

Matthias Grotkopp

## **FORSCHUNGSDATEN QU'EST-CE QUE C'EST?– Matthias Grotkopp über eine ‚Politik der kleinen Schritte‘ auf dem Weg zu open data und qualitativ- empirischer Filmanalyse**

2019

<https://doi.org/10.25969/mediarep/20261>

Veröffentlichungsversion / published version  
Zeitschriftenartikel / journal article

### **Empfohlene Zitierung / Suggested Citation:**

Grotkopp, Matthias: FORSCHUNGSDATEN QU'EST-CE QUE C'EST?– Matthias Grotkopp über eine ‚Politik der kleinen Schritte‘ auf dem Weg zu open data und qualitativ-empirischer Filmanalyse. In: *Open-Media-Studies-Blog*, Jg. 2 (2019). DOI: <https://doi.org/10.25969/mediarep/20261>.

### **Erstmalig hier erschienen / Initial publication here:**

<https://mediastudies.hypotheses.org/1415>

### **Nutzungsbedingungen:**

Dieser Text wird unter einer Creative Commons - Namensnennung - Weitergabe unter gleichen Bedingungen 4.0/ deed.de Lizenz zur Verfügung gestellt. Nähere Auskünfte zu dieser Lizenz finden Sie hier:

<https://creativecommons.org/licenses/by-sa/4.0/deed.de>

### **Terms of use:**

This document is made available under a creative commons - Attribution - Share Alike 4.0/deed.de License. For more information see:

<https://creativecommons.org/licenses/by-sa/4.0/deed.de>

## FORSCHUNGSDATEN QU'EST-CE QUE C'EST?– Matthias Grotkopp über eine ‚Politik der kleinen Schritte‘ auf dem Weg zu open data und qualitativ-empirischer Filmanalyse

VERÖFFENTLICHT 27/11/2019 · AKTUALISIERT 04/06/2020



Schon ein erster vorsichtiger Versuch der Recherche macht jedem deutlich, dass Forschungsdatenmanagement, *open data* und Geistes- und Medienwissenschaften bisher weniger eine erfolgreiche Triangulation als vielmehr ein Bermuda-Dreieck der guten Absichten bilden. Das hat verschiedene Gründe, die bereits Gegenstand dieses Blogs gewesen sind – und zwar unter anderem bei [Pause und Walkowski](#), [Hirsbrunner](#) und [Sondervan](#) – und die an dieser Stelle gar nicht weiter erläutert werden sollen. Eine gesunde Skepsis gegenüber bestimmten Fach- und Sachzwängen derjenigen Wissenschaften, die durch intrinsisch angeschlossenes Anwendungs- und Monetarisierungspotential gekennzeichnet sind, sieht sich auch immer wieder an einzelnen Auswüchsen bestätigt:

*Abb. 1: Screenshot der Seite [Blockchain for Science](#) (einem Link im [Beitrag von Simon Hirsbrunner](#) folgend).*

Das Folgende hat dann auch eher den Charakter eines **Erfahrungsberichts**, in dem es darum gehen soll, wie wir<sup>1</sup> eine filmanalytische Methode – die das Digitale zunächst vor allen Dingen als Arbeitsumgebung betrachte – sukzessive auf die dabei anfallenden Daten und das Potential ihrer systematischen und nachhaltigen Aufarbeitung im Sinne offener Forschungsdaten abgeklopft haben.

### eMAEX als filmanalytische Methode

Im Zentrum des [eMAEX-Verfahrens](#) steht die **qualitativ-empirische Deskription der expressiven Dimension audiovisueller**

**Bewegungsbilder.**<sup>2</sup> Wenn man hier also nach den primären Forschungsdaten fragte, so traf man gleich auf zwei sehr fundamentale Probleme, die sich z.B. in den Sozial- und Naturwissenschaften in dieser Art nie stellen:

- Auf der ersten Ebene sind die **Primärdaten** die Filme selbst, was zum einen die (in der Praxis selten gestellte) Frage nach der Provenienz ihres Digitalisats und viel häufiger die Frage nach den Urheberrechten aufwirft. Ersteres lässt sich nur durch einen [hohen technischen und personellen Aufwand](#) befriedigend einhegen, letzteres höchstens dann völlig umgehen, wenn man sich auf gemeinfreie Medien beschränkt, was nur für sehr bestimmte Fragestellungen (Netzaktivismus oder frühe(re) Filmgeschichte) praktikabel ist.
- Auf einer zweiten Ebene gibt es so etwas wie „die Primärdaten“ gerade nicht als „Daten“, da das primäre Datum eben doch nicht das filmische Bild als Artefakt ist – sei es analog oder digital – sondern der Film als **das gestaltete Wahrnehmungserleben** eine\*r Zuschauer\*in. Denn im Kern zielt der filmanalytische Zugriff auf das affizierende Potential der Verschränkung aus der kompositorischen Ordnung eines Bewegtbildes einerseits mit der Zeit seines Wahrgenommen-Werdens andererseits. Audiovisuelle Daten haben hier nicht den gleichen epistemischen Status wie etwa bei Zeitzeugenaufnahmen in der *Oral History* oder bei Interventionen in der *Visual Anthropology*, wo sie in der Forschungspraxis selbst produziert oder kuratiert und archiviert werden.

Es bleiben also die Sekundärdaten, die bei der Verarbeitung von Filmen und Videos – als digitale, einheitlich formatierte – Sichtungsdateien anfallen können. (Was an dieser Stelle außer Acht gelassen wird, sind all jene Metadaten, für die Filme zunächst eine *black box* sind und die sich auf die Fragen nach dem Wer, Wo, Wann, nach den ökonomischen Kennzahlen oder den [Datenspuren der Filmzirkulation](#) beziehen. Für diese gibt es ganz andere Potentiale und ganz andere Probleme und Problemlösungen.)

Die grundlegenden Formen, in denen die subjektivierende Erfahrungsdimension von audiovisuellen Bewegungsbildern Gegenstand einer Objektivierung werden, sind nur allzu bekannt und auch einem vortheoretischen Verständnis intuitiv zugänglich: Die zeitliche **Segmentierung**, die Aufschlüsselung durch technische Parameter in **Protokollen/Annotationen**, die begrifflich genaue, sprachliche **Beschreibung**. Darauf aufbauen leiteten sich die vier grundlegenden Datentypen herleiten, die wir genutzt haben:

1. Der scheinbar triviale Aspekt der **Segmentierung**, die Einteilung von Filmen in einzelne zeitliche Einheiten steht daher am Ausgangspunkt der eMAEX-Methode. Dass das Konzept der „Szene“ als kompositorische und narrative Sinneinheit dabei ebenso evident ist, wie es sich in ihrer informationstechnologischen Modellierung letztendlich als sehr komplex erweist, ist dabei erstmal nur konstatiert. Wichtiger ist es, dass diese basale Segmentierung der notwendige Ausgangspunkt ist, um die Analyse der zeitlichen Entfaltung des audiovisuellen Gegenstands für weitere Darstellungsformen zugänglich zu machen. Ein Schritt hin zu der Ausgestaltung der Segmentierung zu einem Datensatz qualitativ empirischer Forschung war dann die **Klassifizierung und Qualifizierung** der Szenen. Diese leitet sich jeweils aus den zentralen Fragestellungen der Projekte her, die auch den jeweiligen Korpus begründen. Am stärksten systematisiert und für eine größere Anzahl an Filmen implementiert wurde das [in zwei Projekten zum Hollywood-Kriegsfilm](#), aber auch für das deutsche Genrekino der 1950er oder [den sog. „Deutsch-Türkischen“ Film](#) der 1990er-2000er Jahre wurden Klassifizierungssysteme entwickelt. Das ursprüngliche Datenformat war hier zunächst schlicht eine xls.Datei.
2. Im Rahmen der Segmentierung und Klassifizierung haben wir (auf unorthodoxe Art, zugegebenermaßen) die Intercoder-Reliabilität überprüft. Als Nebenprodukt dieser Überprüfung entstand eine einfache grafische Visualisierung der klassifizierten Segmentierung, die für die weiteren Projektzwecke zu einer der wichtigsten Mittel für die Exploration des Korpus und für die Explikation und Verdeutlichung detailanalytischer Ergebnisse und Vergleichsstudien wurde.

Abb. 2: Arbeitsversion der visuellen Darstellung eines Vergleiches der Klassifizierungen eines Films durch verschiedene Projektmitarbeiter in der Entwicklungsphase von eMAEX.

1. Die **Annotation** der Filme erfolgten auf Basis der ‚handelsüblichen‘, oftmals aus herstellungstechnischen und -praktischen Parametern hergeleiteten Benennungs- und Beschreibungsebenen der multimodalen Komposition audiovisueller Bilder (Einstellungsgrößen, Schnitt, Ton, Figurenchoreographie etc.). Es ist wahrscheinlich kein Zufall, dass diese Ebene als tatsächlich formalisierbare und systematisierbare Beschreibungsmatrix in ihrer klassischen, analogen Erscheinungsform als Einstellungsprotokoll bisher eher der filmwissenschaftlichen Propädeutik zugeschlagen wird, handelt es sich doch um eine Prozedur, bei der schon der minimalste Anspruch auf Vollständigkeit und Intersubjektivität zu sehr hohem Arbeitsaufwand führt. Zwei Gründe gaben für uns den Ausschlag, diesen Arbeitsschritt zu einer zentralen Säule unserer Datenstruktur zu machen: Erstens die Qualitätsänderung, die sich mit **open source Annotationsprogrammen** wie [ELAN](#) und [Advene](#) ergibt, die Annotationen qua *timecode* mit der Videodatei verknüpfen und darstellbar machen. Zweitens sollte auf dieser Datenbasis evident werden, was eine der filmtheoretischen Prämissen war, die in den Projekten untermauert werden sollte: nämlich die kompositorische Binnenstruktur von Filmszenen durch Ausdrucksbewegungseinheiten, die eben nicht mehr narrativ, sondern expressiv und – wenn man so will – formalästhetisch begründet ist. (Davon abgesehen schien es uns ab einem bestimmten Punkt denkbar – auch wenn

die Realität den Erwartungen noch deutlich hinterher hinkt –, dass es für bestimmte Parameter tatsächlich **Potentiale automatischer und semi-automatischer Erkennen** geben könnte, die für zukünftige Projekte den Arbeitsaufwand reduzieren könnten, deren Rahmen sich aber nicht aus dem „bloß machbaren“, sondern aus dem für die filmwissenschaftliche Erfassung sinnvollen herleiten sollte.)

2. Kern- und Endpunkt der Systematik war aber immer noch die Textproduktion in Form einer **dichten Beschreibung** der Szene und ihrer Sub-segmente, ihren Ausdrucksbewegungseinheiten. (Dass in manchen Fällen, das muss man eingestehen, „dicht“ ein Euphemismus ist, gehört zu den Zugeständnissen an die experimentelle, explorative Natur dieser Methodenentwicklung).

## eMAEX als Datenpublikation

Dies also waren – in groben Zügen – die Datenformate, die in eMAEX auf dieser Entwicklungsstufe angefallen sind. Aber was für ein Datenmanagement wurde entwickelt und was für eine Qualität des „open“ haben wir erreicht?

Dazu muss man sagen, dass die primäre Projektarbeit nicht über Webstrukturen sondern über eine Datenserverstruktur organisiert war, die schließlich in eine [Webdarstellung](#) überführt wurde:

- 
- 

Abb. 3-5: Screenshots der Webpräsentation der Datenmatrix „Affektmobilisierung im Kriegsfilm“

Diese Webdarstellung präsentiert die Datenbank des Projekts – aber als eine **fixe CMS-Architektur**, d.h. als eine **open access Publikation**, in der die einzelnen Ergebnisse strukturiert zugänglich gemacht sind, die jedoch nicht für die Nachnutzung, Erweiterung oder Weiterverarbeitung der Daten in strengem Sinne gedacht war. Dahinter stand die Absicht, dass der Zugang zu den Projektergebnissen strukturiert und gerahmt werden sollte durch die entwickelten Konzepte und die methodische Systematik. Die Filme und die Beschreibungen der Szenen sowie der Ausdrucksbewegungseinheiten wurden zugänglich gemacht über ihre zeitliche Segmentierung, über die Klassifizierungen und über die Visualisierungen des zeitlichen Verlaufs. Die Vorgaben des CMS haben allerdings auch dazu geführt, dass die Annotationsdateien, da sie nicht im CMS darstellbar waren, nicht zu dem publizierten Datenset gehörten, sondern retrospektiv Hilfsmittel bei der Mikrosegmentierung und der Erstellung der Beschreibungstexte blieben – die aktuelle Fortentwicklung von eMAEX hat als ein Ziel, genau an dieser Stelle anzusetzen und die Möglichkeiten der Publikation komplexer Annotationsdatensätze zu erproben (mehr dazu weiter unten). Außerdem wurden die Metadaten zu den Filmen und Szenen jeweils händisch in die dafür eigens entwickelten Seitenvorlagen des CMS übertragen, anstatt sie auf Basis eines maschinenlesbaren Schemas ex- und importieren zu können.

Als eine erste Reaktion darauf, dass das CMS zwar die Struktur der Datenbank abbilden konnte, aber nicht selber als semantische Datenstruktur fungiert und nicht für die Ablage beliebiger Dateiformate geeignet war, haben wir als Experimentalfeld die Datenmatrix des Kriegsfilms in ein [Semantic Mediawiki](#) überführt und dabei auch versucht, die Nutzungsmöglichkeiten für weitere Projekte zu erproben. (Das ursprüngliche Nutzungsszenario war in einem interdisziplinären Großprojekt angesiedelt, dessen Nichtverlängerung dem allerdings in die Quere kam.)

Das einfache Grundprinzip der Wiki-Technologie in der Kombination mit maschinenlesbaren, semantischen Verknüpfungen schien uns für die alltägliche Nutzung entgegenzukommen: Jeder Filme, jede Szene, jedes Sub-Segment wird als Seite erstellt und bekommt die relevanten Metadaten und Klassifizierungen als Eigenschaften zugeordnet, durch die sie dann durchsuchbar und sortierbar sind.

Abb. 6: Zwei Screenshots aus dem Semantic Mediawiki unseres Projektverbunds

Bisher sind neben dem Kriegsfilm nur wenige Projekte darin vertreten und aufgrund der Vorläufigkeit der dort erstellten Strukturen noch nicht für die Forschungsöffentlichkeit freigeschaltet. Dies war geplant für diesen Herbst und sollte in diesem Beitrag angekündigt werden, wird sich aber leider noch in seiner Umsetzung verzögern. Dann soll das Wiki als offene, frei zugängliche und frei nutzbare Plattform fungieren.

## Work in progress

Das langfristige Ziel ist es, dieses Prinzip der **Aufbereitung und Veröffentlichung von Forschungsdaten** für eine größere Anzahl an Projekten auszubauen und zur Verfügung zu stellen bzw. auch in andere, geeignete Plattformen zu integrieren. Die Phase einer

weitreichenden **Vernetzung und Kooperation** sollte nun dringend an die Stelle der zunächst auf die eigenen Fragestellungen und Bedürfnisse zugeschnittenen Entwicklungsarbeiten folgen (das adressiert zunächst uns selbst und unsere Arbeitsschritte, ist aber auch ein Aufruf an die Fachkolleg\*innen, die sich in diesem Feld betätigen).

In dem laufenden BMBF-geförderten Projekt [Affektrhetoriken des Audiovisuellen](#) aus unserem Projektverbund wird aktuell auch daran gearbeitet, das Datenformat, das bisher die geringste Formalisierung erhalten hat, in ein **systematisches, operationalisierbares, maschinenlesbares Modell** zu überführen: die Annotation. In der ursprünglichen Anlage in eMAEX war zwar sehr genau vorgegeben, „Was“ an den Filmdateien zu annotieren war, das konkrete „Wie“ aber nicht. Die Annotationen erfolgten als Freitext, der theoretisch (und praktisch) alle möglichen Idiosynkrasien an Abkürzungen und Rechtschreibfehlern zuließ. Nun galt es also eine Ordnung und ein Vokabular zu finden, mit dem perspektivisch auch manuelle und semi-automatische Annotationen verknüpft werden können und das kollaboratives Arbeiten erleichtert wird. So entstand eine „**Filmontologie**“, d.h. ein strukturiertes Netzwerk filmanalytischer Begrifflichkeiten für die Annotation. Der bisherige Stand ist – ganz verknüpft dargestellt –, dass jede Annotation, in dem Moment, in dem sie in der Annotationssoftware Advene erstellt wird, über [GitHub als Linked Open Data](#) publiziert wird und somit unmittelbar verfügbar ist.

*Abb. 7: Screenshot eines Ausschnittes der Arbeitsoberfläche in der Annotationssoftware Advene, wie sie im Projekt Affektrhetoriken des Audiovisuellen eingesetzt wird.*

Diese Verfügbarkeit – auch wenn alle Entwicklungen *open source* sind – ist allerdings immer noch sehr voraussetzungsreich und von der Verbreitung der verwendeten Einstellungen und Werkzeuge abhängig. Für den menschlichen Gegenpart sind mehrere Ansichtsformate in Entstehung, mit denen die Balance aus Daten, wie sie der Computer bearbeiten kann, und Darstellungsweisen, wie sie von menschlichen Akteur\*innen verarbeitet werden können, gewahrt bleibt: das Ziel dieser **dualen Strategie** ist, dass beide epistemischen Operationen, das Zählen, Rechnen und Korrelieren einerseits und das **hypothesengeleitete pattern-seeking** und die in letzter Zeit immer häufiger beschworene serendipity andererseits zu ermöglichen, um empirisch gestützte, spekulative Regeln für den Zusammenhang zwischen einzelnen filmischen Parametern und den affektiven und kognitiven Effekten ihrer Rezeption aufzustellen.

## Fazit

Fakt ist, dass all diese **informationstechnologische Bricolage** dem Umstand geschuldet ist, dass es in der filmwissenschaftlichen Forschung auf dem Gebiet der digital gestützten Analyseverfahren bisher noch **keine standardisierten Datenformate** gibt. Der ausschlaggebende Grund – neben der Tatsache, dass es sich schlicht um ein kleines Fach handelt –, ist dass sich der kulturwissenschaftliche, medienphilosophische und hermeneutische Kern im Selbstverständnis des Faches schwerlich in so etwas wie Datenformate pressen lässt.

Die Erweiterung des filmwissenschaftlichen methodischen Feldes um computergestützte Praktiken kann dabei aber sowohl **Quantifizierungen** und statistische Verfahren mit ihrer notwendigen Komplexitätsreduktion umfassen, als auch die hier beschriebenen **qualitativ-empirischen Verfahren** der Segmentierung, Klassifizierung, Annotation und Visualisierung, die nicht per se „computational“ sind, sondern *computer-gestützt*. D.h. sie machen erstens die Forscher\*in als Akteur\*in zu einem integralen Bestandteil des Datenzyklus, verweisen zweitens die Datenproduktion rekursiv auf **die multimodale, sinnliche Erfahrungsdimension** des Primärdatums und machen es drittens möglich, auch die Nachnutzungsszenarien der Daten eher zu diversifizieren als einzuengen. Voraussetzung für Letzteres ist, dass in den Formaten gleichermaßen Rücksicht auf die Anforderungen der **Maschinenlesbarkeit** und auf die Verarbeitung durch **menschliche Akteure** genommen werden kann – eine Frage der Skalierbarkeit, wie [Walkowski und Pause](#) zu Recht betonen. Dabei ist in Kauf zu nehmen, dass es immer auch einen Anteil unstrukturierter, idiosynkratischer Daten geben wird, der sich der statistischen Auswertung entzieht.

Langfristiges Ziel sollten Plattformen sein, die sowohl eine **Pluralität an tools und Systematiken** zugänglich machen und zugleich ein **Repositorium an Datensätzen** bieten, die mit diesen erstellt wurden und weiterverarbeitet werden können. Dies setzt aber eine kritische Masse sowohl der Zahl an beteiligten Forschungsansätzen als auch der Quantität an zur Verfügung stehenden Gegenständen und Materialien voraus, die bisher nur in letzterem Punkt [Cinematics](#) aufgrund seiner erschütternden Einfachheit erreicht hat. Dazu kommt idealerweise die Bereitstellung kuratierter Ressourcen und Quellen bzw. die Vernetzung mit einem **Fachrepositorium**.<sup>3</sup> (An dieser Stelle kommen dann nicht nur inhaltlich-methodische und informationstechnologische Fragen ins Spiel, sondern die fach- und standortpolitischen sowie ökonomischen Gemengelage: Wo und durch wen werden die Serverkapazitäten bereitgestellt und die langfristige Nachhaltigkeit garantiert?<sup>4</sup>)

Das alles mag aus einer Sicht höherer Weihen des Forschungsdatenmanagements trivial erscheinen. Wenn es uns jedoch Ernst damit sein soll, die Methoden geisteswissenschaftlichen Arbeitens und die Potentiale – sowie die damit einhergehenden Einschränkungen und Notwendigkeiten – informationstechnologischer Tools zu vereinbaren, dann kann dies nur über die [Assemblage multipler](#)

[Zugänge](#) funktionieren, die weder alles Digitale als unzulässige Quantifizierung ablehnt, noch alles, was sich nicht zählen oder messen lässt, als nicht-empirische, nicht-datenfähige Schöngestei abtut.

1. Das wäre das Potential von [media/rep/](#), gesehen 31.10.2019. Vgl. Sophie G. Einwächter: [Forschungsdaten \(in\) der Film- und Medienwissenschaft](#), in: *Open-Media-Studies-Blog*, 2019, gesehen 31.10.2019.

2. Vgl. der Beitrag von Lucas Curstädt und Jana Zündel: [Denn sie wissen nicht, was sie tun? Methodenfrage\(n\) in der Filmwissenschaft](#). Ein Protokoll zum 4. Workshop der Film AG der Gesellschaft für Medienwissenschaft, in: *Open-Media-Studies-Blog*, 2018, gesehen 31.10.2019.

3. „Wir“, das meint in diesem Fall eine Forschergruppe um Hermann Kappelhoff, die in diversen, teils dezidiert interdisziplinären Projektzusammenhängen an der Entwicklung von eMAEX gearbeitet haben bzw. weiterhin an seiner Fortentwicklung arbeiten.

4. Hermann Kappelhoff, Jan-Hendrik Bakels: [Das Zuschauergefühl – Möglichkeiten qualitativer Medienanalyse](#), in: ZfM: Zeitschrift für Medienwissenschaft, Jg. 5, Vol. 2, 2011, 78 – 96.

[Matthias Grotkopp](#)

Teaserbild: [Hans-Peter Gauster](#) auf Unsplash



Schlagwörter: [Filmanalyse](#) [Forschungsdaten](#) [Open Data](#)

## DAS KÖNNTE DICH AUCH INTERESSIEREN ...



PUBLIZIEREN IN DER MEDIENWISSENSCHAFT — Andreas Kirchner über Open Access als Standard

02/12/2022



ROGER ODIN: «KOMMUNIKATIONSRAÜME» (2019). EIN HYBRIDES PUBLIAKTIONSPROJEKT (TEIL 1) — Ein zweiteiliger Bericht von Laura Katharina Mücke

06/01/2020

## SCHREIBE EINEN KOMMENTAR

Kommentar \*

Name \*

E-Mail \*

Website

Kommentar abschicken

Diese Website verwendet Akismet, um Spam zu reduzieren. [Erfahre mehr darüber, wie deine Kommentardaten verarbeitet werden.](#)



{{site\_title}} © {{year}}. Alle Rechte vorbehalten.

Powered by WordPress. Theme by Press Customizr.



Ein Blog präsentiert von Hypotheses - [Zum OpenEdition-Katalogeintrag](#) - [Datenschutzerklärung](#) - [Problem melden](#)  
[Syndication Feed](#) - [Impressum](#) - ISSN 2629-6020