

Vanessa Thiel

## Springer Book Archives

Springer nutzt die digitale Chance, um seinen historischen Buchbestand wieder verfügbar zu machen – gedruckt und digital

### Vom Printverlag zum E-Publisher

Für Springer begann der Start in die digitale Welt im Jahr 1997, als auf der Online-Plattform SpringerLink die ersten elektronischen Zeitschriften gehostet wurden. Seither hat sich der damals noch traditionelle Printverlag zum Pionier in der digitalen Verlagswelt entwickelt. Unter dem Chief Executive Officer (CEO) Derk Haank hat die Fachverlagsgruppe konsequent seine »eStrategie« verfolgt und der Verlag gilt heute als einer der führenden internationalen Wissenschaftsverlage. Haank hatte jedoch nicht nur erkannt, dass die neuesten wissenschaftlichen Erkenntnisse unbedingt elektronisch verfügbar gemacht werden sollten. Durch das zunehmende Interesse an digitalen Inhalten entwickelte sich auch ein Bewusstsein für retrodigitalisierte Medien – der Archivbestand ist ein Schatz, der dank neuer Technologien für alle Zeiten gehoben werden kann und sollte. In der praktischen Übertragung hieß das, alle englischsprachigen Zeitschriften zurück bis zu ihren ersten Ausgaben aufzufinden, zu digitalisieren und im Online Journal Archive neu verfügbar zu machen. Dass dieses Konzept der Retrodigitalisierung 2010 auch für alle Springer-Bücher gelten sollte, war lediglich die konsequente Fortführung der E-Strategie des Verlags.

2006 startete Springer seine ersten E-Books und seither erscheint jeder Titel ab Copyrightjahr 2005 parallel als Print- und Onlineausgabe. Untersuchungen und Nutzersauswertungen belegten auch hier, dass nicht nur das Interesse an aktuellen E-Books groß ist, sondern auch an den älteren Büchern, von denen ein Großteil in der Printausgabe als längst vergriffen galt. Die Springer Book Archives sollen diesen Bestand wieder erschließen: gedruckt und digital, je nach Nutzerwunsch. Bis

Ende 2013, so lautet das Ziel, sollen auf der einst als Zeitschriftenportal gestarteten Plattform SpringerLink mehr als 150.000 E-Books verfügbar sein. Mit diesem Volumen wird es die größte Sammlung an wissenschaftlichen Büchern weltweit sein, allein 100.000 Titel sind retrodigitalisierte Bücher: ein Wissensschatz der Vergangenheit für die Zukunft.

Bis Ende  
2013 mehr als  
150.000 E-Books

### Great minds don't go out-of-print – they go online

Große Geister wie Rudolf Diesel, Werner Siemens, Karl von Frisch und andere herausragende Autoren sollen für Wissenschaftler und Interessierte zugänglich bleiben, ihre Publikationen sollen such- und auffindbar gemacht werden. Kein Buch soll für ewig verschwinden, sondern ab jetzt für immer in elektronischer und auch gedruckter Form verfügbar sein. Dank innovativer Technologien ist es möglich, diese Vision wahr werden zu lassen. Dabei geht Springer zurück bis ins Jahr 1842, als Julius Springer in Berlin eine Buchhandlung gründete und damit den Grundstein für die heutige Fachverlagsgruppe Springer Science+Business Media legte. Aus dem Buchhandel entwickelte sich ein Verlag, der Wissenschaftler wie Niels Bohr, Marie Curie und andere namhafte Forscher verlegte. Heute umfasst die Verlagsgruppe nach einigen Eigentümerwechseln und zahlreichen Zu- und Verkäufen etwa 50 Imprints.

Springer Book Archives bedeutet die Digitalisierung von mehr als 100.000 zum Teil vergriffenen Springer-Büchern. Die Titel sollen künftig nicht nur elektronisch als E-Book, sondern auch wieder als gedrucktes Exemplar verfügbar sein. Der Fokus liegt dabei auf den Titeln, die veröffentlicht wurden, bevor E-Books ein fester Bestandteil des Springer-Portfolios wurden. Dies betrifft somit sämtliche Titel, die zur heutigen Fachverlagsgruppe Springer Science+Business Media gehören und vor 2005 erschienen sind. Dazu gehören Bücher aus dem STM- (Science, Technology, Medicine) und dem

Springer seit  
1842: Vom  
Buchhandel zum  
E-Publisher

Retrodigitalisierung  
in  
großem Stil

Retrodigitalisierter  
Archivbestand  
ist ein Schatz

Imprint
Springer Berlin Heidelberg
Kluwer
Springer New York
Gabler Verlag
VS Verlag
Plenum US
Dt. Universitätsverlag
Vieweg+Teubner
Springer Netherlands
Leske & Budrich
Springer Vienna
Westdeutscher Verlag
Vieweg Verlag
Martinus Nijhoff
Birkhäuser Basel
D. Reidel Publishing Company
Physica-Verlag
Steinkopff-Verlag
Chapman & Hall
J.F. Bergmann-Verlag

Top 20 Springer Imprints für Springer Book Archives

Professional-Bereich. Ebenso sind Bücher von Imprints wie Steinkopff oder Kluwer Academic Publishers, die längst nicht mehr aktiv sind, Teil des Mammutprojekts.

## Identifizieren, konsolidieren, strukturieren

Ein Projekt dieser Größenordnung kann Vollständigkeit immer nur anstreben. Jeden einzelnen publizierten Titel zu identifizieren und zu finden, ist nahezu unmöglich, dennoch angestrebt. Um der Lückenlosigkeit so nah wie möglich zu kommen, war es nötig, eine komplette Bestandsaufnahme aller Imprints zu erstellen. Hierzu wurden alle zur Verfügung stehenden Quellen wie Kataloge, persönliche Aufzeichnungen oder Handarchive gesammelt und ausgewertet. Aufgrund der internationalen Unternehmensstruktur und der langen Geschichte des Verlages war dies kein einfaches Unterfangen.

Vollständigkeit wird angestrebt

Eine weitere Schwierigkeit bilden die bibliografischen Daten. Die Datenlage bei Büchern ohne ISBN ist teilweise verheerend und eine eindeutige Identifizierung und Aufnahme in den Titeltitelkatalog ist eine enorme Herausforderung. Neben Publikationen ohne jegliche Daten gibt es solche mit fehlerhaften, unvollständigen oder widersprüchlichen Angaben. Um auf so verlässlichen Daten wie möglich aufbauen zu können, arbeitete der Verlag verstärkt mit Nationalbibliotheken und Bibliotheksverbänden zusammen und wertete deren Katalogdaten aus. Diese, sowie die Daten von Springer, fließen in einer Datenbank zusammen und werden mittels komplexer Algorithmen und Fuzzy-Matching-Prozessen mit dem Ziel konsolidiert, eine möglichst vollständige, fehlerminimierte Titelliste zu generieren – ein Arbeitsschritt, der nie abgeschlossen sein wird. Regelmäßig werden alle neuen Informationen in die Datenbank importiert und mit dem Ziel ausgewertet, die Datenqualität stetig zu verbessern.

Schwierige Datenlage bei Büchern ohne ISBN

Neben dem allgemeinen Anspruch, das gesamte Archiv zu komplettieren, wurden bestimmte Bereiche definiert, für die eine hundertprozentige Vollständigkeit Pflicht ist. Das sind z. B. 100 Top Buchserien, die in Zusammenarbeit mit den Planungsabteilungen identifiziert wurden. Für diese Buchserien wird Springer Book Archives eine weitgehend lückenlose Dokumentation aller in den Serien erschienenen Titel anbieten. Dies gilt ebenso für die »aktuellsten« Publikationen, d. h. für Bücher, die zwischen den Jahren 2000 und 2004 publiziert wurden. Ein besonderes Augenmerk wird auf den Kerndisziplinen des Verlags liegen. Hier sind beispielsweise Mathematik, Technik oder Naturwissenschaften zu erwähnen.

Top mit 100 Buchserien

## Hunderttausend Titel aus zwölf Themengebieten in zwei Sprachen

Insgesamt umfasst Springer Book Archives Titel aus verschiedensten Disziplinen, die in elf englischen und fünf deutschen Themenpaketen gebündelt werden.

Themenpakete Englisch Springer Book Archives
Behavioral Science
Biomedical and Life Sciences
Business and Economics
Chemistry and Materials Science
Computer Science
Earth and Environmental Science
Engineering
Humanities, Social Sciences and Law
Mathematics and Statistics
Medicine
Physics and Astronomy

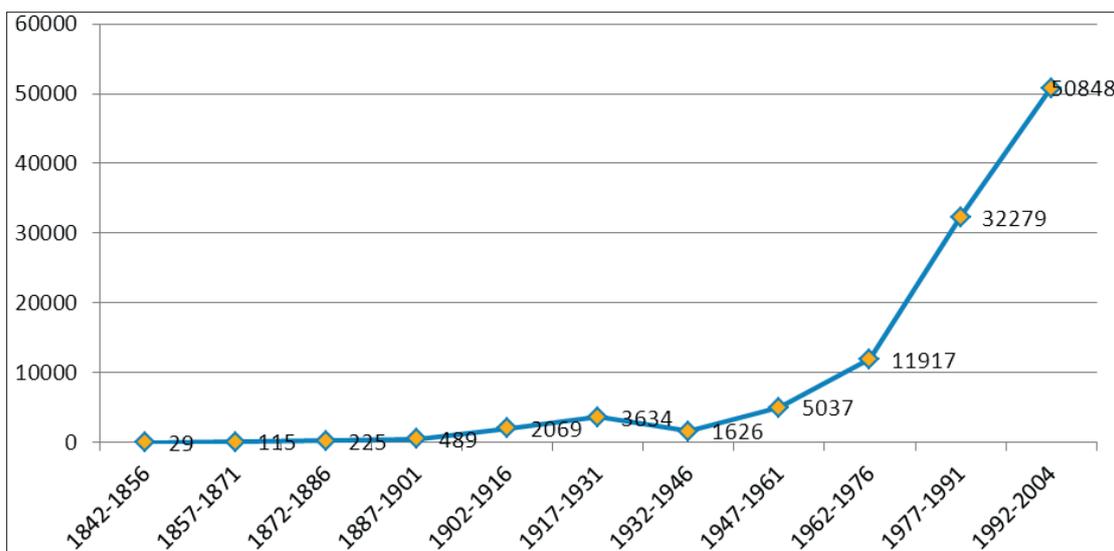
Themenpakete Deutsch Springer Book Archives
Geistes-, Sozial- und Rechtswissenschaften
Medizin
Naturwissenschaften
Technik und Informatik
Wirtschaftswissenschaften

Themenpakete von Springer Book Archives

Der Großteil der bereitgestellten Titel stammt aus den letzten drei Jahrzehnten. Dies erklärt sich zum einen aus der Tatsache des immer größer werdenden Outputs von wissenschaftlicher Literatur und zum anderen aus der gewachsenen Unternehmensstruktur.

Außer der Beschränkung auf die beiden Kernsprachen des Verlags, Deutsch und Englisch, gibt es keinerlei Auswahlkriterien für die Inhalte des Archivs, da ein Archiv grundsätzlich den Anspruch auf Vollständigkeit abbildet. Wurden von einem Werk mehrere Auflagen veröffentlicht, so werden alle Auflagen als Online-Version mit in das Projekt einbezogen und nur die letzte Auflage als Print-on-Demand-Titel über Digitaldruck in gedruckter Form verfügbar sein. Derzeit ist Springer der einzige wissenschaftliche Verlag, der in ein solches, auf Vollständigkeit basierendes Projekt investiert.

Alle Inhalte aus englischen und deutschen Büchern



Springer Book Archives nach Copyright-Jahren

## Auf der Suche nach dem physischen Buch

Nachdem potenzielle Titel identifiziert sind, folgt die Suche nach dem physischen Buch. Neben den eigenen Handarchiven greift Springer dabei auf den antiquarischen Buchhandel zurück. Bei Titeln, die weder in den eigenen Archiven noch über den Markt zu bekommen sind, wird auch der Weg von Kooperationen mit Bibliotheken und Institutionen gegangen. So kam die sehr erfolgreiche Kooperation mit der Deutschen Nationalbibliothek (DNB) zustande. Die Zusammenarbeit begann im Juli 2012 in Frankfurt am Main. Zum Jahresende 2012 waren es bereits knapp 5.000 Bücher (somit mehr als 1 Mio. Seiten), die in der DNB gescannt werden konnten. Dies ist die erste Kooperation in dieser Größenordnung. Aufgrund des Erfolges ein absolut positiver Präzedenzfall in puncto Kooperation.



On-site scanning Kooperation mit der DNB

## Trommeln auf allen Kanälen: Die Klärung der Rechte

Parallel zum Scanprozess findet ein weiterer wichtiger Bestandteil des Projekts statt: Die Bearbeitung der Rechtesituation. Die Klärung der Rechte und Lizenzgebühren hat oberste Priorität im Projekt. Springer unternimmt alle erdenklichen Anstrengungen, jeden einzelnen Autor eines Werkes zu kontaktieren. In verschiedenen Kampagnen, sei es per Post, über E-Mail oder Social Media, arbeitet Springer pro-aktiv daran, rund 150.000 Autoren zu kontaktieren, um über das Projekt zu informieren und die Rechtesituation zu klären. Es gibt kaum

eine Abteilung, die nicht in dieses Projekt eingebunden ist: Von der Programmplanung über die IT, vom Marketing über die Herstellung bis hin zur Personalabteilung. Und auch geografisch ist Springer Book Archives ein Projekt, das jeden Springer Standort einbezieht, ob Heidelberg, London, New York oder Tokio, um nur einige zu nennen.

Rechtklärung ist unabdingbar

## Ein Scan sollte mehr darstellen als eine Kopie

Vor dem Scannen muss jedes physische Buch überprüft werden: Sind alle Seiten konsistent durchnummeriert, fehlen Seiten, müssen handschriftliche Anmerkungen entfernt werden etc.? Alte und fragile Bücher müssen mit besonderer Sorgfalt gescannt werden.



Häufig wird unter Digitalisierung verstanden, dass von einem gedruckten Dokument ein einfacher Scan gemacht wird, sozusagen eine Kopie. Um die Vision Springer Book Archives zu erfüllen, reicht ein Scan des Buches allerdings nicht aus. Denn nur die höchsten Qualitätsstandards finden für die Produkte von Springer Book Archives Anwendung. Die Titel werden nicht nur als komplettes Buch verfügbar sein, sondern auch über die Plattform SpringerLink als durchsuchbares E-Book bereitgestellt werden, das es erlaubt, auf die einzelnen Kapitel Ebenen zuzugreifen. So haben Kunden nicht nur die Möglichkeit, das gesamte Buch zu erwerben, sondern auch einzelne, ausgewählte Kapitel. Darüber hinaus bietet Print-on-Demand dem Interessenten die Option, ein möglicherweise seit Jahr-

Höchste Qualitätsstandards

zehnten vergriffenes Buch wieder physisch als gedrucktes Exemplar in den Händen halten und damit arbeiten zu können. Für englischsprachige Titel wird mit der MyCopy-Option eine zusätzliche Version des ursprünglichen Buches angeboten. Jeder Titel wird also in mindestens zwei neuen Formaten erhältlich sein: In elektronischer und in gedruckter Form. Aufgrund der verschiedenen Produkte, die aus einem Titel generiert werden, ist es unbedingt erforderlich, mit höchsten Qualitätsstandards zu arbeiten. Gescannt wird mit modernsten Buchscannern. Text und schwarz/weiß kompatible Abbildungen werden mit 600 dpi, Farbabbildungen und alle grayscale kompatiblen Tabellen, Graphen etc. mit 300 dpi gescannt. Nach dem Scannen wird jede einzelne Abbildung mittels individueller Bildbearbeitung für die Publikation in den gewünschten Formaten bearbeitet. Ebenfalls werden bibliografische Metadaten wie Autor, Titel, Serientitel etc. extrahiert. Jede Seite wird digital von eventuellen Fehlern wie Staubpartikeln bereinigt und geglättet. Um die Volltextsuche im E-Book zu gewährleisten, wird der Text mittels eines Texterkennungsprogramms, OCR (optical character recognition), analysiert. Aus den individuell überarbeiteten Abbildungen, den Metadaten und dem Text entstehen dann die einzelnen Formate für Springer Book Archives. Die Produktion umfasst die Herstellung eines web-optimierten PDFs sowie eines PDFs für das gedruckte Buch durch Digitaldruck (POD, Print-on-Demand). Ebenfalls werden XML-Metadaten der bibliografi-

schon Daten des Buches produziert. Diese dienen als Basis für Metadatenformate wie MARC, die es Bibliotheken ermöglichen, die bibliografischen Daten direkt automatisiert in ihre Bibliothekskataloge hochzuladen. Für ausgewählte Titel wird zukünftig die Volltext-XML-Produktion angestrebt. Während jedes einzelnen Produktionsschritts gibt es Kontrollen, um die Qualität des endgültigen Produkts zu gewährleisten. Aufgrund ihrer Bedeutung für die Sichtbarkeit und Auffindbarkeit in Bibliothekskatalogen und im Internet stehen die Metadaten unter besonderer Kontrolle.

Qualitätsfokus: Metadaten

Option für Print-on-Demand

### Und wem nutzt das alles ...?

... den Wissenschaftlern  
 Forschern wird durch Springer Book Archives die Möglichkeit geboten, heute und zukünftig auf das Wissen und die Erkenntnisse früherer Publikationen zuzugreifen. Wissenschaftler können nun auf Buch- und oder Kapitelebene auf Inhalte zurückgreifen und aus Primärquellen zitieren, die Ihnen bis dato lange nicht zur Verfügung standen. So können sie historische Präzedenzfälle studieren, die ihre aktuelle Forschung unterstützen und neue bahnbrechende Ergebnisse fördern können. Springer Book Archives bietet ihnen die Möglichkeit, die Entstehungsgeschichte wissenschaftlicher Erkenntnisse oder die Entwicklung von Patenten nachzuvollziehen, und diese auf die aktuelle Forschung zu beziehen.

Output formats					
Features	<ul style="list-style-type: none"> <li>• Print-on-demand (POD)</li> </ul>	<ul style="list-style-type: none"> <li>• Full-text searchable, high-resolution e-book</li> </ul>	<ul style="list-style-type: none"> <li>• Available for selected titles</li> </ul>	<ul style="list-style-type: none"> <li>• Metadata, ONIX, MARC</li> <li>• FullText XML for selected titles</li> </ul>	<ul style="list-style-type: none"> <li>• For English-language titles</li> </ul>

Produkt-Formate von Springer Book Archives

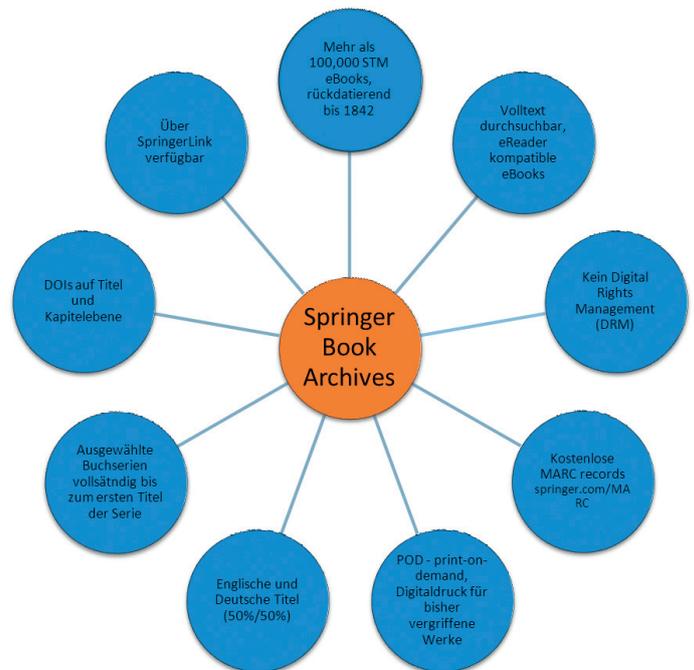
... den Bibliotheken und Institutionen  
 Springer Book Archives bietet Bibliotheken die Möglichkeit, einen Schritt weiter in Richtung virtuelle Bibliothek zu gehen. Alle Fachgebietspakete können gebündelt in einem Archiv erworben werden. Für alle Titel der Springer Book Archives stehen Metadaten in Form von Springer MARC Records und zudem durch OCLC angereicherte MARC Records kostenfrei zur Verfügung. Die MARC Daten können von Bibliotheken direkt in den OPAC geladen werden und bieten mittels einer darin enthaltenen URL zu SpringerLink elektronischen Zugriff auf die Titel. Im Sinne des Bestandschutzes können die bibliothekseigenen Präsenzexemplare so immens geschont werden. Ein weiterer Vorteil ist die physische Platzersparnis – für 100.000 Titel in elektronischer Form muss kein Lagerraum bereitgestellt werden.

Schutz für Präsenzexemplare der Bibliotheken

... den Autoren  
 Autoren, deren Werke bisher vergriffen waren, bekommen durch Springer Book Archives wieder eine größere Sichtbarkeit. Ihre Forschungsergebnisse sind für jeden, überall und jederzeit über SpringerLink elektronisch abrufbar und die Bücher über die POD-Option als gedrucktes Buch wieder bestellbar. Aufgrund der Tatsache, dass jeweils nur die aktuellste Auflage eines Buches als gedruckte Version verfügbar sein wird, besteht für den Autor keine Gefahr des Absatzverlustes von aktuellen Auflagen. Ebenso wird gewährleistet, dass der Kunde immer die aktuellste Version als Printexemplar erhält.

Sichtbarkeit aller Ausgaben

paket mit Titeln von 1842 bis 2004 in Englisch und Deutsch oder eines mit nur einer der beiden Sprachen zu erwerben. Der Preis für die Pakete hängt dabei von der Größe der Institution ab. Individualkunden können über [springer.com](http://springer.com) schon seit August 2012 einzelne POD-Titel in gedruckter Form erwerben. Außerdem wird es die Möglichkeit geben, E-Books über [springer.com](http://springer.com) und andere Online-Vertriebswege wie [amazon](http://amazon.com) zu erwerben. Bereits seit Dezember 2012 sind mehr als 35.000 englische Titel auf SpringerLink verfügbar. Im Laufe des Jahres werden noch rund 65.000 Springer Book Archives Titel hinzukommen. Insgesamt erwarten wir, Ende 2013 rund 50.000 englische und 50.000 deutsche Titel auf SpringerLink zur Verfügung stellen zu können.



## 2013 – ein Archiv geht »live«

Wie auch die aktuellen E-Book-Pakete von Springer wird Springer Book Archives auf dauerhaftem Zugang beruhen. Bibliotheken und Institutionen haben die Möglichkeit, entweder das ganze Archiv-

Anschrift von Vanessa Thiel: Coordinator  
 Springer Book Archives, Springer-Verlag,  
 Tiergartenstraße 17, 69121 Heidelberg,  
 E-Mail: [vanessa.thiel@springer.com](mailto:vanessa.thiel@springer.com)