



- Auditing
- Accounting
- Valuation
- Analysis



FRIEDRICH-ALEXANDER  
UNIVERSITÄT  
ERLANGEN-NÜRNBERG

FACHBEREICH WIRTSCHAFTS-  
UND SOZIALWISSENSCHAFTEN

## Wegweiser Digitalisierung

Lehrstuhl für Rechnungswesen und Prüfungswesen / Prof. Dr. Klaus Henselmann

*Stand Oktober 2019*

Die Digitalisierung in der Wirtschaft hat dazu geführt, dass neben das real existierende Unternehmen ein immer vollständiger werdendes elektronisches Abbild desselben getreten ist. Das Rechnungswesen ist dabei häufig in komplexe ERP-Systeme eingebunden. ES stellt darin neben Funktionen wie z.B. Einkauf, Produktionsplanung, Logistik, Marketing, nur ein Teilsystem dar. Auch regelmäßige und aperiodische Finanzberichte werden per Gesetz zunehmend in speziellen digitalen Datenformaten verlangt (E-Bilanz, ESEF u.a.).

Durch die Verfügbarkeit der Daten in digitaler Form ergeben sich neue Aufgaben, aber auch Chancen. Für deren Bewältigung ist die Beschäftigung mit verschiedenen IT-Systemen und Standards notwendig.

Einen Überblick zum Lehrangebot am Fachbereich verschafft der Menüvorschlag „Digitization & Analytics in FACT“ (<https://www.fact.rw.fau.de/master-fact/studienprofil-und-studienangebot/berufsfeldspezifische-menuevorschlaege/>). Aber auch unabhängig von Bachelor oder Master gibt es außerhalb der FAU eine Reihe von guten und kostenlosen Online-Schulungsangeboten. Der folgende Wegweiser soll hierfür eine Orientierungshilfe bieten.

Der Lehrstuhl beschäftigt sich seit langem mit der Digitalisierung der Finanzberichterstattung in verschiedenen Facetten. Dies gilt sowohl für die Forschung als auch in der Lehre. Für die Bearbeitung bestimmter Themen in Master-, Seminar- oder Bachelorarbeiten sind IT-Kenntnisse, insbesondere über Datenformate und zur Datenanalyse, erforderlich.

### ERP-Systeme

Praxiskenntnisse von ERP-Systemen sind weniger für Abschlussarbeiten als für das spätere Berufsleben relevant.

Das bei großen Unternehmen verbreitetste ERP-System stammt von SAP. Das System SAP S/4HANA (<https://www.sap.com/germany/products/s4hana-erp.html>) ist äußerst komplex. Jedoch sind bereits Grundlagenkenntnisse sehr hilfreich und erleichtern eine spätere Vertiefung in der Praxis.

Am Fachbereich gibt es das Training „TS410 – Integrated Business Process in SAP S/4HANA“ als Blockveranstaltung bei Prof. Matzner (<https://www.is.rw.fau.de/>). Der Kurs schließt mit einer Zertifizierungsprüfung durch SAP ab und führt nach erfolgreichem Bestehen zu dem in der Wirtschaft voll anerkannten und weltweit gültigen SAP-Zertifikat „SAP Certified Business Associate“.

Daneben existieren Angebote zu SAP bei der Virtuellen Hochschule Bayern (<https://www.vhb.org/>) wie die Veranstaltung „SAP-Geschäftsprozesse 1: Grundlagen“ oder andere SAP-Kurse. Bis auf weiteres gilt (solange entsprechende Angebote der vhb Virtuelle Hochschule Bayern verfügbar

sind): Bei erfolgreicher Belegung des Kurses buchen wir im Hinblick auf die erlangte verbesserte IT-Kompetenz ein „bestanden“ unter dem Modul „Schlüsselqualifikationen“ des FACT-Masters direkt ein. Wichtig: Der Kurs kann über die vhb wahlweise mit 3 ECTS oder 5 ECTS studiert werden. Entsprechend wird eine Anrechnung entweder mit 2,5 ECTS oder 5 ECTS vorgenommen. Die Studenten müssen sich für die Anrechnung selbständig für die entsprechenden Module (Schlüsselqualifikation 1 FACT Prüfungsnummer: 66001 bzw. Schlüsselqualifikation 2 FACT Prüfungsnummer: 66002) anmelden und das im vhb-Kurs erworbene Zertifikat vorlegen (Studienleistungen werden im Modul Schlüsselqualifikationen nur unbenotet mit „bestanden“ bewertet). Es wird ggf. zusätzlich seitens der vhb-Partnerhochschule eine Zertifizierung angeboten, die Studenten ihren Bewerbungsunterlagen hinzufügen können.

Ein weiterer Anbieter von Kursen zu SAP ist openSAP (<https://open.sap.com/pages/about>).

Hinsichtlich kleinerer Unternehmen bzw. deren Steuerberatern sind Kenntnisse der DATEV-Software empfehlenswert (z.B. <https://www.steuerverfahren.rwth-aachen.de/lehre/datev-fuehrerschein/>).

## **Datenformate für Finanzberichte und Prüfung**

Die Auswertung digitaler Finanzberichte bildet einen Arbeitsschwerpunkt des Lehrstuhls.

a) Jahresabschlüsse und andere Finanzberichte wurden früher klassisch in Papierform erstellt. Heute sind Exemplare in digitaler Form als pdf oder html-Datei üblich.

Die Inhalte von pdf wie auch html-Dateien können am Bildschirm oder auf einem Ausdruck zwar problemlos von Menschen gelesen werden können. Jedoch sind einer automatisierten Verarbeitung dadurch Grenzen gesetzt, dass ein Computerprogramm für eine bestimmte Zahl wie „380“ aus dem sie umgebenden Text in der Regel nicht zweifelsfrei sagen kann, für welche Bedeutung (z.B. Eigenkapital oder Stammkapital?, für 2018 oder 2019?) diese Zahl steht.

Das Datenformat XML vermeidet diese Problem, da sog. „Tags“ (Etiketten) dem Computer für alle Inhalte die Bedeutung klarmachen. Eine spezielle Form von XML für Finanzberichte stellt XBRL dar (<https://www.xbrl.org>, <https://www.xbrl.eu>, <https://www.xbrl.de>). Existierende Standards in Deutschland sind die E-Bilanz für die Besteuerung (<https://www.estuer.de>) und DiFin für die Kreditwirtschaft (<https://www.digitaler-finanzbericht.de>). Zur Standardisierung der „Tags“ bedient man sich sog. Taxonomien.

Der Text dieser Dateien ist allerdings wiederum für Menschen schlecht zu verstehen. Daher gibt es auch Kombinationsmöglichkeiten in Form von (X)HTML mit eingebetteten XML. (Beispieldokumente eines fiktiven kleinen Unternehmens hier; Videos zur Erläuterung für pdf, docx, xlsx, html, html mit css, xml, xml mit Attributen, xhtml mit xml.)

Bei (X)HTML mit eingebettetem XBRL spricht man von iXBRL (Inline XBRL).

Ab 2020 gilt in der EU für die Konzernabschlüsse kapitalmarktorientierter Unternehmen verbindlich das sog. European Single Electronic Format (ESEF). Hierbei handelt es sich im Prinzip um iXBRL für IFRS-Abschlüsse ([www.esef.eu](http://www.esef.eu)).

Das XBRL-Tagging ist auch Gegenstand der Abschlussprüfung. Die Unterzeichnung der Abschlüsse durch den Vorstand, die Feststellung durch den Aufsichtsrat und der Bestätigungsvermerk des Abschlussprüfers erfolgen durch eine digitale Signatur der Datei statt der Unterschrift auf einem Papierdokument.

b) Das „American Institute of CPAs“ (AICPA) empfiehlt seit 2015 „Audit Data Standards“ zum Einsatz in der Abschlussprüfung (<https://www.aicpa.org/>). Dabei handelt es sich um entweder um eine flache Textdatei mit bestimmten Feldlängen oder um das „XBRL Global Ledger Taxonomy Framework“ (XBRL GL).

## Datenanalysen / Data Analytics

Die Erstellung eigener Datenanalysen steht im Zentrum vieler Abschlussarbeiten.

a) Für die Analyse von Datenbeständen in der Prüfung werden für das sog. Journal Entry Testing meist kommerzielle Produkte wie IDEA und ACL verwendet. Vielfältige grafische Auswertungen ermöglicht Tableau. Für Studierende und deren Dozenten sind Software und Schulungen kostenlos (<https://www.tableau.com/de-de/academic>).

Jedoch kann man zahlreiche Funktionalitäten dieser Spezialsoftware auch in Excel ausprobieren, welches in den Menüs versteckt auch sehr mächtige Funktionen bietet. In Excel sollte man sich mit der Analyse von Daten durch Filter und Pivot-Tabellen sowie mit der Erstellung von Datenabfragen beschäftigen. (Aufgabe und Daten, Video zum Filtern, Video zu Pivot-Tabellen, Video zu überfälligen Forderungen - aber mit englischen Excel-Menüs!)

b) Eine immer beliebtere Form der Datenanalyse stellt das Process Mining dar. Hier werden keine statischen Datenbestände, sondern Ketten von zeitlich aufeinander folgenden Arbeitsschritten analysiert. Ziel kann es beispielsweise sein, Abweichungen von dem eigentlich vorgeschriebenen Ablauf zu finden, z.B. bei Bestellungen die Nutzung eines nicht zugelassenen Lieferanten, eine Zahlung vor der offiziellen Auftragserteilung oder das Unterlassen der Wareneingangskontrolle.

Ein sehr umfangreiches Open Source Programm ist ProM ([www.promtools.org](http://www.promtools.org)) der Eindhoven University of Technology. Für die Ausbildung empfiehlt sich die Version ProM Lite. Es gibt kostenlose Online Kurse, u.a. auf <https://www.futurelearn.com/courses/process-mining>.

Für den FACT Master wird die Veranstaltung „Process Analytics“ von Prof. Matzner angeboten.

c) Eine mächtige Statistiksoftware, die heute Unternehmen in der Praxis einsetzen, ist R. Auch für die allgemeine Programmiersprache Python gibt es spezielle Statistikpakete. Sowohl für R als auch für Python existieren Erweiterungen zu Data Analytics, etwa Maschinellem Lernen (Künstlicher Intelligenz), und grafische Auswertungen. Python ist dabei noch universeller und in der Data Analytics Praxis beliebter. Die Lehrstühle für Statistik und Wirtschaftsinformatik am Fachbereich haben verschiedene Ausbildungsangebote.

d) Daten liegen häufig auch in qualitativer Form, insbesondere als Texte vor (Anhang, Lagebericht, Analyst Report, Ad-hoc-Meldungen, Wertpapierprospekte u.a.). Hier kommen Verfahren des Text Mining zu Einsatz. Es gibt Pakete für R und insbesondere Python. Beispiele für Verfahren des Textmining sind

- das Parsing von Texten zur Extraktion bestimmter Inhalte (Beispiel zur Analyse der Inhalte von Mietverträgen: Aufgabe und Daten, Einführungsvideo, Video zu den Verträgen, Aufgabenvideo, Lösungsvideo);
- die Messung von Sprachtendenzen wie etwa einem mehr oder weniger „positiven“ bzw. „negativen“ Sentiment (Beispiel zum Sentiment von CEO letters: Aufgabe und Daten, Video).

## **Datenanalysen / Data Analytics mit Rapidminer**

Ein Alternative für Studierende, die sich für Datenanalysen und Maschinelles Lernen (Künstlicher Intelligenz) interessieren, aber die von der Einstiegshürde einer Programmiersprache abgeschreckt werden, ist die Nutzung von „Rapidminer Studio“ (<https://rapidminer.com/products/studio/>). RM ist ein führendes Tool für „Data Science and Machine Learning“, das wegen seiner graphischen Benutzeroberfläche sowohl leicht zu bedienen als auch leistungsfähig ist (<https://rapidminer.com/resource/gartner-magic-quadrant-data-science-platforms/>).

Es wird daher bei uns häufig für Abschlussarbeiten genutzt.

RM wird auch von Prof. Bodendorf in der Wirtschaftsinformatik eingesetzt. Es ermöglicht den leichten Einstieg in Data Science Methoden, ohne Kenntnisse einer Programmiersprache wie Python oder R vorauszusetzen. EY, PwC, Deloitte, Warth & Klein u.a. nutzen RM. Die erlernten Methoden können aber später auch in anderen Tools genutzt werden.

Die Grundversion von RM wurde an der Universität Dortmund entwickelt und steht daher jedem frei und kostenlos zur Verfügung (bis 10.000 Datensätze). Kostenlose volle akademische Lizenzen werden Studenten, Professoren und Forschern auf Antrag eingeräumt ([https://rapidminer.com/pricing/#studio\\_pricing](https://rapidminer.com/pricing/#studio_pricing)).

RM umfasst Befehle für Lineare Regression, Logistische Regression, Entscheidungsbäume, Naïve Bayes Klassifikator, (einfache) Neuronale Netze, Deep Learning Neuronale Netze, Support Vector Machines, Diskriminanzanalyse, Clustering, Assoziationsregeln usw. (Video zur Vorhersage von Überlebenden der Titanic mit Verfahren des Maschinellen Lernens.)

Die Fähigkeiten von RM lassen sich durch sog. Extensions für spezielle Anwendungsbereiche erweitern. Dafür gibt es einen sog. Marketplace. Entgegen dieser Bezeichnung sind jedoch die meisten Extensions kostenlos verfügbar. Zu den Erweiterungen gehören

- „Text Processing“, „Web Mining“, „Facebook Extension“ (der Zugriff auf Twitter gehört zur Grundversion);
- „RapidProM“ für Befehle zum Process Mining mit ProM innerhalb von RM;
- „NoSQL Connectors“ für einen Zugriff auf die Datenbank MongoDB (der Zugriff auf SQL-Datenbanken gehört zur Grundversion).

Die Rapidminer Academy (<https://academy.rapidminer.com/>) umfasst von kurzen Beispielvideos über komplette Onlinekurse bis hin zu Zertifizierungen ein kostenloses Ausbildungsangebot.

Das Seminar „Data- und Textmining in Accounting Research“ findet in regelmäßigen Abständen als Lehrauftrag statt. Die Installation der Software auf dem eigenen Laptop (Bring your own device!) und das Studium erster Einführungsvideos sind Teilnahmevoraussetzungen.

## **Datenbanksysteme / DBMS**

Innerhalb von ERP-Systemen und anderer Software werden zur Speicherung Datenbanken (Data Base Management Systems DBMS, DB engines) verwendet. Es gibt unterschiedliche Ansätze und innerhalb dieser wiederum verschiedene Produkte, teils kostenlos als Open Source (siehe <https://db-engines.com/de/>).

Ein Klassiker ist die Datenspeicherung in tabellarischer Form. Man spricht hier von „relationalen“ DBMS. Datenabfragen und Bearbeitung (Einfügen, Verändern, Löschen) erfolgen mit der Sprache SQL (Structured Query Language). Es gibt im Internet zahlreiche Tutorials zu SQL (suche nach „SQL tutorial“). Ein weitgenutztes DBMS ist das schlanke „SQLite“ (<https://www.sqlite.org/index.html>), das sich auch gut zur lokalen Installation auf einem PC eignet. Die maximale Größe der Datenbank beträgt dennoch 140 Terabytes. Es gibt umfangreiche Unterlagen zu SQLite auf der Webseite.

Für nur schwach, uneinheitlich oder gar nicht strukturierte Daten (wie insbesondere Texte) bieten sich DBMS in Form eines „Document Store“ an. Das verbreitete und mächtige MongoDB (<https://www.mongodb.com/>) ist ebenfalls Open Source. Man kann es lokal auf dem eigenen PC installieren, es gibt aber auch eine cloudbasierte Version „MongoDB Atlas“ (<https://www.mongodb.com/cloud/atlas>), die keine eigene Installation erfordert. Zum Lernen und für kleine Datenmengen ist die Nutzung einiger Cloudserver kostenlos. Die MongoDB University (<https://university.mongodb.com/>) bietet außerdem kostenlose Kurse einschließlich Zertifizierungen an.

Künftig an Bedeutung gewinnen werden die Datenspeicherung und der Datenaustausch als RDF-Tripel, denn sie bilden eine Grundlage des semantischen Webs. Die Europäische Union betreibt ein Linked Open Data Portal mit RDF-Tripeln ([data.europa.eu/euodp/en/linking-data](http://data.europa.eu/euodp/en/linking-data)). Semantisches Web bedeutet, dass eine Maschine nicht nur die zur Verfügung gestellten Daten lesen kann (z.B. {2019-10-07, 60.45}, {2019-10-08, 59.95}), sondern dass sie bei entsprechenden Konventionen (in Form von Ontologien) auch deren Bedeutung erkennt. So könnte etwa der Kurs eines Wertpapiers beschrieben werden {"datum":2019-10-07,"schlusskurs":60.45} {"datum":2019-10-08,"schlusskurs":59.95}. SPARQL stellt eine SQL-ähnliche Abfragesprache für RDF-Daten dar.

Zu den Open Source DBMS vom Typ „RDF-Store“ gehört rdf4j (<https://rdf4j.eclipse.org/databases/>). Auf rdf4j beruht „Ontotext GraphDB Free“ (<http://graphdb.ontotext.com/>). Die genannten RDF-Stores stellen eine Untergruppe von „Graphen“ DBMS dar. Ein Open Source Beispiel für Graphen DBMS ist auch Neo4j (<https://neo4j.com/>).

Die „Panama Papers“, „Offshore Leaks“ und andere Datenbestände des International Consortium of Investigative Journalists sind in Neo4j gespeichert (<https://offshoreleaks.icij.org/pages/database>). Der Film „The Spider's Web: Britain's Second Empire“ (Fassung mit deutschen Untertiteln: <https://www.youtube.com/watch?v=1ZZR8vBKqwc>) beschreibt solche Offshorezentren, die auch für illegale Zwecke wie Steuerhinterziehung, Geldwäsche, Terrorfinanzierung u.a. verwendet werden. Er basiert u.a. auf dem Buch „Treasure Islands: Tax Havens and the Men who Stole the World“ von Nicholas Shaxson aus dem Jahr 2011.

Auf „open HPI“ (<https://open.hpi.de/pages/about>), der Online-Kursseite des Hasso-Plattner-Instituts, gibt es kostenlose Kurse u.a. zum semantischen Web – RDF – Linked Data – Verarbeitung von Sprache. Beispiele sind „Linked Data Engineering“ (<https://open.hpi.de/courses/semanticweb2016>) und „Information Service Engineering“ (<https://open.hpi.de/courses/semanticweb2017>).

Am Fachbereich behandelt Prof. Harth die Themen RDF-Daten und Linked Open Data.