

5 Daten erkennen: Visualisierungen als Arbeitsoberfläche

Der Abschnitt 4.2 hat zwei verschiedene Modi von Visualisierungen dargelegt: das Erkennen von Daten (*seeing data*) und das Befestigen von Evidenz (*becoming evident*). Abschnitt 4.3 hat die beiden Modi mit der Heuristik der Aushandlung verbunden. Kapitel 5 untersucht mit diesem Gerüst den erstgenannten Modus in den zwei datenjournalistischen Teams: das Erkennen von Daten auf Visualisierungen¹ – die als Arbeitsoberfläche dienen, um aus den Daten Sinn zu generieren und Erkenntnisse zu produzieren.

Der Abschnitt 5.1 erörtert die vier aus dem empirischen Material resultierenden Arbeitspraktiken: Datafizierung, saubere Daten, Geschichten drehen und Auswertungen plausibilisieren. Durch diese machen sich die beiden Teams ihre drei Expertisen mit ihren materiellen und kulturellen Hintergründen nutzbar. Die folgenden Abschnitte beleuchten die soziale Verwobenheit dieser Expertisen – in den Konversationen (5.2), dem visuellen Vokabular als semiotische Ressource (5.3), den soziotechnischen Werkzeugen (5.4) sowie den Verweisen auf andere Akteur*innen des Feldes.

5.1 Arbeitspraktiken des Erkennens von Daten

Im folgenden, ausführlichen ersten Teil des Kapitels zum ersten Visualisierungsmodus geht es darum, die Arbeitspraktiken zu beschreiben, mit welchen

¹ Ich nutze konsequent den Begriff Datenvisualisierungen, auch wenn im referenzierten empirischen Material dafür Begriffe wie »Grafik«, »Infografik« oder »Chart« verwendet werden. Zwischen diesen Begriffen gibt es theoretische Differenzierungen, die aber im beobachteten Arbeitsalltag der beiden Teams nicht relevant sind.

die beiden Teams das Erkennen von Daten organisieren. Es erfolgt eine Gliederung in vier Praktiken beziehungsweise Sets an Praktiken, die einen analytischen Mehrwert bietet, aber in der Empirie als solche nicht klar abgegrenzt wird. Im ersten Abschnitt folgt ein Schlaglicht auf Datafizierung und einige grundlegende ontologische Überlegungen. Danach folgen Sets an Praktiken zur Datengrundlage, Geschichtenproduktion und -plausibilisierung. Die Gliederung insinuiert auch eine gewisse prozessorientierte Chronologie der Praktiken, die zwar empirisch tatsächlich sinnhaft ist, aber keineswegs immer eingehalten wird. Stattdessen sind die Praktiken und Schritte eng verwoben und der Prozess ist iterativer und abduktiver Natur.²

5.1.1 Datafizierung: Geschichten in Daten übersetzen

Waren es in den Laborstudien natürliche Stoffe, die mathematisiert und geometrisiert wurden, so stellt sich im Datenjournalismus die Frage: Wie werden Daten konstruiert, die auf ›reale Begebenheiten‹ referenzieren? Es handelt sich dabei um einen Prozess der Datafizierung, das heißt, die soziale Welt wird in quantifizierter Form wiedergegeben. Die Sammlung von Daten ist keine neue Praktik, wie die Historie der Statistik aufzeigt. Die Welt in Zahlen auszudrücken ermöglicht es, über Grenzen (beispielsweise von Ländern oder Feldern) hinaus Verbindungen herzustellen (Desrosières 1998), zu vergleichen (Lamont 2012; Heintz 2010) und sich auf Objektivität zu berufen (Porter 1995). Die Datafizierung einer Geschichte ermöglicht eine auf der Logik der Datenbank (Manovich 2001) rationalisierte Sichtweise.

Bei BR Data suchen die Teammitglieder, inspiriert von einer Idee, einfach mal Daten und Indikatoren, um sich ein Thema, eine *Geschichte* mit Daten vorstellen zu können. Während der Recherche nach Daten geht es darum, zu explorieren, welche Variablen man kombinieren könnte; welche Sachverhalte überhaupt quantifiziert existieren, das heißt in Zahlenform vorhanden und mit Skalen erfassbar sind; welche Daten Grundlage kalkulatorischer Praktiken werden können. Für dieses Set an Praktiken ist das Internet die wichtigste Quelle: Fast alle Recherchen passieren online. Es gibt Themenbereiche, in denen BR Data eine stärkere Quantifizierung erwartet. Während Treffen, an denen das Team Geschichten-Ideen diskutiert, erfolgen immer wieder ad hoc

² Wiederum war es das Team von The Pudding, welches Transparenz auch über diesen iterativen Prozess anstrebt und ihn in einem eigenen Artikel aufzuzeichnen versucht: <https://pudding.cool/process/pivot-continue-down/> (27.12.2020)

derartige Abschätzungen: Welche Daten sind zu jenem Thema verfügbar und wie aufwendig wäre die Sammlung und Aufbereitung?

Das ›Finden von Geschichten in Daten‹ ist eigentlich ein ›Geschichten in Daten übersetzen‹. Anderson (2018:138ff) und Coddington (2018) haben in verschiedenen Organisationen geforscht, welche Informationseinheiten (Einzelereignisse beziehungsweise einzelne Nachrichtenmeldungen) sie in Datenbanken versammeln und mittels Geschichten verbinden. Beide interpretieren dies als »computational thinking« (Wing 2006) im Journalismus. Dies ist dadurch charakterisiert, dass die Datenbanken nicht aufgrund von Geschichten und dazugehörigen Fakten (Dokumente, Zitate) aufgebaut werden, sondern dass die journalistische Erkenntnisproduktion so geordnet ist, dass man sich die Realität im Datenbankformat vorstellt – »A fact is true, in other words, if it exists in the database« (Anderson 2018:14) – und die Geschichte auf dieser ontologischen Grundlage imaginiert.

Somit erscheint es logisch, dass das BR-Data-Team keine Geschichte produziert, falls keine digitalen Daten vorhanden sind.³ Basiert die Erkenntnisproduktion nicht auf Daten, so ordnet das Team dies ausserhalb der eigenen Zuständigkeit und Kompetenz ein. BR Data beruft sich damit implizit auf ein »*data a priori*« (Hervorhebung im Original), das Gray (2018:7) beschreibt als »not only designates but also provides the conditions of possibility for seeing and engaging with different aspects of collective life – making possible particular styles of reasoning and particular forms of knowledge and experience.« Diese »Data Worlds« von Gray sind charakteristisch für das Narrativ der neuartigen, datengetriebenen Geschichten (Abschnitt 3.3.1) – die *Datenwelten* zu beherrschen wird in den Teams gerne zelebriert und NZZ-Datenjournalist Yannick kategorisiert Personen mit entsprechender Expertise als »data-savvy«⁴.

Der Imaginationsprozess, eine Geschichte in datafizierte Form zu übersetzen, passiert bei NZZ Storytelling häufig über die Antizipierung geometrischer Verkörperungen der Daten. So sind händische Skizzen, die nicht aus Daten errechnet werden, ein wichtiger Aspekt der Datafizierungspraktik (siehe Walny 2016). Die Teammitglieder beginnen die Datenauswertung sehr oft mit Skizzen auf Papier, mithilfe derer sie eine Visualisierung als Geschichte modellieren (Abschnitt 5.4.3 beleuchtet die Rolle von Papierskizzen als Werkzeug detailliert). Visualisierungen sind insbesondere dann ein Instrument, um die

3 58_20190110_BR-Data-12 (1)

4 22_20180323_NZZStorytelling-10 (11)

Geschichte in Daten zu übersetzen, wenn es darum geht, andere Kolleg*innen mit weniger datenjournalistischer Erfahrung für ein Projekt zu gewinnen. So stossen NZZ Storytelling und das Auslandressort in einer Sitzung viele gemeinsame Geschichten-Ideen an – nicht indem sie überlegen, welche Daten es zu einem Thema gibt, sondern wie die Datenvizualisierung schliesslich aussehen könnte.⁵ Durch die Visualisierung – ob als Vorstellung oder Skizze – werden Daten, die sonst als Zahlenwerte oder in grossen Textmengen stark abstrahiert sind, auf eine Ebene re-konkretisiert, auf der ein Anschluss an die Sinngebung für die Journalist*innen mit spezifischen Themenspezialisierungen möglich ist. Die folgenden Abschnitte werden zeigen, welche datenjournalistischen Praktiken dies in den Teams gewährleisten sollen. Die Produktion von journalistischer Erkenntnis aus Daten basiert damit nicht nur auf journalistischer und programmiertechnischer Expertise, sondern insbesondere die grafische nimmt durch visuelle Verkörperung eine zentrale Rolle ein.

Dieser Abschnitt hat die ontologischen und epistemologischen Grundlagen der datenjournalistischen Arbeitspraktiken in den beiden Teams aufgezeigt: Sie basieren auf einer durch das Datenbankformat rationalisierten Sicht auf reale Ereignisse, also der Datafizierung der Informationen über die Welt.⁶ Gleichzeitig gilt es, die epistemologische Form der Geschichte in Daten zu übersetzen. Dies passiert insbesondere über Imaginationen von visuellen Verkörperungen der Daten. Datafizierung selbst basiert in den beiden Teams demnach oft auf einer visuellen Rationalität, welche »die Strukturlogik sozialer Praxis um den Visualitätsaspekt erweitert« (Burri 2008:354).

Die folgenden Abschnitte vertiefen die erwähnten verschiedenen Elemente der datenjournalistischen Arbeitspraktiken. Der nächste fokussiert auf die Datensätze.

5.1.2 Saubere Daten: Granularität, Ordnung, Interoperabilität

Die zuweilen geäusserte Binsenweisheit, saubere Daten seien die Grundlage für eine saubere Arbeit, reibt sich mit der Feststellung, dass in fast jedem Datensatz Verschmutzungen aller Art vorkommen. Die folgenden Abschnitte beleuchten, anhand welcher Leitlinien die beiden Teams die datenjournalistische

5 21_20180322_NZZStorytelling-09 (4)

6 Die Einordnung des Datafizierungsphänomens bleibt an dieser Stelle knapp. Für weiterführende Überlegungen dazu sei auf die Sammelwerke von Prietl und Houben 2018 und Mämecke et al. 2018 sowie Mayer-Schönberger und Cukier 2013 verwiesen.

Praktik der Herstellung einer analysierbaren Datengrundlage, »sauberer Daten« eben, strukturieren: Granularität, Ordnung und Kategorisierung, Interoperabilität.

Granularität

Ein wichtiges Kriterium in der Herstellung analysierbarer Datensätze ist die Sicherstellung von Granularität. So zum Beispiel in einem Projekt über Strafverfahren zu spezifischen Delikten, für das BR Data einen Datensatz selbst erhebt. Die zwei Teammitglieder Lea und Lukas konzipieren eine Excel-Tabelle, die danach die Grundlage für die Datensammlung und -analyse bilden soll. Insbesondere Lukas betont immer wieder, dass die Tabelle möglichst viele Differenzierungen zulassen muss.⁷ Dies ermöglicht mehr Optionen für Variablen und potenziert entsprechend die Anzahl möglicher Kalkulationen. Und es verringert die Chance, dass in den Daten Verzerrungen und Aggregate unentdeckt bleiben, so eine weitere Argumentationslinie von BR Data.

Das NZZ-Storytelling-Team zieht in einer Situation den Umkehrschluss: Ist ein Phänomen nicht erklärbare, sind die Daten noch zu wenig granular. So lassen sich die Verschiebungen in der politischen Landschaft bei den kurz vor der Forschungsphase stattgefundenen Wahlen auf der Ebene der Wahlkreise nicht anhand bestimmter soziodemografischer Variablen statistisch erklären. Deshalb versucht Datenjournalistin Annalena zusammen mit einem Redaktor aus dem Zürich-Ressort die Wahlstatistiken eines Urnengangs auf einer noch granulareren Ebene als den stadtzürcherischen Wahlkreisen auszuwerten.⁸

Datenjournalistischen Praktiken nutzen grosse Datenmengen, »Big Data«, als Methode, um soziale Phänomene in höherer temporaler und geografischer Granularität zu fassen (Lazer und Radford 2017). Hadley Wickham, Statistiker und Chefentwickler der Programmiersprache R, skizzierte in einem Artikel die Prinzipien hinter den Visualisierungen, die mit dem von ihm entwickelten R-Paket ggplot2 möglich sind (Wickham 2010). Ein besonders wichtiges Prinzip ist dabei die Granularität der Daten. Denn Wickhams Paket basiert wie die meisten Visualisierungswerzeuge darauf, dass verschiedene Schichtungen (»layers«) zusammenkommen, die beispielsweise je eine statistische Transformation oder die Ausgestaltung mittels einer geometrischen Figur umfassen. Und dass Skalen und Aspekte (»facets«) differenzierbar sind, um die Einordnung grafisch darstellen zu können. Die Granularität eines

7 58_20190110_BR-Data-12 (6–8)

8 20_20180321_NZZStorytelling-08 (2)

Datensatzes ist die Voraussetzung der »geschichteten« (»layered«) Visualisierungen, die verschiedene Aspekte abbilden. Dabei wird jede Komponente auf eine unterschiedliche Art ästhetisch umgesetzt: teilweise in geometrischen Formen in einer Skala, oder kategoriale Variablen als Farb-, Größen- oder Kontur-Attribute (siehe ausführlicher in Abschnitt 5.3). Das Befolgen dieser Prinzipien ermöglicht es, einerseits schnell Visualisierungen zu entwickeln und anzupassen; es macht sie replizierbar; es ermöglicht aber vor allem auch die Exploration: Einzelne Komponenten können einfach ausgetauscht und so verschiedene Varianten ausprobiert werden (Healy und Moody 2014:111ff; Tukey 1977). Die Granularität von Daten ermöglicht eine grosse Varianz an explorativen Visualisierungen und stellt damit in den Teams die Vielfältigkeit und Flexibilität der Interpretation sicher.

Daten ordnen und kategorisieren

Für eine Mehrzahl an Projekten nutzen sowohl BR Data wie auch NZZ Storytelling Daten, die bereits in digitaler Form existieren. Sehr oft verwenden die beiden Teams offizielle Statistiken, die in tabellarischer Form vorhanden sind. Durch die vorgegebene Formatierung sind die Daten mit gewissen Kategorisierungen vorstrukturiert. Insbesondere die Teammitglieder von BR Data äussern explizit den Anspruch, diese Kategorisierungen zu erfragen und zu verstehen sowie die Verzerrungen zu antizipieren. BR Data erfragt die Herkunft und die Erhebungsumstände der Datensätze; die Teammitglieder vergleichen dabei Daten mit menschlichen Quellen, die Informationen zur Verfügung stellen und die kritisch hinterfragt werden müssen (Parasie 2014; Ettema und Glasser 1985). Staatlichen Daten begegnen sie oft mit der Attitüde, dass sie lückenhaft und unsystematisch erhoben sind – dies verweist einerseits darauf, dass die Teammitglieder die Vollständigkeit des Datensatzes als sehr wichtiges Kriterium erachten, weil lückenhafte Datensätze die Interpretation stark erschweren. Andererseits wird damit auf den teameigenen Anspruch verwiesen, mächtigen Akteur*innen wie dem Staat kritisch gegenüberzustehen und eine Wachhund-Funktion einzunehmen (siehe Watchdog-Narrativ, Abschnitt 3.3.1). Ein mögliches Spannungsfeld zwischen der enthusiastischen Nutzung von Datensätzen als Quellen sowie dem kritischen Hinterfragen ebendieser, wie es Lesage und Hackett (2014:48) erwähnen, war in den Teams nicht feststellbar.

Neben der Evaluation der Datenquelle verläuft die Einordnung der Qualität eines Datensatzes vor allem über das Kriterium der systematischen Ordnung – bei BR Data wird die positive Bewertung mit dem Adjektiv »sau-

ber« ausgedrückt.⁹ Der Ordnungsaspekt wird in zwei Eigenschaften bewertet: erstens seine Nachvollziehbarkeit und zweitens eine durchgehend konsistente Umsetzung der Ordnung. Nachvollziehbar ist eine Ordnung dann, wenn die abgebildeten Relationen zwischen Beobachtungen und Eigenschaften bzw. kategorisierenden Variablen mit den Vorstellungen möglicher Relationen der Teammitglieder stimmig sind. Der Einordnungsprozess ist ein essenzieller Teil der Sinngebung von Daten – eine Position, die verschiedene Autor*innen aus den Science and Technology Studies untermauert haben: Beispielsweise haben Bowker und Star (1999) Kategorisierungspraktiken als Setzung von ontologischen Grundlagen beschrieben.

Die konsequente Umsetzung des Ordnungsprinzips wurde in der neueren Zeit wiederum von Hadley Wickham und seinem Prinzip der »tidy data« vorangetrieben (Wickham 2014) – Abschnitt 3.2.2 hat es bereits als Teil der programmiertechnischen Expertise im Feld angetippt. Wickham schlägt eine Standardisierung der Verbindung zwischen der Struktur eines Datensatzes und seiner Sinngebung vor – und zwar im folgenden Format: »In tidy data: 1. Each variable forms a column. 2. Each observation forms a row. 3. Each type of observational unit forms a table.« (Wickham 2014:4) Die von Wickham skizzierte Standard-Strukturierung ist heute ein wichtiges Prinzip in der Anwendung der Programmiersprache R, welche sehr oft in datenjournalistischen Praktiken zur Anwendung kommt. Auch in den beiden Teams basieren einige Praktiken auf R, wie die weiteren empirischen Ausführungen zeigen.

Bauen die Datenjournalist*innen einen Datensatz selbst auf oder strukturieren sie ihn wieder oder neu, erfolgt eine zeitaufwendige Phase der Konzeption, während derer die Aspekte Zeitaufwand, Verfügbarkeit, vor allem aber die Einordnung mittels Variablen abgestimmt werden. Dieser Abstimmungsprozess erfolgt vor dem Horizont, dass die angewandte Einordnungsweise der Beobachtungen mittels Variablen oder Metadaten zu spezifischen Sinngebungen und Formen der Visualisierung der Daten einlädt und andere ausschliesst (Flyverbom und Madsen 2015:135).

⁹ beispielsweise 57_20190109_BR-Data-11 (4)

Interoperabilität herstellen

NZZ Storytelling bewertet Datensätze besonders gut, wenn die Daten ohne grossen Bearbeitungsaufwand in ihre Visualisierungstools einfügbar sind. So lobt ein Teammitglied das lokale statistische Amt, welches Daten als Excel-Datenei liefert und über eine *Open-Data*-Strategie verfügt, was das Einspeisen der Daten in die Tools vereinfache.¹⁰ In einem anderen Projekt, für das Wahlresultate aus dem Internet gescrappt werden, die weniger »sauber« sind, unternehmen die Programmierer des Teams einen zusätzlichen Aufwand zur Standardisierung mittels eines Skripts. Dadurch waren die Daten ohne weitere Bearbeitung mit dem Datenvisualisierungstool Q kompatibel und eine Live-Update-Funktion führte dazu, dass die Datenvisualisierungen sich jeweils selbst aktualisierten, sobald neue Daten online gestellt wurden. Ein NZZ-Storytelling-Teammitglied präsentierte den Prozess stolz auf Twitter.

NZZ-Datenjournalist Yannick erzählt, dass er für das Team Prozesse entwickeln soll, um Terroranschläge in ihrem Ausmass schnell grafisch einordnen zu können. Auf die Frage hin, wie der Prozess funktioniert, antwortet er, dass zwei Webseiten die Datengrundlage für die Opferzahlen bieten.

Er habe dann ein Python-Script gebaut, das diese beiden Seiten mittels eines Protokolls schnell absuchen kann. Wir schauen uns das Skript direkt im Python-Interface Jupyter an und lassen es einmal laufen. Gerade heute Mittag hat es in Frankreich einen Terroranschlag gegeben und wir nutzen die dazu bekannten Zahlen und Daten. Diese werden gleich beim Script eingespeist und dann als Vergleichswerte genutzt. Dann, erklärt [Yannick], würden standardmässig sechs bis acht Charts ausgeworfen.¹¹

Die Praktik der Herstellung von Interoperabilität zwischen Datensatz und Tools sichert erst die Einlesbarkeit der Daten in die Tools und ermöglicht Analysen. Das störungslose Einlesen gehört deshalb zur Praktik, über *saubere Daten* zu verfügen. Abschnitt 5.4 geht auf die Organisationsformen ein, welche die Teams anwenden, um die Interoperabilität herzustellen und Dateninfrastrukturen zu ihrer Sicherstellung zu etablieren.

Insgesamt verfolgen die Teammitglieder mit dem Grundsatz der *sauberen Daten* ähnliche Ziele wie beispielsweise Datenanalysten mit »schönen Daten«, nämlich die Sicherung möglichst vieler künftiger Analysemöglichkeiten:

¹⁰ 22_20180323_NZZStorytelling-10 (16)

¹¹ 22_20180323_NZZStorytelling-10 (9)

»Die Schönheit von Daten liegt in ihrer Potentialität. Schöne Daten bzw. das Potential von schönen Daten erlauben es, die spezifischen Ansprüche, Hoffnungen und Unsicherheiten verschiedener AkteurInnen momentan zu absorbieren. Schöne Daten sind verheißungsvoll. Im Falle des untersuchten Unternehmens versprechen schöne Daten eine erfolgreiche Suche nach noch unbekannten Fakten und Erkenntnissen.« (Mützel et al. 2018)

Diese Potenzialität versuchen die Teammitglieder mit einem Set von Praktiken auszuschöpfen, indem die Datengrundlage mithilfe computergestützter Verfahren visualisiert und die Variablen dabei relationiert werden – darum dreht sich der folgende Abschnitt.

5.1.3 Geschichten drehen

Werden Geschichten bei BR Data oder NZZ Storytelling produziert, so nutzen die Teammitglieder häufig die Metapher des »Drehens« von Geschichten.¹² In diesem Set an Praktiken »verdrehen« bzw. »verwickeln« sich besonders ausgeprägt soziale und technologische Elemente: »..., capacities for action are seen to be enacted in practice and the focus is on constitutive entanglements (e.g., configurations, networks, associations, mangles, assemblages etc.) of humans and technologies.« (Orlikowski 2010:135) Die folgenden Abschnitte beleuchteten Praktiken des »Drehens« von Geschichten, wie sie empirisch in den Redaktionen beobachtbar waren, um aus Daten Erkenntnisse und Sinn zu generieren. Die Abschnitte gliedern sich in verschiedene Visualisierungs- und damit Relativerierungstypen sowie die daran angehefteten Diskussionen.

Zuerst aber eingeschoben ein Abschnitt zum aufgespannten methodologischen Raum, in dem diese Praktiken stattfinden: Sie sind verwickelte, iterative und abduktive Vorgehensweisen zwischen explorativer Datenanalyse und Thesen aus Themenwissen – im Argot *data-driven* beziehungsweise *story-driven* genannt. Abschnitt 3.2.3 hat diese epistemologischen Aspekte bereits im Feld thematisiert, auf einer diskursiven Ebene. Die folgenden Ausführungen zeigen auf, was dies in den beiden Teams praktisch bedeutet.

¹² 48_20181211_BR-Data-02 (3); 48_20181211_BR-Data-02 (3); 51_20181214_BR-Data-05 (5); 14_20180313_NZZStorytelling-02 (27).

Die Iteration zwischen *data-driven* und *story-driven*

Als einen der ersten Schritte in einem Projekt explorieren NZZ Storytelling und BR Data die Daten. Sie schliessen dabei an die explorative Datenanalyse (EDA) an, die John Tukey (1977) in den Siebzigerjahren begründete. Tukey erkannte, dass die statistischen Analysen und der Erkenntnisgewinn daraus in engem Zusammenhang mit Möglichkeiten der grafischen Abbildung und Kommunikation durch Visualisierungen standen, obwohl sie sich in Teilen in ihrer Zielsetzung unterscheiden (Gelman und Unwin 2013). Später schloss Edward Tufte an Tukeys Überlegungen an und schrieb: »At their best, graphics are instruments for reasoning about quantitative information« (Tufte 1983:9).

Die EDA provozierte im Feld der Statistik über lange Zeit viel Skepsis (und tut dies immer noch), weil die Exploration ohne oder nur mit schwachen Hypothesen manchen opportunistisch und unstrukturiert erscheint (Healy und Moody 2014:113). Trotzdem hat die EDA einen Aufschwung erlebt durch die Verfügbarkeit von computerbasierten grafischen und statistischen Verfahren, insbesondere *Tools* für interaktive Visualisierung – was auf der Hand liegt, denn »[e]xploratory graphics are all about speed and flexibility and alternative views« (Gelman und Unwin 2013:4). Die Geschwindigkeit und die Flexibilität haben sich durch die technologische Entwicklung der letzten Jahrzehnte zweifellos vergrössert.

Im Zeitalter von Big Data erfolgt eine Annäherung zwischen der explorativen Datenanalyse und konfirmatorischen Ansätzen: Wer mit grossen Datensätzen arbeitet, wird immer – auch wenn sie oder er über eine Hypothese verfügt – explorativ mittels verschiedener Visualisierungsmethoden einen Datensatz erkunden, um ihn zu säubern und seine Validität zu überprüfen (Healy und Moody 2014:113). Verschiedene neue Werkzeuge ermöglichen Interaktivität und damit das Ausprobieren verschiedener Beziehungen, die mit schnell wechselnden Interpretationen einhergehen (Healy und Moody 2014:116). Über ein solches *Tool* verfügt NZZ Storytelling mit Sue. Sue ermöglicht es, Daten explorativ zu visualisieren mittels sehr vieler verschiedener geometrischer Formen der Relationierung. Zum Beispiel zu Beginn des Projekts der Fussballtransfers:

[Kaspar] schiebt und probelt dann auf Excel, wie er die Ligen einteilen und kategorisieren kann. [Kaspar] sagt [zu mir], während er mit den Grafiken rumspielt und seinen Notizblock studiert, er überlege, was man aus der Visualisierung Spannendes herausziehen könnte – in welchen Ländern die Schweizer Liga Ge-

winn mache? Wie liesse sich das herausheben? [Kaspar] probiert einen Area-Chart aus, aber mit nur zwei Punkten: Auf der einen Seite die Einnahmen aus einer Liga, auf der anderen Seite die Ausgaben. Er meint, das sehe spannend aus. Er dreht sich zu Teamleiter David und fragt, ob er kurz vorbeischauen möge. Dieser kommt und fragt gleich, ob [Kaspar] das nun mit Sue gemacht habe. [Kaspar] bejaht. David schaut den Chart an und meint, dieser sei schwierig zu interpretieren. [Kaspar] fragt den Teamleiter, was einzelne Elemente in der Grafik genau bedeuten (zum Beispiel warum die Fläche der Schweizer Liga konstant ist), David erklärt (hierbei handelt es sich um die Transfers liga-intern). [Kaspar] meint, dass er nochmals daran weiterarbeite.¹³

Designer Kaspar tritt ohne starke Hypothesen und ohne viel thematisches Wissen zum Thema Fussballtransfers an die Analyse heran. Er probiert aus und evaluiert die entstehende Datenvisualisierung mit einem Mix aus ästhetischen und journalistischen Kriterien. Einige Stunden später hat Kaspar das Repertoire an Variablen in der Tabelle erweitert, um zusätzliche Visualisierungs- und Geschichtenoptionen zu eröffnen. Er sitzt am Pult und diskutiert mit Yannick, gleich neben ihm sitzend. Ich begebe mich zu ihnen:

Ich blicke auf die Tabelle, die [Kaspar] inzwischen um einige Spalten erweitert hat. Er hat Berechnungen und Sortierungen vorgenommen, die zeigen, ob es Ligen gibt, in die nur verkauft oder aus denen nur gekauft wird. So liessen sich alle Ligen in vier Muster einteilen. Auf dem Notizpapier hat er die vier Muster in Small Multiples aufgezeichnet. [Kaspar] sagt, Muster seien das Spannende. [Yannick] stimmt zu und sagt, dass sei spannend. [Kaspar] hat auch die Differenz berechnet, die aus Ein- und Verkäufen aus verschiedenen Ligen ausmachen, sowieso die Prozentzahl der Steigerungen. [Yannick] zeigt sich kritisch, das könne man nicht so darstellen. Die beiden diskutieren dann während rund zwanzig Minuten, wie und was man visualisieren könne. Sie haben einfach die Daten vorliegen und versuchen, daraus eine Visualisierung zu erreichen. [Yannick] versucht immer wieder, sich zu fragen, was man denn damit aussagen könnte, was die »Geschichte« sei. Er verfügt über etwas Wissen des Fussballgeschäfts und bringt Beispiele. [Kaspar] bringt kaum Fussballwissen-Inputs. Er zeichnet immer wieder Vorschläge auf seinen Notizblock. Auf meine Frage hin,

13 14_20180313_NZZStorytelling-02 (10)

was denn die Sportredaktion für einen Text dazu schreibe, ob denn da noch thematische Inputs oder Vorgaben kämen, verneinen beide. Sie denken nicht. Außerdem lief der Kontakt mit der Sportredaktion über David. [Yannick] sagt, dass sei »manchmal das Problem«, dass man hier selber vor den Daten sitze und eine Geschichte daraus entwickeln müsse. [Kaspar] vermutet, dass NZZ Storytelling die Visualisierung liefert und dann der Sportredaktor noch einen Text/Erklärung dazu liefert. [Yannick] findet, dass das Betrachten einzelner Fussballer-Karrieren spannend wäre. [Kaspar] ist mehr an den Mustern interessiert. Ihre Vorschläge legen immer wieder einen anderen Fokus: Geldflüsse zwischen den Ligen, den Status aus Ausbildungsliga, Management-Strategien, Scouting-Länder, ...¹⁴

In vielen anderen, im Falle von NZZ Storytelling gar den meisten Situationen, ist die datenjournalistische Praktik angeleitet von einer starken Hypothese beziehungsweise Geschichte – wie in einer Situation, als ein Wirtschaftsredaktor nach Vorankündigung beim Teamleiter vorbeikommt und erklärt:

Es gehe um die Firmenbeteiligungen von Pierin Vincenz [ehemaliger CEO der Schweizer Raiffeisenbank]. Der Redaktor legt eine handgezeichnete Skizze einer Visualisierung auf einem gehäuselten Papier auf den Tisch und erklärt seine Vorstellung: Eine Zeitachse von 2005 bis 2015, er erklärt auch die Achsenbeschriftungen. Er erläutert zudem kurz die Sachlage, was Vincenz für Firmenbeteiligungen hatte. Zudem kämen noch »Sub-Firmen« dazu, die visualisiert werden müssten. Es seien noch nicht alle Daten vorhanden, einige Kaufpreise fehlten noch. Der Teamleiter meint, dass einzelne Werte noch nicht entscheidend sind für einen Entwurf, aber der »range« muss klar sein. Der Redaktor schlägt eine logarithmische Skala vor, der vermutliche teuerste Wert ist ihm schon bekannt.¹⁵

Fragt man die Mitglieder der beiden Teams nach dem Kernstück des datenjournalistischen Prozesses, so verweisen sie sehr oft auf die Methodik der Produktion von Geschichten aus Daten, das Zusammenspiel von These und thematischem Wissen. Insgesamt oszilliert die Herangehensweise an die Datenanalyse zwischen den beiden Endpunkten (1) *data-driven* und (2) *story-driven*:

14 14_20180313_NZZStorytelling-02 (20)

15 16_20180315_NZZStorytelling-04 (10)

- (1) *data-driven*: Die Geschichte wird erst *aus den Daten* und deren Analyse heraus entwickelt (NZZ Storytelling) bzw. es gibt noch keine Vorstellung, *was in den Daten steckt* (BR Data). Obwohl in den beiden Redaktionen etwas andere Begriffe verwendet werden, referenzieren sie dieselbe Herangehensweise.
- (2) *story-driven*: Es gibt eine Geschichte oder eine klare Idee dafür, die mit Daten als zusätzlicher Perspektive *ergänzt* wird (NZZ Storytelling), bzw. es gibt eine Fragestellung oder Hypothese, welche an einem Datensatz *überprüft* wird (BR Data). Auch hier unterscheiden sich die gewählten Begriffe, aber es wird dieselbe Herangehensweise referenziert.

Im redaktionellen Alltag sind diese beiden Herangehensweisen selten in reiner Form zu finden. Viel eher ist es plausibel, sich die beiden Typen als Endpunkte eines Spannungsfeldes vorzustellen. Die Erkenntnisproduktion bewegt sich spiralförmig in diesem Spannungsfeld und ist folglich abduktiv (Kitchin 2014a; Halford und Savage 2017): In einem iterativen Vorgehen wechseln sich das Erheben und Analysieren von quantitativen Daten und das Sammeln von Kontextwissen mit qualitativen Methoden ab, sich stetig zu einer Geschichte hin entwickelnd – Abschnitt 5.1.4 geht darauf ein, wie diese unterschiedlichen Perspektiven passend gemacht werden. Zuerst folgt ein Blick in die Erkenntnismaschinerie: Praktiken des *Geschichten-Drehens* aus den Redaktionsräumen von BR Data und NZZ Storytelling. Diese Relationierungspraktiken sind implizit strukturiert durch das Iterieren zwischen *data-driven* und *story-driven* und operieren im dadurch aufgespannten Raum.

Suchen von Mustern und Auffälligkeiten

Man schaut sich die Datensätze an, sucht nach Auffälligkeiten, Peaks, Verteilungen; mit den statistischen Methoden kann man dann Geschichten herausholen – so beschreibt BR-Data-Teammitglied Johannes, wie man aus Daten Geschichten »macht«.¹⁶ Die Suche nach Auffälligkeiten beschrieb bereits John Dewey (1927) als Charakteristikum von *news*, dessen Überlegungen Wihbey (2019:151) wie folgt auf den Punkt bringt: »The search for discontinuity is what makes news news.« Für diese Suche nach Diskontinuitäten in Datensätzen, die einem *auffallen*, nutzen die beiden Teams Datenvisualisierungen. Das Überstülpen eines visuellen Vokabulars produziert visuell erkennbare Muster und Auffälligkeiten:

16 BR_Fragebogen_Mitarbeitende_10

[Brigitte] hat nun auf ihrem Computer html-Dokumente mit Säulengrafiken der Anzahl Berichte nach Quartal, dabei ist in der Säule der Anteil der Berichte mit Herkunftsennnung ausgewiesen. [Brigitte] und [Lea] vergleichen verschiedene Städte und Regionen (jeweils eigenes html-File nach Polizeistation – teilweise sind dies Städte, Landkreise, ganze Bundesländer, die Einteilung ist sehr unterschiedlich). Die Zahl der Berichte aus den einzelnen Stationen ist sehr unterschiedlich, auch der Anteil der Herkunftsennnung. [Brigitte] und [Lea] finden den Vergleich »spannend«, vor allem zwischen Städten ähnlicher Grösse, die aber einen sehr unterschiedlichen Anteil haben. Dieser schwankt teilweise auch über die Zeit.¹⁷

Der Ausschnitt aus der Feldnotiz beleuchtet ein Projekt von BR Data, von dem die Daten – kurze Polizeimitteilungen (»Berichte«) in Textform, verteilt auf fünf Jahre, insgesamt rund eine Million – gesammelt sind. Seitens des Teams besteht die Vermutung, dass sich die Herkunftsennnungen der Personen in diesen Berichten verändert haben – über die Zeit, möglich wären aber auch Differenzen zwischen verschiedenen Polizeistationen. Mittels Textanalyse hat die Programmiererin Brigitte deshalb extrahiert, wie oft die Berichte die Herkunft nennen. Der Anteilsquotient wird Variablen wie dem Jahr oder der Absenderin (Polizeistation einer Stadt, Landkreis, Bundesland) zugeordnet und dann diese Relation über die Zeit oder die Verteilung nach Absenderin verglichen. Die Produktion der Geschichte folgt dem Kriterium der Differenz der Anteile und eine hohe Differenz bewerten Brigitte und Lea mit »spannend«.

Der Vergleich erfolgt mittels einer Datenvisualisierung in Form eines Säulendiagramms. Grosse Differenzen widerspiegeln sich in sehr unterschiedlich hohen Säulen und sind für Brigitte und Lea schnell visuell erkennbar. In dieser Phase des Projekts sind noch viele *Geschichten-Drehs* offen, das heisst, der Möglichkeitsraum der Erkenntnisproduktion wurde durch die granulare Datenerhebung und diverse Metadaten der Berichte – die nun zu Variablen für die Auswertung werden – offen gehalten. Aus der hohen Anzahl an Vergleichsmöglichkeiten selektieren die beiden Teammitglieder visuell anhand der erkennbaren Höhendifferenzen der Säulengrafiken, welche Muster und Auffälligkeiten relevant sind für die Erkenntnisproduktion.

17 56_20190108_BR-Data-10 (8)

Ausschlagende Kurven

Die Kurve ist bei BR Data keine eng definierte mathematische Gleichung, sondern wird ganz generell als Beschreibung des Graphen eines Liniendiagramms genutzt. Der Verlauf der Kurve ist während der Phase der Erkenntnisproduktion ein wichtiger visueller Indikator für den *Geschichten-Dreh*:

Ein Programmierer hat dann mittels eines R-Scripts diverse Auswertungen vorgenommen, er stand dafür in Austausch mit [Martin]. Diese Auswertungen hat er in ein [html-]Dokument eingefügt, in dem der Datensatz mittels R ausgewertet wird, sichtbar in diversen grafischen Darstellungen. Der Code dazu ist ein-/ausklappbar. Es wurden die Anzahl der Meldungen, deren Verteilung auf die Bundesländer und Polizeistellen, deren zeitliche Entwicklung, die Aufschlüsselung auf einzelne Delikte sowie mittels bestimmter Wörter auch auf Nationalitäts- und ethnische Zuschreibungen vorgenommen. Diese Explorations wurden mittels Linien- und Balken-Diagrammen (produziert mit R) abgebildet. [Lea] erwähnt, dass es darum ging, Auffälligkeiten, »Peaks« in Kurven und ungewöhnliche Entwicklungen über die Zeit zu finden.¹⁸

Das Finden von Abweichungen bedingt eine Vorstellung der Normalität. Beides wird in diesem Projekt aus den Datenvisualisierungen heraus konstruiert. Denn was ein normaler Anteil an Herkunftsennungen sei, davon hat BR Data keine klare Vorstellung: Man geht davon aus, dass es keine kollektiven Leitlinien für Polizeistationen gibt, wann die Herkunft zu nennen ist – und deshalb ist auch nicht klar, welche Verteilung ›normal‹ wäre.¹⁹ Wenn die Kurve ausschlägt, die Richtung ändert, dann ist das ›nicht normal‹. Dann ist es ›spannend‹. Dann begibt man sich auf die Suche nach der Variablen, welche die Veränderung verursacht hat.

Adrian untersucht für ein gemeinsames Projekt mit einer Redaktorin aus einem anderen Ressort Verschreibungszahlen von verschiedenen Medikamenten derselben Sorte. Zusammen mit Martin versucht er, Auffälligkeiten in den Kurven zu finden.

[Martin] und [Adrian] wechseln zu Tabelle und Linien-Diagramm, die den allgemeinen Verbrauch (Verschreibungszahl) anzeigen. Die Kurven gehen in der

¹⁸ 48_20181211_BR-Data-02 (5)

¹⁹ 48_20181211_BR-Data-02 (5)

Tendenz zurück. [Martin] meint, man könne [der Redaktorin aus einem anderen Ressort] Daten anbieten, die zeigten, dass der Verbrauch eher zurückgehe. Auf einem weiteren Liniendiagramm zeigt sich, dass die Verschreibungszahl eines Medikaments entgegen dem sinkenden Trend stark ansteigt. [Martin] fragt [Adrian], was das für ein Medikament sei. [Adrian] erklärt kurz, [Martin] findet diese Entwicklung spannend.²⁰

Der mehrheitlich sinkende Trend als Normalfall dient als Abgrenzungsfolie für das eine Medikament mit steigendem Trend, welcher das Interesse weckt: »spannend« – die Spannung baut sich auf zwischen der konstruierten Normalität und der mindestens für den Moment unerklärbaren Abweichung (wobei sich diese Spannung wenig später auflöst, weil man sich den gegenläufigen Trend des einen Medikaments mit dem abgelaufenen Patentschutz erklärt).

Der um die Kurve tanzende *Geschichten-Dreh* findet seine Fortsetzung im nächsten Excel-Tabellenblatt. Dort gibt es zu einem verwandten Aspekt des Themas (Resistenzen) Daten in Tabellenform, Balken- und Liniendiagramme. Adrian und Martin betrachten das Tabellenblatt:

[Adrian] zeigt auf die Diagramme und meint, dass die Resistenzen etwas zunehmen, aber nicht allzu stark. [Martin] meint, dass es aus Datensicht »nichts gebe, dem man nachgehen sollte«. [Martin] und [Adrian] interpretieren zusammen die Kurve und wie sie sich verändert – [Martin] schlägt auch vor, sich die Kurve vorzustellen, würde man die ausschlagenden Werte einiger Jahre weglassen. [Martin] meint, man könne dies [der Redaktorin aus dem anderen Ressort] mitteilen, dass man in den Daten nichts sehe – was aber nicht heisse, dass das Problem nicht existiere.²¹

Adrian und Martin erscheint die Entwicklung der Kurve aufgrund des Hintergrundwissens plausibel. Auch das gedankliche Experiment, sich einige (offenbar plausibel begründet) ausschlagende Werte ›wegzudenken‹, führt nicht zu Auffälligkeiten, denen man nachgehen sollte. Der etwas erfahrene Martin grenzt die eigene Perspektive als »Datensicht« ab, in der man ›nichts sieht‹ – oder von der anderen Seite betrachtet: *seeing data* heisst für die beiden Datenjournalisten, mittels Auffälligkeiten sich potenzielle Geschichten imaginieren zu können.

²⁰ 51_20181214_BR-Data-05 (2)

²¹ 51_20181214_BR-Data-05 (2)

Bemerkenswert an der Situation ist der reflexive Moment, der zum Schluss auftaucht (aber schliesslich nicht weiter ausgeführt wird). Martin grenzt die »Datensicht« nicht nur ab, sondern relativiert deren Aussagekraft: Ein von der Redaktorin mit ihren journalistischen Werkzeugen postulierter Sachverhalt widerspiegelt sich nicht zwingend »in den Daten«. Damit anerkennt er nicht nur die Grenzen der eigenen Perspektive, sondern beugt vor allem einer konfliktbeladenen Situation im darauf geplanten Gespräch mit der Redaktorin vor: Es braucht kein Streit darüber ausgetragen werden, welches Werkzeug nun die validere Erkenntnis hervorbringt.

Ausreisser

Neben sich »ungewöhnlich« verhaltenden Kurven wecken auch Ausreisser das Interesse. Als Ausreisser wird gemeinhin eine Beobachtung bezeichnet, die nicht in einen erwarteten Bereich fällt. Diese Definition teilen Datenjournalist*innen mit Statistiker*innen, welche sich seit Bernoulli 1777 mit dem Phänomen beschäftigen (Beckman und Cook 1983). Treffen die Praktiker*innen aus den jeweiligen Bereichen auf einen Ausreisser, so unterscheiden sich jedoch ihre Schlussfolgerungen: Während die Statistik nach Wegen sucht, trotz Ausreissern ihre Resultate sinnvoll interpretieren zu können (Zimek und Filzmoser 2018), so bezeichnet das Datenjournalismus-Handbuch Ausreisser entweder als Fehler – oder als potenzielle »good stories« (Gray et al. 2012:186). Ausreisser und starke Abweichungen sind wichtige Attribute, welche die Aufmerksamkeit für eine Datenvisualisierung erhöhen (Segel und Heer 2010:1140).²² Murray Dick bestätigte in seiner Forschung über die Infografik-Produktion in englischen Medienorganisationen die vorhandene, wenn auch nicht unhinterfragte Attraktivität von Ausreissern (Dick 2014:500).

Lea schlägt für das Polizeibericht-Projekt vor, Ausschau nach solchen zu halten:

[Lea] schlägt vor, die Ausreisser in den Polizeistationen zu finden mit [Brigitte] zusammen. [Martin] meint auf jeden Fall, das sei ja auch spannend, sich dann

²² Welche Attraktivität die Figur des Ausreissers auf die Datenvisualisierungs-Community ausübt, zeigt sich daran, dass die Data Visualization Society – Fussnote 4 (Kapitel 2) hat bereits auf sie hingewiesen – ihr Konferenzformat »Outlier« nennt. <https://www.outlierconf.com/> (21.10.2020)/<https://medium.com/nightingale/introducing-outlier-a-conference-hosted-by-the-data-visualization-society-801ca8ff87ff> (21.10.2020)

genau diese Stationen anzuschauen. Dann könne man auch schauen, wie die Mitteilungen sich aus diesen Stationen verhalten.²³

Im Gegensatz zu den für statistische Analysen entwickelten Tests (Beckman und Cook 1983), welche Ausreisser mathematisch definieren, dient der Ausreisser in der datenjournalistischen Praktik als vage Zuschreibung für einen gegenläufigen Trend oder Extremwert, hier in einem Projekt von NZZ Storytelling:

[Susanne] erzählt [Annalena], dass sie gerade die Reportagen für das Gemeinde-Projekt im Wallis vorbereitet und eine Reportage aus Naters eine Option sei, sie sei gerade am Abklären mit der Gemeinde. Naters sei genau das Gegenteil der Tendenz und gerade darum spannend. Es gebe diverse Gründe, wieso die Leute aus den Seitentälern nach Naters zögen. Solche, die aus ähnlichen Fällen bestens bekannt seien – aber es werde auch ein neues Altersheim gebaut. [Annalena] meint, sie fände das sehr spannend, super.²⁴

Die Sinngebung der Datenanalyse und Produktion von Erkenntnis aus der Visualisierung fällt Susanne und Annalena in diesem Beispiel besonders leicht, weil sie sowohl die Grundtendenz (Abwanderung aus den Alpentälern des Wallis) wie auch der Ausreisser mit der gegenläufigen Tendenz (Naters) mithilfe des sonstigen Wissens (Altersheim) plausibilisieren können.

Die Beispiele verweisen auf drei Eigenschaften, welche die Attraktivität des Ausreissers ausmachen: Erstens ist er visuell auffällig und weckt damit Aufmerksamkeit. Zweitens sind Ausreisser eng verknüpft mit der Idee, Neuheiten zu ›entdecken‹ – einerseits durch die statistischen Verfahren, welche für die Feststellung von Ausreissern sowie für Neuheit die gleichen sind (Pimentel et al. 2014), andererseits bergen sie, gespiegelt auf den Journalismus, das Versprechen, dass Neuheiten *Neuigkeiten* und damit *news* sind. Drittens bieten Ausreisser, die per Definition auf einzelne Fälle referenzieren, eine ideale Ausgangslage für die Ergänzung von qualitativen und quantitativen Erkenntnissen und deren Passendmachung – Abschnitt 5.1.4 geht auf diese Praktiken der Plausibilisierung näher ein.

²³ 56_20190108_BR-Data-10 (4)

²⁴ 24_20180327_NZZStorytelling-12 (4)

Daten geografisch verteilen

Das Verteilen von Daten nach geografischen Variablen, das *Mappen*²⁵ bzw. Kartieren, ist eine oft gewählte Vorgehensweise zur Erkenntnisgewinnung. Sie schafft Übersicht durch ein vertrautes Ordnungssystem (Nohr 2004). Beide Achsen sind mit vertrauten Variablen besetzt: Breiten- und Längengrad. Bei diesem Wiedererkennen handelt es sich um wichtiges methodisch-technisches Wissen, das dem Publikum erst erlaubt, eine Datenvisualisierung zu verstehen (Passmann 2013). Deshalb schreiben Datenjournalist*innen ihm zu, dass es Visualisierungen auf Karten versteht – und gleichzeitig ist die Karte auch für Redaktionsmitglieder ausserhalb der datenjournalistischen Teams aus den gleichen Gründen die zugänglichste Datenvisualisierungsform. Sie wird bei der NZZ mit am häufigsten gebraucht, sei es als ›Auftrag‹ für das Infografik-Team oder produziert mittels Q.

Jedes dritte NZZ-Storytelling-Mitglied wünscht, sich Fähigkeiten in QGIS anzueignen oder diese zu verbessern – wobei es sich dabei um diejenige Software handelt, mit welcher im Team sogenannte Choroplethenkarten produziert werden.²⁶ Einige Monate nach meiner Feldforschung bei der NZZ ist es dann auch eine Choroplethenkarte, welche es als (vermutlich) erste auf die Meinungsseite der NZZ-Printausgabe schafft²⁷:

Die Geschichte dreht sich hier um die Emigration aus osteuropäischen Staaten, was durch die Einfärbung *augenfällig* wird. Begonnen hat das Projekt an einer Sitzung während meiner Forschungsphase, wobei damals erst von Daten zu (demografisch) ›schrumpfenden‹ osteuropäischen Ländern die Rede war²⁸ – woraus schliesslich in Kooperation mit einem Auslandredaktor, ›mit dem wir immer gut zusammenarbeiten‹²⁹, ein Beitrag über die Emigration und deren Gründe wird.

-
- 25 Ist im deutschsprachigen Datenjournalismus von »Mapping« die Rede, so meist im Sinne des Kartografierens auf einem geografischen Koordinatensystem – auch wenn der englische Begriff »mapping« ganz generell das Abbilden auf einem Koordinatensystem (welcher Art auch immer) meinen kann.
- 26 13_20180312_NZZStorytelling-01 (28) – Choroplethen sind Karten, auf denen die Einheiten (beispielsweise Länder) auf der Ausprägung einer Variablen eingefärbt oder konturiert sind.
- 27 Twitter-Austausch mit David Bauer vom 2. August 2018.
- 28 21_20180322_NZZStorytelling-09 (4)
- 29 Twitter-Austausch mit David Bauer vom 2. August 2018.

Abbildung 8: Choroplethenkarte des Meinungsbeitrags »Europas verdrängte Revolution wirkt nach«.



NZZ vom 2. August 2018, Seite 10

Mittels Karten wird das geografische Koordinatensystem überlagert mit einer ordinalen oder metrischen Variable (in Abbildung 8 mit Zu- bzw. Abnahme der Bevölkerung). Das Koordinatensystem wird dafür in der Regel in politischen Gebietskörperschaften kategorisiert: Kommunen, Regionen, Bezirke, im Beispiel hier Länder, oder Kontinente. Diese Relationierung der Daten mit politischen Entitäten ermöglicht einerseits einen schnell augenschein-

lichen Vergleich. Andererseits korrelieren mit politischen Entitäten und deren Grenzziehungen weitere strukturelle Merkmale. So lassen sich an politische und gesellschaftliche Diskussionen schnell anschlussfähige Erkenntnisse treffen und die politischen Verantwortlichkeiten klar benennen – wie im Meinungsbeitrag der NZZ. Dies macht Choroplethenkarten im Journalismus besonders attraktiv.

Karten als Visualisierungen weisen ihre Tücken auf (siehe Cairo 2019). Bei BR Data wird ihre Aussagekraft kritisch reflektiert, als Johannes an einem Teammeeting von einem Projekt mit Daten über die Funktionstüchtigkeit der Aufzüge an den deutschen Bahnhöfen erzählt:

[Mario] habe ihm [Johannes] SQL so eingestellt, dass man die defekten Aufzüge auf einer Karte ansehen und verorten könne. Es gebe keine Überraschungen, viele defekte Aufzüge befänden sich in den Ballungszentren. Allenfalls müsse man eine Verbindung zur Bevölkerungsdichte machen, um die Intensität der Nutzung abzuschätzen. [Martin] meint, das könnte schwierig sein. [Moritz] meint, es sei eigentlich egal, ob ein Aufzug vielbenutzt sei. [Thomas] sagt, es sei eigentlich räumlich nur spannend, wenn es irgendwo besonders schlimm sei – sonst erscheine ihm eine Karte nicht sinnvoll, auch nicht der Abgleich mit der Bevölkerungsdichte. Es wäre besser, den Anteil an defekten Aufzügen auszuweisen. [Johannes] meint, auch die Dauer der Störung wäre spannend, er müsse mit [Mario] absprechen ob das gehe. [Martin] meint, es wäre auch gut, Einzelfälle aufzugreifen.³⁰

Trotz der verbreitet kritischen Einschätzung von Karten – auch bei der NZZ³¹ – gehören Karten in den beiden Teams zu den meistgenutzten Datenvisualisierungsformaten, weil die damit produzierten Erkenntnisse schnell auf die politische Ebene spiegelbar sind.

Auf den Abschnitt zurückblickend lässt sich zusammenfassen, dass die Metapher des *Geschichten-Drehens* soziotechnisch verwinkelte Praktiken beschreibt, durch welche die Teams Relationierungen von Variablen vornehmen. Die Praktiken verorten sich in einem Spannungsfeld von *data-driven* und *story-driven*, zwischen denen die Teammitglieder iterieren.

Für die Relationierungen setzen die Teams Verrechnungen durch geometrische Verkörperungen ein. Die *Geschichten*, das heißt aus journalistischer

³⁰ 58_20190110_BR-Data-12 (1)

³¹ 13_20180312_NZZStorytelling-01 (29)

Perspektive relevante Erkenntnisse, entstehen durch journalistisch anschlussfähige Zuschreibungen, welche die Praktiken provozieren. Der Abschnitt hat beschrieben, wie beispielsweise Säulen-, Balken- und Liniendiagramme Auffälligkeiten und (Dis-)Kontinuitäten sichtbar machen, wie Normalitäts- und Abnormalitäts-Einteilungen aktiviert werden, wie Ausreisser Neuheit und Neuigkeit versprechen, wie Karten die Verbindung von Mustern mit politischen Entitäten und ihren Strukturen ermöglichen.

Die Datenvisualisierungen sind Objekte, welche die analysierenden Aushandlungen in Konversationsform strukturieren und so eine Wissensordnung zugrunde legen (Burri 2008). Gleichzeitig erfolgt eine ständige Verarbeitung und Anpassung der Objekte, um sie zu Geschichten zu drehen.

5.1.4 Auswertungen plausibilisieren

Der folgende Abschnitt beleuchtet datenjournalistische Praktiken, welche die Datenauswertungen und die Sinngebung daraus zu Erkenntnissen plausibilisieren. Zuerst werden die Voraussetzungen eingeführt, welche für die Plausibilisierung erst erforderlich sind.

Die Differenzierung von ›Daten‹ und ›Realität‹

Die beiden Teams von NZZ Storytelling und – akzentuierter – BR Data differenzieren zwischen zwei Welten: die »echte Welt« und die »Datenwelt«. Roman von BR Data erzählt begeistert über ein vergangenes Projekt, das eine Recherche vor Ort beinhaltete: Es sei toll gewesen, vom Bildschirm wegzukommen und auch mal Verbindungen zur »realen Welt« hinter den Daten zu machen.³² Und sein Teamkollege Thomas blickt folgendermassen auf eine vergangene investigative Recherche zurück:

Bei der Datenanalyse selber hätten sich dann die Hypothesen teilweise bestätigt, teilweise nicht. Er habe eher die journalistische Seite abgedeckt, und bei den Analysen hätte er sich schon stark auf [Lukas] und die [andere Medienorganisation]-Leute verlassen, die seien da wirklich »tief drin« gewesen. Den Code habe er selten angeschaut, manchmal habe er die Notebooks herausgefordert, kritische Fragen gestellt. Er habe viel Wissen über den Bereich gesammelt, mit Experten gesprochen, Ergebnisse »in die echte Welt« übertragen und geprüft.

³² 47_20181210_BR-Data-01 (12)

Auch habe er geschaut, dass die Nachrichtenwerte drin seien – er habe fast schon eine Redakteurs-Rolle gehabt.³³

Thomas zählt sich innerhalb von BR Data eher zu den ›Autoren‹, also denjenigen, welche aus dem journalistischen Bereich kommen und entsprechende Expertise mitbringen. Er kann zwar programmieren, aber meistens ist er es, der die Geschichten schreibt. Er sieht seine Aufgabe darin, die »Verbindung [der Daten] zur echten Welt« sicherzustellen.³⁴

Besonders für Personen mit grosser journalistischer Expertise ist die Differenzierung der Welten ein Reibungspunkt in der datenjournalistischen Praktik: Daten können ein schlechtes Abbild der Welt sein und man kann sich nicht nur darauf verlassen³⁵; man muss die Daten mit der Realität abgleichen³⁶; man blickt für eine Geschichte auch darauf, was hinter den Daten steckt³⁷ – so widerspiegelt sich die Differenzierung in einigen Aussagen der Datenjournalist*innen. Sie hängt eng zusammen mit der bereits erläuterten epistemologischen Unterschiedlichkeit von digitalen Daten und Methoden der Erkenntnisproduktion (Abschnitt 3.2.3) und manifestiert sich besonders häufig im iterativen Prozess von *data-driven* und *story-driven* – nun mit einem Bezug zur Sinngebung der hergestellten Relationierung: Was bilden die Daten *wirklich* ab?

Dass sich diese Frage stellt, insbesondere bei Teammitgliedern mit grosser journalistischer Expertise, verweist auf die Unterschiedlichkeit des datengetriebenen epistemologischen Zugangs im Vergleich zum ›klassischen‹ – Abschnitt 3.2.3 hat diese Unterschiedlichkeit bereits erläutert und insbesondere zwei Punkte herausgehoben: den abstrahierten Zugang zu einem Phänomen sowie die Annahme, dass Fakten real und stabil sind. Während der letztere Punkt eine von datenjournalistischen Analysen geteilte Annahme ist, gilt es den ersten Punkt, den abstrahierten Zugang zu einem Phänomen, mit den anderen Zugängen abzulegen, *passend* zu machen.

33 57_20190109_BR-Data-11 (5)

34 BR_Fragebogen_Mitarbeitende_08

35 BR_Fragebogen_Mitarbeitende_08

36 BR_Fragebogen_Mitarbeitende_01

37 BR_Fragebogen_Mitarbeitende_02

Die Passendmachung von qualitativen und quantitativen Zugängen

Konflikte auf epistemologischer Ebene bei investigativen, datengetriebenen Projekten hat Parasie empirisch ausführlich erforscht. Während die Studie aus Chicago (Parasie und Dagiral 2012) in dieser Arbeit zur Einordnung der epistemologischen Grundlagen im Feld diente, bietet Parasies Folgeprojekt (2014) einen geeigneten Anschlusspunkt zur Untersuchung in einem Team auf Projektebene. Parasie untersuchte die Produktion einer investigativen Recherche (»On Shaky Ground«) 2011/12 in Kalifornien mithilfe der Perspektive der »epistemischen Kulturen« (Knorr-Cetina 2002). Er beschreibt, wie die Spannung zwischen der datengetriebenen und der ›klassischen‹ Epistemologie des investigativen Journalismus – es gibt eine aufzudeckende Wahrheit; dafür werden verschiedene Darstellungen gesammelt und abgestimmt; die ›Fakten‹ und die Geschichte entwickeln sich simultan – nur unter hohem Ressourcen-Aufwand aufgelöst werden kann (Parasie 2014:373): »Adjusting the artifacts to the established epistemologies was a demanding process for the organization. [...] But it ended up producing collective confidence in the data.« Das Vertrauen basiert auf einer geteilten epistemischen Kultur, die dazu führt, dass Akteur*innen, die eigentlich verschiedene epistemologische Grundannahmen vertreten, das-selbe in den Daten *sehen*.

»Adjusting the artifacts« meint in Parasies (2014:372f) Beitrag das Hinzufügen von »real-life evidence« – womit genau der oben beschriebene qualitative Zugang durch Methoden wie Interviews gemeint ist – zu Datenbanken und -visualisierungen. Im Beispiel wird auf einer im Projekt entstehenden Karte eine zusätzliche Ebene hinzugefügt. Diese Ebene soll einen bestehenden Unsicherheitsbereich anzeigen, weil das Recherche-Team herausfand, dass sie die Präzision der Kartendaten nicht garantieren konnten. Gleichzeitig verpflichteten sie die Standards der investigativen Recherche dazu, wahrheitsgemäß über die Verortung der Institutionen auf der Karte zu berichten. Schliesslich war die interaktive Karte (Parasie 2014:369) nicht nur ein zentrales Resultat der Recherche, sondern ermöglichte es, die quantitativ und qualitativ gesammelten Daten zusammenzubringen und abzustimmen.

Je grösser ein Projekt, desto umfangreicher sammeln die beiden untersuchten Teams weiteres Material mittels ›klassischer‹ Methoden: Man spricht mit mindestens einer*m Expert*in, sammelt wissenschaftliche Literatur zum Thema, recherchiert bereits erschienene Berichterstattung.³⁸ Aus diesen

³⁸ In einem Projekt von BR benötigt es gar eine *readme*-Datei, um die Übersicht in der Recherche-Dokumentation zu behalten, siehe 52_20181217_BR-Data-06 (5).

qualitativen Recherchen gewinnen die Teams Wissen über das Thema, treffen Annahmen und evaluieren mögliche Betrachtungsperspektiven (»angles«) auf die Daten – wie Lea anhand des Polizeiberichte-Projekts erläutert:

Dort habe sie [Lea] ein Dossier angelegt mit einem Datenordner (alle Daten in CSV-Format), einem Literatur-Ordner (sie habe viel gelesen, auch Studien über psychologische Hintergründe), ein Daten-Exposé, wo die Ergebnisse der Datenanalyse niedergeschrieben seien, die vertraulich geführten Hintergrundgespräche unter anderem mit Polizei-Angehörigen – diese hätten viel erzählt, sobald sie wüssten, dass es nicht veröffentlicht würde, wobei es einen dann wieder vor Schwierigkeiten stelle, wenn man Zitate brauche –, dann ein sehr detailliertes Recherche-Protokoll, das alle Schritte auflistet, die im Projekt gemacht wurden, und ein Dokument namens »Background« ([Lea]: »eine kleine Master-Arbeit«), in dem sie die Hintergründe, Hypothesen, die möglichen Recherche-Richtungen (»angles«) festhalte – Möglichkeiten seien hier, die Polizei als Publizisten zu sehen und sie an entsprechenden Vorgaben zu messen, oder die Geschichte mit der »gefühlten Kriminalität« zu rahmen.³⁹

Mit welchen Praktiken versuchen die beiden Teams, die Zugänge passend zu machen, ›gemeinsam zu sehen‹ und das ›Vertrauen in die Daten‹ zu sichern? Es handelt sich um Praktiken der Plausibilisierung: Sample ziehen, Einzelfälle untersuchen, vergleichen.

Sample ziehen

Im Projekt über die Polizeimeldungen besteht eine der Annahmen darin, dass sich diejenigen Meldungen mit Nationalitätsnennung viel öfter auf fremdenfeindlichen Plattformen und Blogs im Netz wiederfinden als diejenigen ohne. Um dies zu prüfen, arbeitet BR Data mit dem externen Experten Gabriel zusammen, der ein *Tool* entwickelt hat, welches die Weiterverbreitung von Text im Internet nachzeichnen kann. Die Projektverantwortliche Lea versuchte bereits vor der Zusammenarbeit mit Gabriel, mittels eines kleinen Samples aus der gesamten Menge an Meldungen herauszufinden, ob sich diese im Internet wiederfinden lassen – und sie diskreditiert später Gabriels (intransparente) Methode, weil sich ihr Anteil an manuell wiedergefundenen Meldungen als

39 48_20181211_BR-Data-02 (5)

weit höher herausstellt als derjenige Gabriels. Sie macht Gabriels Resultate damit unplausibel – zuerst implizit⁴⁰, etwas später dann noch explizit, indem sie ihre manuellen Resultate höher bewertet als die (fehlenden) Resultate Gabriels: Ihre manuelle Sample-Analyse habe ergeben, dass »da etwas ist«, und man müsse nach weiteren Methoden suchen, um dies zu automatisieren.⁴¹

Kurz darauf beraten Lea und Martin über weitere quantitative Methoden. Um diese zu explorieren, soll die Programmiererin Brigitte Ähnlichkeitssuchen mit einem Sample von Meldungen vornehmen.⁴² Einige Stunden später, das Büro hat sich bereits geleert, sitzen Lea und Martin am frühen Abend gemeinsam vor dem Bildschirm und probieren eine weitere Methode aus. Tatsächlich ergibt die Suche von einigen Textbausteinen über Google Resultate – die beiden schalten über Skype Brigitte dazu:

[Martin] erklärt [Brigitte], er habe bei den Polizei-Meldungen jeweils das Kürzel am Schluss und die Ortsbezeichnung am Anfang weggenommen und diese dann in eine Google-Suche eingefügt – Google finde dann verschiedene Portale mit derselben Meldung. Es gebe aber auch Treffer bei Google, die aber nur in einigen (fett markierten) Begriffen übereinstimmten, aber eigentlich nicht die gleiche Meldung seien – dies wären gemäss [Martin] dann vermutlich False Positives, die man also anhand der fett markierten Schlagwörter identifizieren könnte. Er würde dann nur die wortgleichen Meldungen anschauen, die auf der ersten Seite der Google-Suche erscheinen. [Brigitte] hat in der Zeit ebenfalls die gleiche Meldung – [Martin] hat sie ihr über Slack zugestellt – über Google gesucht und teilweise die gleichen Ergebnisse erhalten, teilweise nicht. [Lea] meint, das sei ja auch schon nur zu verdauen, dass das so unterschiedlich sei... [Brigitte] meint, normalerweise suche sie Artikel über einzelne Schlagwörter, bei ganzem Text gebe es jeweils doch recht viele False Positives. [Martin] meint, wenn er in der Struktur der Google-Webseite der Ergebnisse schaue, dann sehe er in den Kurztexten, dass diese in einem »<div>« steckten, und dort könne man die fettgedruckten Wörter mittels »« automatisiert identifizieren (und so die False Positives aussortieren). Bei »schlechten Meldungen«, also eher »Allerwelts-Meldungen«, so [Martin], würde Google mehr Treffer ausspucken, habe er nun bei einigen Versuchen gemerkt. Aber man könne ja mal die html-Seite der Google-Seite wegspeichern und dann die <div>s anschauen. [Brigitte]

40 57_20190109_BR-Data-11 (3)

41 57_20190109_BR-Data-11 (4)

42 57_20190109_BR-Data-11 (3)

meint, sie könnte das mal versuchen zu automatisieren. [Lea] meint, so müssten sie halt die Idee, auch die umgeschriebenen Meldungen erfassen zu können, verabschieden.⁴³

Kurz darauf wird Martin beim Gedanken an die Transparenz und Nachvollziehbarkeit etwas bange⁴⁴ und am Tag darauf ergeben sich bei der Automatisierung weitere Probleme mit falschen Treffern.⁴⁵ Obwohl erwogen wird, nur ein Sample an Polizeistationen zu untersuchen oder nur ein Sample an Medienplattformen, welche die Berichte weiterverbreiten, wird schliesslich keine Geschichte zu den Polizeimeldungen veröffentlicht. Die möglichen Perspektiven auf die Daten aus den qualitativen Recherchen liessen sich nie mit der Sinngebung aus den automatisierten Datenanalysen passend machen – auch weil die Untersuchung von Samples die Nachvollziehbarkeit immer wieder in Frage stellte.⁴⁶ Die genutzten Werkzeuge (*Gabriels Tool* beziehungsweise Google) waren extern und die Transparenz über deren Funktionsweise nicht gewährleistet. Für die beiden Teams ist die Legitimität einer datenjournalistischen Erkenntnis eng mit Transparenz und Nachvollziehbarkeit der Vorgehensweise verbunden. Transparenz und Nachvollziehbarkeit im Datenjournalismus wurden in der Forschung bisher vor allem als Glaubwürdigkeitszuschreibung von aussen diskutiert (siehe Abschnitt 6.1.3) – hier zeigt sich, dass es bereits für die Sinngebung in der Erkenntnisproduktion unabdingbar ist, dass die Beteiligten die automatisierten Schritte nachvollziehen können. Das Changieren zwischen manuell untersuchten Stichproben und automatisierten Analysen grosser Datensätze gehört in beiden Redaktionen zur datenjournalistischen Praktik der Erkenntnisproduktion. Ein Sample zu untersuchen dient dazu, die automatisierten Analyseschritte, welche sonst vom Computer übernommen werden, selbst zu absolvieren. Manuelles Nachvollziehen der Vorgehensweisen ermöglicht, die Validität der Automatisierung zu überprüfen und so die Datenauswertungen zu plausibilisieren – oder wie im Falle von *Gabriels* Methode zu diskreditieren.

43 57_20190109_BR-Data-11 (6)

44 57_20190109_BR-Data-11 (6)

45 58_20190110_BR-Data-12 (9)

46 Im Follow-Up-Interview rund ein halbes Jahr später meint Martin zum Polizeimeldungen-Projekt, dass sich die Thesen in den Daten nicht bestätigt und die Datenquellen für die Nachverfolgung der Meldungen im Internet sich immer wieder verändert hätten – schliesslich sei es »zu kompliziert« geworden. (Follow-Up-Interview mit Martin von BR Data vom 10. Juli 2019)

Einzelfälle untersuchen

Eine weitere Praktik zur Plausibilisierung ist die Untersuchung von Einzelfällen. Abschnitt 5.1.3 hat bereits erläutert, wie die Suche nach Ausreisern und Extremfällen wichtige Ziele der explorativen Datenanalyse sind. Beides sind Klassifizierungen, die per Definition nur auf wenige Fälle zutreffen. Die Praktik verweist darauf, dass beide Teams neben der Analyse von grossen Datenmengen ergänzend einzelne Fälle beleuchten. NZZ-Datenjournalistin Susanne erklärt die Vorgehensweise in ihrem Projekt, welches die Bevölkerungsstruktur der Schweizer Gemeinden untersucht:

Auf einem Scatterplot würde man sehen, welche Gemeinden wachsen würden. Es seien unglaublich viele Daten und es sei nun die Frage, was genau sie zeigen würden. Die Standard-Annahme des Bundesamts für Raumentwicklung, wo von sie die Daten hätte, sei halt, dass wenn eine Gemeinde wachse, sie weniger stark altere. Es gebe diese Korrelation schon, aber sie sei nicht so stark. [Emma] habe nun die gute Idee gehabt, einen Scatterplot aufzustellen, der auf der x-Achse aufzeige, wie viel eine Gemeinde gewachsen sei, auf der y-Achse dann wie viel älter eine Gemeinde geworden sei. Diese Grafik sei relativ komplex im Verständnis, aber wenn man sie mehrmals bringe und gut einbette, würde es funktionieren. Ich frage [Susanne], welche Aspekte der Geschichte sie denn nun betonen werde und wieso. [Susanne] meint, dass man Sachen wie die »jüngste« und die »älteste« Gemeinde schon kenne, es gebe aber noch viele Aspekte und spezielle Fälle, die noch nicht gezeigt wurden. Die Geschichte werde zudem ergänzt mit verschiedenen Reportagen, das sei gut, so könne man die »Menschen hinter den Daten« zeigen.⁴⁷

Die Datenvisualisierung dient in diesem Projekt dazu, in der Phase der explorativen Analyse *spannende* Einzelfälle zu identifizieren. Im Feldnotiz-Ausschnitt oben wie in demjenigen in Fussnote 24, Kapitel 5, (aus demselben Projekt) verläuft die Sinngebung über *spannende* Einzelfälle. Das heisst, für die Beteiligten geht es darum, dass die sinnhafte Erschliessung der Relationierung der Variablen – in diesem Falle Durchschnittsalter und Bevölkerungswachstum – über deren Untersuchung am Einzelfall passiert.

Auch in der Fortsetzung von Adrians Datenanalyse zu den Medikamenten (siehe Fussnoten 20 und 21 des Kapitel 5) erfolgt die Erkenntnisproduktion über die »Extremfälle«. Er hat die in die Recherche bisher nicht involvierte Lea

47 25_20180328_NZZStorytelling-13 (7)

für eine Einschätzung angefragt. Diese schlägt sofort vor, über die beiden *Extremfälle* die Beziehungen zwischen den Variablen Land und Anteil Antibiotika qualitativ zu erschliessen, mittels eines Interviews mit einem Experten:

[Lea] und [Adrian] sitzen gemeinsam vor dem Bildschirm, vor sich eine Excel-Datei, in der die sehr simple Tabelle und ein gestapeltes Säulendiagramm angezeigt sind. Die Tabelle zeigt den Anteil der Antibiotika, die auf der »Watchlist« sind (also die Reserve), und zwei weitere Kategorien. [Lea] stellt fest, es sei ja in den Ländern sehr unterschiedlich. Italien sei der Extremfall auf der einen Seite. Schweden und Island dagegen auf der anderen Seite seien »super«. [Adrian] solle doch die »Extremfälle rauspicken« und den Experten (ein Uni-Forscher zu Allgemeinmedizin, der sich mit der Verschreibung von Antibiotika befasst hat), den er noch befragen wird, darauf ansprechen, was das heisse und was die Auswirkung sei.⁴⁸

Diverse weitere Situationen würden die Praktik illustrieren können, quantitative Auswertungen mittels einer Vertiefung von Einzelfällen zu plausibilisieren⁴⁹ – Datenjournalist Roman bezeichnet diese Vorgehensweise denn auch als »klassisches« datenjournalistisches Konzept.⁵⁰ In den Sozialwissenschaften würde man von einem Mixed-Methods-Design sprechen, wobei diese Bezeichnung im datenjournalistischen Feld wenig gebräuchlich ist.

Das Ziel der Teams ist es, durch die Untersuchung eines Einzelfalls die durch Visualisierung aktivierte Relationierung nachzuvollziehen und zu verstehen. Geometrische Verkörperungen bilden als »visuelles Vokabular« die Syntax (siehe Abschnitt 5.3), sind aber nicht in der Lage, die Qualität oder Modalität der Relationierung zu spezifizieren (siehe Abschnitt 4.2.2). Um mittels einer Datenvisualisierung Erkenntnisse zu produzieren, untersuchen die Teammitglieder deshalb Einzelfälle mit qualitativen Methoden – beispielsweise wie in der erwähnten Situation mit der Befragung eines Experten –, um die Art der Relationierung nachzuvollziehen.

48 53_20181218_BR-Data-07 (5)

49 siehe bei BR Data weitere Praktiken und Aussagen: 47_20181210_BR-Data-01 (12); 47_20181210_BR-Data-01 (8)

50 47_20181210_BR-Data-01 (12)

Vergleichen

Um Datenanalysen zu plausibilisieren, nutzen die Teammitglieder von BR Data und NZZ Storytelling sehr oft die Praktik des Vergleichs. Für NZZ-Storytelling-Teammitglied Christoph gehört der Vergleich ganz essenziell zur Arbeit mit Daten im Journalismus und er definiert die Praktik mittels des Inhalts: Es [Datenjournalismus] ist der Versuch, Daten zu vergleichen, um eine Aussage zu machen, Trends und Strukturen zu erkennen.⁵¹

Bei BR Data wird fast in jedem Projekt während der Datenanalyse verglichen, um aus den Daten Sinn zu generieren. Beliebt sind insbesondere Vergleiche unter geografischen Entitäten wie Bundesländern oder Ländern⁵² – bereits die Ausführungen zur geografischen Verteilung in Abschnitt 5.1.3 haben gezeigt, dass die Überlappung von Datenkategorisierungen mit politischen Entitäten das Drehen von Geschichten vereinfacht, weil die Sinngebung anhand polit-struktureller Gegebenheiten erfolgen kann. Die schnörkellosen Balken- oder auch Liniendiagramme der explorativen Phase laden das Auge zu einem sofortigen visuellen Vergleich ein – die Feldnotiz (Fussnote 17, Kapitel 5) zu den Polizeimeldungen hat bereits eine solche Praktik beschrieben. Für Barlow (2015) ist es eine der Hauptfunktionen von Datenvisualisierungen, schnell Vergleiche zu ermöglichen – Brooker et al. (2018) haben dies bereits empirisch für Datenvisualisierungen in der Astrophysik aufzeigt, Burri (2001) für Visualisierungen in der Medizin. Lowrey und Hou (2018:7) stellen den (visuellen, quantitativen) Vergleich gar ins Zentrum ihrer inhaltsorientierten Definition von Datenjournalismus: »informational, graphical accounts of current public affairs for which data sets offering quantitative comparison are central to the information provided.«

Ob mittels Visualisierungen oder nicht: Vergleiche können einerseits ad hoc und simpel passieren – wie in der Besprechung zu einem Projekt der Nutzungsbedingungen und Datenschutzerklärungen, als Lea energisch darauf hinweist, dass Microsofts Nutzungsbedingungen 49'000 Wörter umfassten, ihre BA-Arbeit aber nur 10'000.⁵³ Der Vergleich kann aber auch ein sehr aufwendiger Schritt sein. So schlägt NZZ-Datenjournalist Yannick in einer Gruppendiskussion zum Fussballtransfer-Projekt vor, dass die Aussage der Datenvisualisierung sein könnte, dass die Schweizer Fussballliga eine

⁵¹ NZZ_Fragebogen_Mitarbeitende_08

⁵² 47_20181210_BR-Data-01 (13); 50_20181213_BR-Data-04 (3); 51_20181214_BR-Data-05 (2); 59_20190111_BR-Data-13 (5)

⁵³ 48_20181211_BR-Data-02 (3)

Ausbildungsliga sei – man müsse dazu die Daten aus der Schweizer Liga mit einer anderen *vergleichen* und dafür deren Daten scrapen.⁵⁴ An diesem Beispiel zeigt sich, wie sich die Teammitglieder die Operationsweise von Vergleichen nutzbar machen: die Kombination von Gleichheitsunterstellung und Differenzbeobachtung (Heintz 2010:164). Das heisst, durch den *Vergleich* der Ligen soll die *Differenz* bezüglich der Kategorie Ausbildungsliga verifiziert werden. So verläuft die Sinngebung durch eine Vergleichspraktik.

Neben dem Vergleich über verschiedene Kategorien wie Länder hinweg ist die Relationierung verschiedenartiger Datensätze eine weitere Form der Vergleichspraktik. Diese bringt meistens die Notwendigkeit mit sich, Datensätze auf ihre Kategorienbildung und Variablensetzung hin zu untersuchen, sie allenfalls zu transformieren, um dann einen Vergleich anstellen zu können. Sie müssen kommensurabel gemacht werden: »Commensuration is characterized by the transformation of qualities into quantities that share a metric, a process that is fundamental to measurement« (Espeland und Sauder 2007:16). Lea und Martin diskutieren, ob sie die Herkunftsennnung in den Polizeimeldungen mit einer Variable (Anteil an ausländischen Straftäter*innen) aus der Kriminal-Statistik korrelieren wollen – oder, anders ausgedrückt, ob die beiden Statistiken überhaupt kammensurabel sind:

[Lea] meint, es sei ja auch spannend, der Standard [österreichische Zeitung] habe untersucht, ob sich die Anweisung des Innenministers ausgewirkt habe. Und das dann auch mit der Polizei-Kriminal-Statistik verglichen. [Lea] zeigt [Martin] den Online-Artikel und die Grafik mit Über-/Unterrepräsentation. [Lea] sagt, der Vergleich mit der Statistik sei »völlig ohne Hemmungen« passiert, vielleicht müsste BR Data auch nochmals darüber diskutieren. [Martin] meint, ein Vergleich sei halt schwierig, die Datenerhebungen seien sehr verschieden. Ein Ereignis könne mehrere Meldungen [von Straftaten] auslösen. Deshalb würde er es nicht machen. [Lea] meint, es habe sie verunsichert, dass der Standard das einfach so mache.⁵⁵

Es trifft hier Leas Perspektive, sich durch eine Verbindung mit einem anderen Datensatz und einem damit ermöglichten Vergleich eine zusätzliche Interpretationsoption zu verschaffen, auf diejenige Martins, der auf die Fallstri-

54 14_20180313_NZZStorytelling-02 (27)

55 55_20190107_BR-Data-09 (8)

cke verschiedenartiger Datenerhebungskriterien hinweist und argumentiert, dies verunmögliche einen sinnhaften Vergleich.

Der Vergleich durch das Übereinanderlagern von Variablen aus anderen Datensätzen stellt stets eine reizvolle Möglichkeit dar, Korrelationen zu produzieren und sie als sinnhafte Kausalitäten zu plausibilisieren. Martin erzählt an einem Teammeeting von einer Projektidee zum Thema Bildung, zu der Johannes bereits einige Recherchen vorgenommen hat:

BR Data selber habe ja schon mal das Thema Schulsubventionen angedacht und die These gehabt, die Verteilung dieser sei in Bayern sehr unklar. [Johannes] sagt, er habe ein Pilotprojekt aus München gefunden, das gute Erfahrungen mit gezielten Subventionen gebracht habe. [Martin] sagt, das Thema würde auch in anderen Städten diskutiert. [Johannes] meint, er werde weitere Städte recherchieren. [Lukas] sagt, man müsste eine Datenbank aufbauen, die über Einzelfälle hinausgehe, und dann das Verhältnis zum Sozialindex anschauen.⁵⁶

»For statisticians, there always have to be comparisons; numbers on their own are not enough« (Gelman und Unwin 2013:5). Vergleiche sind also seit langem ein bewährtes Mittel, um aus Zahlen Sinn zu generieren und so Auswertungen zu plausibilisieren. Durch die Quantifizierung und Datafizierung unserer Lebenswelt – auch im Datenjournalismus – sind Vergleichspraktiken ubiquitär geworden. Visualisierungen unterstützen sie durch spezifische Formen wie Balkendiagramme oder Small Multiples⁵⁷ (Healy und Moody 2014:110).

Wenn Daten nicht abbilden

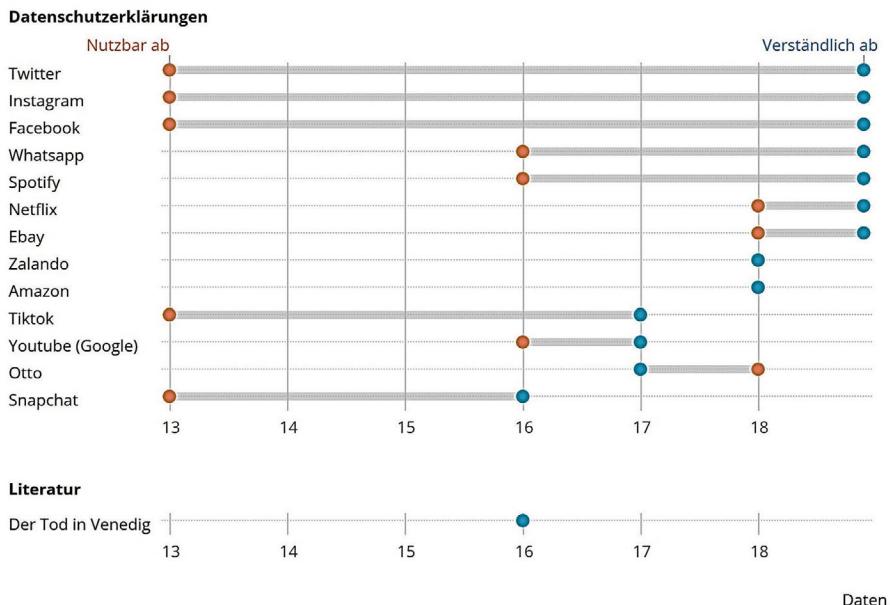
Wie bereits der erste Abschnitt dieses Kapitels gezeigt hat, drehen sich viele Praktiken in beiden Teams um die Beziehung zwischen ›Daten‹ und ›Realität‹ und deren Passendmachung. Die darauffolgenden Abschnitte zeigten ein Set an Praktiken der Passendmachung auf, welche Auswertungen plausibilisieren und so die Produktion von sinnhaften Erkenntnissen ermöglichen. Den Abschluss dieses Kapitels bilden folgend Ausführungen darüber, was in den beiden Teams passiert, wenn Daten nicht abbilden beziehungsweise die Passendmachung nicht erfolgreich ist.

56 56_20190108_BR-Data-10 (2)

57 Small Multiples referenziert auf die Visualisierung in Form vieler kleiner, simpler Diagramme, in welche die Visualisierung aufgeteilt wird (faceting). Small Multiples sollen einen schnellen Vergleich verschiedener Variablen ermöglichen (Tufte 1990:67).

Die Frage, was Daten abbilden können, ist für BR Data eine Abgrenzungspraktik gegenüber anderen Redaktionsteilen: Das Team sieht sich zuständig für Fragen (*Thesen*), welche man mittels Daten untersuchen (*prüfen*) kann. Die Mitglieder sehen ihre Expertise (auch) darin, dass sie die »Einschränkungen« der Abbildungen durch Zahlen reflektierten⁵⁸ und sie den »nicht daten-affinen« Kolleg*innen aus anderen Redaktionen erklären müssten, dass sich vieles halt nicht in den Daten zeige.⁵⁹

Abbildung 9: Visualisierung des Verständlichkeitsniveaus von Datenschutzerklärungen aus dem BR-Data-Artikel »Komplizierte Datenschutzerklärungen: Der Haken mit dem Häkchen«.



<https://web.br.de/interaktiv/datenschutzerklaerungen/> (25.02.2020)

58 47_20181210_BR-Data-01 (5)

59 49_20181212_BR-Data-03 (9)

In einem eigenen Projekt, in dem BR Data die Komplexität von Nutzungsbedingungen und Datenschutzerklärungen aufzeigen und insbesondere auf die Diskrepanz zwischen dem für das Verstehen eigentlich erforderlichen Bildungsniveau und dem Mindestalter von Apps und Medienplattformen hinweisen will⁶⁰, kann nur ein Teil der Annahmen mit der Datenauswertung passend gemacht werden: Man beschränkt sich auf Datenschutzerklärungen und lässt die Nutzungsbedingungen weg, weil es sonst »zu kompliziert« geworden wäre, wie Martin im Rückblick erklärt.⁶¹ Durch die Datenvisualisierungen ergab sich nur bei den Datenschutzerklärungen ein plausibles Abbild einer (auch) qualitativ bestätigten Annahme, eine Geschichte, die sich *sehen* lässt (Abbildung 9). Die Nutzungsbedingungen kommen schliesslich im Artikel gar nicht mehr vor, weil die Auswertungen unplausibel (»zu kompliziert«) blieben.

Auch in einem gemeinsamen Projekt zwischen NZZ Storytelling und der Lokalredaktion über Wahlergebnisse erfolgt eine Abgrenzung der Expertise darüber, was die Zahlen »sagen«.⁶² Die im Raum stehende These der Lokalredaktion, dass die Mobilisierung bestimmter politischer Lager und junger Frauen insgesamt in einem Zusammenhang mit gleichzeitig stattfindenden Abstimmungen sowie mit der Quartierentwicklung stehe, sorgt bei Annalena für Stirnrunzeln:

Die These des [Lokal-]Redaktors ist es, dass neuere Wohnangebote auch eher links-liberale Leute anzögen. [Annalena] mahnt zur Vorsicht. Der Redaktor entgegnet, es gebe aber schon wissenschaftliche Studien, die eindeutig sagten, dass Junge generell links-liberal, Alte eher konservativ und Studierte nicht SVP wählten – das seien signifikante Ergebnisse. [Annalena] meint, man sollte trotzdem »ultra-vorsichtig sein«. Der Redaktor meint, er fange es [die These] qualitativ auf, er werde die Reportage mit qualitativen Elementen aus einem eben-solchen Quartier begleiten.⁶³

Die These, dass die Mobilisierung mit dem Wohnangebot zusammenhängt, lässt sich mittels Datenauswertungen und -visualisierungen nicht prüfen und passend machen. Annalena besteht deshalb auf einer Abgrenzung von einer

⁶⁰ 48_20181211_BR-Data-02 (3)

⁶¹ Follow-Up-Interview mit Martin von BR Data vom 10. Juli 2019.

⁶² 20_20180321_NZZStorytelling-08 (2)

⁶³ 22_20180323_NZZStorytelling-10 (1)

quantitativen Auswertung über die Mobilisierung. Im Artikel bleiben schliesslich der datengestützte Teil mit der Datenvizualisierung (siehe Abbildung 19) und ein Reportage-Teil mit qualitativ gestützten Annahmen getrennt.⁶⁴

So lässt sich das Fazit ziehen, dass die Plausibilisierung von Datenvizualisierungen ein Prozess der Passendmachung ist, der nicht zwingend erfolgreich endet. Insbesondere gegenüber von ›Common-Sense‹-gestützten Annahmen, welche andere Redaktionsteile einbringen und überzeugt vertreten, grenzen sich die beiden Teams ab.

Die datenjournalistischen Arbeitspraktiken des Erkennens von Daten – Geschichten in Daten übersetzen (Datafizierung), die Herstellung sauberer Daten, das Drehen von Geschichten und die Plausibilisierung von Auswertungen – hat dieser Abschnitt beschrieben und mit theoretischen Überlegungen ergänzt. Durch diese Arbeitspraktiken verfügen die beiden Teams über ein in Datensätzen eingepresstes Abbild der Welt, welches verrechenbar und visualisierbar ist. Die visuelle Verkörperung der Daten durch geometrische Formen auf Ordnungssystemen produziert Anschlusspunkte für Geschichten wie beispielsweise Diskontinuitäten, Abnormalitäten, Neuheit und die Verbindung mit politischen Strukturen. Die hervorgerufenen Geschichten-Ideen der quantitativen Auswertungen plausibilisieren die Teams in einem iterativen Prozess mit qualitativen Zugängen zum Thema, indem sie Samples manuell nachprüfen, Einzelfälle beleuchten und vergleichen.

Diese Arbeitspraktiken sind angeheftet an materielle Objekte wie die Datenbank und Datenvizualisierungen. Um mithilfe dieser Objekte Daten zu erkennen und Sinn zu generieren, braucht es ein gemeinschaftliches Sehen. Diesem Aspekt widmet sich der nächste Abschnitt.

5.2 Erkenntnisproduktion als Aushandlung

Seeing data, das Erkennen von Daten, ist eine geteilte, kollaborative Praktik. Sie erfolgt mittels Aushandlungen in Diskussionen, anhand eines visuellen Vokabulars, auf formalisierter Ebene mittels Werkzeugen und durch das Imitieren von Ideen. Die folgenden Abschnitte zeigen auf, wie die Teams ihre Expertisen koordinieren, um aus den erläuterten Arbeitspraktiken sinnhafte datenjournalistische Erkenntnisse zu produzieren. Dies geschieht auf der Basis der in den datenjournalistischen Teams vorhandenen Expertisen, welche sich in

64 Zitat aus dem Artikel »No City for Old Men«, NZZ vom 27. März 2018, Seite 21.

einem Dreieck der Triaden aufspannen lassen. Abschnitt 3.2 hat die Expertisen als Schema eingeführt und anhand der Feld-Untersuchungen definiert, in jeweils den drei Aspekten Materialität, epistemologische Kultur und Professionskultur (Abbildung 2).

In den im Abschnitt 5.1 beschriebenen datenjournalistischen Arbeitspraktiken des Erkennens von Daten scheinen die verschiedenen Expertisen immer wieder durch. Vorerst lassen sie sich folgendermassen skizzieren:

- Die technische Expertise orientiert sich daran, welche Variablen überhaupt datafizierbar und kalkulierbar sind, indem sie in Datenbanken aufbewahrt und verschiedenseitig anschliessbar für Verrechnungen gemacht werden.
- Die journalistische Expertise orientiert sich daran, welche Relationen interessant als Geschichte (*story*) zu erzählen sind, indem sie mit thematischem Wissen und weiteren qualitativ gewonnenen Erkenntnissen verbunden und damit plausibilisiert werden können.
- Die grafische Expertise orientiert sich daran, welche Relationen durch verschiedene geometrische Formen und auf welchem Ordnungssystem sinnhaft verkörpert und damit sichtbar gemacht werden.

Das allgemein gehaltene erste Schema lässt sich damit auf den Visualisierungsmodus des Erkennens von Daten beziehen und mit den spezifischen Eigenschaften auffüllen, welche die Aushandlungen ausmachen, wenn Datenvisualisierungen als Arbeitsoberfläche dienen. Zuerst gilt es aber, diese Aushandlungen und Koordinationsmuster zwischen den Expertisen zu beschreiben und einzuordnen. Dafür hat der Abschnitt 4.1 die Heuristik der *trading zone* eingeführt.

Wichtig ist die Prämisse, dass sich die einzelnen Teammitglieder nicht eindeutig und ausschliesslich jeweils einer Expertise zuordnen. Im Gegenteil: Die meisten sind aufgrund ihrer Ausbildungen und/oder langjähriger Arbeitserfahrungen mit mehreren Expertisen in Berührung gekommen und haben sich diese angeeignet. Die hier vorgenommene analytische Trennung der Expertisen ist für sie im Arbeitsalltag kaum je relevant. Stattdessen sind sie eng verwoben nicht nur innerhalb des Teams, sondern oft auch in den persönlichen Rollen. Diejenigen, welche persönlich fähig sind, verschiedene Expertisen besonders gut zu koordinieren, übernehmen gar die wichtige Funktion als »traders, strategically coordinating parts of interpreted systems against parts of others«

(Galison 1997:816), siehe Abschnitt 5.2.2. Nun zeichnen aber zuerst einige empirische Situationen die Aushandlungen und Koordinationsmuster nach.

5.2.1 Koordinationsmuster in den Arbeitspraktiken

Der folgende Abschnitt beleuchtet Situationen, in denen die verschiedenen Expertisen und ihre materiellen, epistemologischen und professionellen Dimensionen für das Erkennen von Daten in den Teams koordiniert werden.

Wie Abschnitt 5.1.2 gezeigt hat, bildet die Herstellung *sauberer Daten*, das heißt granularer, geordneter und interoperabler Datensätze, die Grundlage für die Produktion von Erkenntnissen. Ein Ausschnitt aus einer Feldnotiz von BR Data lässt die dabei ablaufenden Koordinationsprozesse nachvollziehen. Im darin verhandelten Projekt wollen die Verantwortlichen Lea und Lukas Daten zur Bearbeitung von bestimmten Strafverfahren bei Staatsanwaltschaften untersuchen. Die journalistische Potenzial-Abschätzung für eine Geschichte rund um das Thema Befangenheit der Staatsanwaltschaft wird koordiniert mit den Datafizierungsprozessen: Welche Daten können erhoben werden und wie können sie geordnet werden, um sie danach verrechnen und das Phänomen mit Daten erkennbar machen zu können?

In einem ersten Schritt geht es darum, überhaupt an die Staatsanwaltschaften zu gelangen und bei ihnen die Daten für die Analyse abzufragen – für dieses Schreiben sind die beiden in telefonischem Kontakt mit einer BR-eigenen Juristin. Es ist unklar, in welcher Form die Staatsanwaltschaften die Daten bereits erheben. Bevor BR Data die Stellen anschreibt, diskutieren die beiden Teammitglieder über die Tabellen-Strukturierung und damit darüber, mit welchen Variablen und mit welcher Granularität sie die Daten erheben wollen. Sie sitzen dafür gemeinsam am Pult und blicken auf einen Tabellenentwurf auf dem Bildschirm vor ihnen:

[Lukas] fragt, ob die Tabelle als Vorlage alle nötigen Optionen der Differenzierung offen lasse für die Analyse. [Lea] fragt, ob die Zuordnung zur Einstellungs-Instanz wohl möglich sei. Sie will die Tabelle noch gemeinsam durchgehen, ihr sei die Systematik nicht immer ganz klar. [Lukas] sagt, dass sie sich ja auf die Staatsanwaltschaften konzentrierten, und ob diese als Einstellungs-Instanz immer klar zuordenbar seien? [Lea] und [Lukas] lesen einen der Artikel der Strafprozeßordnung zu Einstellungen im Internet nach. Sie diskutieren über die Deutung der Aussagen der Juristin. [Lukas] meint, sie sollten in der Tabelle eine zusätzliche Spalte einfügen, um differenzieren zu können. [Lea] meint,

sie werde die Tabelle nochmals durchgehen. [...] [Lukas] fragt, ob man die Differenzierung zwischen Einstellungen der Staatsanwaltschaften und der Gerichte aus den Zahlen lesen könne oder ob man sie abfragen müsste im Schreiben? Falls man sie abfrage, bestehe die Gefahr, dass es Widersprüche gebe zu anderen Zahlen. [Lea] meint, sie werde die Tabelle durchgehen und versuchen, diese zwischen den Instanzen auszudifferenzieren, um dies herauszufinden. [...] [Lukas] fragt, ob man eigentlich nicht gleich die vorhandene Tabelle des Ministeriums nutzen könne, um Auswertungen zu machen – [Lea] erklärt ihm dann aber, dass in dieser noch viele weitere Straftatbestände zu finden seien, weshalb man die genauen Daten erfragen müsse. Man wolle sich auf einen [Straf-]Artikel beschränken. [Lukas] fragt, wie viele Staatsanwaltschaften es eigentlich seien, die sie anfragten. [Lea] antwortet: 117. [Lukas] ist dann entschlossen, dass sie eine Vorlage mitschicken müssten, sonst würden sie wahnsinnig, wenn es darum gehe, die Daten zu sammeln und zu analysieren, das sei zu viel Arbeit. [Lea] meint, dass sie davon ausgehe, dass andere Bundesländer eine ähnliche Struktur zur Datensammlung hätten wie Bayern. [Lukas] betont nochmals, dass sie gut überlegen müssten, wie die einzelnen Zeilen zu interpretieren und den Instanzen zugeordnet seien. [Lea] meint, sie fände es auch bemerkenswert, wie unterschiedlich von den Gerichten teilweise bestraft werde und Massnahmen gesprochen würden. [Lukas] meint, das sollten sie sich auch anschauen, wenn sie das Gefühl hätten, »da steckt was drin«. [Lea] sagt, sie habe die These, dass die Staatsanwaltschaften schon viel Ermessensspielraum hätten. [Lukas] meint, sie sollten die Tabelle nochmals mit dem Bundesjustizministerium besprechen und nachfragen, ob es ähnliche Schemata in allen Bundesländern so gebe – und falls ja, ob sie die eigenen Punkte separat abfragen müssten. [Lea] schaut sich die Tabelle vom Bundesjustizministerium an. [Lukas] gleicht die Zahlen für Bayern darin mit den Zahlen aus dem bayerischen Ministerium ab – es gibt einen grossen Unterschied. [Lukas] meint, sie sollten die Juristin dann beim Telefonat auch noch um eine Einschätzung bitten, wie auf Bundesebene die Daten erhoben würden und wie diese vergleichbar seien.⁶⁵

Dieses Gespräch zeigt, was mit *sauberen Daten* gemeint ist: durchgehend konsistent strukturierte Daten, deren Eigenschaften eindeutig einzelnen Variablen zuordenbar sind. Diese Variablen sollen dann wiederum konsistent mit Kategorien aus dem Thema übereinstimmen – hier beispielsweise Art

65 58_20100110_BR-Data-12 (8)

der Straftat (operationalisiert mit dem Strafartikel) oder Instanz, welche das Verfahren einstellt. Diese Abstimmung zwischen Kategorien aus der thematischen Recherche, einer daraus generierten vagen Hypothese (in Bayern ist der Anteil der Verfahrenseinstellungen besonders hoch) und der Strukturierung in Tabellenform erfolgt zu zweit am Pult, auf Basis des Tabellenentwurfs auf dem Bildschirm sowie weiterer Materialien auf dem Tisch oder auf dem Bildschirm. Die so vorhandenen Informationen stimmen Lea und Lukas mittels Diskussion ab. Lukas verweist immer wieder auf die programmiertechnische Konvention einer hohen Granularität der Datenbank. Aufgrund der noch sehr offenen Fragestellung tippt die Konversation viele verschiedene mögliche Thematiken im Bereich der Staatsanwaltschaften an, wobei die beiden diese immer auf ihre mögliche *saubere* Datafizierung hin abklopfen – die tabellarische Ordnung muss sinnhaft sein, die Interoperabilität ist durch die Excel-Tabelle gegeben – und dann im Raum stehen lassen. Die Praktik der Herstellung *sauberer Daten* ist stark geprägt durch die programmiertechnische Expertise, ihre Vorstellung von granularen, vollständigen, systematischen und interoperablen Datensätzen – und sie wird dergestalt mit der journalistischen Expertise koordiniert, dass die Ordnung in der Datenbank journalistische Sinngebung ermöglicht. Und zwar zu Beginn, wenn die explorative Auswertung eine grosse Varianz an Fragestellungen ermöglicht.

Die Existenz verschiedener Ansatzpunkte und offener thematischer Fragen scheint insofern wünschenswert, als damit verschiedene *Geschichten-Drehs* offen sind – die Potenzialität wirkt als Versprechen: Im Hintergrund schwingt die Vorstellung mit, dass sich bei einer guten, eben *sauberen* Datengrundlage mittels explorativer Auswertung Muster finden lassen – ein Paradigma aus dem Big-Data-Zeitalter (Kitchin 2014a; Mützel et al. 2018). Mögliche Schwierigkeiten in der Interpretation von Daten antizipieren die beiden und lassen diese stehen mit dem Verweis auf zukünftige Abklärungen mit weiterem Material, der Juristin oder einem Ministerium.

Bei BR Data liess sich das Erkennen von Daten häufig dann beobachten, wenn zwei Teammitglieder gemeinsam vor einem Bildschirm am selben Pult sass. Gemeinsam betrachtet man Datenanalysen und -visualisierungen, um aus den Daten Erkenntnisse zu schliessen, Annahmen zu bilden und zu hinterfragen. Die Tandem-Konstellation strebt BR Data explizit für alle grösseren Projekte an.⁶⁶ Die Legitimation von Erkenntnisansprüchen bedingt

66 49_20181212_BR-Data-03 (1); 50_20181213_BR-Data-04 (3); 53_20181218_BR-Da-
ta-07 (1)

für das Team einen interpersonalen Abgleich in Form einer Diskussion der Instrumente und der produzierten Erkenntnisse. Damit verbunden ist stets die Koordination der verschiedenen Expertisen: Bei BR Data fliesst die programmiertechnische Expertise durch viele Personen mit entsprechendem Hintergrund ein und die Idee, Datenbanken als »wahre« Perspektive auf »die Realität« als neue Quellen zu erschliessen, bietet einen geteilten Koordinationspunkt mit der (investigativ-)journalistischen Kultur. Evident wird dies auch in den geführten Kurz-Interviews: Auf die Frage nach dem Nutzen von Datenjournalismus antworten viele Teammitglieder von BR Data mit einer Kombination von Verweisen auf dadurch neu ermöglichte Geschichten wie auch die unbedingt nötige technische Kompetenz, um Praktiken von mächtigen Digital-Konzernen und datengestützte Prozesse in der Gesellschaft insgesamt als *Watchdog* hinterfragen zu können – sie schliessen damit an die entsprechenden Narrative des Feldes an.

Nicht immer werden dabei alle Details verbalisiert wie exemplarisch in der Situation der Datensammlung über die Staatsanwaltschaften. In der im Abschnitt zum Drehen von Geschichten geschilderten Situation über die Analyse von Herkunftsennungen in Polizeimeldungen (Fussnote 17, Kapitel 5), in der Brigitte und Lea Säulenografiken am Bildschirm betrachten, werden wenige Schritte der Erkenntnisproduktion expliziert. Durch die gewählte geometrische Verkörperung (Säulen) erscheint beiden implizit klar, dass die Analyse auf einen Vergleich hinausläuft. Treten Unterschiede im Anteil der Herkunftsennungen zutage, sichtbar durch die unterschiedliche Höhe der Säulen, so ist dies »spannend« – eine weitere Explikation der Annahme, dass auffällige Unterschiede auf ein journalistisch interessantes Muster deuten, bleibt aus. Die grafische Expertise ist bei beiden Teammitgliedern verinnerlicht und bleibt implizit. Zur Exploration werden einfache grafische Formen, deren Variablen-Relationierung wenig komplex ist, genutzt – denn die grafische Expertise bei BR Data überlappt mit der technischen Expertise. Oder anders gesagt: Erstens stützt man sich auf Auswertungen mittels Programmiersprachen, welche ein bestimmtes, eingeschränktes Set an visuellem Vokabular für explorative Auswertungen zur Verfügung stellen. Und zweitens gibt es keine ausgebildeten Designer*innen im Team und es sind meist die Programmierer*innen, welche die Datenvisualisierungen produzieren.

So sind die Expertisen in den weiteren Situationen und Besprechungen des Polizeimeldungen-Projekts klar zugewiesen: Brigitte programmiert und rechnet und visualisiert, Lea verfügt über die Hauptrolle in der Sinngebung der Daten. So entscheidet sie mit journalistischer Expertise, dass die visuellen

Muster noch keine sinnhaften Auffälligkeiten ergeben und dass deshalb noch mehr Daten gesammelt werden sollen, um weitere Muster zu produzieren – die Situation endet wie folgt:

[Lea] fragt [Brigitte], ob es ein Problem sei, nochmals nachzuscrepen bis Ende 2018 – [Brigitte] meint nein, der Scraper würde wohl noch »irgendwo rumliegen«. Der Scraper ist gespeichert in Github, der Code und beschreibender Text dazu sind in einem html-File sauber abgelegt und abrufbar.⁶⁷

In den Situationen, in denen Adrian und Martin gemeinsam am Bildschirm die Excel-Liniendiagramme über Medikamentenverschreibungen interpretieren (siehe Feldnotizen mit Fussnoten 20 und 21 des Kapitels 5), verfügt Adrian über das grössere thematische Wissen, weil er in den Tagen zuvor die Daten online gesammelt und sich ins Thema eingelesen hat. Die Daten sind in Balken- und Liniendiagrammen abgebildet, womit Relationen zwischen Variablen oder verschiedenen Zeitpunkten sichtbar vorliegen. Dem journalistisch routinierteren Martin obliegt es dann, nach thematischer Erläuterung durch Adrian die vorliegenden Datenvisualisierungen zu koordinieren und mit der journalistischen Expertise zu bewerten – das heisst, ob eine der sichtbar gemachten Relationen als Geschichte erzählenswert ist. Martin schliesst aus den Kurvenverläufen, dass es »nichts gebe, dem man nachgehen sollte«, und dass man in den Daten nichts sehe.⁶⁸ Er schlägt während der Diskussion vor, sich den Kurvenverlauf visuell vorzustellen ohne einige ausschlagende Werte – aber auch diese imaginierte Kurve verbleibt ohne Auffälligkeiten, die Martin als relevant erachtet (siehe Ausschnitt Fussnote 21, Kapitel 5).

Die grafische Expertise koordiniert in diesem Sinne mittels ihrer Materialität, in der Regel geometrische Formen und ein Koordinatensystem, die Sinngebung und journalistische Bewertung – wie dies in den verschiedenen Aspekten des *Geschichten-Drehens* (Abschnitt 5.1.3) deutlich wird. Das vorhandene visuelle Vokabular bestimmt, mit welchen Relationierungen Daten überhaupt verkörpert und in diesen Formen von den Datenjournalist*innen *erkannt* werden können. Abschnitt 5.3 erläutert das visuelle Vokabular als koordinierende Organisationsform.

67 56_20190108_BR-Data-10 (8)

68 51_20181214_BR-Data-05 (2)

Im Team von NZZ Storytelling, einerseits auf die Kollaboration mit anderen Ressorts ausgerichtet und andererseits eigene grössere Projekte verfolgend, fungieren andere Koordinationspunkte für die verschiedenen Expertisen. Die bei vielen Teammitgliedern ausgeprägte grafische Expertise führt dazu, dass Datenvisualisierungen oft hinterfragt und diskutiert werden – begründet darin, dass die Konvention der Prüfung der akkuraten Abbildung von Informationen nicht nur im Journalismus verankert (Parasie 2014), sondern auch eine in der grafischen Professionskultur verbreitete Vorstellung ist (Kosminsky et al. 2019:50).

Die Teammitglieder holen dabei Perspektiven nicht nur von denjenigen ein, die am Projekt mitarbeiten, sondern auch von mehreren anderen Teammitgliedern. Diese Feedbacks passieren ad hoc, manchmal sogar ungefragt. Zum Beispiel über die Fussballtransfers-Datenvisualisierung: Kaspar verfügt zum Diskussionszeitpunkt über einen Entwurf in Form einer Sankey-Grafik (auch die schliesslich veröffentlichte Datenvisualisierung wird eine Sankey-Grafik sein, siehe Abbildung 17). Aufgrund der Idee aus der Sportredaktion ist es zwar klar, dass es um die Ein- und Verkäufe der Schweizer Fussballklubs gehen soll, über präzise journalistische Hypothesen für die Datenauswertung verfügt Kaspar aber nicht. Er praktiziert einen datengetriebenen Ansatz und interessiert sich primär für Muster, um überhaupt eine interessante Relatiorierung konstruieren zu können.⁶⁹ Nun entspinnt sich an seinem Pult, vor dem Bildschirm, eine Diskussion zwischen diversen Teammitgliedern, die über eine halbe Stunde dauert – hier ein Ausschnitt aus dem abschliessenden Teil der Diskussion:

[Yannick] versucht allen zu erklären, wieso die jetzige Darstellung problematisch ist: Nicht nur wegen der Durchschnittswerte, sondern auch, weil der Chart Punkte verbinde, die aber in der Realität nicht die gleichen seien (nicht die gleichen Spieler werden in der einen Liga gekauft und wieder in dieselbe zurückverkauft). Der Chart suggeriere aber genau das, dabei seien doch nicht die einzelnen Spieler-Flüsse, sondern die Geldflüsse zwischen Ligen visualisiert. [Annalena] betont, sie fände das nicht problematisch. [Emma] gibt sich zurückhaltend. [Yannick] sagt, Spielerflüsse müsste man mit einer Grafik visualisieren, die man auch für Wählerflüsse braucht. [Annalena] belehrt ihn, dass auch diese nicht auf tatsächlichen Wanderungen, sondern Regressionen beruhen. [Yannick] ist überrascht, das wusste er nicht. Bisher ging er davon aus, es handelte

69 14_20180313_NZZStorytelling-02 (20)

sich bei solchen Grafiken immer um tatsächliche Wanderungen. [...] [Yannick] bringt nochmals seine Idee vom frühen Nachmittag ein (ein Bar-Chart mit einer Null-Linie). Die Diskutanten fragen nach meiner Meinung, ich halte mich aber zurück. Mehrmals steht die Frage im Raum, ob sich denn nun die Geschichte um die einzelnen Spieler, das ländervielfältige Scouting oder einen Geldfluss-Vergleich drehen soll. Was für eine Aussage kann getroffen werden? Die Diskussion landet am Schluss aber bei ziemlich genau jener Grafik, die bereits David am Morgen vorgeschlagen hat. [Yannick] betont, dass die eine Aussage, die man treffen könne, wohl die sei, dass die Schweizer Liga eine Ausbildungsliga sei. Von [Annalena] und [Kaspar] kommt das Argument, dass man dies nur sagen könne, wenn man einen Vergleich anstelle. Am besten mit einem der Top-5-Ligen. Alle finden das eine spannende Idee. [Yannick] schlägt England vor. Er sagt, dass sein Scraper einfach umgebaut werden kann. Man müsse halt einfach bei David nochmals »einen halben Tag rausreden«. Die Diskussion kommt zu einem Ende. [Kaspar] wird morgen mit David sprechen und sagt, die anderen könnten sicher auch dazustossen. [Kaspar] muss gehen. [Yannick] sagt ihm noch halb-fragend, ob er denn nun schon mal den Scraper umbauen solle – [Kaspar] meint ja.⁷⁰

Yannick versucht, die in Flussform (Sankey) relationierten Daten mit journalistischer Expertise über Fussball zu interpretieren (er ist kein Sportredaktor, sondern aktiviert sein Allgemeinwissen darüber). Die von ihm vorgeschlagenen *Geschichten* sind aber nicht mit der vorliegenden Darstellung sinnhaft, beispielsweise die Verbindung von Punkten durch die Verkörperung als Fluss. Sein Vorschlag für die Nutzung eines Balkendiagramms (»bar chart«) ist der Versuch, eine andere Art der Relationierung der Daten (Mengen- statt Flussdarstellung) zu visualisieren. Es gelingt in der Diskussion aber lange Zeit nicht, eine journalistisch relevante Erkenntnis in Form einer Hypothese mit einer akkurate, sinnhaften geometrischen Form und ihrem impliziten Relationierungstyp übereinstimmend zu machen. Solche Aushandlungsprozesse, in denen sowohl aus journalistischer wie auch aus grafischer Perspektive eine grosse Palette an Optionen und somit grosse Kontingenz besteht, erfordern einen grösseren Abstimmungsaufwand für das Erkennen von Daten.

Schliesslich erfolgt eine Koordination anhand des Vorschlags, eine Aussage aus einem Vergleich mit einer anderen Fussballliga heraus zu treffen. Yannick

70 14_20180313_NZZStorytelling-02 (27)

proklamiert aufgrund seines thematischen Wissens, dass dies die englische Liga sein sollte. Dieser Vorschlag muss abgestimmt werden mit den technischen Möglichkeiten: Die Frage, ob Informationen über die englische Liga datafiziert verfügbar und genauso kalkulierbar sind wie die Informationen über die Schweizer Liga, beantwortet Yannick positiv, weil dasselbe methodische Vorgehen (Scraping) anwendbar ist. Die Koordination haben die Akteure dadurch erreicht, dass sie auf die programmiertechnische Perspektive zurückgreifen und den Datensatz als nicht vollständig einstufen – dies impliziert, dass sie nach der Vervollständigung eine sinnhafte Erkenntnis durch einen Vergleich produzieren können.

Solche Einigungen sind stets temporär und können im abduktiven Prozess lange folgenlos revidiert werden. So verfolgt der Verantwortliche Kaspar am nächsten Vormittag eine Visualisierungsidee, die Emma ihm auf ein Blatt Papier gezeichnet hat.⁷¹ (Abschnitt 5.4.3 vertieft den Aspekt von Handzeichnungen für die Konstruktion von datenjournalistischen Erkenntnissen.) Auch für das Projekt zur Alterung der Bevölkerung in den Gemeinden hat Emma eine Visualisierungsidee eingebracht, welche die Daten auf ein Koordinatensystem spiegelt und welche die Verantwortliche Susanne deshalb überzeugt, weil sie anschlussfähig an die journalistische Konvention ist, eine Erkenntnis sowohl als generellen Trend in Form eines Musters sichtbar zu machen als auch dieses mit Einzelfällen zu plausibilisieren (siehe Ausschnitt der Feldnotiz mit Fussnote 47, Kapitel 5).

Die Praktiken der Plausibilisierung (Abschnitt 5.1.4) koordinieren die mit technischer und grafischer Expertise hergestellten Datenvisualisierungen mit journalistischen Konventionen wie der Zuspritzung auf eine ›Wahrheit‹, der Prüfung der Nachvollziehbarkeit durch Ziehen eines Samples, der Sinngebung und Narration durch die Beleuchtung von Einzelfällen und dem Herstellen von Vergleichen. In beiden Teams prägt die eingebrachte journalistische Expertise diese Praktiken, denn für die Plausibilisierung der Erkenntnisse verlässt man sich primär auf qualitativ gewonnenes Wissen aus der Recherche, Allgemeinwissen oder Themenwissen aus Fachbereichen. Dabei adressiert die journalistische Expertise die Differenz zwischen der ›Zahlenwelt‹ und der ›echten Welt‹, die sie mittels Plausibilisierungen zu überbrücken versucht. Diese Passendmachung soll mittels grafischer Expertise und ihres Ideals der

⁷¹ 15_20180314_NZZStorytelling-03 (3) – leider lässt sich aus den Feldnotizen nicht nachvollziehen, in welcher Rahmung Emmas Idee zu Kaspar gelangte und welche Aushandlungen dabei stattfanden.

akkuraten Abbildung von realen Phänomenen koordiniert werden. Scheitert diese Passendmachung, so bleibt die Erkenntnisproduktion stehen – das Projekt zu den Polizeiberichten hat dies gezeigt.

5.2.2 *Traders I*

Die beschriebene Passendmachung kann mittels einer verbalen Einigung wie in der Diskussion über die Fussballtransfers-Visualisierung passieren. Aber auch die Vorschläge einzelner Teammitglieder, welche die verschiedenen Expertisen überbrücken, ermöglichen Koordination. Bei NZZ Storytelling wie auch etwas weniger ausgeprägt bei BR Data lässt sich feststellen, welche Personen mit ihren Hintergründen besonders oft als Vermittler*innen in Koordinationssituationen fungieren. Es sind Personen, welche sich in der Selbsteinschätzung im Expertisen-Dreieck (siehe Tabelle) nicht in einer der drei Ecken verorten, sondern deren Expertisen mehrdimensional sind. Es handelt sich um Teammitglieder wie Martin (bei BR Data) oder Emma (NZZ Storytelling), die in den Situationen eine koordinierende Stellung einnehmen, indem sie unterschiedliche Expertisen an bestimmten Punkten beidseitig anschlussfähig machen. In besonderem Masse gilt das für David, von welchem sehr oft eine Bewertung eingeholt wird, um eine Datavisualisierung zu evaluieren. Sowohl gemäss eigener Verortung wie auch aufgrund seiner Teamleiterfunktion verbinden sich bei ihm Expertisen aus dem journalistischen, grafischen und dem technischen Bereich. Die Wertschätzung seiner Bewertungen hat auch mit seiner institutionellen Position als Teamleiter zu tun, geht aber über die hierarchische Komponente hinaus: Seine Vorschläge werden meist als sinnhaft aufgefasst und ihm attestiert ein langjähriges Teammitglied, die meisten guten Ideen zu haben.⁷²

Galison (1997:816) hat Personen, welche in den Zwischenräumen verschiedener Subkulturen durch die Entwicklung von vereinigenden Sprachen (»interlanguages«) koordinieren, »traders« genannt. Diese Rolle wird je nach Situation neu verteilt, fällt aber in der Empirie gehäuft denjenigen Teammitgliedern zu, welche mindestens zwei Expertisen sinnhaft koordinieren – sei es explizit oder durch ein analytisch schwer fassbares, implizites Fingerspitzengefühl oder »tacit knowledge«. Dabei zeigt ein Blick in die jeweiligen persönlichen Hintergründe, dass es keine Rolle spielt, ob die Multiperspektivität auf verschiedenartige Berufs- oder Ausbildungserfahrung oder eine

72 20_20180321_NZZStorytelling-08 (11)

hohe Routine in datenjournalistischen Praktiken zurückzuführen ist. Dabei geht es nicht um reine Übersetzungsleistungen – Galison weist den Begriff »translation« explizit zurück (Galison 1997:47) –, sondern um die Performance lokaler Koordinationssprachen, verkörpert durch geteilte Praktiken und Interpretationen. Während Galison die Rolle nicht weiter beschreibt, hat Stark (2009:186f) die Vermittlung als einen konstruktivistischen Prozess von »actively making a new type of association« beschrieben, welche das Erkennen von Daten ermöglicht. Das deckt sich mit den in den beiden Teams beobachteten Situationen. Die koordinierenden Relationierungen entstehen insbesondere im Zusammenhang mit der Produktion oder Interpretation von Datenvizualisierungen. Stark weist darauf hin, dass der Vermittlungsprozess immer eines interaktionalen, geteilten Raumes der Sinngebung bedingt (Stark 2009:187). Abschnitt 5.5 wird zeigen, dass diese koordinierenden Relationierungen auch über Referenzierungen und Imitation anderer datenjournalistischer Geschichten verlaufen – wobei es wiederum die *traders* sind, welche diese in die Teams einbringen.

In beiden Teams befinden sich am Rande dieses Raumes diejenigen, welche Expertise nur aus einem Bereich mitbringen. Diese Mitglieder sind unterdurchschnittlich oft Teil von Aushandlungssituationen beziehungsweise befinden sich darin in einer peripheren Rolle. Bei BR Data trifft das auf die meisten Programmierer*innen zu: Sie erfüllen oft klar abgegrenzte Aufgaben in ihrem spezifischen Zuständigkeitsbereich. Dies steht in einem sich gegenseitig verstärkenden Zusammenhang damit, dass diese Teammitglieder selbst ihre journalistische oder grafische Expertise als schwach einschätzen. Bei der NZZ ist es vor allem das »Subteam⁷³ der Infografik – ihre periphere Rolle in datenjournalistischen Praktiken gründet auf einer mehrfachen Bruchlinie: Neben der Selbsteinschätzung von fast ausschliesslich grafischer Expertise sowie der traditionell verankerten Zuschreibung einer Service-Rolle in der Redaktion unterscheiden sie sich auch durch eine eigene Organisationsstruktur, welche nicht derselben Zeittaktung folgt wie die Projekte der restlichen Teammitglieder. Die Infografiker*innen folgen einem printorientierten Tages-Nachrichten-Zyklus, während die meisten anderen mehrtägige oder -wöchige Projekte verfolgen. NZZ Storytelling unternimmt zwar vielfältige und aufwendige

⁷³ Bei diesem Begriff handelt es sich um eine Kategorisierung aus dem empirischen Material – besonders oft manifestiert in der Abgrenzung des Infografik-Teams, beispielsweise 15_20180314_NZZStorytelling-03 (10–14).

Versuche, den »Brückenschlag«⁷⁴ zum Infografik-Subteam so oft als möglich zu vollziehen und die Organisationsprozesse und Gestaltungsstandards aller Datenvisualisierungen zu vereinheitlichen.⁷⁵ Trotzdem brechen die professionalkulturellen wie organisationalen Bruchlinien in Situationen der Erkenntnisproduktion wie auch in weiteren Aspekten immer wieder auf, wie sich in der weiteren Arbeit zeigen wird.

Neben der konzeptionellen Fassung als koordinierende Vermittler*innen oder Randfiguren liessen sich die verschiedenen Rollen auch typisieren, wie dies Baack (2018) vorschlägt. Der Autor fasst die datenjournalistische Community in Berlin als Überlappung von Journalismus und »civic tech« und identifiziert vier Typen darin, welche die beiden Konventionen »gatekeeping« (traditionelle journalistische Filterung von Information vor der Präsentation) und »facilitating« (civic tech, andere ermächtigen, selbst die Daten zu nutzen) auf einem geteilten Kontinuum ausbalancieren. Gemünzt auf die vorliegende Empirie liesse sich mit der Orientierung an Ästhetik durch die grafische Expertise eine weitere Dimension hinzufügen und ermöglichte eine Typen-Verortung. Dafür ist die Anzahl der in den beiden Teams vorliegenden Personen allerdings zu klein und dies muss deshalb der Gegenstand weiterer Forschung sein.

5.2.3 Fazit: koordinierte Expertisen für das Erkennen von Daten

Der Abschnitt hat gezeigt, wie die beiden Teams den datenjournalistischen Erkenntnisprozess innerhalb der drei Expertisen – technisch, journalistisch, grafisch – aushandeln und dadurch koordinieren. Sinnhaftes Erkennen von Daten entsteht durch die Abstimmung der grundlegenden Materialität der Datenbank und ihrer quantitativen Auswertung, deren Visualisierung mittels spezifischer geometrischer Formen und dadurch verkörperter Relationierungen, deren Passendmachung mit qualitativen Recherchen durch Hypothesen und Plausibilisierungen.

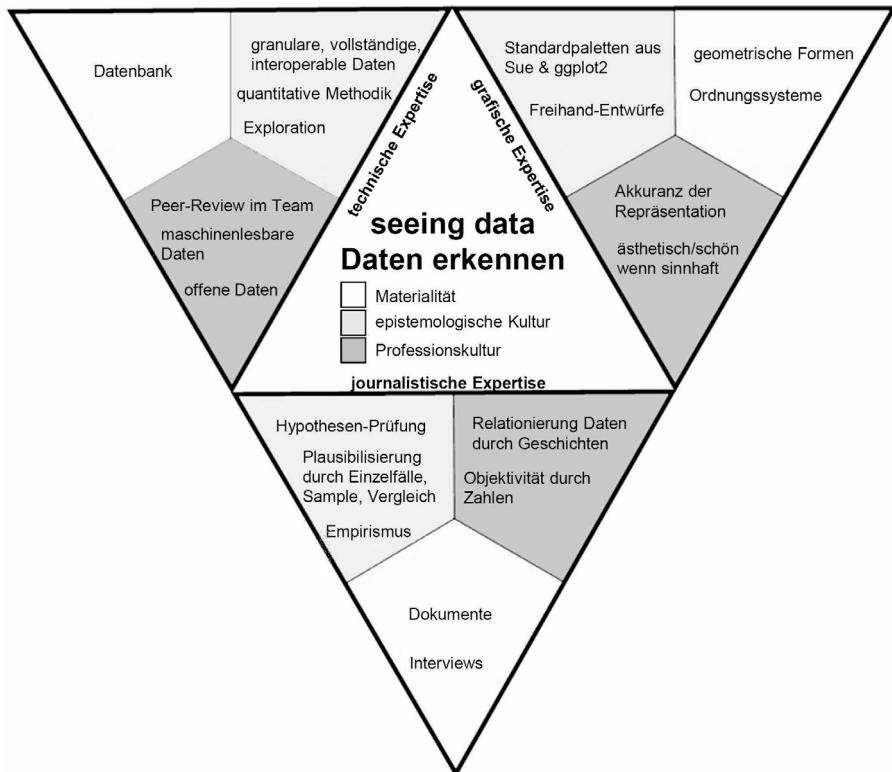
Die Abstimmung geschieht durch Aushandlungen, oft mittels mündlicher Diskussionen direkt vor einem Bildschirm, gebunden und strukturiert durch vorliegende Datenvisualisierungen – es handelt sich um kollaboratives, interaktives ›Erkennen‹, wie es Amann und Knorr-Cetina (1988:138ff) in der Beschreibung der »machinery of seeing« skizziert haben. Diese konversatio-

74 13_20180312_NZZStorytelling-01 (4)

75 15_20180314_NZZStorytelling-03 (10–14)

nale Praktik verbindet mittels der Visualisierung die Daten mit der ›realen‹ Welt, indem koordiniert wird, was als akkurate und objektive Erkenntnis gilt.

Abbildung 10: Koordinierte Expertisen im Modus des Erkennens von Daten.



(eigene Darstellung)

In die Aushandlungssituationen fliessen die epistemologischen Vorgehensweisen, Standards und professionellen Konventionen dreier Expertisen ein. Um die Koordination zu ermöglichen, sind vermittelnde Diskussionen oder Interpretationsleistungen von *traders*, welche mehrere Perspektiven sinnhaft zu verbinden vermögen, nötig. In den abgefragten Weiterbildungswünschen aller Teammitglieder spiegelt sich dieses Bedürfnis nach Multi-

perspektivität darin, dass die meisten sich mehr Expertise aus den jeweils anderen Bereichen wünschen.

Die unterschiedlichen Ausrichtungen – BR Data als eher investigativ-journalistische Einheit, NZZ Storytelling mit der dominierenden Aufgabe der Entwicklung von visuellem Journalismus auf der Basis von Daten – und entsprechenden Teamkompositionen widerspiegeln sich in den Gewichtungen der jeweiligen Expertisen. Und doch finden sich dieselben datenjournalistischen Praktiken zur Erkenntnisproduktion (siehe Abschnitt 5.1) schliesslich in beiden Teams. Bezuglich des Erkennens von Daten lässt sich das aufgespannte allgemeine Schema über die drei Expertisen (Abbildung 2) folgendermassen anpassen:

Das andauernde Auftreten von Aushandlungssituationen – aufgrund der Teamkomposition und der Arbeitsprozesse – hat zwei Konsequenzen: Einerseits verwenden die Teams Formalisierungen in Form von *Tools*, um den ständigen Koordinationsaufwand durch internalisierte Regeln und Standards zu senken – darauf gehen die folgenden Abschnitte ein. Andererseits bilden sich neue Wissensbestände und Subdisziplinen (Burri und Dumit 2008:301ff) – das Feld des Datenjournalismus, auf das Kapitel 3 einging.

5.3 Visuelles Vokabular

Die datenjournalistischen Arbeitspraktiken, insbesondere das *Geschichten-Drehen* (Abschnitt 5.1.3), haben die soziotechnische Verwicklung von (sozialer) Sinngebung und (technischer) Visualisierung gezeigt. Die Grundoperation der Sinngebung besteht dabei darin, die Variablen aus den Datensätzen zu relationieren – durch spezifische geometrische Verkörperungen, beispielsweise durch Diagramme oder durch Karten. Bei den geometrischen Verkörperungen handelt es sich um ein Symbol-System, das in einer Reihe mit verbalen und anderen visuellen Sprachen (zum Beispiel Gesten, Fotografie) steht (Kosminsky et al. 2019:45). Abschnitt 4.2.2 hat gezeigt, dass Datenvisualisierungen und ihre Elemente ihre eigenen semiotischen Eigenschaften, Ausformungen und Restriktionen mitbringen – sie bestimmen also durch ihre »jeweiligen Leistungen und Grenzen [...] als auch die verschiedenen Operationsarten der ›Übersetzung‹ und ›Transposition‹ selber, wobei die Frage entscheidend ist, wie diese das in ihnen aufgezeichnete oder dargestellte Wissen jeweils verändern [Hervorhebungen im Original]« (Heßler und Mersch 2009:34). Um diese Veränderungsprozesse zu untersuchen, steht hier nicht

eine medienwissenschaftliche Perspektive, sondern die Kontexte, in denen die Interpretationsleistungen passieren, im Zentrum – insbesondere die Frage nach den aktivierten Expertisen und kulturellen Perspektiven (siehe auch Aiello 2020).

Die Relationierung durch eine geometrische Verkörperung zielt auf den Kern der grafischen Expertise. *Seeing data* bedeutet, eine Visualisierungsform der Daten zu finden, welcher Akkuranz in der Abbildung eines Phänomens zugeschrieben wird – insbesondere die Diskussion rund um die Fussballtransfers (siehe Ausschnitt mit Fussnote 70 in Kapitel 5) hat dies illustriert. Das Abbilden bedingt, zuerst eine Relationierung zu konstruieren. Basierend auf der Begrifflichkeit der ›Aussage‹, die es aus den Daten zu produzieren gelte, erscheint hierfür die Bezeichnung des ›visuellen Vokabulars‹ treffend. Es erzeugt als Formalisierung eine Syntax, die Semantiken der Daten reduziert, sie aber andererseits in logische Strukturen einbettet (Heßler und Mersch 2009:35). Am Ende von Abschnitt 4.2.2 wurde bereits darauf verwiesen, dass theoretische Überlegungen das ›visuelle Vokabular‹ einbetten in eine Sprachlehre (Grammatik), welche die fundamentalen Prinzipien und Regeln für eine spezifische Fertigkeit vorgibt.

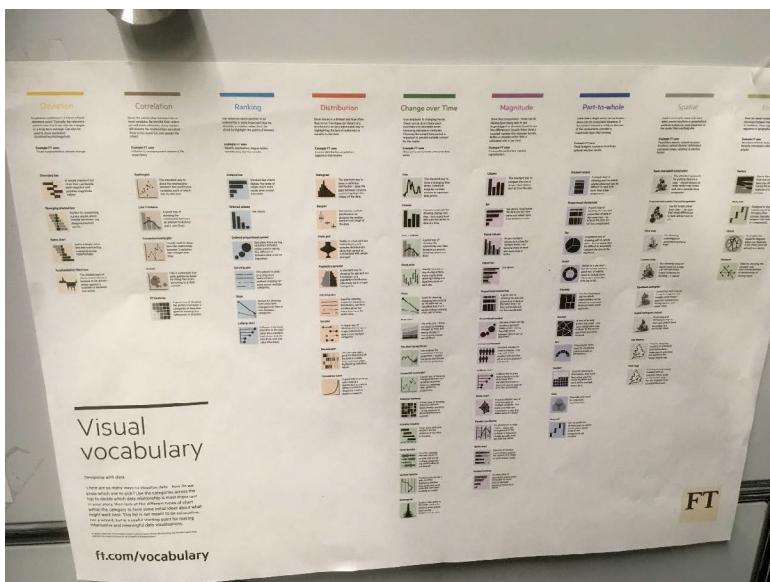
Das visuelle Vokabular referenziert dabei auf eine im Grundsatz unbegrenzte Palette an möglichen Relationierungen von Variablen durch geometrische Verkörperungen. Als grundlegende Materialität der grafischen Expertise wird sie durch alle Infografiker*innen und Designer*innen bzw. Personen mit solchem Bildungswissen in die beiden Teams gebracht. Das visuelle Vokabular fließt so implizit durch die grafische Expertise in die Aushandlungssituationen ein. Es hat aber auch bereits mehrere Stufen der Formalisierung und Institutionalisierung durchlaufen – einerseits durch verschiedene Zusammenstellungen aus dem Feld der Datenvisualisierung, andererseits dient es in verschiedenen Tools als Grundlage der Visualisierungsoptionen. Während der folgende Abschnitt 5.4 sich den Tools widmet, zuerst im Folgenden ein Blick auf die zugrundeliegende Palette.

Wird im Feld auf das visuelle Vokabular verwiesen, so führt diese Referenz sehr oft auf das »Visual Vocabulary«⁷⁶ der englischen Medienorganisation Financial Times (FT) – so hängt beispielsweise ein kleines Plakat mit dieser Palette in den Redaktionsräumen von NZZ Storytelling (siehe Abbildung 11). Die Palette der FT hat sich als Standard in der Community etabliert –

⁷⁶ <https://github.com/ft-interactive/chart-doctor/tree/master/visual-vocabulary> (20.06.2020)

auch die Teammitglieder der NZZ verweisen immer wieder auf die hohe grafische Qualität der FT-Publikationen, und einige sind gar in direktem Kontakt mit FT-Datenjournalist*innen. Die Palette gruppiert verschiedene Visualisierungsformen nach ihrem Relationierungstyp: »Deviation, Correlation, Ranking, Distribution, Change over Time, Magnitude, Part-to-whole, Spatial, Flow« – also Abweichung, Korrelation, Rangordnung, Verteilung, Veränderungen über die Zeit, Größenordnung, Teil-zum-Ganzen-Verhältnis, räumliche Kartierungen, Flüsse (Übersetzungen RE).

Abbildung 11: Aufstellung des »Visual Vocabulary« der Financial Times in den Redaktionsräumen von NZZ Storytelling.



Das Plakat der FT verweist auf die Inspiration aus dem Feld der Datenvisualisierung: Das »Graphic Continuum by Jon Schwabish and Severino Ricca«⁷⁷, bei beiden handelt es sich um Datenvisualisierungsexperten. Vergleicht

77 <https://policyviz.com/2014/09/09/graphic-continuum/> (20.06.2020) Die beiden Autoren berufen sich auf verschiedene Inspirationen für ihre Klassifikation – darunter

man die beiden Paletten, so fallen die zahlreichen Gemeinsamkeiten auf; einzig die gruppierenden Relationierungstypen sind etwas breiter gefasst und die Visualisierungsformen spärlicher beschrieben. So lässt sich festhalten, dass grafische Expertise fast ohne weitere Formung durch journalistische Konventionen in Form des »Visual Vocabulary« in die Redaktionen einfließt.

Das »Visual Vocabulary« bietet einen Ansatz für eine optische Konsistenz und eine vereinheitlichte Perspektive, wie sie Latour theoretisierte, oder eine kulturelle Sinnstruktur, wie sie Burri bezeichnete – Abschnitt 4.2.2 hat dies ausgeführt. In den beiden Teams prägt sie, wie Daten verkörpert und dadurch gesehen werden können. Das Vokabular bestimmt in den datenjournalistischen Praktiken die validen Kombinationsmöglichkeiten und den Variationsraum der Sinngebung – die Kapitel zu den Praktiken des Geschichtendrehens und der Plausibilisierung von Auswertungen (insbesondere durch Vergleiche) haben die Rolle des visuellen Vokabulars bereits implizit beschrieben. Es entspricht damit dem, was Knorr-Cetina (2002) die Technologien beziehungsweise Maschinerien zur Produktion von Erkenntnis in bestimmten Wissenskulturen genannt hat – »[they] refer to the social and cultural structures that channel, constrain, define, and enable the production and evaluation of knowledge—indeed, such structures are both preconditions and constraints for the latter«, schreibt Lamont (2012:211) über die Bedeutung dieser Maschinerien (Abschnitt 4.2.1 hat dieses Konzept bereits ausführlich theoretisch erläutert).

Je vertiefter dabei die grafische Expertise, desto vielfältiger das Vokabular, das aus der Palette heraus auch zur Anwendung gebracht wird. Die datenjournalistischen Arbeitspraktiken in beiden Teams zeigen, dass sie im Modus der Erkenntnisproduktion hauptsächlich grundlegende Formen wie Balken-/Säulen-Vergleiche, Liniendiagramme und Kartierungen nutzen. Weil diese bereits seit längerem eingeübt sind und deshalb auch für Personen mit wenig grafischer Expertise eine schnelle Sinngebung ermöglichen, sind sie stimmig mit den schnellen und flexiblen explorativen Datenanalysen. Visuell weniger gewohnte Visualisierungsformen erfordern einen erhöhten Aufwand, um Daten sinnhaft zu erkennen. Insbesondere in der Koordination mit der journalistischen Expertise und deren Relevanzkriterien zur Produktion von Geschichten verläuft die Sinngebung anhand von visuell unmittelbar erkennbaren Auffälligkeiten: ausschlagende Kurven (Link 2001:88), Ausreisser,

Bertins »Grafische Semiolegie« von 1974, siehe https://policyviz.com/2014/10/06/graphic_continuum_inspiration/ (20.06.2020).

sich farblich abhebende Entitäten auf Karten. Die Sinngebung mit diesem Standardvokabular weckt gleichzeitig das Bedürfnis, dieses zu erweitern: In beiden Teams äussern viele Mitglieder den Wunsch, sich zusätzliche Visualisierungsexpertise anzueignen – sei es durch spezifische *Tools* oder das vertiefte Beherrschung von Programmiersprachen.

Diese Sinngebung passiert sehr oft im Stillen, sich nur in einer reibungslosen Produktion von Datenvisualisierungen manifestierend. Sinngebungsprozesse brechen in beiden Teams meist erst dann hervor und werden explizit, wenn sie mit Reibung verbunden sind – beispielsweise im Fussballtransfer-Projekt, als die Koordination zwischen grafischer Formgebung und journalistischer Geschichte Diskussionen provoziert (siehe Feldnotiz mit Fussnote 70 in Kapitel 5). Oder dann, wenn die *story* eben nicht mit dem gängigen Muster einer ausschlagenden Kurve verbunden werden kann, sondern NZZ Storytelling spezifische Massnahmen ergreift, um eine »Nicht-Kurve« zu einer Geschichte zu machen:

[Susanne] geht zu [Yannick] und [Annalena]: Sie habe Daten zur Namenswahl der Frauen bei Heirat – es verändere sich nichts, die Frauen nähmen immer noch genau gleich oft den Namen des Mannes an! Das sei doch echt erstaunlich. [Yannick] und [Annalena] stimmen zu – und meinen das wäre es doch wert, gerade deshalb eine Grafik zum Artikel zu machen. [Susanne] meint, sie müsste mit den Verantwortlichen auf der Redaktion schauen, ob eine Grafik zum Artikel dazukommen soll. Es wäre aber schon viel Aufwand, und sie sei wirklich nicht sicher, ob sich das lohne. [Annalena] und [Yannick] machen Vorschläge, wie man diese »Nicht-Kurve« umsetzen und anschreiben könnte. [Susanne] meint ja, dann würde sie nochmals insistieren. Es sei wohl ihre erste Grafik ohne wirkliche Kurve. [Annalena] meint, sie solle doch einen Vorschlag machen, auf Q mal entwerfen, dann sei es oft einfacher, dass sie darauf einsteigen würden. [Susanne] sagt, dann schaue sie nochmals.⁷⁸

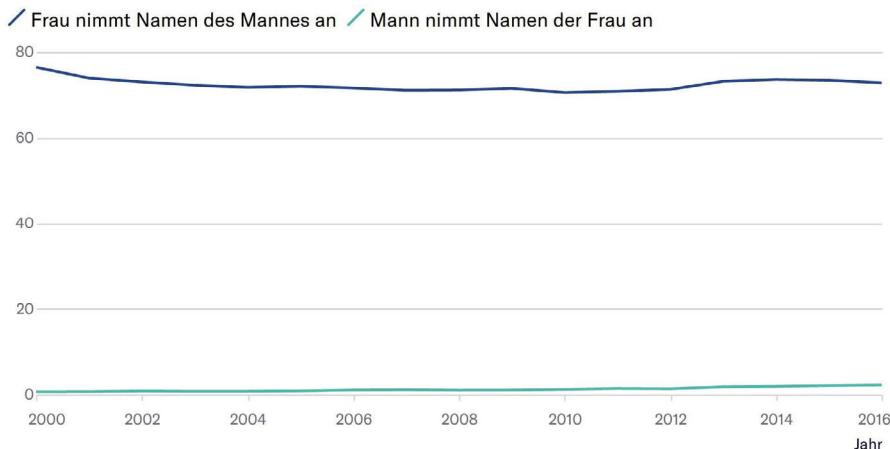
Obwohl die Kurve aus der Perspektive der drei Teammitglieder abnormal verläuft – dass sie nämlich den vermeintlichen Trend nicht widerspiegelt, Frauen würden bei der Heirat ihren Nachnamen behalten –, es sich also um eine Auffälligkeit handelt, zweifelt vor allem die federführende Susanne daran, ob sich hier eine Grafik »lohnt«. Ist plötzlich das konstant Verlaufende das Auffällige, verweigert die Datenvisualisierung als »Auflösungsmaschine«, welche

78 25_20180328_NZZStorytelling-13 (10)

die Daten mittels kognitiv wahrnehmbarer Differenz (Abweichung) zu einer Geschichte transformiert, ihren Dienst. Schliesslich erscheint der Artikel in der Print-Ausgabe ohne Datenvisualisierung⁷⁹, während sich im Online-Artikel die folgende Grafik wiederfindet:

Abbildung 12: Grafik aus dem Artikel »Die meisten Frauen nehmen noch immer den Namen ihres Mannes an«.

Nach wie vor geben meistens die Frauen bei der Heirat ihren Namen ab Namensführung, in Prozent



Seit dem 1. Januar 2013 gilt das neue Namensrecht: Ehepartner können bei der Heirat ihre Namen behalten, die Wahl eines gemeinsamen Familiennamens ist freiwillig. Den Doppelnamen ohne Bindestrich gibt es nicht mehr.

<https://www.nzz.ch/gesellschaft/im-namen-meines-mannes-ld.1371279> (17.02.2020)

5.4 Tools

Die Datenvisualisierungswerzeuge der beiden Teams formalisieren bestimmte Paletten des visuellen Vokabulars und betten sie dadurch in die Arbeitsprozesse ein. Der erste Abschnitt zeichnet die Traditionslinien dieser

79 NZZ vom 6. April 2018, Seite 51.

Paletten nach und erläutert, wie sich diese mit den technischen Infrastrukturen verwickeln. Darauf folgen zwei Abschnitte, welche sich mit spezifischen Werkzeugen auseinandersetzen – *computational notebooks* und Papier – und wie diese die Koordinationssituationen zwischen den Expertisen formen und institutionalisieren.

5.4.1 Formalisierte Visualisierungssysteme

BR Data nutzt für seine visuellen Datenauswertungen die Programmiersprache R, insbesondere das *Package ggplot2*. Wie der Hauptentwickler (Wickham 2010:4) festhält, beruft sich dieses hauptsächlich auf Wilkinsons »The Grammar of Graphics« (Wilkinson 2005) sowie auf Bertins »Graphische Semiologie« (Bertin 1974). Bertin entwickelte seine grafischen Grundsätze lange bevor computergestützte Visualisierungen grosse Verbreitung erfuhren. Wilkinson schloss an Bertins Konzeptionen an, betrachtete dann Visualisierungen aber als datengestütztes System mit verschiedenen Komponenten, die variabel kombinierbar sind.⁸⁰ Das macht sein System anschlussfähig für die Entwicklung einer programmiersprache-gestützten Grammatik der Datenvisualisierungen – eben das *ggplot2-Package* in R. Es entwickelt Wilkinsons System weiter, lehnt sich aber immer noch eng an dessen Komponenten für die Produktion von Datenvisualisierungen an (Wickham 2010): *data, mapping, statistics, scales, geometries, facets, coordinates, theme*. Die Möglichkeit der Zergliederung in diese Komponenten ergänzt sich dabei mit dem *tidy-data*-Ordnungsprinzip in R (Wickham 2014, siehe Abschnitt 5.1.2).

Auch NZZ Storytelling nutzt R, bedient sich aber zusätzlicher Software zur explorativen Datenauswertung: das eigenentwickelte *Tool Sue*. Es orientiert sich am *Visual Vocabulary* der FT⁸¹ und basiert für die technische Umsetzung auf »Vega – A Visualization Grammar«⁸², welches sich auf verschiedene Wurzeln aus dem Visualisierungsfeld beruft, darunter wiederum Bertin (siehe Satyanarayan et al. 2017). Vega offeriert einen Baukasten für die Gestaltung einer

⁸⁰ Bei diesen Komponenten handelt es sich um: *algebra, scales, statistics, geometry, coordinates, aesthetics*. Sie alle bauen auf *data* und *data sets* sowie *variables* und *variable sets* – siehe Schema von Wilkinson 2005:24.

⁸¹ 14_20180313_NZZStorytelling-02 (7)

⁸² <https://vega.github.io/vega/> (27.06.2020)

Palette von Visualisierungen, welchen Designer Kaspar während zweieinhalb Monaten an die Bildsprache der NZZ angepasst hat.⁸³

Ohne hier tiefer in die Details einzutauchen, lässt sich festhalten, dass in beiden Teams die jeweilige Software für die Datenauswertung sich auf bestimmte Visualisierungspaletten und -systeme aus dem Feld der Visualisierung stützt. Es handelt sich um eine Traditionslinie von Bertin über Wilkinson bis hin zu Umsetzungen in R oder Vega – und damit um jene Systeme, welche sich durch ihre Zergliederung in Komponenten besonders gut mit der Expertise der Programmierer*innen und ihren Konventionen der Granularität und der Materialität der Datenbanken koordiniert. Um die Transponierung der Daten in die Visualisierungswerzeuge sicherzustellen, ist diese Sicherung von Interoperabilität essenziell – insbesondere für NZZ Storytelling, welches mit eigener Visualisierungssoftware und -standards arbeitet. Programmierer Jonas erklärt, wieso sich das Team für Vega als »Motor«⁸⁴, welcher die verschiedenen Tools integriert, entschieden hat:

Es nehme die »Grammar of Charts, quasi die ›Bibel der Visualisierungen«, auf und habe somit viele Visualisierungskonzepte schon drin. Man brauche nur noch die Parameter zu setzen, beispielsweise die Grösse und die Achsenbeschriftungen, und Vega rechne und gruppere dann vieles selber. Zudem könne Vega-Code in verschiedenen Programmiersprachen geschrieben werden (bzw. wird mit Packages übersetzt). Derzeit machen viele NZZ-Storytelling-Mitglieder mit Code ihre Grafiken – aber die Sprachen seien nicht einheitlich, es kommen verschiedene zur Anwendung. Man hätte auch sagen können, man konzentriert sich nun auf eine Sprache, das hätte sicher einiges einfacher gemacht. Aber man würde so die Fähigkeiten einiger einfach aussen vor lassen, und einige müssten neue Sprachen lernen. So habe man als Kompromiss Vega gewählt, das alle integriere.⁸⁵

Die Ziele der Infrastruktur-Wartung durch die Programmierer des Teams besteht also darin, Interoperabilität mit den gängigen Datenbankformaten und Programmiersprachen zu sichern und so eine Vielfalt an Daten für die Visualisierungssoftware maschinenlesbar zu machen. In dieser Software wiederum bilden die in der Visualisierungscommunity entwickelten Paletten

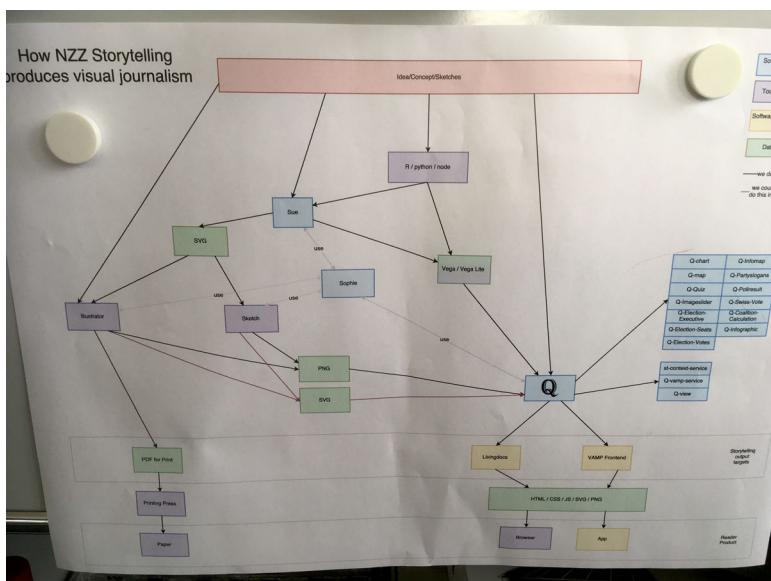
⁸³ 14_20180313_NZZStorytelling-02 (7)

⁸⁴ 14_20180313_NZZStorytelling-02 (7)

⁸⁵ 19_20180320_NZZStorytelling-07 (12)

an *grammars* oder *vocabularies* die Formung, um Daten zu erkennen und zu relationieren. Es lässt sich eine Institutionalisierung bestimmter Paletten – wie das »Visual Vocabulary« der FT sowie Wilkinsons »Grammar of Graphics« – beobachten, welche die Formalisierung durch Software prägt und gleichzeitig von ihr geprägt wird (Star 1999).

Abbildung 13: Von NZZ Storytelling erstelltes Schema der informatik-technischen Infrastruktur des Teams: Zusammenspiel von Tools, Dateiformaten und Arbeitsprozessen in der Produktion von Visualisierungen.



Die Formalisierung in Software beeinflusst die Koordination mit der journalistischen Expertise in dem Sinne, dass sie die Exploration mittels verschiedener Visualisierungstypen beschleunigt: Mit Sue kann NZZ Storytelling verschiedene Visualisierungsformen innerhalb von kurzer Zeit ausprobieren (siehe Projekt zu den Fussballtransfers, Feldnotiz mit Fussnote 13 in Kapitel 5). Designer Kaspar spricht dabei von einem »intuitiveren Zugang« zu den Daten.⁸⁶

Abbildung 14: Screenshot der Visualisierungsoptionen im Menü der NZZ-Software Sue.



NZZ Storytelling eröffnen sich noch weitere Automatisierungs- und damit Beschleunigungsmöglichkeiten der explorativen Datenanalyse, die Programmierer Fabian an einem Teammeeting vorstellt:

[Fabian] zeigt Vega Voyager, ein Interface, wo man einfach Daten einfügen kann, und das dann direkt Vorschläge für Visualisierungsarten liefert. Wenn man eine Grafik anwählt, werden direkt »related views« vorgeschlagen, also Abwandlungen der gewählten Grafik, aber mit noch mehr in die Visualisierung integrierten Daten (zum Beispiel die gewählte Darstellung des Geschlechts mittels Balken nun noch gestapelt nach Gebietseinteilung). Von der angewählten Grafik kann man auch direkt den Code rauskopieren.⁸⁷

Die Präsentation schliesst mit wohlwollendem Applaus – und obwohl ich in meiner Forschungsphase den Einsatz von Voyager nicht beobachten konnte, so zeigt sich, wie hoch der Stellenwert einer flexiblen und schnellen Datenvizualisierung für die Exploration von potenziellen Geschichten ist.⁸⁸ Durch die

⁸⁷ 15_20180314_NZZStorytelling-03 (6)

⁸⁸ Einer der Hauptentwickler von Voyager, Jeffrey Heer, betont diesen Punkt in einem Podcast: »So the idea with Voyager is: How do you get this more broad exposure

verschiedenen möglichen Visualisierungsformen bleiben – innerhalb der angebotenen Palette an Relationierungstypen – möglichst viele Optionen offen, um sie mit den Geschichtenideen und Hypothesen aus den begleitenden Recherchen abzugleichen.

5.4.2 Computational notebooks

Vega Voyager ist ein *computational notebook*⁸⁹ und leitet über zu jenem *Tool*, auf welchem die journalistische Expertise mit der technischen und der grafischen koordiniert wird, um datenjournalistische Erkenntnisse zu produzieren. Bei Notebooks handelt es sich um eine Programmierumgebung für »literate programming«, also Programmierabläufe, die auch für Nicht-Programmierer*innen lesbar sind. Notebooks bieten ein ›Interface‹⁹⁰ für die Passendmachung von qualitativen und quantitativen Zugängen. Es existieren für weitverbreitete Programmiersprachen wie R, Python oder JavaScript Notebook-Programmierumgebungen: R Markdown, Pythons Jupyter Notebook oder ObservableHQ für JavaScript.

Bei BR Data werden Notebooks im html-Format ausgegeben, das alle Teammitglieder auf ihrem Computer problemlos anzeigen können. Die Feldnotiz mit Fussnote 18 (Kapitel 5) beschreibt eine Situation, in der ein solches Notebook zum Einsatz kommt: In einem html-Dokument sind diverse Datenvisualisierungen abgebildet. Das Notebook vereinigt dabei zwei semiotische Ebenen: eine Code-Ebene, die direkt das Schreiben in und die Ausführung

to your dataset early on?« Dies solle nicht nur mögliche Qualitätsprobleme in einzelnen Variablen aufzeigen und deren Verständnis verbessern, sondern auch verhindern, dass man sich zu schnell auf *einen Geschichten-Dreh* fokussiert. »What are the different variables I might combine?« – solche Varianten solle Voyager durch automatisierte Visualisierungsproduktion und dadurch weniger Aufwand aufzeigen. Voyager soll sogar Visualisierungsvarianten empfehlen. Dafür baut Voyager auf einem fein austarierten soziotechnischen Zusammenspiel auf, »a guided tour«. Für die Empfehlung bewertet das Programm die Datenvisualisierungen nach »perceptual effectiveness principles«, welche Variante »present[s] the data in the most effective way«. Dabei sollen auch verschiedene »visual encodings« vergleichbar sein. Siehe: <https://datastori.es/72-jeff-heer>, (30.08.2020), Minuten 19 bis 28.

- 89 Die vollständige und präzise englische Bezeichnung lautet *computational notebook*. In der beobachteten Empirie war jedoch immer von *Notebooks* die Rede.
- 90 Im Englischen wird es sogar explizit ›notebook interface‹ genannt, um es von anderen Notebook- bzw. Notizbuch-Formen abzugrenzen, siehe auch https://en.wikipedia.org/wiki/Notebook_interface (04.10.2019).

von Code erlaubt. Auf dieser Ebene lassen sich Kalkulationen direkt ausführen und Datenvizualisierungen abilden. Die Code-Ebene kann man zu- und ausklappen. Man kann sie damit der anderen Ebene unterordnen. Diese andere Ebene ist die Text-Ebene dank *markdown*-Sprache (die sich sinnigerweise von der *markup language*, der maschinenlesbaren Sprache, abgrenzt⁹¹). Sie erlaubt es, Wörter und einfache Formatierungen in gewohnter Weise zu schreiben – also *rich text* zu produzieren, das für Journalist*innen bisher gewohnte Werkzeug, um Erkenntnisse zu rationalisieren. Da ein Notebook eine Computer-Software ist, basiert auch diese Text-Ebene in letzter Konsequenz auf Code. Dieser bleibt aber verborgen. Schliesslich lässt sich noch auf die zellartige Anordnung der Elemente in Notebooks verweisen: »This new generation of notebooks is based on a linear collection of cells, each of which contains rich text or code that can be executed to compute results or generate visualizations. These cells are linearly arranged, but can be reorganized, reshuffled, and executed in any order.« (Rule 2018:14)

Notebooks bieten mit ihrer auf Code basierenden Struktur ein interoperables, flexibles Instrument, um die auf verschiedene Weise generierten Erkenntnisse und Recherchen eines Projekts zusammenzubringen. Die Notebooks lösen so die Spannung zwischen den verschiedenen epistemologischen Kulturen auf beziehungsweise liefern eine Grundlage, um diese in einer mündlichen Aushandlung zu koordinieren, um Daten zu erkennen.

Die anspruchsvolle Koordination der drei Expertisen gelingt aber nicht immer. Die Teams greifen deshalb auf ein weiteres Werkzeug zurück, wie der folgende Abschnitt zeigt.

5.4.3 Papier

Auf allen Pulten der Teammitglieder von NZZ Storytelling liegt Notizzettel immer bereit. Ab und zu setzen sie es ein, um der anwesenden Forscherin einen komplizierten Sachverhalt zu erklären. In den meisten Fällen nutzen sie es aber, um Visualisierungen zu entwerfen. Obwohl das Team über diverse Werkzeuge verfügt, die schnell computergestützt visualisieren, ist Notizzettel immer noch in fast jeder Geschichte irgendwo involviert.

⁹¹ Die Zentralität des *Markup*-Begriffs für den Datenjournalismus zeigt sich darin, dass eines der meistdiskutierten Projekte des Feldes, das investigativen Journalismus in und über die digitale Welt betreiben will, sich »The Markup« nennt: <https://themarup.org/> (04.10.2019)

Die Teammitglieder setzen dabei schon ganz zu Beginn handgezeichnete Notizen ein – zu einem Zeitpunkt, an dem ihnen die vorliegenden Daten noch nicht im Detail bekannt sind. Es dient dazu, von Anfang an *das Hauptargument* ins Zentrum zu stellen, es entsprechend visuell herauszustellen.

[Caroline] zeichnet auf vor ihr liegendes Notizzettel erste Skizzen. [Die Auslandkorrespondentin] meint, der rote Faden der Geschichte sei ungefähr, dass das Wassermanagement der Stadt einfach unfähig sei. Die Stadt sinke ab, das Problem sei multifaktoriell. [Caroline] schlägt vor, eine grosse Darstellung mit einem Querschnitt zu machen. Und dazu »Small Multiples« mit verschiedenen Zukunftsszenarien, dargestellt als Skizzen.⁹²

Handgezeichnete Entwürfe begleiten auch die explorative Datenanalyse im Fussballtransferprojekt: Als David den Auftrag für die Datenanalyse und -visualisierung im Rahmen eines kleinen Meetings an Kaspar und Caroline überträgt, liefert er gleich einen Visualisierungsentwurf mit:

Es ist eine »Sankey«-Grafik. Es sei ja auch spannend, woher die Spieler kämen [meint David]. Außerdem schlage er vor, dann einzelne Spieler herauszupicken, um »es« greifbar zu machen. David präsentiert auf einem Excel-Sheet direkt auf seinem Laptop gleich eine Auswahl von fünf Spielern, die teuersten Transfers aus der Schweiz. Die Datensätze hat er bereits vorsortiert und präsentierte sie ebenfalls auf dem Laptop. [Caroline] nimmt dann einen Notizzettel, den sie mitgebracht hat, und zeichnet einen eigenen Visualisierungsvorschlag. Sie findet, dass eine Bar-Chart-Variante bessere Vergleichbarkeit biete.⁹³

Etwas später am selben Tag bespricht sich Kaspar mit seinem Kollegen Yannick. Die beiden diskutieren die möglichen Geschichten sowie Datenvisualisierungen zum Fussballtransferprojekt intensiv. Das iterative Datenerkennen geschieht dabei nicht nur anhand der explorativen Visualisierungen mit Sue und dem thematischen Wissen der beiden, sondern bezieht auch die handgezeichneten Visualisierungsentwürfe von Kaspar mit ein. Die Produktion der Erkenntnis richtet sich danach aus, wie sie sich *griffig*, das heisst schnell erfassbar, visualisieren lässt. Das heisst, die Daten müssen sofort *als Geschichte* erkennbar sein. Diese Evidenzmontage (siehe Kapitel 6) wirkt in den beiden

92 24_20180327_NZZStorytelling-12 (9)

93 14_20180313_NZZStorytelling-02 (3)

Teams bereits in die Erkenntnisproduktion hinein. Die Zeichnung auf Papier während der Datenanalyse sorgt für eine Selektion derjenigen Analysemöglichkeiten, die erlauben, dass am Schluss die »Visualisierung als Geschichte« taugt (Segel und Heer 2010).

Traders mit viel grafischer Expertise nutzen besonders häufig Papier, um die journalistische mit der grafischen Expertise zu koordinieren. So liegt einen Tag nach der Diskussion mit Yannick auf dem Pult von Kaspar, neben seiner Tastatur, ein Entwurf auf einem Notizblock. Er stammt von Emma, die bisher nicht in das Projekt involviert war und auch wenig Wissen über das Fussballgeschäft mitbringt. Der Entwurf ähnelt stark demjenigen, welcher bereits am initialen Meeting von David vorgelegt wurde, und Kaspar versucht nun, diesen mit der explorativen Grafiksoftware Sue nachzustellen.⁹⁴ Indem also die technischen Formalisierungen mit ihren Begrenzungen (temporär) ausgeschaltet werden, vereinfacht sich die Koordination einer Visualisierung mit einer Geschichte. Ist demnach die Abstimmung zwischen den drei Expertisen besonders komplex, entstehen oftmals datenjournalistische Erkenntnisse ohne formalisierte Datengrundlage.

5.4.4 Fazit

Der vorangegangene Abschnitt über *Tools* hat gezeigt, wie Werkzeuge zur explorativen Datenauswertung in den beiden Teams die Koordinationssituationen zwischen den Expertisen formen. Software zur Datenvisualisierung wie das R-Package ggplot2 oder NZZ Storytellings Sue und Vega bauen auf Paletten von Visualisierungsformaten auf, welche sich in Komponenten zergliedern lassen. Sie ermöglichen so im Zusammenspiel mit der Datenbank-Materialität und der technischen Expertise eine hohe Varianz an Visualisierungsoptionen und eine teilweise Automatisierung von explorativen Methoden. Die Interoperabilität zwischen den Programmen ist dabei eine Voraussetzung für eine vernetzte technische Infrastruktur, wie sie NZZ Storytelling aufgebaut hat. Diese Grundlagen ermöglichen ein flexibles, schnelles und variables Erkennen von Daten mit computergestützten *Tools*. Das Werkzeug des computational notebooks bietet dann ein *interface* zwischen den grafisch-technischen Datenauswertungen und der journalistischen Expertise: In der Verwicklung von Code und Text werden unterschiedliche Relevanzbewertungen und epistemologische Zugänge abgeglichen, um aus Daten Geschichten zu erkennen.

94 15_20180314_NZZStorytelling-03 (3)

Für dasselbe Ziel kommt dabei häufig Papier als Werkzeug zum Zuge. Dabei werden die computertechnischen Restriktionen temporär ausgeschaltet und die Expertisen ausserhalb von Software-Struktur koordiniert. Dadurch fokussiert die Koordination stärker auf die visuelle Verkörperung als Geschichte, indem die Visualisierung bereits in der explorativen Phase der Konstruktion einer Hauptaussage untergeordnet wird.

5.5 Ideen imitieren

[Lea] scrollt durch den BBC-Artikel und meint, so eine griffige Grafik wie die erste aus dem Artikel zu haben wäre schon gut.⁹⁵

Als Lea mit Roman und Adrian das Projekt zu den Nutzungsbedingungen und Datenschutzerklärungen diskutiert, referenziert sie auf einen Artikel der BBC⁹⁶, welcher ein halbes Jahr vorher erschien. Er basiert auf den (englischen) Nutzungsbedingungen verschiedener Internet-Plattformen und ordnet ihre Verständlichkeit im Vergleich mit klassischer englischer Literatur ein. Nach einem Austausch mit Personen der BBC war BR Data gewillt, die Geschichte auf Deutschland zu adaptieren. Lea referenziert primär auf die Grafik (so im Ausschnitt aus der Feldnotiz) sowie die Illustration mittels Kindern, welche die für sie viel zu komplizierten Texte vorlesen.

Dies ist ein Beispiel für die Praktik, dass BR-Data-Teammitglieder andere Medien dann referenzieren, wenn sie eine Geschichte adaptieren wollen. Referenziert werden dabei sehr oft Visualisierungen – so erfolgt während eines Ideen-Brainstormings bei BR Data die Bewertung eines anderen Projekts⁹⁷ anhand des Kriteriums, etwas *sichtbar* gemacht zu haben, wie eine Teilnehmerin ausführt:

⁹⁵ 48_20181211_BR-Data-02 (3)

⁹⁶ <https://www.bbc.com/news/business-44599968> (10.03.2020)

⁹⁷ <https://interaktiv.tagesspiegel.de/radmesser/> (10.03.2020)

[...], die Radmesser-Recherche habe schon gewirkt, auch wenn man vorher schon wusste, dass Radfahrer eng überholt werden. Es habe einen schon beeindruckt, das nochmals zu sehen.⁹⁸

Im Projekt zu den Polizeiberichten wiederum führt eine mitten in der Datenanalyse erscheinende Grafik des österreichischen Der Standard dazu, dass die eigene Entscheidung, einen Vergleich mit einer anderen Statistik nicht vorzunehmen, nochmals evaluiert wird (siehe Feldnotiz Fussnote 55 in Kapitel 5): Liesse sich die Erkenntnis, welche Der Standard durch den Vergleich ermöglicht hat, auch für das eigene Projekt übernehmen?

Das Referenzieren auf Visualisierungen als Sichtbarmachungen oder auf die Auswertungsmethoden – wie es in anderen Situationen in beiden Teams vorkam – ist ein Referenzieren auf Sinngebungen von Daten: Die Datengrundlage ändert sich in der Regel bei der Adaption auf den lokalen Kontext, aber es wird die Erkenntnis, die Aussage aus den Daten (und damit auch das methodische Vorgehen, mit dem diese produziert wurde) übernommen.

NZZ Storytelling verfügt über einen »Inspirations-Channel« auf Slack, wo einige Teammitglieder pro Woche rund fünf bis zehn Einträge platzieren und darauf hinweisen, wieso sie jene Geschichte »inspirierend« finden. Designer Kaspar hat die Einträge gesammelt und sie auf die Organisationsplattform trello hochgeladen. Er zeigt die Sammlung:

Er hat selbst Kategorien erstellt (ca. 20), differenziert nach Formen und Formaten von Visualisierungen. Es gibt zudem eine Kategorie, die sich ›Einschweizer‹ bzw. ›Einzürchern‹ nennt: Diese beinhaltet Ideen, die man lokal adaptieren könnte. Die Einteilung funktioniere manuell, erklärt [Kaspar], und er habe die Plattform schon länger nicht mehr nachgeführt. Es brauche sie derzeit auch niemand. Der Inspirationschannel werde schon gepflegt, aber die Dinge dann wieder zu finden sei natürlich mühsam. Ich frage [Kaspar], welche Medien oder Plattformen besonders oft vertreten seien im Channel – er meint, das seien so ›die Üblichen‹, wie Financial Times und New York Times. Er geht etwas durch die Quellen und meint dann, es seien aber auch viele kleine dabei. Ich frage ihn, wie sie zu den Quellen kämen – [Kaspar] meint, vieles laufe wohl über Twitter. Als er die Einträge der Kategorie »Scrollytelling« anschaut, meint er, das sei auch etwas ›vorbei‹. Ich frage ihn, ob es denn Trends gebe – das bejaht er.⁹⁹

Die Sammlung ist kategorisiert nach Visualisierungsformaten, was den Relationierungstyp als auch das Präsentationsformat – beispielsweise »Scrollytelling«¹⁰⁰ – als für die Teammitglieder relevantes Ordnungskriterium ausweist. Ein weiteres Beispiel ist die Grafik-Sammlung der Designerin Emma auf Pinterest. Als sie überlegt, wie NZZ Storytelling die vielen Wechsel in der Trump-Administration beleuchten könnte, erklärt sie mir ihr Vorgehen am Pult:

Ausserdem hätten die New York Times und die Washington Post schon Artikel dazu gemacht, die schaue sie sich an, was halt andere schon gemacht haben. Sie sammle nun Ideen. Ich frage sie, wie sie diese Ideen nun sucht. Sie meint, das sei eine gute Frage – sie schaue sich halt viele Beispiele und Grafiken an, in diesem Falle Ideen für eine Timeline. Sie habe dazu auf Pinterest Sammlungen angelegt – geordnet je nach Typen. Sie zeigt mir die Alben gleich auf dem Bildschirm. Es sind sehr viele für alle möglichen Sorten von Grafiken. Sie sammle diese selber zusammen, schaue sich Beispiele im Internet an.¹⁰¹

In die datenjournalistischen Praktiken beider Teams fliessen die Ideen der datenjournalistischen Geschichten anderer ein. Die Imitation verläuft über Datenvisualisierungen und deren Sinnzusammenhänge. Damit übernehmen die Teams ein bereits ausgehandeltes Schema für die Produktion von datenjournalistischen Erkenntnissen. Das Imitieren nimmt insofern die gleiche Rolle ein wie die neuen Relationierungstypen, welche die *traders* einbringen: Sie schlagen ein aus Sicht der Teams sinnhaftes Expertisen-Koordinationsmuster (in Form einer geometrischen Verkörperung) vor. So überrascht es nicht, dass in beiden Teams Personen, welche überdurchschnittlich oft die *trader*-Rolle einnehmen, genauso überdurchschnittlich oft datenjournalistische Geschichten aus anderen Redaktionen referenzieren, diskutieren und Sammlungen darüber anlegen.

Dieser Abschnitt zeigt auf einer zusätzlichen Ebene, wie sich die datenjournalistische Community gegenseitig beobachtet und welche Prozesse dies lokal, in den Teams, antreibt. Imitation bedeutet also nicht nur ein Editieren von Narrativen über die Potenziale von Datenjournalismus, wie es Abschnitt

99 26_20180329_NZZStorytelling-14 (11)

100 Unter »Scrollytelling« versteht man gemeinhin eine Erzählweise mit sich abwechselnden Text- und Visualisierungselementen, durch welche man sich einfach »durchscrollen« kann. Detaillierter: <https://medium.com/nightingale/from-storytelling-to-scrollytelling-a-short-introduction-and-beyond-fbda32066964> (31.12.2020)

101 20_20180321_NZZStorytelling-08 (5)

3.3 illustriert hat. Imitation bezieht sich *auch* auf konkrete Projekte, gar auf einzelne Datenvisualisierungen, die besonders *griffig* erscheinen in der Erkenntnisproduktion. Der Imitationsprozess ist geprägt durch die Expertisen und Kulturen in den Teams: So re-evaluerte BR Data zwar aufgrund des österreichischen Berichts zu den Polizeimeldungen die eigene Auswertungsmethode, verzichtet aber aufgrund der eigenen Ansichten über statistische Qualität auf die gleiche Vorgehensweise.¹⁰²

5.6 Fazit

Das Kapitel 5 hat empirisch nachgezeichnet, welche datenjournalistischen Arbeitspraktiken den ersten Modus der Datenvisualisierungen, das Erkennen von Daten (*seeing data*), ausmachen, welche Aushandlungs- und Koordinationsprozesse sich in diesem Zusammenhang in den beiden untersuchten Teams abspielen und welche Rolle dabei Visualisierungswerkzeuge einnehmen.

»Other researchers present in the laboratory are attracted by such events, gather around the visual materials, finger the documents and gaze about their surfaces. As they examine the film, scientists begin a series of verbal exchanges. [...] When embedded in talk, ›seeing‹ is interactively accomplished« (Amann und Knorr-Cetina 1988:138).

Genauso wie sich die Forscher*innen um ihre molekulargenetischen Laborproben versammeln, sie betrachten und diskutieren, um kollektiv Erkenntnisse zu produzieren, nutzen die beiden datenjournalistischen Teams Arbeitspraktiken, um Daten zu erkennen: Visualisierungen dienen als Arbeitsoberfläche für die Praktiken der Datafizierung, der Herstellung sauberer Daten, des Drehens von Geschichten und der Plausibilisierung der Auswertungen. Der Abschnitt 5.1 hat diese Arbeitspraktiken nachgezeichnet:

- (1) Den Praktiken zugrunde liegen die ontologischen und epistemologischen Annahmen, dass sich in datafizierter Information reale Ereignisse widerspiegeln und sich das Geschichten-Format in Daten übersetzen lässt. Dies passiert insbesondere über die Vorstellungen von visuellen Verkörperungen der Daten.

¹⁰² 56_20190108_BR-Data-10 (4)

- (2) Die Herstellung »guter« datenjournalistischer Geschichten bedingt *saubere Daten* als deren Grundlage: Die Datensätze sollen granular, systematisch geordnet und kategorisiert sowie interoperabel sein. Diese Konventionen bringt vor allem die programmiertechnische Expertise ein. Die Eigenschaften sind für die journalistische Expertise besonders anschlussfähig, weil sie ein Maximum an Optionen für mögliche Geschichten offenhalten.
- (3) Die Relationierungen der Variablen in den Datensätzen erfolgen unter dem Begriff des *Geschichten-Drehens*. Die Teams bewegen sich dabei iterativ in einem Spannungsfeld von datengetriebener (*data-driven*) und hypothesengetriebener (*story-driven*) Vorgehensweise. Geometrische Verkörperungen wie Säulen-, Balken und Liniendiagramme, Ausreisser oder Karten sind das visuelle Vokabular, um Daten zu erkennen, und diejenigen Objekte, welche die Aushandlungen strukturieren. Die dabei produzierten Auffälligkeiten, (Dis-)Kontinuitäten, Normalitätseinteilungen und Neuheiten bieten sinnhafte Anschlusspunkte für die Relevanzzuschreibungen der journalistischen Expertise.
- (4) Mittels Praktiken der Plausibilisierung machen die Teams dann die Datenvisualisierungen, deren Grundlage in der Regel eine quantitative Auswertung ist, mit den für den »klassischen« Journalismus typischen Common-Sense-Annahmen passend. Für die Plausibilisierung werden Samples und Einzelfälle vertieft qualitativ untersucht oder Sinnzusammenhänge durch Vergleiche konstruiert.

Diese Praktiken des Erkennens von Daten handeln die Teams in einem Spannungsfeld dreier Expertisen aus: der journalistischen, der programmiertechnischen sowie der grafischen. Viele Teammitglieder sind dabei in der Lage, als *traders* mehrere Expertisen zu überbrücken, und sie nehmen dadurch Vermittlungspositionen ein, welche die Koordination ermöglichen.

Die Sinngebung steht dabei in einem engen Zusammenhang mit dem visuellen Vokabular, das über die grafische Expertise in die Teams einfliest. Die Koordination kann einerseits situativ, das heißt für jede Datenvisualisierung neu erfolgen, was meist über mündliche Diskussionen passiert. Oder die entsprechenden Arbeitspraktiken sind bereits als Visualisierungssysteme formalisiert in Programmiersprachen und Werkzeugen zur explorativen Datenauswertung. Als besonders geeignete Arbeitsoberflächen zur Sinngebung haben sich das *computational notebook* und Skizzen auf Papier herausgestellt: Während Ersteres den Abgleich verschiedener Expertisen und ihrer epistemologischen Zugänge erlaubt, indem es sie in einem Werkzeug kombiniert, ermöglicht Pa-

pier das Ausschalten der technischen Beschränkungen und dadurch eine erleichterte Koordination von journalistischer und grafischer Bewertung.