorennachname im SSCI-Record fast richtig getroffen, abgesehen vom letzten Buchstaben, der dafür zwischen die zwei richtigen Initialen gequetscht wird. Aus dem weltbekannten „Mega-Journal“

<sup>14</sup> Harvard Law Review, <http://harvardlawreview.org/submissions/> [10.5.2014].

Nature wird jedoch ein Phantomjournal: Aufgrund eines OCR-Fehlers wird „NATURE“ zu „NXRURE“ und aufgrund eines Parsingfehlers wird aus der in der Fußnote folgenden Referenz der Autorennamen inklusive eines OCR-Fehlers (FT statt F.P.) zum zweiten Teil des Phantomjournals „NXRURE FT BUNDY“. Es folgt die Jahrgangszahl der zweiten Referenz (die mit Bovenkerk 1959 nichts zu tun hat), und eine Seitenzahl, die offensichtlich mittels OCR- und Parsingfehler aus dem Publikationsjahr von Bovenkerk gezogen wurde (P959 aus 1959 im Original wird P959 im Record).

Diese drei Beispiele (3 von 197 Phantomeinträgen im Record) mögen hier genügen. Das ernüchternde Ergebnis des Vergleichs Original – Record ist in Tabelle 4 zusammengefasst.

**Tabelle 4:** Indexierungsfehler im SSCI-Record zu Beebe 2010 nach Fehlerstatus. Quelle: Eigene Berechnung, Tüür-Fröhlich 2016, 69.

Fehlerstatus	Referenzen im SSCI-Record nach Fehlerstatus, n=	Referenzen im SSCI-Record nach Fehlerstatus, % von allen Originalreferenzen in Beebe 2010, n=493
Fehlende Einträge insgesamt	277	56,2
Phantome/Mutationen/Zombies	197	40,0
Kleinere Fehler	14	2,8
Korrekte Einträge	5	1,0
Literaturverweise bei Beebe 2010 insgesamt	493 (n=)	100 (%)

Über die Hälfte der Originalreferenzen sind auf dem Weg in den SSCI-Record spurlos verschwunden („Lost in Indexing“, würde die Regisseurin Sofia Coppola betiteln). 40 Prozent sind nur in Form von Phantomeinträgen ‚erhalten‘ geblieben. Neben ca. drei Prozent mit kleineren Fehlern indexierten Einträgen, sind *nur ein Prozent* der im Original korrekten Referenzen auch im SSCI-Record korrekt repräsentiert.

Zusammenfassend: Referenzen in Fußnoten laufen große Gefahr, in Phantomautoren, Phantomwerke und Phantomreferenzen verwandelt zu werden, durch irrtümliche „Interpretation“ von Zitattextteilen, Zitatsymbolen und Fußnotenfragmenten als bibliographische Informationen.

## 7.1 Detailanalysen: Struktur der Totalverluste

Sehen wir uns die Totalverluste bei der Indexierung von Beebe in Tabelle 5 genauer an.

**Tabelle 5:** Fehlende Referenzen (Totalverluste) nach Dokumentenkategorie. Quelle: Tüür-Fröhlich 2016, 69.

Dokumentenkategorie	Zitiert in Beebe (n=)	Fehlende Referenzen im SSCI-record (n=)	Fehlende Referenzen im SSCI (%)
Juristische Dokumente, z. B. Rechtsakten, Gerichtsverfahren, internationale Abkommen	186	158	85 %
Bücher: Monographien, Sammelbände	129	37	29 %
Journalartikel	120	47	39 %
Zeitungsberichte, z. B. der NYT	39	21	54 %
Andere, z. B. Internetquellen	19	14	74 %
GESAMT	493	277	56 %

Die Anzahl aller fehlenden Referenzen im SSCI-Record festzustellen, war keine einfache Aufgabe. Der naheliegende Ansatz, um die Anzahl fehlender Referenzen zu bestimmen, wäre die Subtraktion der n=256 SSCI-indexierten Quellenangaben von den n=493 Quellenangaben des Originals – und dem Ergebnis n=237 zu vertrauen. Ein schneller und oberflächlicher Vergleich von SSCI-Record und Original zeigt oft dieselbe Zahl an Einträgen. Dann könnte man annehmen: Alles in Ordnung. Aber die „Sachlage“ ist komplexer. Wie oben erwähnt, enthält der SSCI-Record zahlreiche Dubletten und Triplets. Diese täuschen eine höhere Vollständigkeit der Indexierung von Originalreferenzen vor. Hervorzuheben ist hier auch, dass selbst bei der Indexierung der Zeitschriftenaufsätze, man könnte sie als Kernkompetenz des SSCI ansehen, 39 Prozent Totalverluste zu bemerken waren.

## 7.2 Detailanalysen: Vermutliche Genese von Phantomautorinnen/-referenzen

Zum besseren Verständnis soll im Folgenden die mutmaßliche Genese von Phantomautorinnen und -autoren bzw. Phantomreferenzen skizziert werden und mit einigen Beispielen illustriert werden. Nahezu jedes Zeichen in den

Fußnoten im Original von Beebe (inkl. Fußnotensymbolen) kann sich im SSCI-Record als Autorinnen- oder Werkname bzw. bibliographische Angabe finden. Die Verwechslung von Datenfeldern ist dabei häufig zu beobachten, wie die Beispiele in den folgenden Tabellen illustrieren.

**Tabelle 6:** Patenntitel als Autorennamen indexiert. Quelle: Tüür-Fröhlich 2016, 81.

ORIGINAL	SSCI-RECORD
(Beebe 2010, 839, Fußnote 148, fett TTF) “Long-Sleeved Garment with Wristwatch Accommodations, U. S. Patent No. 7,380,287 (filed July 1, 2005)”	<b>LONGSLEEVED G</b> , 2005, US PATENT <b>NO FILED J</b> , V287, P1

Kommentar „LONGSLEEVED G“ (Tabelle 6):

Wie wir sehen, wurde ein Wort unter Anführungszeichen aus einem Patenntitel als vorgeblicher Autorennamen angeführt, der erste Buchstabe aus dem nächsten Wort aus dem Patenntitel als Initiale eines angeblichen Autorennamens verwendet. Der Quellentitel ist teilweise falsch. Aufgrund des angefügten „J“, welches normalerweise als Abkürzung für „Journal“ steht, und der angeblichen Volume Number wird so ein Phantomjournal generiert. Die angebliche Jahrgangszahl ist allerdings falsch, sie wurde ganz offensichtlich aus den letzten drei Ziffern der Patentnummer generiert, die Seitenzahl aus dem Datum. Das Publikationsjahr ist aber immerhin korrekt.

**Tabelle 7:** Indexierung von Worten und Wortfragmenten aus Gerichtsverfahren als vermeintlicher Autorennamen. Quelle: Tüür-Fröhlich 2016, 83.

ORIGINAL	SSCI-RECORD
(Beebe 2010, 854, Fußnote 243, fett TTF) “Bd. of Supervisors of La. State Univ. v. Smack Apparel Co., 438 F. Supp. 2d 653, 660 (E.D. La. 2006) (finding the defendant’s novelty apparel <b>items to be infringing</b> ); <b>Grey v. Campbell Soup Co.</b> , 650 F. Supp. <b>1166, 1174–75</b> (C.D. Cal. 1986) (...);”	<b>GREY V, 2001, ITEMS BE INFRINGING</b> , V166, P174

Kommentar „GREY V“ (Tabelle 7):

„Grey V“ ist ein völlig fiktiver Autor. Der vorgebliche Nachname des Autors ist falsch (er stammt von dem Kläger „Grey“ in der Rechtssache „Grey v. Campbell Soup Co.“ in

der zweiten Referenz des Originals. Die Initiale des Vornamens ist falsch. Das „V“ stammt aus dem Titel des gleichen Gerichtsverfahrens („v“ ist die Abkürzung von „versus“ nach den Regeln des Bluebook). Der angebliche Journaltitel ist falsch. Er leitet sich aus dem Fußnotentext ab, einem Kommentar von Beebe. Die Jahrgangszahl ist falsch – wahrscheinlich ist sie aus der zweiten Referenz abgeleitet und verliert eine Zahl „1“ vermutlich aufgrund eines OCR-Fehlers (166 statt 1166). Die Seitennummer ist falsch – wahrscheinlich wird sie von der zweiten Referenz abgeleitet und verliert eine Zahl „1“ vermutlich aufgrund eines OCR-Fehlers (174 anstelle von 1174). Diese Phantomreferenz resultiert aus Datenfeldverwechslungen und zerhackten Bruchstücken aus dem originalen Fußnotentext und könnte auch als „Zombie-Referenz“ bezeichnet werden.

**Tabelle 8:** Interpretation von Fragmenten aus Fußnotentexten als Autorennamen. Quelle: Tüür-Fröhlich 2016, 85.

ORIGINAL	SSCI-RECORD
(Beebe 2010, 875, Fußnote 369 fett TTF) <i>See, e.g., Model Provisions, supra note 367, ¶ 36 (“‘Characteristic elements’ of the traditional artistic heritage, of which the production must consist in order to qualify as a protected ‘expression of folklore,’ means in the given context that the element must be generally recognized as representing a distinct traditional heritage of a community.”).</i>	MODEL P, CHARACTERISTIC See, e.g., <b>ELEME</b> , V367, P136

Kommentar „MODEL P“ (Tabelle 8):

Der Eintrag im SSCI-Record sieht wie ein ganz normaler Referenzeintrag aus, der einen Zeitschriftenartikel indexiert. Doch der Name des angeblichen Autors ist falsch: Der angebliche Vor- und Zuname des Vornamens ergibt sich aus dem Kurzzitat in Fußnote 369 (nach den Regeln von Bluebook). Diese Kurzform bezieht sich auf die vollständige Fundstelle in Fußnote 367 unter „Fußnote 367“. Der Journaltitel ist falsch. Er ist einer Anmerkung Beebes im Fußnotentext entnommen. Die Volume-Nummer ist falsch. Es gibt keine Bandnummer in Beebes Verweis auf dieses UNESCO / WIPO-Dokument. Die angebliche Bandnummer ergibt sich aus der Fußnote 367, dem Ort der vollständigen Referenz.

Auch die Seitenzahl ist falsch (136 statt 36). Sie ist abgeleitet von der Seitenzahl in der Kurzangabe (¶ 36), die einen OCR-Fehler enthält („¶“ wird als „1“ interpretiert). Das Zeichen „¶“ wird normalerweise als Layoutmarkierung in Textverarbeitungsprogrammen verwendet. Doch

die Untersuchung aller Vorkommen dieses Zeichens im Beebe-PDF zeigt, dass „¶“ hier fälschlicherweise in Original für „at“ verwendet wird und eine Seitennummer signalisiert. Daher resultiert die falsche Seitenzahl aus einer Kombination von „trivialen“ Fehlern, die dem PDF-Produzenten und dem SSCI unterlaufen sind.

Zusammenfassend: Es ist unklar, wie die Datenerfassung des SSCI erfolgt. Doch lassen einige Äußerungen höherer Mitarbeiter (vgl. Adam 2002) darauf schließen, dass bei der Datenerfassung ganz oder teilweise weiterhin auf die Papierausgaben der Journale zurückgegriffen wird und diese mit SCAN- und OCR- und Parsingprogrammen ausgewertet werden (was offensichtlich die verwendete Software überfordert). Unklar ist, inwieweit menschliche Indexierer, soweit wir wissen in Südostasien (cheap labour), eingreifen und inwieweit eine Qualitätskontrolle überhaupt vorgenommen wird.

## 8 Kritisches Resümee der Fallstudien

Die hier skizzierten Fallstudien<sup>15</sup> zeigen eine hohe Anzahl *endogener* Datenbankfehler wie Auslassungen und Fehler in Form schwerer Mutationen, Verstümmelungen, Zusammenklumpungen, die noch nie in der Fachliteratur erwähnt wurden. Moed (2005,175) gibt für „defekte“ oder „diskrepante“ Referenzen nur einen Prozentsatz von zwei bis neun Prozent im SCI an.<sup>16</sup> In meinen bisher durchgeführten Fallstudien im SSCI liegt der Prozentsatz der schwerwiegenden Defekte und verstümmelten Referenzen weitaus höher, nämlich zwischen 93 und 99 Prozent.

Die Fehlerraten der Phantom-Autorennamen (Nachnamen und Vornamen getauscht) sind in den Fällen Sofio 2008a und Beebe 2010 ähnlich hoch. Dies ist merkwürdig: Denn Sofio 2010 enthält eine nach den Nachnamen alphabetisch geordneten Literaturliste am Ende ihres Artikels. Beebe hingegen verwendet den Bluebook-Stil für Zitierungen, mit den Literaturangaben in Fußnoten. Im Fall von Sofio sollte es also keine große Herausforderung für die Produzenten von Zitatdatenbanken sein, diese Literaturliste korrekt zu übernehmen und damit den Artikel kor-

rekt zu indexieren. Wie es zur fast durchgängigen Verwechslung von Vor- und Nachnamen kommt, bleibt daher ein Rätsel.

Im Bluebook-Stil für juristische US-Publikationen werden die vollständigen Referenzen in Fußnoten jeweils nur bei ihrer erstmaligen Zitierung im Text angeführt. Diese Fußnoten sind nummeriert, bei weiteren Zitierungen desselben Werkes wird in abgekürzter Form auf diese Fußnote verwiesen. Die Indexierung von Referenzen in Fußnoten scheint also stark fehleranfällig zu sein und überfordert offenbar die Textparser (Programme zum „Lesen“, zur inhaltlichen Interpretation der Dateien) und die (menschlichen) SSCI-Indexierer: Aufgrund fehlerhafter „Interpretationen“ werden beliebige Wörter und Wortfragmente aus Fußnoten zu angeblichen Autorennamen oder Werktiteln, Zahlen bzw. Zahlenfragmente zu angeblichen Jahrgangs- oder Seitenzahlen.

Die Vergleiche von Original und Datensatz zeigte: Die ursprünglichen Referenzen waren fehlerfrei. Die Verluste (Missings), Mutationen, Verstümmelungen müssen im Laufe der Dateneingabe bzw. Datenverarbeitung bei der Bearbeitung des SSCI entstanden sein. Es handelt sich hier also eindeutig um *endogene Datenbankfehler*.

Artikel, Bücher und andere Publikationsformen, die fälschlicherweise gar nicht indexiert oder nur stark mutiert und verstümmelt werden, sind nicht suchbar und werden in Zitatstatistiken nicht mitgezählt. Der Oxford-Stil beim Zitieren ist jedoch eine Gemeinsamkeit von Rechts- Sozial-, Kultur- und Geisteswissenschaften, vor allem bei Monographien und Sammelbänden vieler traditioneller Verlage.

Es ist daher davon auszugehen, dass diese Fächer bei der Zitatindexierung benachteiligt werden. Die Phantomwerke führen, so die These dieses Beitrags, zu gering oder gar nicht zitierten Publikationen im SSCI-Pool. Behauptungen, die Sozial- und Kulturwissenschaften hätten kaum Resonanz (siehe Hamilton 1990, 1991, Pendlebury 1991), sind daher sehr fragwürdig. Es könnte sich bei den Nichtzitierungen um die Effekte der schweren Indexierungsfehler handeln, weil die tatsächlich vorgenommenen Zitierungen nicht von der Datenbank registriert wurden.

Der Anspruch von SSCI ist es, die „relevante“ *globale* Sozialwissenschaft abzubilden, für *alle* im SSCI erfassten Disziplinen. Diesem Anspruch wird der SSCI in der derzeitigen Form nicht gerecht.

<sup>15</sup> Siehe zu Methodik und Ergebnissen, wie bereits erwähnt, ausführlich Tüür-Fröhlich 2016.

<sup>16</sup> Für US Law Journals erwähnt Moed (2005, 176) allerdings einen Prozentsatz von 50 Prozent „mismatches“ in Bezug auf indexierte Seitenzahlen.

## 9 Empfehlungen für betroffene Autorinnen und Autoren

Abschließend noch fünf Empfehlungen:

- Überwachen Sie Ihre eigenen Zitierungen! Geben Sie Ihren Namen und Ihre Publikationstiteln auch absichtlich fehlerhaft ein, um auf Fehlervarianten zu stoßen!
- Ändern Sie niemals Ihren Namen! Im Falle eines unvermeidlichen Namenswechsels: vermerken Sie unbedingt Ihre früheren Namen als „früher bekannt als ...“ im Autorenfeld (und in ihren Lebensläufen)!
- Übermitteln Sie aktiv Fehlermeldungen an die Datenbankproduzenten, Zeitschriftenredaktionen und akademischen Verwaltungen!
- Diskutieren Sie öffentlich über gefundene Fehler und veröffentlichen Sie diese (in Briefen an die Herausgeber, in Blogs, in Massenmedien)!
- Unterzeichnen sie DORA, die *San Francisco Declaration on Research Assessment*! Bewegen Sie Ihre wissenschaftliche Fachgesellschaft, ihr Institut, ihre Universität dazu, DORA ebenfalls zu unterzeichnen!

DORA fordert eine Reform der derzeit gängigen Evaluationspraktiken und wendet sich gegen die Verwendung des JIF<sup>17</sup> zur Evaluation persönlicher Forschungsleistungen (denn die durchschnittliche Zitierung einer Zeitschrift hat wenig mit der Zitierung einzelner Beiträge zu tun). DORA kritisiert die intransparente Praxis der Zitatdatenbanken SCI, SSCI, AHCI (zum Zeitpunkt der Formulierung von DORA war dies der Medienkonzern Thomson Reuters). DORA stellt verschiedene sinnvolle Forderungen auf, z. B.

<sup>17</sup> Der „Journal Impact Factor“ (JIF) soll die durchschnittliche Anzahl der Zitate von Artikeln einer bestimmten Zeitschrift in einem sehr begrenzten Zeitraum ermitteln. Nach der Grundidee aus den 1920er-Jahren entwickelten Irving Sher und Eugene Garfield in den 1960er-Jahren die Formel. In einem Beispiel erläutert: Der JIF 2016 der Zeitschrift ABC wird errechnet, indem alle Zitate, die ABC in den Jahren 2014/2015 (nur von Zeitschriften aus dem Pool von SCI oder SSCI) erhalten hat, dividiert wird durch die Zahl nur jener Artikel der Zeitschrift ABC in 2014/15, die von Indexierern des SCI bzw. SSCI als „zitierfähig“ eingestuft wurden (Originalreports, Reviews). Diese „Gummibestimmung“ (Fröhlich 2008, 73) bevorteilt gemischte Zeitschriften wie „Nature“ und „Science“, die neben einem referierten wissenschaftlichen Teil in einem großen redaktionellen Teil mit kontroversen Themen und Kommentaren berühmter Personen (z. B. von Nobelpreisträger(inne)n) „kostenlos“ viele weitere Zitate erwirtschaften. Diese Gummibestimmung fürchten andere Zeitschriftenherausgeber, denn je mehr Artikel (z. B. Briefe) die Indexierer als „zitierfähig“ einstufen, umso niedriger wird der JIF (siehe auch Tüür-Fröhlich 2016, 19 und dort zitierte Literatur).

dass Zitatlisten der einzelnen Artikel nicht mehr Geschäftsgeheimnisse sein sollten, sondern open access zugänglich sind, damit Forschende diese Zitate überprüfen, aber auch für eigene Auswertungen nützen können.<sup>18</sup> Es ist wichtig, dass wir Wissenschaftlerinnen und Wissenschaftler uns nicht wie die Lemminge allen Evaluationspraktiken fügen, sondern unsere kritischen Stimmen erheben. Evaluation ist keine Naturgewalt, sondern ein menschengemachtes Phänomen, welches der Kritik unterliegen muss, und das veränderbar ist.

**Danksagung:** Dank an Claus Harringer für die Übertragung englischsprachiger Rohmanuskriptteile ins Deutsche.

## Literatur

- Adam, David (2002): The counting house. *Nature*, 415, 726–729.
- Beebe, Barton (2010): Intellectual property law and the sumptuary code. *Harvard Law Review* 123 (4), 809–889.
- Franceschini, Fiorenzo; Maisano, Domenico & Mastrogiacomio, Luca (2013): A novel approach for estimating the omitted-citation rate of bibliometric databases with an application to the field of bibliometrics. *JASIST* 64, 2149–2156.
- Franceschini, Fiorenzo; Maisano, Domenico & Mastrogiacomio, Luca (2016): The museum of errors/horrors in Scopus. *Journal of Informetrics* Vol. 10 (1), 174–182.
- Fröhlich, Gerhard (2008): Wissenschaftskommunikation und ihre Dysfunktionen: Wissenschaftsjournale, Peer Review, Impact Faktoren, in: Hettwer, Holger et al. (Hg.): *WissensWelten*. Gütersloh: Verlag der Bertelsmann Stiftung, 64–80.
- Garfield, Eugene (2005): The Agony and the Ecstasy – The History and Meaning of the Journal Impact Factor <http://garfield.library.upenn.edu/papers/jifchicago2005.pdf> [09.01.2013].
- Hamilton, David P. (1990): Publishing by – and for? – the Numbers. *Science* 250, 1331–2.
- Hamilton, David P. (1991): Research Papers: Who's Uncited Now? *Science* 251, 25.
- Moed, Henk F. (2005): *Citation Analysis in Research Evaluation*. Dordrecht/NL: Springer.
- Pendlebury, David A. (1991): Science, Citation, and Funding. *Science* 251, 410–411.
- Simkin, Mikhail V./Roychowdhury, V.P. (2002): Read before you cite ! <https://arxiv.org/abs/cond-mat/0212043> [17.09.2018].
- Sofio, Séverine (2008a): Les Vertus de la reproduction. *Travail, genre et sociétés* 19, 23–39.
- Sofio, Séverine (2008b): Des discours aux pratiques, comment approcher la réalité des rapports de sexe ? *Genre et professions*

<sup>18</sup> Siehe zu DORAs Transparenzforderungen Tüür-Fröhlich 2018; zu DORAs Unterzeichnerinnen und Unterzeichnern und dem Resolutionstext von DORA <https://sfdora.org/> DORA wurde bis jetzt (Stichtag 14.8.2018) von insgesamt 12508 Einzelpersonen und 505 wissenschaftlichen Organisationen (Forschungsfonds, Universitäten, wissenschaftliche Gesellschaften, Journale) unterzeichnet.

artistiques au XIXe siècle. Sociétés et représentations 24, 177–193.

- Tüür-Fröhlich, Terje (2016): The Non-trivial Effects of Trivial Errors in Scientific Communication and Evaluation. Glückstadt/D.: VWH, Schriften zur Informationswissenschaft, Band 69 (zugl. Doctoral Thesis, Johannes Kepler Universität Linz, Institut für Philosophie & Wissenschaftstheorie).
- Tüür-Fröhlich, Terje (2017): Eine „autoritative“ Datenbank auf dem Prüfstand: Der Social Sciences Citation Index (SSCI) und seine Datenqualität. UNILEX 2017, 3–9.
- Tüür-Fröhlich, Terje (2018): Open Citations – Die Transparenzforderungen der San Francisco Declaration on Research Assessment (DORA). Information – Wissenschaft und Praxis (IWP), 69 (4), 183–189.
- Unver, Bayram et al. (2009): Referency accuracy in four rehabilitation journals. Clinical Rehabilitation 23, 741–5.
- Veblen, Thorstein (1912): The Theory of the Leisure Class. N.Y.: Macmillan (2nd Ed.).



**Dr.in phil. Terje Tüür-Fröhlich, MSSc.**

Johannes Kepler Universität Linz

Altenberger Straße 69

4040 Linz

Österreich

**terje.tuur@jku.at**

Foto: Christo Christoph

Dr.in phil. Terje Tüür-Fröhlich ist Lektorin am Institut für Philosophie und Wissenschaftstheorie an der Johannes Kepler Universität (JKU) Linz/Österreich (Wissenschaftsforschung, Kultur- und Medientheorie, Informationskompetenz und Informationsethik). Ihre Forschungsinteressen sind Datenqualität und endogene Fehler in Zitatdatenbanken und deren Auswirkungen v. a. auf Wissenschaftlerinnen und Wissenschaftler, aus non-ascii- bzw. LOTE (languages other than English)-Sprachräumen, wie auch für deren Institutionen in der Ära permanenter Evaluation.