

ars digitalis

Peter Klimczak
Christer Petersen
Samuel Schilling *Hrsg.*

Maschinen der Kommunikation

Interdisziplinäre Perspektiven auf Technik
und Gesellschaft im digitalen Zeitalter

EBOOK INSIDE



Springer Vieweg

ars digitalis

Reihe herausgegeben von

Peter Klimczak, FG Angewandte Medienwissenschaften, Brandenburgische Technische
Universität, Cottbus, Deutschland

Die Reihe ars digitalis wird herausgegeben von PD Dr. Dr. Peter Klimczak.

Sollen technische und kulturelle Dispositionen des Digitalen nicht aus dem Blickfeld der sie Erforschenden, Entwickelnden und Nutzenden geraten, verlangt dies einen Dialog zwischen den IT- und den Kulturwissenschaften. Ausgewählte Themen werden daher jeweils gleichberechtigt aus beiden Blickrichtungen diskutiert. Dieser interdisziplinäre Austausch soll einerseits die Kulturwissenschaften für technische Grundlagen, andererseits Entwickler derselben für kulturwissenschaftliche Perspektiven auf ihre Arbeit sensibilisieren und den Fokus auf gemeinsame Problemfelder schärfen sowie eine gemeinsame ‚Sprache‘ jenseits der Fachbereichsgrenzen fördern. Notwendig ist eine solche interdisziplinäre Auseinandersetzung nicht zuletzt deshalb, um den vielfältigen technischen Herausforderungen an Mensch, Kultur und Gesellschaft ebenso informiert wie reflektiert zu begegnen.

In dieser Reihe finden nicht nur Akteure aus Wissenschaft, Forschung und Studierende aktuelle Themen der Digitalisierung fundiert aufbereitet und begutachtet, auch interessierte Personen aus der Praxis werden durch die interdisziplinäre Herangehensweise angesprochen.

Peter Klimczak, Dr. phil. et Dr. rer. nat. habil., ist Privatdozent an der Brandenburgischen Technischen Universität.

Weitere Bände in der Reihe <http://www.springer.com/series/16465>

Peter Klimczak · Christer Petersen ·
Samuel Schilling
(Hrsg.)

Maschinen der Kommunikation

Interdisziplinäre Perspektiven auf Technik
und Gesellschaft im digitalen Zeitalter

Hrsg.

Peter Klimczak
FG Angewandte Medienwissenschaften
Brandenburgische Technische Universität
Cottbus, Deutschland

Christer Petersen
FG Angewandte Medienwissenschaften
Brandenburgische Technische Universität
Cottbus, Deutschland

Samuel Schilling
Brandenburgische Technische Universität
Cottbus, Deutschland

ISSN 2662-5970

ISSN 2662-5989 (electronic)

ars digitalis

ISBN 978-3-658-27851-9

ISBN 978-3-658-27852-6 (eBook)

<https://doi.org/10.1007/978-3-658-27852-6>

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Springer Vieweg

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2020

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag, noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Springer Vieweg ist ein Imprint der eingetragenen Gesellschaft Springer Fachmedien Wiesbaden GmbH und ist ein Teil von Springer Nature.

Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Vorwort

Waren vor nicht einmal einer Generation Computer, Digitales und das Internet noch die Domäne eines versierten Spartenpublikums, sind heute digitale Endgeräte in jeden Winkel der Gesellschaft vorgedrungen und aus dem Leben nicht nur der jüngeren Generationen kaum mehr wegzudenken. Diese Entwicklung geht mit einer immer stärkeren Maskierung der Komplexität der Apparatur einher: Ehemals kompliziert zu handhabende Hard- und Software wird durch Simplifikation und Reduktion für die breite Masse als Medium erfahrbar; gleichzeitig treten (informations-)technische Mechanismen digitaler Medien durch die opaken Oberflächen immer nutzerfreundlicherer Interfaces kaum mehr ins Bewusstsein ihrer Anwender.¹

Digitalisierung sowie maschinenvermittelte und maschinelle Kommunikation präformieren dabei nicht nur kulturelle Praktiken, sondern wirken sich genauso auf industrielle, wirtschaftliche und soziale Prozesse aus. In Form humanoider Roboter oder digitaler Agenten, als Gesprächspartner zum Zweck des Fremdsprachenlernens oder als Gegenüber im Spiel etwa können diese künstlichen Akteure zum Bestandteil sozial anmutender Situationen werden. Online-Dienste wie Wikipedia oder Google prägen unseren Zugang zu Informationen. Zugleich erzeugen die Handlungen Einzelner im World Wide Web Daten, welche eine Klassifizierung von Nutzern erlauben, die wiederum in personalisierten Informationen resultiert. All das sind Entwicklungen, die bestehende Machtbeziehungen neu bestimmen und damit von höchster politischer Brisanz sind. Das Medium wird hier nicht zuletzt auch zur politischen Botschaft. So finden sich Aussagen wie „Ohne die sozialen Medien wären die Brexit-Abstimmung

¹Ob wie in unserer kurzen Einführung das tradierte generische Maskulinum, wie in einigen der Artikel das generische Femininum oder andere Formen wie der Gendergap gewählt werden, obliegt den Autorinnen und Autoren. Wir greifen hier editorisch nicht ein, sondern plädieren für Vielfalt, auch dann, wenn dies in einigen der Artikel zur hybriden Verwendung der Formen führt.

und die letzten US-Wahlen anders verlaufen“ und die Sorge darüber, dass nur mehr „diejenigen die Wahlen gewinnen, die über die beste Technologie und die meisten Daten verfügen,“² im wissenschaftlichen und journalistischen Diskurs heute allorts. Differenzierte Analysen und Lösungsansätze bleiben dagegen rar.³

Zugleich eröffnen technische Methoden der Informatik den Geisteswissenschaften derzeit eine Vielzahl neuer Zugänge zu ihren Untersuchungsgegenständen. Diese neuen wissenschaftlichen Werkzeuge erlauben es, zu bereits bekannten Phänomenen große Datensätze zu aggregieren und diese mit völlig neuen Fragestellungen zu untersuchen. Ebenso produzieren die Nutzer des World Wide Web wie auch die Sensorik technischer Apparaturen eine Vielzahl an Daten. Die Existenz dieser Daten wiederum suggeriert der wissenschaftlichen Forschung eine Transparenz technischer Systeme und deren Nutzung, wobei der Entstehungshintergrund und die Verarbeitungsprozesse der Daten jedoch die aus ihnen gewonnenen Ergebnisse maßgeblich beeinflussen und oftmals im ‚blinden Fleck‘ der Forschung bleiben. In manchen Formen des maschinellen Lernens, den künstlichen neuronalen Netzen etwa, bleiben die jeweils konkreten Ergebnisse konstruktionsbedingt selbst ihren Entwicklern verborgen, da diese aufgrund der Vielzahl von Neuronenverbindungen – einer Black Box gleich – Outputs generieren, die für den menschlichen Beobachter nicht mehr nachvollziehbar sind.

Dieser erste kursorische Überblick über die unterschiedlichen Aspekte maschineller Kommunikation und der Kommunikation mit Maschinen zeigt die vielzähligen Arten auf, in denen uns heute in der Gesellschaft digitale technische Systeme begegnen und auf welche unterschiedliche Weise diese auf jene einwirken. In ihren unterschiedlichen Erscheinungsformen – von technischen Anwendungen für Endverbraucher bis hin zum industriellen Einsatz oder auch in Form scheinbar sozialer Akteure – beeinflussen diese technischen Apparaturen heute die Gesellschaft und Wirtschaft, unseren Zugriff auf und Umgang mit Information und Wissen und nicht zuletzt die Wissenschaften.

Sollen die technischen und gesellschaftlichen Dispositionen der neuen digitalen Medien dabei nicht aus dem Blickfeld der sie Erforschenden geraten, verlangt dies einen Dialog von Kultur- und Sozialwissenschaftlern mit Computer- und Ingenieurwissenschaftlern. Die einzelnen Themen dieses Sammelbandes wurden daher im Rahmen

²Wir zitieren exemplarisch Toby Walsh, Professor of Artificial Intelligence an der University New of South Wales, in einem Interview mit Christiane Peitz im *Tagesspiegel* vom 20. Juli 2019: „Mikrotargeting zersetzt die Demokratie“ (<https://www.tagesspiegel.de/kultur/ki-forscher-toby-walsh-mikrotargeting-zersetzt-die-demokratie/24679352.html>) [letzter Zugriff: 24.07.2019].

³Siehe beispielweise den Datenjournalisten Michael Kreil (<https://data.info.graphics/blog/2018/12/21/social-bot-research-is-flawed/> [letzter Zugriff am 30.07.2019]), der an der gegenwärtigen Forschung zu und Bestimmung von Social Bots mittels künstlicher neuronaler Netze darlegt, dass die so gewonnenen Ergebnisse zweifelhaft sind und die Methoden nicht transparent gemacht werden.

eines dreitägigen Kolloquiums an der Brandenburgischen Technischen Universität jeweils aus einer geisteswissenschaftlichen und einer informationstechnischen Blickrichtung diskutiert.⁴ Der interdisziplinäre Austausch sollte einerseits Geisteswissenschaftler für die technischen Grundlagen ihrer Untersuchungsgegenstände, andererseits die Entwickler dieser Technologien für kultur- und sozialwissenschaftliche Perspektiven auf ihre Arbeit sensibilisieren und so den Fokus auf gemeinsame Problemfelder schärfen. Trotz der Vielfalt der in diesem Band beschriebenen Phänomene zeigen sich dabei immer wieder Überschneidungen und gemeinsame Fragestellungen sowie vor allem die Notwendigkeit einer interdisziplinären Auseinandersetzung mit diesen Untersuchungsgegenständen. Notwendig ist eine solche Auseinandersetzung nicht zuletzt deshalb, weil es darum geht, den vielfältigen Herausforderungen durch diese neuen Techniken an Mensch, Kultur und Gesellschaft ebenso informiert wie reflektiert zu begegnen.

im Juli 2019

Peter Klimczak
Christer Petersen
Samuel Schilling

⁴Wir danken bei dieser Gelegenheit dem Forschungscluster Kognitive Systeme an der Brandenburgischen Technischen Universität für die Unterstützung des im Juli 2018 veranstalteten Kolloquiums.

Inhaltsverzeichnis

Teil I Metadaten

1 Metadaten im Kontext	3
Karoline Krenn und Jens Tiemann	
2 Maschinen in der visuellen Kommunikation	29
Harald Klinke	

Teil II Wissen und Suchmaschinen

3 Gefangen in der Filterblase?	45
Carsten Hartmann	
4 Wissen im Zeitalter von Google, Fake News und alternativen Fakten	63
Thomas Zoglauer	

Teil III Personalisierung

5 Wir unterrichten die Maschinen, die Maschinen unterrichten uns	87
Sviatlana Höhn	
6 Personalisierter Wahlkampf 2.0?	105
Isabel Kusche	

Teil IV Mensch und Maschine

7 Mit Computern spielen	125
Jonathan Harth	
8 Die Spur des simulierten Anderen	143
Thomas Christian Bächle	

Teil V Maschinenlernen

9 KI im Verlagswesen	167
Christoph Bläsi	

10 Künstliche neuronale Netzwerke als Black Box.	189
Andreas Sudmann	

Teil VI Normativität

11 Technikentstehung und Determination.	203
Jens Schröter	

12 Algorithmic Governance der Wikipedia.	215
Claudia Müller-Birn	

Stichwortverzeichnis.	241
-----------------------------------	------------

Teil I

Metadaten

Metadaten im Kontext

1

Warum wir eine neue Datenkunde brauchen

Karoline Krenn und Jens Tiemann

Zusammenfassung

Metadaten sind ein Schlüssel zur Nutzung und Deutung von Daten. Im Zusammenhang mit dem in digitalen Prozessen anfallenden Datenvolumen gilt das umso mehr. Der Beitrag diskutiert Metadaten in der digital vermittelten Kommunikation aus verschiedenen disziplinären Perspektiven: im Betrieb von Maschinen, bei der Aufbereitung und Verwertung von Daten und schließlich aus einer soziologischen Beobachtungsperspektive. Metadaten sind in allen Fällen ein kontextabhängiges Element der Wissensrepräsentation. Metadaten zu erzeugen und zu nutzen, heißt immer, Daten zu deuten und Deutungen zu materialisieren. Am Beispiel eines Community-basierten Funknetzes für das Internet der Dinge (IoT) erläutern wir die Bedeutung von Domänenwissen. Wir argumentieren, dass die Nachnutzung von (Meta-)Daten eine Rekonstruktion ihres Entstehungskontexts erfordert. Mit der sozialwissenschaftlichen Datenkunde stellen wir abschließend ein solches Werkzeug zur Rekonstruktion von Kontext vor.

Schlüsselwörter

Metadaten · Kontext · Datenqualität · Wissensrepräsentation · Nachnutzung · Datenkunde

K. Krenn (✉) · J. Tiemann

Fraunhofer-Institut für offene Kommunikationssysteme, Berlin, Deutschland

E-Mail: karoline.krenn@fokus.fraunhofer.de

J. Tiemann

E-Mail: jens.tiemann@fokus.fraunhofer.de

1.1 Einleitung

In Deutschland ist der Begriff „Metadaten“ einer breiteren Öffentlichkeit seit den Snowden-Enthüllungen bekannt. Vorher gab es in Deutschland allerdings schon eine öffentliche Debatte über die Vorratsdatenspeicherung, dabei standen noch die Begriffe „Verbindungsdaten“ und „Verkehrsdaten“ im Vordergrund. Metadaten haben seitdem nicht nur im Kontext staatlicher Überwachung durch Behörden und Geheimdienste an öffentlichem Interesse gewonnen. Der Aufschwung datengetriebener Geschäftsmodelle weist ihnen eine fundamentale ökonomische Bedeutung zu. Der kommerzielle Einsatz von Metadaten beispielsweise für personalisierte Werbung oder Ansprache ist mittlerweile zu einem zentralen Wertschöpfungsfaktor von Unternehmen geworden. Vor dem Hintergrund von Antidiskriminierungsprinzipien und demokratischer Teilhabe geraten diese klassifizierenden Praktiken allerdings zunehmend in die Kritik.

Metadaten sind dabei aber schon immer ein Schlüssel für die Nutzung von Daten gewesen. Einfach ausgedrückt sind Metadaten Daten über Daten. Sie entstehen bei digitalen Prozessen gemeinsam mit Inhaltsdaten sowie bei der Verarbeitung von Inhaltsdaten. Sie reichern Daten mit Informationen an und werden konserviert (oder nicht).

Dieser Beitrag unternimmt den Versuch, eine sozial- und eine ingenieurwissenschaftliche Blickrichtung auf Metadaten in einem Beitrag zu integrieren. Dass dieses Unterfangen nicht so ganz einfach ist, hat sich schon bei der grundlegenden Frage offenbart, was eigentlich unter Metadaten zu verstehen ist. Da begriffliche Probleme im interdisziplinären Kontext nicht ungewöhnlich sind und ihre Beleuchtung unserer Meinung nach gerade dem Ziel des Sammelbandes folgt, wollen wir den ersten Abschnitt unseres Beitrages (den Herausforderungen) einer Begriffsklärung widmen. Bei unseren Betrachtungen stehen dabei zunächst Systeme digital vermittelter Kommunikation im Vordergrund. Unsere Sichtweise auf Metadaten lässt sich ohne Weiteres auf Systeme zur Verarbeitung von Daten verallgemeinern, wie überhaupt die Trennung zwischen Übertragung und Verarbeitung von Informationen in einer komplex digitalisierten Welt immer mehr verschwimmt.

Im weiteren Verlauf unseres Beitrages zeichnen wir an einem Beispiel aus dem Bereich Smart City einen konkreten Prozess der Metadatenbestimmung nach. Am Beispiel eines Community-basierten Funknetzes für das Internet der Dinge (IoT) zeigen wir, dass sich die Bestimmung von Metadaten und Daten im Prozess der Datenverarbeitung verschiebt, dass diese Daten durch Fiktionen generiert werden und dass für die Deutung von Metadaten Domänenwissen erforderlich ist. Wir argumentieren schließlich, dass die Nachnutzung von (Meta-)Daten eine Rekonstruktion ihres Entstehungskontexts erfordert und dass Metadaten diese Kontextualisierung nicht leisten können, sodass externe Werkzeuge hierfür notwendig werden. Mit der sozialwissenschaftlichen Datenkunde wollen wir abschließend ein solches Werkzeug zur Rekonstruktion von Kontext vorstellen. Jedoch kann selbst Datenkunde Verzerrungen nur identifizieren, letztlich kann sie aber das Problem der Modellifiziertheit von Daten nicht grundlegend lösen.

1.2 Eine interdisziplinäre Verortung von Metadaten

Wir beschäftigen uns im Engeren mit Metadaten, die digital in der Kommunikation mittels Maschinen erzeugt werden. Diese Metadaten können aus vier analytischen Blickwinkeln betrachtet werden. Die ersten drei Blickwinkel beschreiben sich bedingende Prozesse, die man sich behelfshalber als Abfolge vorstellen kann. Sie nehmen Metadaten von ihrer Rolle bei der technischen Hervorbringung von Daten über ihre Rolle bei der Aufbereitung von Daten zur Speicherung und schließlich bis zu ihrer Rolle bei der Verwertung von Daten ins Visier. Die abschließende vierte Perspektive nimmt eine soziologische Beobachtungsperspektive zu diesen Prozessen ein.

In (nicht notwendiger) Abfolge betrachtet entstehen Metadaten zunächst in den technischen Systemen („Maschinen“), welche Daten generieren und verarbeiten. Zum Zweiten sind sie ein zentrales Element bei der Datenaufbereitung innerhalb von Systemen, wo Daten für ihre weitere Nutzung abgelegt werden. Und drittens sind sie aus einer Verwertungsperspektive relevant für die zentralen Selektionsprozesse in allen weiteren Verwertungskontexten, in Wissenschaft und Forschung ebenso wie in praktischen Anwendungen.

Aus einer soziologischen Beobachtungsperspektive sind Metadaten in allen drei Phasen ein soziotechnisches Phänomen, also das Ergebnis soziotechnischer Interaktionen, die festlegen, was im Weiteren als ein (Meta-)Datum Geltung hat.

Je nachdem, nach welchem Filter man das Bild nun scharf stellt, lässt sich die Frage nach Metadaten unterschiedlich beantworten. Metadaten zu erzeugen, bedeutet in jeder Phase etwas anderes. Die analytische Trennung dieser Blickwinkel ist dabei ein hermeneutisches Werkzeug, um die Begriffsverwendung zu schärfen und Bedeutungsverschiebungen beobachtbar zu machen. Im konkreten Prozess der Herstellung, im Betrieb von Maschinen und bei der Verwendung von Metadaten greifen diese verschiedenen Perspektiven beharrlich ineinander. Wir wollen sie zunächst im Einzelnen betrachten, um anschließend klarzustellen, was wir im weiteren Verlauf unseres Beitrages unter Metadaten verstehen wollen.

Vorweg wollen wir noch darauf hinweisen, dass wir in der Darstellung auf den Gebrauch der Bezeichnungen „Rohdaten“ und „unstrukturierte Daten“ verzichten. Diese werden häufig verwendet, um den durch Metadaten angezeigten Prozess der Transformation von Rohdaten in strukturierte Daten zu beschreiben. Diese Bezeichnungen suggerieren dabei, dass Daten zu einem Zeitpunkt einen „reinen“ Ursprungszustand vollständiger Information abbilden. Wir teilen diese Ansicht nicht und folgen hier der These der Modellifiziertheit von Daten (vgl. Pohle 2016; Steinmüller 1993). Daten sind immer schon das Ergebnis von theoriegeleiteten Transformationen (vgl. Gitelman 2013).

1.2.1 Metadaten beim Betrieb technischer Systeme – die ingenieurwissenschaftliche Sicht

Aus einer technischen Perspektive entstehen Metadaten durch „Maschinen (in) der Kommunikation“, also durch die Übertragung bzw. allgemeiner durch die Verarbeitung von Daten mittels digitaler Maschinen. Zum Zweck der Übertragung von Daten müssen den eigentlichen Inhalten der Kommunikation weitere Daten hinzugefügt werden, beispielsweise Adressen der Kommunikationspartner, Daten für Fehlerprüfung oder -korrektur und Zustandsinformationen über gegenwärtige technische Kommunikationsbeziehungen. In Abgrenzung zu den eigentlichen Nutzdaten werden diese zusätzlichen Daten von Technikern zusammenfassend als Overhead bezeichnet. Es handelt sich dabei um eine Quelle für Metadaten, da sie den Zustand von Maschinen repräsentieren und Informationen zu Datenströmen geben. Diese Daten sind stark von den Eigenschaften des genutzten technischen Systems abhängig. Techniker entwerfen ein Kommunikationssystem anhand von definierten Anforderungen, haben dabei allerdings auch einen Gestaltungsspielraum für die konkrete Realisierung des Systems. Geht man vom Vorhandensein einer breit verfügbaren Kommunikationsinfrastruktur wie dem Internet aus, so ist dadurch für ein darauf aufbauendes Kommunikationssystem bereits eine Reihe von Eigenschaften festgelegt. Die eigentlichen Inhalte können auf Basis der vorhandenen Infrastruktur allerdings mithilfe von unterschiedlichen Mechanismen übertragen werden, die jeweils während der Übertragung weiteren Overhead erzeugen und damit eine zusätzliche Quelle für Metadaten darstellen. Durch die Gestaltung der Übertragung und Weiterverarbeitung von Nutzdaten wird also von Technikern bereits im Kommunikationssystem angelegt, welcher und wie viel Overhead bei jeder digitalen Kommunikation zwangsläufig anfällt.

Wir nutzen im Weiteren den Begriff „Overhead“ für diese durch Digitaltechnik entstehenden Daten, um ganz unterschiedliche Datenarten von fundamental verschiedenen Kommunikationsinfrastrukturen (bspw. leitungsvermittelterm Telefonnetz, paketvermittelterm Internet) und aus sich stark unterscheidenden Protokollschichten mit ihren jeweils speziellen Aufgaben zusammenzufassen. Ein zentrales Merkmal des Overheads ist, dass er zunächst in Maschinen entsteht und auch zum großen Teil in den verteilten Systemen der Kommunikation verbleibt. Beispielsweise werden Strukturinformationen zur Abgrenzung der einzelnen Nutzdaten oder Prüfsummen erzeugt, die nur für die konkrete Übertragung von Bedeutung sind und danach wieder von den technischen Mechanismen entfernt werden. Andere Teile des Overheads, wie Adressen der Kommunikationsteilnehmer oder der Zustand von Kommunikationskanälen, können für eine spätere Auswertung auch außerhalb des Systems von Interesse sein (bspw. Speicherung in Logfiles). Die Techniker entscheiden dabei darüber, welche Metadaten nutzbar gemacht werden, an andere Teile des Systems weitergegeben oder aufgezeichnet werden (vgl. Mayernik und Acker 2018).

Der Overhead ist eine notwendige Voraussetzung für die Übertragung der Nutzdaten und enthält daher Informationen über die Kommunikation, die eigenständig und ohne die

Kenntnis der Nutzdaten innerhalb der Kommunikationssysteme zur Steuerung einzelner Funktionen genutzt werden, aber auch als Kontextinformation etwas über die Nutzdaten aussagen und so als Metadaten genutzt werden können. Aus technischer Sicht können Metadaten zur Diagnose zunehmend komplexerer technischer Systeme dienen, sie können auch zukünftig für die Selbststeuerung technischer Systeme genutzt werden (vgl. Kephart und Chess 2003).

1.2.2 Metadaten in der Datenaufbereitung – eine informationswissenschaftliche Sicht

Für ihre Nutzung müssen Daten gespeichert und aufbereitet werden. Durch den Einsatz computergestützter Erhebungsinstrumente werden oft bereits automatisch einheitlich formatierte und aufbereitete Datenbanken erstellt. Metadaten sind eine wesentliche Komponente in diesen Prozessen (vgl. Riley 2017). Die Beschäftigung mit der strukturierten Beschreibung von Daten gehört zur Kernaufgabe der Informationswissenschaften, die dazu mit verschiedenen Begriffen operiert. Die Strukturierung von Daten erfolgt nach bestimmten Systematiken, welche die Erst- und insbesondere die Nachnutzung von Daten erleichtern. Daten werden dafür nach ausgewählten Metadatenkonzepten unterschieden, verglichen, gelistet und gruppiert. Solche Metadatenkonzepte beinhalten die Klassifikationsprinzipien, nach denen Systeme der Datenspeicherung eingerichtet werden. Ein Metadatenschema ist ein übergeordnetes Beschreibungsmuster, das diese Konzepte in einen Zusammenhang setzt. Die Informationswissenschaft unterscheidet dabei verschiedene Metadatenarten: deskriptive (zum Dateninhalt), strukturelle (zum inneren Aufbau der Datenobjekte und ihrer Beziehungen) und administrative Metadaten (rechtliche, technische und die Provenienz betreffende Daten) (vgl. Riley 2017).

Die sozial- und humanwissenschaftliche Methodenliteratur beschäftigt sich zwar mit der Datenaufbereitung, jedoch nur wenig explizit mit dem Metadatenkonzept. Allgemein wird im Umgang mit Daten darauf hingewiesen, dass jede Datenaufbereitung einen Eingriff in die Daten darstellt und daher besondere Sorgfalt und Kontrolle erfordert (vgl. z. B. Döring und Bortz 2016). Durch ihre strukturierende Rolle gilt das unserer Ansicht nach für Metadaten in besonderem Maße. Wir lehnen uns dabei an die neu entfachte Diskussion über Datenqualität und Vergleichbarkeit von prozessproduzierten Massendaten im Kontext neuer Technologien an (vgl. Baur 2009).

In der Informationswissenschaft hat die Evaluierung von Metadaten eine lange Tradition, jedoch ohne dass sich die Disziplin über Gütekriterien einig wäre (vgl. Lee et al. 2015). Als abstrakte Kriterien werden in der Literatur am häufigsten Vollständigkeit, Akkuratheit und Konsistenz der Klassifikation genannt (vgl. Park 2009), andere Ansätze argumentieren dafür, dass die Qualität von Metadaten abhängig von der Zwecktauglichkeit sei (vgl. Guy et al. 2004). Im Weiteren herrscht Uneinigkeit darüber, wie die Güte von Metadaten methodisch überprüft werden könne. Die Entwicklung und Evaluation eines Metadatenschemas ist nämlich ein höchst komplexer Prozess (vgl. Lee et al. 2015).

Die Metadatenqualität wird besonders durch fehlende Datenquellen, Diskrepanz in der Kategorisierung unter verschiedenen Fachkräften und divergierende Granularitätstiefen beeinträchtigt. Wichtig im Zusammenhang mit der Datenaufbereitung ist also die Dokumentation von relevanten Informationen (vgl. Bose und Frew 2005). Qualitätssicherung wird letztendlich im Zusammenhang mit der Verwertung der Daten relevant und stellt damit eine Schnittstelle zur nächsten Perspektive her.

1.2.3 Metadaten in der Verwertung und Analyse

Bei der Datenverwertung steht die zielgerichtete Deutung der Daten im Vordergrund. Metadaten dienen dabei der inhaltlichen Begründung für die Verknüpfung von Variablen und die daraus gezogenen Schlussfolgerungen. Dies geschieht (im Idealfall) in Auseinandersetzung mit der Datenerhebung und Datenaufbereitung. Begibt man sich auf die Suche nach dem Metadatenkonzept in der sozialwissenschaftlichen Methodenliteratur, findet man wenig. In der Praxis wird diesen abgeschlossenen Prozessen offenbar wenig Aufmerksamkeit zuteil, und die Datenqualität wird häufig als selbstverständlich betrachtet. Dies gilt vor allem in nichtwissenschaftlichen Anwendungskontexten, aber auch für die wissenschaftliche Praxis. Diese Nichtthematisierung steht in Widerspruch zur wichtigen Legitimationsfunktion von Metadaten für Entscheidungen im Prozess der Verwertung. Diese zentralen Aspekte können hier nur schlaglichtartig beleuchtet werden.

Datenverwertung folgt in der Regel einem kontextabhängigen Zweck. In praktischen Anwendungskontexten in Wirtschaft, Politik und Verwaltung sowie dem sozialen Leben sind auf Metadaten basierende Klassifikationen die Grundlage für diverse operative Prozesse. Seien es die Gewährung von Leistungen, die Anerkennung von Rechten, die Verteilung von Ressourcen, betriebswirtschaftliche Entscheidungen oder Steuerungsprozesse innerhalb von smarten Infrastrukturen, alle diese Prozesse verwenden metadatenbasierte Zuordnungen.

Die wissenschaftliche Verwertung von Daten verfolgt als Ziel den Erkenntnisgewinn und unterliegt dabei strengen Gütekriterien (vgl. Döring und Bortz 2016). Dieses Ziel erfordert methodische Strenge beim Untersuchungsdesign sowie bei der Operationalisierung, also der Übersetzung theoretischer Konzepte in empirisch messbare Größen. Die Qualitätssicherung für die Entwicklung von Messinstrumenten ist hierbei besonders wichtig.

Die Datenauswertung selbst baut auf den im Prozess der Datenaufbereitung getroffenen Reduktionen und Klassifikationen auf. Es können auch weitere Transformationen der Daten vorgenommen werden, beispielsweise die Heranziehung von Metadaten als Untersuchungsdaten. Bei der wissenschaftlichen Datenanalyse geht es entweder um eine Beschreibung, Erklärung oder Vorhersage empirischer Phänomene, um damit die Menge gesicherten Wissens zu steigern, oder auch um die Lösung praktischer Probleme (vgl. Döring und Bortz 2016). Dazu werden Daten mithilfe statistischer Verfahren in einen systematischen Zusammenhang gebracht.

Metadaten zu den erhobenen Daten spielen für die Deutung von Daten sowie die Zusammenführung mit den spezifizierten theoretischen Konzepten und damit für die Konstruktvalidität eine maßgebliche Rolle. Die Passung von Metadaten mit der Fragestellung bzw. dem Untersuchungsgegenstand ist Grundlage für die Datenqualität. In Methodenbüchern wird dieser Aspekt jedoch noch wenig oder nur im Zusammenhang mit der Datenerhebung expliziert, deren Güte über die methodische Strenge des Untersuchungsdesigns geprüft wird. Die Nutzung der aufbereiteten Daten ist dabei ein Randthema. Eine Ausnahme bildet hier die sozialwissenschaftliche Datenkunde, die sich in den 1980er Jahren systematisch mit Ursachen von Fehlern in prozessproduzierten Massendaten beschäftigt und dazu den Entstehungskontext der Daten untersucht hat (vgl. Baur 2009; Bick und Müller 1984). Allgemein weist die Methodenliteratur jedoch auf den invasiven Charakter der Datenaufbereitung hin und ihr daher besondere Sorgfalt und Kontrolle zu (vgl. Döring und Bortz 2016). Vorsätzliche Manipulationen am Datenmaterial zur Herstellung von „Wunschergebnissen“ werden generell als „gravierende Verletzung der Wissenschaftsethik“ betrachtet (Döring und Bortz 2016, S. 582).

1.2.4 Metadaten als Artefakte – eine (technik-)soziologische Sicht

Die (technik-)soziologische Sicht auf Metadaten nimmt den Prozess ihrer Fabrikation und ihrer Verwendung in der Datenverarbeitung in den Blick. Durch die Wahl einer beobachtenden und befragenden Methode rekonstruiert sie die sozialen Sinnzusammenhänge der epistemischen Praktiken des Messens (vgl. Heintz 2007), Quantifizierens (vgl. Diaz-Bone und Didier 2016), Kalkulierens und Klassifizierens (vgl. Bowker und Star 2000) und damit die soziale Konstruktion von Wissen und die Materialisierung von Daten. Zur systematischen Darstellung der mittlerweile höchst umfangreichen Literatur muss auf andere Quellen verwiesen werden. Neben der Soziologie der Konvention (vgl. Diaz-Bone 2015) und den Science and Technology Studies (vgl. Bauer et al. 2017) beschäftigen sich auch die Accounting Studies (vgl. Mennicken und Vollmer 2007) mit diesen Themen.

Über verschiedene Richtungen der Soziologie hinweg gibt es einen Konsens darüber, dass eine Verfügbarmachung und Objektivierung von Wissen durch die Produktion von Daten und Zahlen nicht vom Prozess der Bedeutungsgebung zu trennen ist. Die Entscheidung darüber, was ein Datum ist, wie es gesichert und aufbereitet und schließlich analysiert werden soll, ist demgemäß nicht eindeutig, sondern unterliegt Selektions- und Transformationsprozessen. Motive und Ziele der Messenden, die Auswahl der Messparameter und die Randbedingungen der Messung, wie Zeit, Ort, Umfang und Regelmäßigkeit, beeinflussen die Herstellung von Daten (vgl. Duttweiler 2018). Die Darstellung von Daten, Messungen und Berechnungen findet innerhalb gesellschaftlicher Interpretationsrahmen statt (vgl. Diaz-Bone und Didier 2016). Klassifikationsverfahren verdecken Heterogenität innerhalb von Kategorien und überbetonen sie zwischen Kategorien (vgl. Zerubavel 1991, 1996). Eindeutigkeit ist eine epistemische Fiktion, um Daten passend

zu machen. Die Verwendung kulturell vertrauter Steuerungs- und Bewertungssysteme (und Symbole) erzeugt dabei (eine selten hinterfragte) Evidenz. Die soziologische Kritik wendet sich hier vor allem gegen die Ausblendung der Komplexität und Kontingenz von Mess- und Rechenoperationen (vgl. Heintz 2007).

Der soziologische Blick richtet sich auf die Praxis der Erzeugung und Nutzung von Metadaten bei der Strukturierung von Daten (vgl. Passoth und Wehner 2018). Daten lassen sich in der einen oder anderen Weise unterscheiden, filtern, vergleichen und gruppieren. Wie eine Identifikation als Informationsobjekt vonstattengeht und welche Reduktionen diese Prozesse begleiten, liegt nicht allein an den intrinsischen Eigenschaften von Daten, sondern ist sozial (und kulturell) eingebettet. Die zur Beschreibung von Daten gewählten Begriffe stellen die Verbindungen zwischen so auch zugleich konstruierten empirischen Sachverhalten her und machen die Welt auf diese Art verstehbar (vgl. Baecker 2013). Wenn in Bezug auf Metadaten somit von Ordnungsvermutungen gesprochen wird (ebd.), ist die Vorstellung von Gesellschaft also immer Referenzobjekt. Um Metadaten soziologisch zu verstehen, ist die interpretative Leistung hinter diesem strukturierenden Vorgang zu rekonstruieren. Dazu wird angenommen, dass diese Prozesse von technischen und sozialen Fiktionen begleitet werden, durch welche Daten passend gemacht werden (vgl. Kalthoff 2011).

Über die sozial eingebetteten epistemischen Praktiken der Erzeugung von Metadaten hinaus wird der Einfluss von Metadatenkonzepten auf die Technikgestaltung betrachtet (vgl. Bowker und Star 2006; Star 1999). Die Festlegung von Metadaten strukturiert digitale Infrastrukturen und Netzwerkarchitekturen, noch bevor sie zum Standard geworden sind (vgl. Fidler und Acker 2017). Durch technische Infrastruktur wird Anschlussfähigkeit für die weitere Nutzung von Daten hergestellt (vgl. Manhart 2008). Damit schließt sich der Kreis zur technischen Perspektive auf Metadaten. Was allerdings für den Techniker ein unhinterfragter Output der Maschine ist, ist aus techniksoziologischer Sicht soziale Konstruktion, Produkt von Fiktionen und Ergebnis soziotechnischer Aushandlungsprozesse.

1.2.5 Transdisziplinäre Zusammenschau und Synthese

Maschinenvermittelte Kommunikation ist in der digitalen Gesellschaft ein alltägliches Phänomen geworden (vgl. Houben und Prietl 2018). Die prozessuale Komplexität dieser Vorgänge gerät dabei leicht aus dem Blickfeld. Sowohl aus ingenieur- wie auch aus sozial- und geisteswissenschaftlicher Sicht spielen Metadaten für die Informations- und Datenprozesse eine zentrale Rolle. Wir haben Metadaten aus verschiedenen Betrachtungswinkeln diskutiert, was hier noch einmal grafisch zusammengefasst wird.

Metadaten erzeugen, heißt aus jeder der beschriebenen Perspektiven und Phasen etwas anderes: 1) Die Gestaltung und der Betrieb von Maschinen erzeugen Metadaten. 2) Daten aufzubereiten, erzeugt Metadaten. 3) Metadatenerzeugen begründet Deutungszusammenhänge. 4) Metadaten sind Artefakte. Ihre Erzeugung ist das Ergebnis von Fiktionen innerhalb sozial und kulturell eingebetteter Prozesse, durch die Daten passend gemacht werden (Abb. 1.1).

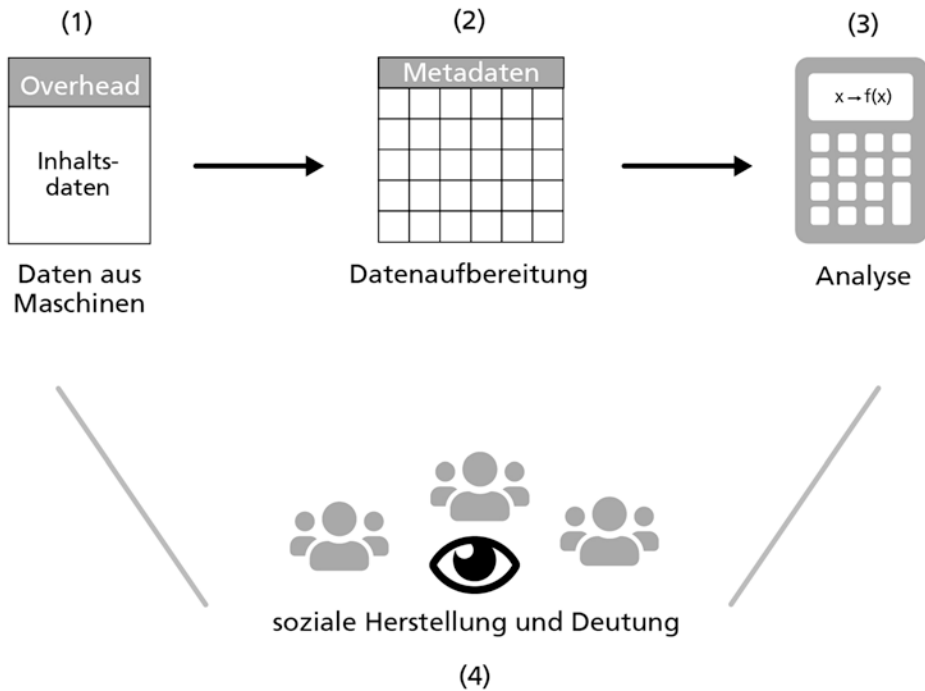


Abb. 1.1 Perspektiven auf Metadatenerzeugung. (Eigene Darstellung)

Eine multiperspektivische Betrachtung, wie wir sie hier entwickelt haben, lässt sich in drei Einsichten synthetisieren.

Metadaten sind ein zentrales Element der Wissensrepräsentation

Metadaten zu erzeugen, heißt Daten zu deuten und Deutungen zu materialisieren. Metadaten repräsentieren den Zustand von technischen Prozessen und stellen überhaupt erst die Kategorien bereit, nach denen Informationen über Daten gewonnen und aufbewahrt werden. Mit den Metadaten variiert das darin repräsentierte Wissen. Informationen können verloren gehen oder nachträglich hinzugefügt werden. Der technische Overhead, der aus Ingenieurperspektive als Nebenprodukt des Gerätebetriebs anfällt, kann zu einem Messdatum werden. Ebenso können inhaltliche Daten als ein Proxy für Metainformationen verwendet werden. Technische Infrastrukturen und Metadatenkonzepte verschmelzen in der Praxis (vgl. Fidler und Acker 2017). Wissensrepräsentation ist damit auch eine Frage der Technikgestaltung, und die Verantwortlichkeit dafür ist auch dort zu verorten. Aus soziologischer Perspektive ist Wissensrepräsentation nur eine soziale Konstruktion.

Die Unterscheidung zwischen Daten und Metadaten ist kontextabhängig

Wenn von Daten gesprochen wird, wird in der Regel auf den informativen Gehalt Bezug genommen, während Metadaten auf die Beschreibung von Daten abzielen. Allerdings

besteht kein grundsätzlicher Unterschied zwischen beiden. Die Unterscheidung kann bei der Erzeugung, Verarbeitung und Verwertung unterschiedlich ausfallen und ist abhängig vom Betrachtungswinkel derjenigen, welche die (Meta-)Daten für bestimmte Ziele nutzen (vgl. Mayernik und Acker 2018; Boellstorff 2013). Die Unterscheidung zwischen Daten und Metadaten in einem sozialen Verwendungszusammenhang ist „immer nur fallweise, vorübergehend und beobachterabhängig“ (Baecker 2013, S. 161). Eine Metadatenbestimmung nach Typen, wie wir sie in der informationswissenschaftlichen Literatur finden (vgl. Riley 2017; Mayernik und Acker 2018), ist nicht ausreichend, um die Bedeutungsverschiebungen und Transformationen, die Metadaten durchlaufen, einzufangen.

Metadaten sind mehrfach gedeutete „Deuter“

Metadaten erfahren im Prozess der Datenerzeugung, -aufbereitung und -verwertung mehrfache (Um-)Deutungen. Aus der Zusammenschau der Perspektiven scheinen Metadaten begründungsabhängig wie Daten selbst. Um diese in der digitalen Gesellschaft immer komplexeren Prozesse zu verstehen, ist umfangreiches technisches und prozessuales Domänenwissen notwendig, das die Eindeutigkeit und Treffsicherheit von Einordnungen beurteilen kann. Dieses Domänenwissen geht verloren, wenn es nicht in irgendeiner Form konserviert wird. Das wird vor allem in der Nachnutzung von Daten relevant, wo Metadaten der Schlüssel für ihre Deutung sind. Mit ihnen steht und fällt die Aussagekraft der Datenverwertung.

Im weiteren Verlauf dieses Textes wollen wir unter „Metadaten“ kontextabhängige Metainformationen zu Inhaltsdaten bzw. zu digitalen Prozessen verstehen, welche die Datengewinnung dokumentieren, die Erstellung von Datensätzen strukturieren und Daten kommentieren. Metadaten halten dabei den Daten zugeschriebene Bedeutungen fest.

Sie sind die Voraussetzung für eine Produktion und Interpretation von Daten, zugleich jedoch selbst Datum. Wie diese Verschiebungen zwischen Datum und Metadatum in einem Datenzyklus aussehen, wollen wir uns nun an einem Fallbeispiel anschauen.

1.3 Metadatenerzeugung am Fallbeispiel

Im folgenden Abschnitt möchten wir unsere bisher allgemeinen Betrachtungen zu Metadaten an einem konkreten Beispiel illustrieren. Dabei wollen wir aufzeigen, wie die verschiedenen Perspektiven auf Metadaten in einer konkreten Analyse aussehen. Unser Beispiel stammt aus dem Bereich Smart City, wo exemplarisch die zunehmende Durchdringung der physischen Welt mit Informationstechnik zu beobachten ist. Durch diese Entwicklung entstehen neue Datenquellen für praktische Steuerungsanwendungen

ebenso wie für wissenschaftliche Untersuchungen, wobei die Daten aber in einem komplexen Kontext entstehen. Das wollen wir am gewählten Beispiel aufzeigen.

Seit Kurzem etabliert sich eine neue Klasse von Funknetzen (LPWAN=Low Power Wide Area Network) speziell für das Internet der Dinge (IoT), angepasst an die besonderen Anforderungen in diesem Bereich (vgl. Tiemann und Mancke 2018). Diese Netze eignen sich ausgezeichnet zur Anbindung drahtloser Sensoren aller Art, mit denen (vorwiegend) urbane Räume ausgestattet werden. Eine wesentliche Besonderheit ist, dass es erstmals auf unkomplizierte Weise möglich wird, (regional) flächendeckende Mobilfunkinfrastrukturen durch kleinere Unternehmen oder sogar durch private Initiativen aufzubauen. Die Konsequenzen für die Gewinnung von (Forschungs-) Daten sind vielfältig: Grundsätzlich steigt die Zahl möglicher Datenquellen durch das Vorhandensein von umfangreicher Sensorik an, gleichzeitig ist es aufgrund der Vielzahl von Netzen schwer, ein vollständiges Bild zu gewinnen (zumindest während der initialen Wachstumsphase bis zu einer Konsolidierung). Dafür finden sich unter den verschiedenen Infrastrukturen auch Netze und Plattformen, die für unmittelbare Anwendungszwecke und diverse Nachnutzungen von Daten einen einfachen Zugang gewähren.

Unser Beispiel veranschaulicht die Erzeugung von Metadaten im The Things Network (TTN), einer Community-basierten Initiative mit dem Ziel eines globalen LPWAN-IoT-Netzwerks (vgl. Riemann 2019; <https://www.thethingsnetwork.org/>). Die aufwendige Abdeckung mit Funkbasisstationen (Gateways) wird hierbei von einzelnen Privatpersonen, Hochschulen und Unternehmen übernommen. Den Möglichkeiten der kooperativen Nutzung eines flächendeckenden Netzes stehen nur geringe Hardware-Kosten und geringe Anforderungen an die Mitnutzung des meist bereits vorhandenen Internet-Anschlusses gegenüber.

Vor der Nutzung einer vernetzten Anwendung muss sichergestellt sein, dass eine Netzverfügbarkeit gegeben ist. Im Gegensatz zu kommerziellen Mobilfunknetzen, deren Anbieter detaillierte Karten zur Netzabdeckung bereitstellen, sind Aussagen über die lokale Verfügbarkeit eines Community-basierten Netzes nur begrenzt möglich. Während des Aufbaus der Infrastruktur ist offen, ob und wann an einem Ort das Funknetz verfügbar wird, zudem können jederzeit einzelne Komponenten wieder abgeschaltet werden. Der Dienst TTN-Mapper (<https://ttnmapper.org>) stellt die verfügbare Netzabdeckung auf Basis von Messungen dar. Bei dem Dienst handelt es sich wiederum um eine private Initiative, die mit Datenspenden auf freiwilliger Basis und mithilfe von unkalibrierten Messinstrumenten arbeitet.

Aus dem Kontext von Funkübertragungen entstehen im The Things Network Metadaten, die der Dienst TTN-Mapper dazu nutzt, eine Karte der Netzabdeckung zu erzeugen. Die Erzeugung von Metadaten soll anhand der in Abb. 1.1 herausgearbeiteten Perspektiven anschaulich gemacht werden.

1.3.1 Metadaten durch den Betrieb von Maschinen

Abb. 1.2 (unten) zeigt ein LoRaWAN-Datenpaket, wie es vom The Things Network empfangen wurde. Das Datenpaket enthält eine Nutzdateninformation („Payload“) von 15 Byte, die hier sowohl kodiert („9E 1B ...“) als auch nach verschiedenen Inhalten dekodiert („Fields“) dargestellt wird. Die Nutzdaten dieses Datenpakets enthalten Ortsinformationen eines GPS-Empfängers („latitude“, „longitude“), aber auch Informationen zum Zustand der Stromversorgung („battery_level“) und zur Temperatur innerhalb des Geräts.

time	counter	port	dev id: ftd-abp	payload: 9E 1B 52 31 57 70 01 31 89 10 24 E5 00 0F 7E
14:47:09	230	1		

Uplink

Payload

9E 1B 52 31 57 70 01 31 89 10 24 E5 00 0F 7E
📄

Fields

```
{
  "battery_level": 3966,
  "dl_counter": 0,
  "gps_quality": "Average",
  "hdop": 2,
  "lati_hemisphere": "North",
  "latitude": 52.52628333333333,
  "latitude_dmm": "N 52°31.577",
  "long_hemisphere": "East",
  "longitude": 13.315166666666666,
  "longitude_dmm": "E 013°18.91",
  "payload": "9E1B523157700131891024E5000F7E",
  "sats": 4,
  "temperature": 27,
  "ul_counter": 229
}
```

Metadata

```
{
  "time": "2019-01-17T13:47:09.389374266Z",
  "frequency": 868.3,
  "modulation": "LORA",
  "data_rate": "SF7BW125",
  "coding_rate": "4/5",
  "gateways": [
    {
      "gtw_id": "eui-7276ff00b030c96",
      "timestamp": 3914042092,
      "time": "",
      "channel": 6,
      "rssi": -94,
      "snr": 8.2,
      "rf_chain": 1,
      "latitude": 52.52603,
      "longitude": 13.314334,
      "altitude": 65
    }
  ]
}
```

Abb. 1.2 LoRaWAN-Datenpaket mit Metadaten. (Ausschnitt aus TTN Console, <https://console.thethingsnetwork.org/>)

Damit diese Nutzdaten überhaupt übertragen werden konnten, mussten verschiedene technische Mechanismen genutzt werden, die in Form von Metadaten („metadata“) ihre Spuren hinterlassen. Der Satz von Metadaten besteht zunächst aus einem Zeitpunkt („time“), zu dem das Paket übertragen wurde, der Frequenz des Übertragungskanals („frequency“) sowie verschiedenen Angaben zum Übertragungsverfahren. Diese Metadaten sind im Aussenden des Datenpakets begründet. Das Paket kann dann von mehreren Gateways empfangen werden, was jeweils wieder zur Generierung von Metadaten führt, diesmal einem Satz von Metadaten für jedes Gateway (innerhalb der Struktur von „gateways“). Zu diesen Metadaten eines Gateways gehört neben der Identität („gtw-id“) bspw. der Ort des Gateways („latitude“, „longitude“) und die Signalstärke („rssi“) beim Empfangen des Datenpakets. Da sich aus diesen Metainformationen der Aufenthaltsort des Senders abschätzen lässt, werden diese technischen Informationen ggf. innerhalb der Netzinfrastruktur analysiert sowie dem Nutzer der Übertragung zur Verfügung gestellt.

1.3.2 Datenaufbereitung und Nutzung eines Metadatenschemas

Um eine Karte der Netzabdeckung zu generieren, müssen die oben dargestellten Nutz- und Metadaten zuerst in eine Datenbank überführt werden. Im Rahmen der Datenaufbereitung werden dazu die Nutzdaten eines Positionssensors (GPS-Empfängers) mit den Metadaten der Signalstärke an den umliegenden Gateways in einer Datenstruktur gespeichert. Weitere Datenfelder eines Datensatzes sind bspw. die Identität des Senders bzw. einer Messkampagne oder zeitliche Angaben, damit man Messungen aufgrund ihres Alters bewerten oder im Nachhinein als fehlerhaft erkannte Messungen entfernen kann.

Vom Dienst TTN-Mapper wird auch die Genauigkeit der Positionsmessung erfasst. Dafür sind verschiedene Möglichkeiten zugelassen, die Streuung der Positionswerte („hdop“, vgl. Abb. 1.2), die Genauigkeit in Metern (wird vom obigen Sensor nicht erfasst, er bietet als weiteres Maß „gps_quality“ an) oder als minimale Option die Anzahl der sichtbaren Satelliten („sats“) (<https://ttnmapper.org/faq.php>).

Um die Messungen zu speichern, braucht es ein Ordnungsschema, das einerseits aus technischen Gründen neue Metadaten erzeugt (wie Nummerierungen, Zeitstempel), andererseits als Verweis auf Datensätze selbst Metadaten darstellt. Abb. 1.3 zeigt

ID	Zeit („time“)	...	Gateway („gtw_id“)	Signalstärke („rssi“)	...	Ort („latitude“, „longitude“)	Genauigkeit („hdop“, „sats“)	...
...	2019-01-17 14:47:09		eui-7276ff000b030c96	-94.00		52.526283, 13.315167	2, 4	

Abb. 1.3 Beispielhafte Speicherung der Messwerte in einem Datenschema. (Eigene Darstellung)

dieses Ordnungsschema beispielhaft als Datentabelle zur Speicherung der Kontextinformationen einer Messung. Voraussetzung für die spätere Nutzung der Daten ist damit die Übereinkunft, welche Daten gespeichert werden und welche Information auf diese Weise durch die Datenstruktur impliziert im Datensatz enthalten sind.

1.3.3 Analyse und Verwertung

Eine mögliche Verwertung dieser Daten ist die Erzeugung von Karten. Dabei wird der Datensatz in eine grafische Darstellung transformiert, bei der komplexe Messergebnisse leicht erfassbar werden. Es werden neue Variablen gebildet, in denen (Mess-)Daten gruppiert werden. Beispielsweise enthält der Datensatz zu den einzelnen Messreihen Informationen zur Signalstärke, mit dem ein Messsignal an einem Gateway empfangen wurde. Im Zuge der Auswertung werden nun verschiedene Abstufungen von Signalstärke in Farben übersetzt. Die Darstellung eines Messpunkts ist aber nur sinnvoll, wenn die Ortsinformation aus der Messung einer erwarteten Genauigkeit entspricht und das empfangende Gateway noch an seinem angestammten Ort verfügbar ist. Durch Verknüpfungen auf der Basis der in der Datenaufbereitung erhaltenen Metadaten werden neue Daten erzeugt. Diese Metadaten entscheiden, ob und mit welcher Gewichtung die Messinformation in die finale Darstellung einfließen soll.

Die erstellte Karte kann nun genutzt werden, um die Netzabdeckung am Ort einer geplanten oder möglicherweise gestörten Nutzung einer vernetzten Anwendung abzuschätzen. In vielen Fällen wird die Karte in einem Bereich nicht flächendeckend sein, sondern weiße Flecken enthalten. Die Kartendarstellung hilft mit ihrer räumlichen Darstellung dem Betrachter bei der Interpolation, von umliegenden bekannten Netzzuständen auf einen bisher nicht vermessenen Ort zu schließen.

Abb. 1.4 zeigt in Form einer Heatmap die Netzabdeckung durch das The Things Network in einem Teil von Berlin-Charlottenburg. Eine gute Funkversorgung ist durch rote Farbe gekennzeichnet, über Farbabstufungen bis zu Dunkelblau werden schlechtere Netzabdeckungen dargestellt. Alle derartigen Punkte der Heatmap beruhen auf Messungen an diesen Orten. Kleine Kreise mit dem TTN-Logo markieren die Standorte von Gateways, zu erkennen sind rechts im Bild drei Gateways bei Fraunhofer FOKUS in der Kaiserin-Augusta-Allee.

1.3.4 Soziologische Beobachtung des The Things Networks

Hier geht es um die Fiktionen, welche die beschriebene Erzeugung und Deutung von Metadaten begleiten. Wir betrachten dazu den sozialen Entstehungskontext der TTN-Daten. Da Fraunhofer FOKUS und einzelne Mitarbeiter Teil der TTN-Community sind,