

Kapitel 7 -

Die Konstruktion der Datenwissenschaften im akademischen Feld durch Begriffsarbeit und *boundary work*

»We are witnessing an Abbott (1988) style disciplinary turf war, with computer science, statistics, mathematics, information science, and various collaborations with scientific, state, and industrial sectors each making claims to the expertise, funding, and regulatory sphere of data science. No single configuration seems to hold, and at each academic institution I have visited or inspected, I have found a different arrangement of interlinked disciplines and sectors (and those left out or cut out) of the game« (Ribes 2019: 515f.).

7.1 Einleitung

Das akademische Feld ist eine zentrale gesellschaftliche Instanz für die Genese und Konstituierung neuer Wissensfelder. Sowohl individuelle als auch kollektive Akteur*innen entwerfen unterschiedliche disziplinäre, epistemische, organisationale und weitere Perspektiven auf entstehende Wissensgebiete. Die dabei entstehenden vielfältigen Bedeutungen rahmen und strukturieren einerseits Forschungsschwerpunkte, Professuren oder Studiengänge und positionieren diese andererseits im akademischen Feld. So materialisieren sich in der Struktur und im Aufbau neuer Curricula, aber auch in organisationalen Implementationen wie Forschungsinitiativen, -zentren oder -netzwerken insofern je nach epistemologischer, disziplinärer und institutioneller Verortung multiple Verständnisse der jeweiligen Wissensformation.

Daran anknüpfend untersuche ich, wie die Datenwissenschaften im akademischen Feld eingeführt werden und in welchem Verhältnis sie zu verwandten Wissensfeldern und Disziplinen stehen. Der Fokus der Analyse liegt dabei neben den Akteur*innen auf den Diskursen und Praktiken, die die Kategorie »Data Science« im akademischen Feld konstruieren. Zur Beantwortung der Fragestellungen stütze ich mich auf die inhaltsanalytische Auswertung qualitativer Interviews mit Lehrenden an Schweizer Hochschulen und Universitäten und stelle sie in Relation zu akademischen Diskursen über die Emergenz der Datenwissenschaften. Die Analyse identifiziert zwei zentrale Praxismodi in der Konstruktion der Datenwissenschaften im akademischen Feld. Erstens stellt Begriffsarbeit eine zentrale diskursive Praxis dar: Datenwissenschaften werden im akademischen Feld durch Definitionen, Kategorisierungen und Narrative

eingeführt und hergestellt. In den angeführten Begriffen artikulieren sich einerseits Unklarheiten bzw. divergierende Ansichten darüber, ob es sich um einen fundamental neuen Modus wissenschaftlicher Erkenntnisproduktion handelt, der alle Disziplinen transformieren wird, oder aber um eine neue Disziplin, die sich zunächst gegenüber anderen Disziplinen behaupten muss und sich schliesslich ins disziplinäre Gefüge einpassen wird. Andererseits manifestieren sich darin Suchprozesse nach den ›richtigen‹ Begriffen zur Beschreibung der beobachtbaren Veränderungen im akademischen Feld.

Zweitens sind Praktiken von *boundary work*, d. h. Grenzziehungen und -überschreitungen, fundierend in der Genese der Datenwissenschaften im akademischen Feld. Sie finden sich im empirischen Material in zwei unterschiedlichen Dimensionen, die für die involvierten Disziplinen und Akteur*innen eine wichtige Orientierungsfunktion haben. Zentral ist in einer ersten Dimension das Verständnis als Wissensfeld, das sich aus unterschiedlichen wissenschaftlichen Disziplinen zusammensetzt. Dabei wird ein interdisziplinärer »Kern« in den Computerwissenschaften, Statistik, Mathematik und Engineering verortet, der sich auf Methoden, Technologien und andere Formalisierungen abstützt. Wie ich bereits im Forschungsstand gezeigt habe, erheben diese – und einige weitere – Disziplinen Anspruch auf die epistemische Autorität für das Feld der Datenwissenschaften. Im empirischen Material artikuliert sich dies in intensiven Konflikten und Grenzziehungen zwischen den involvierten Disziplinen, insbesondere zwischen der Informatik und der Statistik. Gleichzeitig äussern sich darin allerdings auch Praktiken der Grenzüberschreitung, indem ein interdisziplinärer Kern, der auf jedes Forschungsfeld oder -problem anwendbar und in diesem Sinne universal ist, durch verschiedene Anwendungsbereiche, sogenannte *domains*, in der Wissenschaft, der Industrie, im Non-Profit-Sektor, in der staatlichen Verwaltung und in weiteren Feldern komplementiert wird. Die Differenzierung in eine interdisziplinär zusammengesetzte methodisch-technische Expertise und eine domänenspezifische Fachexpertise stellt das zentrale Organisationsprinzip vieler Institutionen der Datenwissenschaften dar und transzendent somit aktiv etablierte Grenzen zwischen verschiedenen Disziplinen.

Eine zweite Dimension betrifft die grundlegende Unterscheidung zwischen »Wissenschaft« und »Industrie«,¹ die auf unterschiedliche Handlungslogiken in sozialen Feldern im Sinne gesellschaftlicher Differenzierung abzielt. Im empirischen Material findet die Dichotomie eine Übersetzung in akademische Grundlagenforschung einerseits und angewandte Forschung in der Praxis andererseits. Obwohl die Grenzziehung historisch stets porös war (Ribes et al. 2019: 297), bleibt sie für viele interviewte Datenwissenschaftler*innen eine wichtige Referenz. Gleichzeitig wird durchaus anerkannt, dass die Differenzierung aufgrund transversaler Forschungskollaborationen und Karrierewege sowie eines beschleunigten Transfers von Wissen, das in Methoden, Instrumenten und Technologien kodifiziert ist, über die unterschiedlichen Felder hinweg noch instabiler wird. Zu beobachten sind insofern unterschiedliche Spielarten von Anwendungsorientierung, die sich in der Unterscheidung von echten und konstruierten Daten und Problemen einerseits und einer an ausserwissenschaftlichen Referenzen orientierten »operativen Exzellenz« andererseits manifestieren.

¹ »Industrie« wird dabei über die wirtschaftsstatistische Einteilung hinaus breit verstanden und umfasst sowohl traditionelle Technologiekonzerne (wie ABB) als auch Unternehmen der IT- und Softwarebranche.

Das Kapitel ist entlang der beiden zentralen Praktiken strukturiert, die Datenwissenschaften im akademischen Feld konstruieren: Zunächst untersuche ich, inwiefern sich am Begriff des Paradigmenwechsels multiple Verständnisse von Datenwissenschaften zwischen Universalisierung und Disziplinierung dokumentieren (Kap. 7.2). Anschliessend zeigt die Suche nach den richtigen Kategorien, wie unterschiedliche Begriffe angeführt werden, um die veränderten Modi der Wissensproduktion adäquat zu beschreiben (Kap. 7.3). »Data Science« als inklusiver Kategorie kommt dabei die Funktion der Verknüpfung unterschiedlicher Disziplinen und Bezugsfelder zu. Die nachfolgenden Teilkapitel behandeln den zweiten identifizierten Praxismodus: Praktiken von *boundary work* äussern sich zum einen in disziplinären Konfliktlinien, die Datenwissenschaften als Wissensfeld durchziehen und strukturieren (Kap. 7.4). Zum anderen werden durch die kontinuierliche Ausrufung von Interdisziplinarität und die Logik der Domänen aber auch etablierte Grenzen überschritten (Kap. 7.5). Schliesslich tragen die verschiedenen Spielarten von Anwendungsorientierung zur Differenzierung zwischen unterschiedlichen Positionen im akademischen Feld bei (Kap. 7.6). In einer abschliessenden Diskussion werden die zentralen Erkenntnisse der Analyse zusammengefasst und in Bezug zu den übergeordneten Fragestellungen der Arbeit gesetzt (Kap. 7.7).

7.2 Paradigmenwechsel? »Data Science« zwischen Universalisierung und Disziplinierung

Ein erstes zentrales Moment in der Konstruktion der Datenwissenschaften ist die Frage nach dem grundlegenden Verständnis der Wissensformation. Eine Vielzahl an Definitionen, Positionsnahmen und Begriffsverständnissen kursiert in wissenschaftlichen, ökonomischen, medialen und weiteren Diskursen, die alle spezifische Perspektiven auf die epistemische Formation werfen (vgl. für einen Überblick Kap. 3.1). Ein wichtiger Teil der Diskussion bezieht sich auf die veränderten Bedingungen der Wissensproduktion, die mit der Verfügbarkeit grosser Datenmengen und neuen Analysemöglichkeiten einhergehen. Diese kristallisieren sich anhand des Begriffs Paradigmenwechsel heraus, der multiple Verständnisse von Datenwissenschaften zwischen Universalisierung und Disziplinierung prägt.

Alle Interviewpartner*innen teilen die sehr allgemein gehaltene Definition, wonach sich »Data Science« vorrangig mit »Methoden und Techniken zur Analyse von Daten in verschiedenen Anwendungsgebieten« beschäftigt, wie es ein Professor formuliert (vgl. unten). Die Befragten identifizieren die dahinterstehenden soziotechnischen Veränderungen wie die permanente Verfügbarkeit von grossen Datenmengen aus unterschiedlichen Quellen sowie neue Systeme zu deren Speicherung, Prozessierung, Analyse oder Visualisierung als die Treiber einer Entwicklung, die gesellschaftsdiagnostisch als Datafizierung (Prietl & Houben 2018; Schäfer & van Es 2017; van Dijck 2014) beschrieben wird. Manche der befragten Personen waren bereits vor Jahrzehnten mit grossen Datenmengen konfrontiert, lange bevor Begriffe wie »Big Data« oder »Data Science« überhaupt bekannt waren:

»Als ich an der [Grossforschungsinstitution] war, haben wir brutal viele Daten produziert [...]. Heute produziert nicht mehr nur die Forschung, sondern überall werden sehr

viele Daten produziert. Unternehmen möchten diese nutzen können« (Prof_FH_A: 8:26).

Die Verfügbarkeit von Daten, neuen Technologien zu deren Speicherung und Prozessierung und methodische Innovationen kennzeichnen jene veränderte Epistemologie, d. h. Art und Weise, wie Wissen und Erkenntnisse produziert werden. Die Diagnose eines »computational turn« (Berry 2011) beschreibt nicht in erster Linie die Genese einer neuen Disziplin oder eines neuen Feldes, sondern eine umfassende Transformation wissenschaftlicher Erkenntnisproduktion an sich.² Im empirischen Material wird diese Entwicklung vor allem durch den Begriff des Paradigmenwechsels auf den Punkt gebracht: Er bezeichnet den Übergang zu einem neuen Modus der Wissensproduktion, der auf der Verfügbarkeit von »Big Data« und neuen Möglichkeiten zu deren Analyse basiert (Gray 2009: xviii; Kitchin 2014).

Ein Professor für Mathematik umschreibt den Paradigmenwechsel anhand seines eigenen Werdegangs wie folgt:

»[Ich] könnte sagen, es geht um Methoden und Techniken zur Analyse von Daten in verschiedenen Anwendungsgebieten. Das gab es früher nicht, es ist eine extreme Beschleunigung von mathematischen und statistischen Innovationen, erstmals seit Jahrzehnten. Plötzlich [gibt es] sehr grosses Interesse einer riesigen Zahl von Leuten, [das] habe ich in meiner 25-jährigen Tätigkeit in diesem Feld noch nie erlebt, vor allem in der Mathematik und deren Verknüpfung mit technologischen Aspekten und Fragen. Heute beschäftigen sich Armeen von Leuten damit und schaffen völlig neue Angebote, nicht nur in der Forschung, sondern vor allem in der Lehre, und das grösstenteils gratis! Ich denke wirklich, dass wir in aussergewöhnlichen Zeiten leben, es ist nicht nur ein weiterer Hype! [lacht] [...]. Es ermöglicht ganz viele neue Fragestellungen in verschiedenen Gebieten und bietet Möglichkeiten für Anwender in der Industrie, aber auch Akademiker« (Prof_UH_C: 3:44).

Kennzeichnend für den Paradigmenwechsel ist nicht nur die Art und Weise, wie wissenschaftliches Wissen heutzutage produziert wird, die der Mathematikprofessor im Zitat euphorisch beschreibt, sondern die artikulierte Verknüpfung von wissenschaftlicher Forschung, Methoden der Datenanalyse und Anwendungen »in der Industrie« bzw. in »Unternehmen«, die in beiden Zitaten deutlich wird. Die Immanenz feldüberschreitender Bezüge ist ein wiederholtes Charakteristikum der Konstruktion der Datenwissenschaften, sowohl in den Interviews als auch im feldspezifischen Diskurs: Methoden und Techniken zur Analyse von Daten sind zwar im wissenschaftlichen Feld entwickelt worden, ihre eigentliche Bedeutung erhalten sie allerdings erst in Anwendungsmöglichkeiten der *domains* – wissenschaftliche und vor allem industrielle Anwendungsgebiete,³ in denen es »echte Probleme« zu lösen gilt. Indem der datenintensive Modus der Erkenntnisproduktion zur Grundlage des Paradigmenwechsels gemacht wird, wird diesem gleichzeitig eine universelle Relevanz für alle Disziplinen

2 Der Wandel äussert sich unterschiedlich in wissenschaftlichen Feldern wie den Sozialwissenschaften (Lazer et al. 2009; Chang et al. 2014), der Biologie (Leonelli 2016) oder der Medizin (Trajanoski 2012).

3 So beantworten diverse Studienprogramme die Frage nach der Verortung des Paradigmenwechsels dahingehend, dass »Data Science« in erster Linie »aus den Unternehmen« komme (EPFL 2017: 3).

attribuiert. Damit positionieren sich die Datenwissenschaften in diesem Sinne als »universelle Wissenschaft« (Ribes 2019; Ribes et al. 2019), die nicht nur für die Wissenschaften, sondern für *alle* gesellschaftlichen Bereiche fundamental ist.

Die Repräsentation von »Data Science« als einem Paradigmenwechsel und – daran anknüpfend – einer »universellen Wissenschaft« ist eine wichtige Lesart des Feldes. Parallel dazu begleitet die Datenwissenschaften von Beginn weg die Frage, wie sich ein transversales Wissensgebiet in ein disziplinär strukturiertes Feld einpassen lässt. Akademische Diskussionen über die Inklusion und Exklusion bestimmter Disziplinen in bzw. aus dem Kernbereich der Datenwissenschaften führen zu intensiven epistemischen und organisationalen Konfliktlinien. Neben dem akademischen Feld beabsichtigen in erster Linie politische und industrielle Akteur*innen, durch eine Differenzierung in verschiedene Subgruppen und Kompetenzprofile die Konturen des Feldes zu schärfen und die Bedürfnisse des Arbeitsmarktes mit den verfügbaren Kompetenzen in Übereinstimmung zu bringen (BHEF 2016; Demchenko et al. 2016; PwC & BHEF 2017). In den Interviews äußert sich dies in Form wiederholter Desiderata nach »Landkarten« oder »Mappings« der verschiedenen Kompetenzprofile – nicht zuletzt, um die Ausrichtung der eigenen Studiengänge zu justieren und Kopplungen zwischen Universitäten und der Nachfrage in der Industrie herzustellen.

Solche Bestrebungen nach einer Differenzierung in spezifische Subgruppen und klaren Grenzziehungen stehen einer diskursiv als universal konstruierten Wissensformation entgegen. Sie führen letztlich zu einer Disziplinierung »im Sinne einer Unterwerfung unter Regeln« und »mit einer deutlichen Einschränkung der individuellen Gestaltungsfreiheit zu Gunsten kollektiver Autonomie« (Schultheis 2005: 67f.). Als disziplinierte Wissensformation stehen die Datenwissenschaften in Widerspruch zur permanenten Akklamation von Interdisziplinarität, die diesen Raum diskursiv als einen mit offenen Grenzen herstellt.

7.3 Die Suche nach den richtigen Kategorien

Mit den Debatten über Universalisierung und Disziplinierung der Datenwissenschaften wird die Frage nach den passenden Kategorisierungen und Begrifflichkeiten viрulent, um die fundamentalen epistemologischen Veränderungen zu beschreiben, die sich in Diagnosen eines Paradigmenwechsels artikulieren. Es stellt sich deshalb die Frage, welche Begriffe zur Charakterisierung des neuen Wissensfeldes verwendet werden, inwiefern sie sich unterscheiden und weshalb sich »Data Science« schliesslich als Kategorie durchsetzen kann.

Verschiedene Begriffe werden wiederholt angeführt, um die veränderten Bedingungen der Wissensproduktion zu fassen. Dazu zählen neben »Data Science« insbesondere »Big Data«, »Machine Learning« sowie »künstliche Intelligenz«. Die Kategorien repräsentieren nicht nur divergente Wissensformationen, sondern weisen je eigene Historizitäten und Konjunkturen auf (Aronova et al. 2017; Nilsson 2009). Für die interviewten Lehrenden ist »Data Science« als Kategorie in dieser Hinsicht von zweitrangiger Bedeutung: Einerseits wird der Begriff als »Google-fähig« betrachtet, was heutzutage wichtig sei, und »[von] dem her ist das Label eigentlich nicht schlecht« (Prof_FH_D: 12:66). Zudem sei »Data Science« hilfreich für die »beruflische Positionierung« der Studierenden in einem stark expandierenden Sektor des Arbeitsmark-

tes. Andererseits wird der Begriff aber von manchen eher kontrovers betrachtet im Sinne eines diffusen »Container for everything« (Prof_UH_B: 3:45) oder gar als »ein Marketinginstrument, das nichts über die Inhalte aussagt« (Prof_UH_A: 6:20), abgestempelt.⁴ Wieder andere schliesslich sehen die Begriffswahl als pragmatisches Mittel zum Zweck, worauf ein Interviewer hinweist: »Vielleicht ist Data Science [einfach] ein besserer Begriff dafür, dass man irgendwie an der Schnittstelle von Informatik, Statistik, Mathematik und Business arbeitet« (Prof_FH_A: 8:29).

Diverse Befragte weisen darauf hin, dass »Data Science« bis 2014 kaum verbreitet war. Ein Professor an einer Fachhochschule beschreibt die Situation in jener Zeit so:

»Für die [erste Swiss Data Science] Konferenz 2014 fand ich auf LinkedIn keine Schweizer Data Scientists, die ich hätte anfragen können. [Wir] haben mit 40 Leuten gerechnet, es kamen 120 Leute« (Prof_FH).

Das neue Label stiess auf deutlich grösseres Interesse, als die Organisator*innen der Konferenz erwartet hatten. Dies wirkte sich in den folgenden Jahren auch auf die Verfügbarkeit von Bildungsangeboten aus: Gab es vor 2014 noch kein Programm unter diesem Titel, können heute an fast allen Schweizer Universitäten und Fachhochschulen Studiengänge, Vertiefungen oder Weiterbildungen in Datenwissenschaften belegt werden (vgl. Tabelle 13 im Anhang). Angesichts des zeitlichen Planungsvorlaufs von zwei bis drei Jahren indiziert das Angebot eine rapide Diffusion und Implementation der Kategorie »Data Science« innerhalb des schweizerischen Hochschulfeldes.

Die Vieldeutigkeit der zirkulierenden Begriffe zur Beschreibung des Feldes verdeutlicht sich anhand des feldübergreifenden Korpus der Stellenanzeigen, auf das ich im Detail im Kapitel 5 eingehende. Hier greife ich darauf zurück, um zu zeigen, inwiefern das untersuchte Wissenschaftsgebiet durch unterschiedliche Kategorien repräsentiert wird. Als Kategorien weisen »Data Science« und »Machine Learning« relativ ähnliche Inhalte und Verteilungen auf: »Machine Learning« wird enger gefasst und bezieht sich in erster Linie auf ein breites Methodenspektrum, das von linearen Regressionsanalysen bis zu neuronalen Netzwerken reicht. Gleichzeitig wird »Machine Learning« in nicht-technischen Diskussionen oft als Synonym für »künstliche Intelligenz« verwendet (Engemann & Sudmann 2017; Krause 2019). »Data Science« selbst enthält sowohl methodische als auch disziplinspezifische Begriffe und findet Anwendung auf eine Vielfalt unterschiedlicher wissenschaftlicher und ausserwissenschaftlicher Felder.

Demgegenüber unterscheiden sich die Anzeigen für den Suchbegriff »Big Data« erheblich von jenen für »Data Science« und »Machine Learning«. Sie beziehen sich explizit auf ein technisch-ingenieurorientiertes Vokabular und stellen neben Softwarelösungen Design und Architektur von Environments zur Prozessierung von Daten in den Vordergrund. Disziplinäres Vokabular ist mit Ausnahme computerwissenschaftlicher Begriffe selten. Die Inserrate für »Big Data« weisen bis Anfang 2018 ein starkes Wachstum auf und sind seither rückläufig (vgl. Abbildung 4). Auf der diskursiven Ebene deutet sich ein Wandel weg von »Big Data« hin zu »Data Science« und »Machine Learning« an: Infolge der zunehmend negativ konnotierten öffentlichen Wahrnehmung

4 Die Debatte über die Nützlichkeit des Begriffs ist so alt wie die jüngere Verwendungsweise von »Data Science« selbst, vgl. den frühen Text von Pete Warden (2011): Why the term »data science« is flawed but useful. Online: <http://radar.oreilly.com/2011/05/data-science-terminology.html> (Zugriff: 03.02.2022).

von »Big Data« im Nachgang zu Überwachungs- und Spionageaffären (»Big Brother«) verzichten Unternehmen und Organisationen zunehmend auf den Begriff und verwenden andere Kategorien (Elish & boyd 2018).⁵

Die Verschiebungen auf der Makroebene finden ihr Äquivalent in Studiengängen, die im Entwicklungsverlauf ihre Bezeichnung geändert bzw. »präzisiert« haben, wie es ein Befragter ausdrückt. Diverse Befragte attribuieren dem Begriff »Big Data« nur geringe Wissenschaftlichkeit und taxieren diesen als nicht praktikabel. So grenzt ein Fachhochschulprofessor die Wahl für »Data Science« als neue Studiengangbezeichnung explizit von »Big Data« ab:

»Data Science scheint nicht schlecht als Begriff, besser als Big Data, weil damit noch mehr Themen assoziiert werden. Data Science impliziert Systematik und Umgang mit Technologie. Mit Big Data assoziiert man gesellschaftliche Daten, gesellschaftliche Veränderung insgesamt, also ein sehr weites Feld« (Prof_FH_C: 9:35).

Die Gegenüberstellung von »Systematik und Umgang mit Technologie«, d. h. einer wissenschaftlich-technischen Dimension, und der »gesellschaftliche[n]« Dimension, d. h. der politischen, ökonomischen, rechtlichen oder sozialen Implikationen der Datennutzung, erfährt im empirischen Material nur wenig Resonanz. Die zwei Dimensionen koexistieren mehr oder weniger unabhängig voneinander und werden in den Curricula selten in Beziehung gebracht. Indem der Bedeutungsgehalt von »Big Data« nur schwierig zu fixieren ist (Kitchin & McArdle 2016), manifestiert sich in der Verwendung der Kategorie eine Ambiguität bzw. »konzeptuelle Vagheit« (Favaretto et al. 2020).

»Machine Learning« wiederum wird zwar feldübergreifend verwendet, dient aber insbesondere zur Charakterisierung eines breiten Spektrums methodischer Praktiken. Die Kategorie findet in einer Vielzahl von Beschreibungen methodenbezogener Kurse Verwendung, aber auch in Stelleninseraten, die disziplinenübergreifend angeboten bzw. gesucht werden. Viele Disziplinen bedienen sich eines Konvoluts an statistischen Methoden, ohne allerdings damit ein neues Feld zu begründen; vielmehr erfolgt die Transformation der methodischen Praxis – etwa die »mathematischen und statistischen Innovationen« bzw. »ganz viele[n] neue[n] Fragestellungen in verschiedenen Gebieten« (vgl. Zitat oben) – innerhalb der existierenden wissenschaftlichen Disziplinen bzw. institutionellen Gefässe.

»Data Science« kommt insofern die Funktion der Verknüpfung unterschiedlicher Bezugsfelder und Disziplinen zu: Neue Forschungszentren oder Studiengänge werden mit »Data Science« gelabelt, weil andere Kategorien als zu eng (»Machine Learning«) bzw. als zu weit (»Big Data«, »Künstliche Intelligenz«) taxiert werden. Gleichzeitig wurde »Data Science« als Wissenskategorie in den letzten zwei Jahrzehnten massgeblich durch Protagonisten der US-amerikanischen Tech- und Software-Industrie (Hammerbacher 2009; Varian 2009) gerahmt und symbolisch aufgeladen (vgl. Kap. 10.7). Die Attraktivität des Labels trifft darüber hinaus auf die Bereitschaft einer Vielzahl von Akteur*innen (vgl. das Zitat des Mathematikprofessors), nicht nur wissenschaftliche Forschungsergebnisse, sondern auch Daten, methodische Innovationen oder Software zur öffentlichen Verfügung zu stellen, was wiederum im Einklang steht

5 Um die These zu überprüfen, ist allerdings eine Betrachtung über einen längeren Zeitraum notwendig.

mit wissenschaftsinternen Entwicklungen wie der Reproduzierbarkeitskrise und der *open-science*-Bewegung. Vor diesem Hintergrund kann »Data Science« in der Suche nach den richtigen Begriffen folglich als offene, transversale Kategorie interpretiert werden: Sie erlaubt es Akteur*innen aus verschiedenen Feldern und Disziplinen, sich unter diesem Label zu positionieren, aber dennoch im Herkunftsgebiet tätig zu bleiben – was oft als die Interdisziplinarität der Datenwissenschaften umschrieben wird. Die Inklusivität der Kategorie ermöglicht es legitimerweise, multiple Rollen (als Forschende, Unternehmerinnen, Hacker etc.) in verschiedenen Feldern wahrzunehmen und darüber hinaus an jenen materiellen und symbolischen Ressourcen teilzuhaben, die im Rahmen der gegenwärtigen epistemologischen Transformation verteilt werden.

7.4 Innerwissenschaftliche Konfliktlinien

Neben der Begriffsarbeit und der Suche nach den passenden Kategorien bilden Praktiken von *boundary work* einen zweiten zentralen Praxismodus in der Konstruktion der Datenwissenschaften im akademischen Feld. Grenzziehungs- und Grenzüberschreitungsprozesse stellen ein konstituierendes Moment in der Herausbildung der Datenwissenschaften und insbesondere ihrer feldspezifischen Konstruktion in der Wissenschaft dar. Eine wesentliche Herausforderung der Datenwissenschaften seit ihrer Formierung besteht darin, wie ein Wissensgebiet, das auf unterschiedlichen technischen, methodischen und inhaltlichen Expertisen basiert, in ein disziplinär strukturiertes Feld überführt werden kann. In den Interviews wird oft problematisiert, dass solche Disziplinierungsbemühungen bestehende disziplinäre Konfliktlinien im akademischen Feld aktualisieren würden. Ein Professor an der ZHAW schildert den Konflikt der zwei »Kerndisziplinen« Statistik und Informatik anlässlich der Gründung eines interdisziplinären Forschungsnetzwerks in »Data Science«, des Datalab⁶, wie folgt:

»[Wir] haben an der ZHAW schnell gemerkt, dass es Zeit ist, etwas Neues zu gründen, interdisziplinär unter dem Stichwort Data Science: Wie können wir aus Daten Mehrwert schaffen, indem wir mit unterschiedlichen Brillen darauf schauen? [...] [Die] Kerngruppe waren Statistiker und Informatiker, die ähnliche Methoden verwenden, aber sich sonst spinnefeind sind. Später kamen Life Sciences, Juristen, Business-Leute oder auch quantitativ arbeitende Soziologen, Linguisten und Psychologen hinzu. [Es ist] ein interdisziplinär aufgestelltes Netzwerk, aber kollaborieren tun echte Menschen. Ziel ist es, dass sich echte Menschen zusammensetzen und gemeinsam Projekte entwickeln, Anträge schreiben und so weiter.« (Prof_FH)

Die zitierte Passage verweist auf die bisweilen widersprüchlichen Verschränkungen der Konfliktlinien zwischen Universalisierung und Disziplinierung: Darin äussert sich zum einen die Erkenntnis, dass Forschung und Wissensproduktion in Datenwissenschaften stets in Auseinandersetzung mit anderen Disziplinen durchgeführt werden. Die Offenheit und Multiperspektivität auf denselben Gegenstand, der hier hinter

6 Online: <https://www.zhaw.ch/de/forschung/departementsuebergreifende-kooperationen/datalab/> (Zugriff: 03.02.2022).

den »Daten« verschwindet, eröffnet somit einen Raum, damit sich das Forschungsnetzwerk entfalten kann.⁷ Die Inklusion weiterer Disziplinen über die »Kerngruppe« hinaus ermöglicht es, die im Netzwerk zirkulierende Expertise – im Zitat verkürzt als »quantitatives Arbeiten mit Daten« gefasst – zu stabilisieren, ohne dass dadurch eine neue Entität entsteht. Im Sinne einer Universalwissenschaft können so Forscher*innen der beteiligten Disziplinen teilhaben, ohne ihre jeweilige Fachidentität aufzugeben zu müssen.

Zum anderen wird darin gleichzeitig die Problematik manifest, dass zwar verschiedene Disziplinen mit »ähnlichen« Methoden Daten bearbeiten, aber den jeweiligen Bestandteilen des Forschungsprozesses wie Daten, Modellen oder Theorien sehr unterschiedliche Bedeutungen zuschreiben. Das Netzwerk unterschiedlicher Expertisen droht dabei durch disziplinäre Konflikte und »Feindschaften« jederzeit instabil zu werden. Über die epistemologischen Unterschiede, »die Konkurrenz der beiden Herangehensweisen« hinaus geht es in der Auseinandersetzung letztlich auch um Drittmittel, Forschungsprojekte sowie das »grössere Stück vom Big Data Kuchen« (Kauermann & Küchenhoff 2016: 143). Obwohl mit dem Datalab eine neue organisationale Kollaborationsform zur Akkumulation von wissenschaftlichem und ökonomischem Kapital geschaffen wurde, bleiben die unterschiedlichen zur Verfügung stehenden Ressourcen also Gegenstand für soziale Konfliktlinien und Distinktionsprozesse.

Allen Betonungen von Kollaboration und Interdisziplinarität zum Trotz werden also wiederkehrend epistemologische Grenzen zwischen den Disziplinen artikuliert (Gieryn 1983, 1999), die auf den zentralen Elementen wissenschaftlicher Forschungsprozesse wie Methoden und Daten beruhen. Explizit wird dies in den Interviews – auch aufgrund der einseitigen Verteilung der Interviewpartner*innen – vor allem anhand der Disziplinen der Informatik und Statistik, die sich »spinnefeind« (vgl. oben) seien. In einem weiteren sehr prägnanten Beispiel schreibt ein Informatikprofessor der Statistik »eine sehr enge Sicht auf Daten« zu:

»Ob es dann Data Science heisst, ist mir egal. Aber es ist eben auch nicht Statistik! Weil die Statistiker haben eine sehr enge Sicht auf Daten und was man mit Daten tun darf und was nicht. Aber wir tun es trotzdem, weil es funktioniert. Weil wir so viele Daten haben. Viele der statistischen Verfahren, die Sie in Ihrem Studium gelernt haben, basieren auf Datensätzen mit 50 Datenpunkten. Aber bei 50'000 Datenpunkten sind andere Verfahren besser geeignet« (Prof_UH_A: 6:69).

Der springende Punkt ist hier weniger die effektive Grösse des Datensatzes, sondern »was man mit Daten tun darf und was nicht«. Angesprochen werden hier grundsätzliche epistemologische Differenzen zwischen den involvierten Disziplinen, die nun unter der Prämisse der interdisziplinären Zusammenarbeit reartikuliert werden: Das Narrativ einer »normativen, modellgebundenen Statistik wird dabei einer »explorativen, datengetriebenen Computerwissenschaft gegenübergestellt. Indem andere Disziplinen Anspruch auf einen traditionellen Hoheitsbereich des Feldes wie die (Weiter-)Entwicklung von statistischen Methoden zur Datenanalyse erheben, stellen die

⁷ »Datalab connects several departments and institutes at the ZHAW, creating an internal collaboration network for researchers and also an interface for external partners and organisations« (ZHAW 2016: 7).

Datenwissenschaften insofern für die Statistik eine spezifische Herausforderung dar (Grommé et al. 2018).

Umgekehrt sind epistemische Praktiken, die die Computer- und Informationswissenschaften für sich in Anspruch nehmen oder die diesen zugeschrieben werden, unter Statistiker*innen verpönt. So distanzieren sich der Statistik zugehörige Befragte beispielsweise explizit von »Data Mining« aufgrund »mangelnder Systematik« (Prof_UH_E: 13:2): Eine kontrollierte wissenschaftliche Arbeitsweise durchforstet ihrer Ansicht nach nicht nur Daten und sucht nach Korrelationen, wie dies prominent Chris Anderson (2008) postuliert hat, sondern formuliert gezielt Fragestellungen an die Daten, überprüft Hypothesen und kommuniziert Ergebnisse (Schutt & O'Neil 2013: 44) – eine Selbstbeschreibung, der sich auch die »Data Science«-Community regelmäßig bedient (Bowne-Anderson 2018; Strachnyi 2017).

An diese Grenzziehung – nun von der entgegengesetzten Seite her – anknüpfend, verwehrt sich ein Professor einer Fachhochschule, der sich selbst nicht als Statistiker beschreibt, sondern im »Analytics-Bereich« positioniert, einer Vereinnahmung der Datenwissenschaften durch die Informatik ebenfalls sehr pointiert:

»Noch heute glauben Informatiker, Data Science habe etwas mit Informatik zu tun. [Das] Verhältnis zwischen Informatik und Data Science ist wie das zwischen Limonade und einer Flasche: [...] Der Flasche kann man Verschiedenes abfüllen, der ist das egal. Der Limonade ist es auch egal, wo sie aufbewahrt wird, oder? Und [...] es gibt natürlich eine Schnittstelle: Ein Data Scientist muss eigentlich sagen können, was er will: Welche Daten, in welcher Form und so weiter. All diese Parameter, die wichtig sind. Er muss es auch gegenüber dem Informatiker kommunizieren können, aber er ist nicht derjenige, der das System kennt. Er muss auch nur begrenzt Prozesse kennen. Er muss einfach diese Schnittstelle beherrschen, oder?« (Prof_FH_D: 12:63)

Trotz der anfangs scharfen Grenzziehung gegenüber der Informatik verortet der Interviewte »Data Science« im Zitat weniger (inter-)disziplinär, sondern als ein Wissensfeld, das quer zu bestehenden wissenschaftlichen Einteilungen liegt und sich in verschiedenen »Flaschen«, d. h. Disziplinen, wiederfinden kann. Der Begriff der »Schnittstelle«, der sowohl in den Interviews als auch datenwissenschaftlichen Diskursen allgemein angeführt wird, verdeutlicht die räumliche Positionierung vieler Befragter, die ihre wissenschaftliche Praxis nicht in abgeschlossenen Bereichen, sondern an Übergangszonen situieren, wo sich verschiedene Felder überlappen und miteinander in Kommunikation stehen. Solche diskursiven Praktiken sind insofern Ausdruck jener Bemühungen, die Grenzen einer stark disziplinär strukturierten Wissenslandschaft zu transzendifzieren und gleichzeitig die nach wie vor existierenden internen Konfliktlinien latent zu halten.

Trotz der vielfältigen interdisziplinären Kollaborationen und Arrangements in Forschung und Lehre erweisen sich feldspezifische Grenzziehungen und Markierungen in solchen Settings als äußerst persistent. Die theoretisch-konzeptuelle Annahme von Interdisziplinarität, Zusammenarbeit und Austausch existiert parallel zu konflikthaften Beziehungen, die durch die Existenz gemeinsamer Kategorien und institutioneller Gefässe (wie interdisziplinärer Institute, Netzwerke oder Studiengänge) nicht einfach verschwinden. Wie insbesondere die Analyse zum Datalab zeigt, halten die involvierten Akteur*innen die Grenzziehungen latent: Sie beteiligen sich am

Aufbau eines interdisziplinären Forschungsnetzwerks, ohne zwingend die Positionen und Perspektiven der anderen Beteiligten zu teilen. Das resultierende Netzwerk von Expertisen ist sodann durch die Kopräsenz unterschiedlicher disziplinärer Wissensbestände, aber auch Interessen, Strategien und Machtverhältnisse charakterisiert. So mit münden die identifizierten gemeinsamen Kategorien und institutionellen Gefässe in synchrone Relationen von Konkurrenz und Kollaboration.

7.5 Die Logik der Domänen: Interdisziplinarität im Kernbereich und *domain knowledge*

Trotz der wechselseitigen Grenzziehungen teilen beide Positionen – sowohl Informatik als auch Statistik – die Feststellung, dass die nach wie vor »stark disziplinär getriebene« Verortung der Datenwissenschaften ein grosses Problem darstellt. Darin würden sich Kompetenzbereiche und Methoden (wie beispielsweise Datenmanagement, Datenanalyse, Scripting, Programmierung etc.) ergänzen, die sowohl aus der Statistik als auch aus der Informatik stammten. Obwohl an der Schnittstelle der beiden wichtigsten Bezugsfelder seit längerem praktiziert, existierten bis vor wenigen Jahren an den meisten Universitäten kaum explizite curriculare Gefässe und Strukturen dafür. Die Implementation der Datenwissenschaften in Forschungszentren, Netzwerke oder Studienprogramme kann in diesem Sinne als *boundary crossing* betrachtet werden, als Übertreten der etablierten Grenzziehung existierender Disziplinen.⁸ Im Sample manifestiert sich dies ferner auch auf personaler Ebene: Bei mehreren Befragten changieren die beruflichen Karrieren zwischen dem akademischen und industriellen Feld. Datenwissenschaftler*innen können zwischen den involvierten Feldern hin und her wechseln und multiple Rollen einnehmen (Safavi et al. 2018). Grenzüberschreitende Rollen haben eine lange Kontinuität in den besagten Feldern (Ribes 2019) und erfahren gegenwärtig in der Schnittstellenfunktion von Data Scientists eine erneuerte Konjunktur.

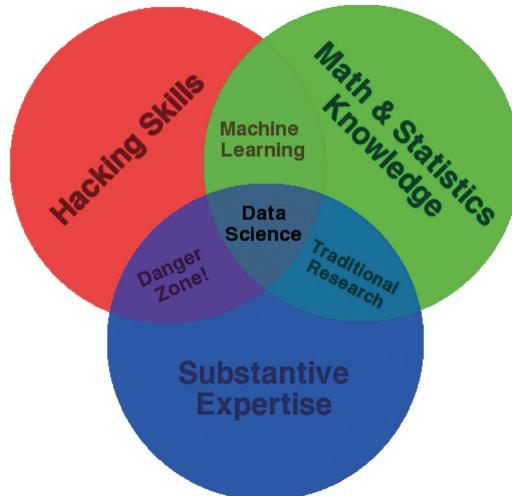
Interdisziplinarität zwischen den genannten informatischen und statistisch-mathematischen Kompetenzbereichen und Methoden alleine reicht für die Praxis als Data Scientist nicht aus: Es gilt sich in bestimmten Anwendungsbereichen, sogenannten *domains*, zu vertiefen und zu fokussieren. Domänen werden sehr breit definiert und können prinzipiell jede wissenschaftliche Disziplin oder gesellschaftliches Feld umfassen. Das »Data-Science-Venn-Diagramm« (vgl. Abbildung 10) repräsentiert die Datenwissenschaften als Schnittfläche einer interdisziplinär zusammengesetzten mathematisch-statistischen (»Math & Statistics Knowledge«), methodisch-technischen (»Hacking Skills«) sowie einer inhaltlichen Fachexpertise (»Substantive Expertise«).

Ribes et al. (2019) bezeichnen die Differenzierung in einen universellen, interdisziplinär zusammengesetzten Kern und spezifische Anwendungsdomeänen als »the logic of domains«. Sie hat eine lange Geschichte in den Computerwissenschaften, im Engineering sowie der Forschung zu künstlicher Intelligenz (Dreyfus & Dreyfus 2005;

⁸ Andere Beispiele aus jüngerer Zeit sind Neuro- und Klimawissenschaften oder Cultural Studies (Brint et al. 2009; Holley 2009; Knight et al. 2013), wobei keines dieser Felder eine ähnlich rasche Institutionalisierung erfahren hat wie die Datenwissenschaften.

Malazita 2019; Ribes 2019; Ribes et al. 2019). Weil den Datenwissenschaften ein eigenes Anwendungsgebiet grösstenteils fehlt, positioniert sich das Feld durch die Verwendung der Logik der Domänen – analog den Computerwissenschaften im letzten Viertel des 20. Jahrhunderts – als allgemeine, universelle Wissenschaft, die für jeden gesellschaftlichen Bereich relevant ist.

Abbildung 10: Das Data-Science-Venn-Diagramm



Quelle: Drew Conway (2010) – eigenes Werk, CC BY-NC 3.0,
<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>

In den Interviews präsentiert sich das Denken in der Logik der Domänen als zentrales Ordnungs- und Klassifizierungsprinzip von Lehrenden und Forschenden der Datenwissenschaften, das sich in unterschiedlichen Ausprägungen artikuliert. Ein Professor beschreibt dies wie folgt:

»Ich bin überzeugt, dass Interdisziplinarität ein Kern ist, wenn wir jetzt auf Data Science zurückkommen. Deshalb haben wir das Ganze so strukturiert, dass man Data Science entweder als Haupt- oder als Nebenfach wählen kann, das war eigentlich unser Ziel. Es geht um domain knowledge, die man entweder schon mitbringt, oder daneben als weitere Kompetenz sich aneignen kann« (Prof_UH_A: 6:70)

Interdisziplinarität wird nicht nur als Kombination der genannten Kompetenzbereiche, sondern in Verknüpfung mit einem Anwendungsfeld konzipiert. Indem »domain knowledge« in den Vordergrund rückt, wird die methodisch-technische Expertise, die gemeinhin den »Kern« von Definitionen und Curricula der Datenwissenschaften ausmacht, unsichtbar und universal zugleich. Zudem sind – wie ich noch zeigen werde – die untersuchten Studienprogramme in Konzeption und Aufbau (vgl. Kap. 9.5) als auch die Forschungsförderung in der Logik der Domänen strukturiert.

Die Konstruktion der Datenwissenschaften als universal und domänenunabhängig erlaubt es ansatzweise, die vorgängig diskutierten Limitationen von Disziplinie-

rung, adäquaten Begrifflichkeiten und innerwissenschaftlichen Konfliktlinien zu überwinden und zu neutralisieren. Indem der Schwerpunkt auf die Verknüpfung der unterschiedlichen Bezugsfelder gelegt wird, überschreitet das Denken in der Logik der Domänen existierende disziplinäre und soziale Grenzziehungen und präsentiert sich somit als pragmatisch und anwendungsorientiert.

7.6 Spielarten von Anwendungsorientierung: »Echte Probleme« und »operative Exzellenz«

Die in der Logik der Domänen implizierte Anwendungsorientierung wird als ein Ansatzpunkt betrachtet, um die beschriebenen Spannungen zwischen quantitativ-methodischer Expertise (Universalisierung) einerseits und spezifischen Themenfeldern (*domains*) andererseits zu überbrücken. In den konkreten Anwendungsmöglichkeiten der Datenbanktechnologien, Algorithmen und Methoden in den Unternehmen wird – wie bereits erwähnt – »das eigentlich Neue« des Feldes verortet. Große Datenmengen würden keine Fragestellungen an sich beinhalten; solche könnten erst durch »Anwendungsorientierung und konkrete Projekte« gefunden werden. Je nach Verortung und Profil der untersuchten akademischen Institutionen äußern sich unterschiedliche Spielarten von Anwendungsorientierung: Im Falle der technischen Hochschulen sind Kollaborationen der Daten- und Computerwissenschaften mit spezifischen Domänen keineswegs neu. So bestehen etwa an der ETHZ datenbezogene Kooperationen mit Unternehmen, aber auch zwischen den involvierten Departementen über den neu geschaffenen Masterstudiengang hinaus schon sehr viel länger:

»Das passiert laufend, dass andere Projekte und Leute auf uns zukommen mit spezifischen Problemen bezüglich der Verarbeitung grosser Daten. [...] An der ETH wird täglich mit Firmen zusammen gearbeitet. Ich habe das in meiner Promotion selbst gemacht, das wird nicht spezifisch sein. Denn die Unternehmen bringen die echten Probleme. Aber wir arbeiten auch mit anderen Departementen zusammen, die ebenfalls mit Datenproblemen zu uns kommen« (Prof_ETH_B: 2:9).

In dieser ersten Ausprägung von Anwendungsorientierung, die nicht nur bei Professor*innen an den ETH zu beobachten ist, sind es Forschungspartner*innen aus anderen wissenschaftlichen Disziplinen, vor allem aber aus Unternehmen, die die »echten Probleme« mitbringen. Das Narrativ der echten oder realweltlichen Probleme (*real-world problems*), die es im Rahmen der praxisorientierten Lehre zu bearbeiten gilt, findet eine Entsprechung in echten Daten (*real-world data*), die in unzähligen Kursausschreibungen, Webseiten und Werbematerialien von Studiengängen angeführt werden. Das evozierte Bild referiert einerseits auf einen Typus von Daten, die unabhängig von ihrer Konstruktion durch soziotechnische Systeme existieren und die es entsprechend bloss zu erheben und zu analysieren gilt, um unmittelbare Erkenntnisse für die eigene Forschung, das wissenschaftliche Feld oder »die Gesellschaft« zu ziehen. Die Metapher der »Roh-Daten« liegt hier nahe (Gitelman 2013) und markiert ein essentialistisches Datenverständnis, wie es insbesondere den politischen Diskurs

kennzeichnet.⁹ Zum anderen impliziert die Rahmung, die sonstige wissenschaftliche Arbeit mit Daten sei nicht echt, d. h. konstruiert, praxisfern oder sonst wie weniger bedeutend (Dreyfus & Dreyfus 2005). Die Betonung des angewandten, realweltlichen Bezugs von akademischer Forschung und Lehre ist eng gekoppelt an historisch gewachsene Vorstellungen und Ideale der Engineering-Ausbildung (Malazita 2019), innerhalb derer sich die meisten Studiengänge in Datenwissenschaften positionieren.

Die Grenzziehung zwischen konstruierten wissenschaftlichen und echten realweltlichen Daten und Problemen wurde in den letzten drei Jahrzehnten zu einem zentralen Organisationsprinzip der Forschungsförderung (Ribes et al. 2019). In der Schweiz folgen die Förderprogramme zu »Big Data« (SNF 2015) und »Data Science« (SDSC 2017) dieser Logik. Die Projektausschreibung des SDSC (2017) macht dies explizit, indem der Ausschreibungstext die zwei Module »Data science meets domain science« und »Data science methods for the real-world [sic!]« (ebd.: 5f.) einander gegenüberstellt: Bei datenwissenschaftlichen Anwendungen und Methoden in *domain science* handelt es sich um spezifische Probleme, die lediglich innerhalb des wissenschaftlichen Feldes relevant sind. Demgegenüber weisen Technologien und Methoden der Datenwissenschaften einen Bezug zur »real world« auf und ermöglichen Anwendungen über die Wissenschaft hinaus, sind also von universeller Relevanz. Die darin artikulierte Betonung des gesellschaftlichen »Impacts« von »Data Science« evoziert wiederum eine Trennung in wissenschaftliche und nichtwissenschaftliche Anwendungsgebiete.

Eine zweite Ausprägung von Anwendungsorientierung, die der ersten gegenübergestellt wird, zeigt sich in einer divergierenden Verortung der relevanten Probleme und deren Verknüpfung mit bestimmten methodischen Vorgehensweisen innerhalb der Datenwissenschaften: Während an den ETH und Universitäten die »echten Probleme« durch Kooperationen mit anderen Disziplinen oder externen Partner*innen adressiert werden, sind es an den Fachhochschulen die Studierenden, die »echte Probleme« aus ihrer beruflichen Praxis mitbringen. Ein Professor einer Fachhochschule verknüpft die unmittelbare Anwendungsorientierung im beruflichen Umfeld mit den notwendigen Kompetenzen und Vorgehensweisen von Data Scientists wie folgt:

»Ein Grundstock an klassischer sowie deskriptiver und schliessender Statistik ist notwendig, damit sie fundiert arbeiten können. Algorithmen- oder Methodenentwicklung als solche sind aber nicht Ziel des Studiengangs. Es geht vor allem um die Optimierung von Arbeits- und Betriebsabläufen in ihren Unternehmen, beispielsweise anhand von Logfiles die Performance von Systemen zu verbessern. Es geht nicht darum, Hypothesentests durchzuführen oder Forschungsfragen zu beantworten, sondern darum, operative Exzellenz zu schaffen« (Prof_FH_C: 9:29).

Die befragte Person markiert hier Differenzen zwischen verschiedenen epistemischen Praktiken innerhalb der Datenwissenschaften. Dabei assoziiert sie bestimmte

⁹ Die Konzeption als »real-world data« unterschlägt zudem, dass Forschung – sowohl in wissenschaftlichen als auch anderen Kontexten – Probleme, Modelle oder Daten immer zuerst bestimmen, auswählen und bearbeiten, d. h. als solche *konstruieren* muss: Nicht bloss die Daten sind nicht einfach gegeben, sondern auch Fragestellungen und zu lösende Probleme müssen als solche anerkannt und bestimmt werden (Mützel et al. 2018).

Vorgehensweisen (wie »Algorithmen- oder Methodenentwicklung«, »Hypothesentests durchführen oder Forschungsfragen zu beantworten«) mit den beiden ETH, während sie für die Fachhochschulen eine angewandte Form der Datenwissenschaften (wie beispielsweise die »Optimierung von Arbeits- und Betriebsabläufen«) reklamiert, bei der »operative Exzellenz« geschaffen werden soll. In den Zuschreibungen unterschiedlicher Problembearbeitung im Rahmen der Anwendungsorientierung manifestieren sich divergierende Konzeptionen der Datenwissenschaften, die wiederum unterschiedliche Anknüpfungspunkte für feldexterne Akteur*innen ermöglichen.

Die Anwendungsorientierung infolge der postulierten Universalität datenwissenschaftlicher Expertise stellt einen zentralen *Modus Operandi* dar. Daraus resultiert eine gewisse Notwendigkeit für alle Akteur*innen im Feld, die eigenen Lehr- und vor allem Forschungsaktivitäten als anwendungsorientiert zu markieren: Alle Akteur*innen kollaborieren mit internen und externen Partner*innen und verstehen sich als angewandt und praxisorientiert. Trotzdem fordert die Analyse des empirischen Materials unterschiedliche Spielarten von Anwendungsorientierung zutage: Zum einen zeigt sich eine Varietät von Anwendungsorientierung, die sich am realweltlichen Bezug von Forschungsproblemen und Daten – gegenüber »rein wissenschaftlich« induzierten – unabhängig von der feldspezifischen Verortung (Wissenschaft, Industrie, Verwaltung etc.) orientiert. Zum anderen manifestiert sich eine Spielart von Anwendungsorientierung als »operative Exzellenz«, die Praxisbezüge ausschliesslich in ausserwissenschaftlichen Referenzen verortet. Auch wenn beide Ausprägungen unter dem gemeinsamen Label »Data Science« koexistieren können, zeigen sich unterschiedliche institutionelle Verortungen im Feld (Erstere vor allem an den ETH, Letztere an den Fachhochschulen). Sie äussern sich in Zuschreibungen divergierender methodischer Vorgehensweisen und epistemischer Praktiken, die somit die jeweiligen Positionen im akademischen Feld charakterisieren.

7.7 Diskussion

Die Analyse hat die Bedeutung von zwei zentralen Praktiken, diskursiver Begriffsarbeit einerseits und *boundary work* andererseits, in der Konstruktion der Datenwissenschaften sichtbar gemacht. Neben disziplinären Wissensbeständen, Daten, Methoden und Instrumenten sind auch Definitionen, Kategorien und Narrative fundierend für die Herausbildung eines neuen Wissensfeldes. So repräsentiert das akademische Feld die Datenwissenschaften als ein Wissensfeld, das durch technologische und epistemologische Verschiebungen an der Schnittstelle unterschiedlicher Disziplinen entstanden ist und gleichzeitig verschiedene Anwendungsbereiche transformiert. Dem Begriff des Paradigmenwechsels und der Universalität der Datenwissenschaften stehen Differenzierungen in spezifische Subgruppen und Desiderata nach klaren Grenzziehungen gegenüber anderen Disziplinen und Feldern entgegen. Betonen Erstere die Offenheit und Transversalität des Wissensgebiets, so zielen Letztere auf die strukturelle Einpassung und inhaltliche Engführung in ein disziplinäres Gefüge.

In der Suche nach den adäquaten Kategorien zur Charakterisierung der identifizierten epistemologischen Verschiebungen hat sich im Untersuchungsfeld die Bezeichnung »Data Science« gegenüber anderen Begrifflichkeiten (wie »Big Data« oder »Machine Learning«) etabliert und durchgesetzt. Werden Letztere als zu weit bzw. zu

eng taxiert, stellt »Data Science« eine inklusive Kategorie dar, die es Akteur*innen erlaubt, multiple Rollen in unterschiedlichen Feldern und Organisationen wahrzunehmen und somit existierende Grenzen zu überschreiten. Solche Suchprozesse haben indes auch sinnstiftenden Charakter, indem sie zur Kohäsion und zum Zusammengehörigkeitsgefühl der unterschiedlichen Akteur*innen beitragen und womöglich gar die Herausbildung einer distinkten feldübergreifenden Identität befördern.

Neben der konstitutiven Rolle von Begriffen und Kategorien bilden synchrone Praktiken von *boundary work*, d. h. Grenzziehungen und -überschreitungen, den zweiten Praxismodus, der in die Konstruktionsweise der Datenwissenschaften im akademischen Feld eingeschrieben ist: Durch die Konzeption der Datenwissenschaften als ein interdisziplinär zusammengesetztes Wissensgebiet erheben diverse Disziplinen Ansprüche auf die epistemische Autorität für das Feld der wissenschaftlichen Datenbearbeitung, was sich im empirischen Material in intensiven Konfliktlinien und Grenzziehungen zwischen einzelnen Disziplinen artikuliert.

Umgekehrt werden durch Interdisziplinarität und die Logik der Domänen aber auch etablierte Grenzen überschritten: So produziert die permanente Formulierung von Interdisziplinarität, Anwendungsorientierung oder »echten Problemen« einen diskursiven Raum zwischen disziplinären und sozialen Feldern, der vielfältige Schnittstellen und Anknüpfungspunkte schafft. Indem die Datenwissenschaften als universal und domänenunabhängig konstruiert werden, ermöglicht dies ansatzweise die Limitationen, die sich durch Disziplinierung sowie innerwissenschaftliche Konfliktlinien ergeben, zu überwinden und zu neutralisieren. Dies gelingt den Akteur*innen vor allem durch diskursive Strategien der Grenzüberschreitung, indem sie die gemeinsamen Interessen am Aufbau des neuen Wissensfeldes betonen und gleichzeitig interne Konflikte latent halten.

Entgegen den Bestrebungen des akademischen Feldes um Autonomisierung und Immunisierung, die sich auf Prinzipien wie Domänenunabhängigkeit, Reproduzierbarkeit oder der universellen Anwendbarkeit von Methoden und Technologien abstützen, ist das Soziale stets in Datenwissenschaften eingeschrieben: Es äussert sich in sozialen Kategorien (wie Paradigmenwechsel, Domänen, *real-worldness*, Anwendungsorientierung etc.), disziplinären Konfliktlinien sowie Grenzüberschreitungen, die für die Sinnhaftigkeit des Phänomens Datenwissenschaften ebenso fundierend sind wie die Praxis mit Daten, Methoden und Algorithmen selbst. Die Synchronizität und die Kontingenzen von Suchprozessen einerseits, von Abgrenzung und Zusammenarbeit andererseits generieren in diesem Sinne äusserst widersprüchliche Dynamiken, die neben wissenschaftlich-technischen Artefakten (Projekte, Publikationen, Patente etc.) auch auf der sozialen und symbolischen Ebene produktive Potenziale freisetzen können. Die Datenwissenschaften stellen insofern einen Fall dar, bei dem es den beteiligten Disziplinen und Akteur*innen gelang, einen solchen Raum der Möglichkeiten zwischen etablierten disziplinären Feldern zu schaffen sowie – zumindest bis zum jetzigen Zeitpunkt – offen und durchlässig zu halten.