Provenienzforschung und ihre Quellenbestände. Aktuelle Nutzungsszenarien zwischen Open Access und Inaccessibility

Hopp, Meike

meike.hopp@tu-berlin.de TU Berlin, Deutschland

von dem Bussche, Ruth

rbussche@fotostoria.de Fotostoria

Die Frage nach einer spezifisch deutschen Verantwortung für historisches Unrecht und die Kritik an der Realisierbarkeit und Verhältnismäßigkeit von "Wiedergutmachung" an Opfern der Verfolgung, Entrechtung und Enteignung durch den nationalsozialistischen Unrechtsstaat in der Nachkriegszeit, beschäftigte in den vergangenen Jahrzehnten vor allem Forscher_innen aus den Bereichen der Zeitgeschichte, der Philosophie und der Rechtswissenschaften. In der kunsthistorischen Forschung haben die Washington Principles on Nazi-Confiscated Art von 1998 ein Umdenken herbeigeführt. In der Folge hat sich die Provenienzforschung zunächst mit der Koordinierungsstelle für Kulturgutverluste in Magdeburg und der 2008 eingerichteten Arbeitsstelle für Provenienzrecherche/-forschung beim Institut für Museumsforschung, Stiftung Preußischer Kulturbesitz in Berlin - seit 2015 zusammengeschlossen zum Deutschen Zentrum für Kulturgutverluste (DZK) in Magdeburg - institutionalisiert. Die Provenienzforschung stellt inzwischen eines der virulentesten geisteswissenschaftlichen Forschungsfelder dar, welches von großem öffentlichem und medialem Interesse begleitet wird und inzwischen an einigen, v.a. kunsthistorischen Instituten zum Ausbildungs- bzw. zum Lehrangebot gehört. Seit 2015 wurden an fünf Universitäten in Bonn, Hamburg, München, Berlin und Lüneburg (Junior-)Professuren dieser Denomination eingerichtet, auch an anderen Universitäten werden Seminare oder spezialisierte Masterstudiengänge angeboten.

Dabei produziert die Provenienzforschung in erheblichem Maße objekt- und personenbezogene Daten: Die seit 2020 vom DZK betriebene Forschungsdatenbank PROVEANA bündelt Entitäten aus 467 von der Stiftung geförderten Projekte, d.h. Daten zu Akteur_innen, Objekten, Archivbeständen und Sekundärquellen, die den Forschenden aber ebenso betroffenen Nachfahren von Verfolgten oder aber Interessierten zur Verfügung zu stehen, darin enthalten auch die Such- und Fundmeldungen

der seit 2000 betriebenen Datenbank LostArt.¹ Parallel entstanden in den vergangenen zwei Jahrzehnten weitere unabhängige Datenbankprojekte, etwa die am Deutschen Historischen Museum in Berlin angesiedelte Datenbank zum Central Collecting Point (CCP) in München, welcher unter der Leitung amerikanischen Militärbehörden in der in der unmittelbaren Nachkriegszeit Wesentliches für die Rückführung des im Nationalsozialismus enteigneten Kulturguts geleistet hat², oder die Datenbank zum Kulturgutraub in Frankreich durch den Einsatzstabs Reichsleiter Rosenberg (ERR).³

Auch innerhalb der vom DZK geförderten Projekte werden Provenienzdaten dokumentiert, dies in vielfältiger Form, seien es Excelsheets, kleinere selbstgestrickte Datenbanklösungen oder auch zusätzliche Provenienzfelder, die in bestehende Museumsdatenbanken integriert wurden. Diese hausintern erzeugten Daten bleiben aber in der Regel nicht projekt- oder länderübergreifend nutzbar (s. Hopp 2018). Die Frage der Langzeitarchivierung dieser lokal gehaltenen Projektdaten ist vielfach noch gar nicht angegangen worden. Das mag auch mit daran liegen, dass wir es mit einer föderalen Arbeitsstruktur, mangelnden personellen Kontinuitäten (bedingt durch Drittmittelförderungen und befristete Verträge) sowie Heterogenitäten von Datenmodellierungen auf der räumlichen wie auf der zeitlichen Achse zu tun haben, was die Koordination des Datenrückflusses an Datenzentren zeitaufwendig und schwierig macht.

Ein Blick auf die Provenienzforschung in Museen zeigt, wie vielfältig Provenienzdaten zu einem Objekt sein können. Es handelt sich um Metadaten zu Erwerbsumständen (Zugangsdaten, Rückseitenbefunde, etc.) aber auch um solche, die aus externen Quellen und Überlieferungen in privaten oder öffentlichen Archiven (Beschlagnahme-, Wiedergutmachungsakten, etc.) stammen. Hinzu kommen Daten aus Forschungseinrichtungen, die sich mit den Mechanismen der Verlagerung von Kulturgütern und Akteursnetzwerken sowie dem Kunstmarkt befassen.⁴ Die Herausforderung besteht darin, diese Erkenntnisse und Informationen zu bündeln, wobei die erhobenen Daten immer auch eine Rückbindung an die jeweiligen archivalischen Belege benötigen, die ähnlich vielfältig wie die durch sie beschriebenen Objekte sind.

Wie ist es nun um die Zugänglichkeit zur Archivmaterialien bestellt? Die Archive selbst leisten seit Jahren wichtige Unterstützung für die Provenienzforschung. Die Zielvorgaben von Politik und Forschung waren dabei in der Regel bislang Quellen oder archivalische Bestände schnell und/oder unkompliziert als Digitalisate oder Datenbanken zur Verfügung zu stellen, um die direkte Zugänglichkeit für die Forschung zu gewährleisten, wobei auf spezielle bzw. individuelle Abfragemöglichkeiten weniger Wert gelegt wurde. Aufgrund der schieren Menge und Fülle des vorhandenen Materials können die bisherigen Online-Angebote bzw. oben genannte Datenbankprojekte zwar einen wichtigen aber eben auch nur sehr kleinen Ausschnitt der Informationen zum NS-Kunstund Kulturgutraub wiedergeben. Zudem scheinen bei den digitalen Zugriffsmöglichkeiten auf die für die Provenienzforschung relevanten Archivbestände, große Unterschiede auf.

Das Transparenzgebot der Washingtoner Prinzipien von 1998 steht dabei noch immer im Widerspruch zu den archivrechtlichen Bestimmungen einzelner Einrichtungen, ganz gleich ob auf bundes-, landes- oder kommunaler Ebene. Während die Ende 2016 in Kraft getretene DGSVO mit dem in ihr enthaltenen Erwägungsgrund 158 einen transparenten Umgang mit Daten empfiehlt, die "im Zusammenhang mit dem politischen Verhalten unter ehemaligen totalitären Regimen, Völkermord, Verbrechen gegen die Menschlichkeit, insbesondere dem Holocaust, und Kriegsverbrechen" stehen,⁵ bleibt sie bezogen auf die praktische Umsetzung der Verarbeitung und Publikation der Daten weiterhin auslegbar, bzw. wird in der Praxis in Archiven und Verwaltungen sehr viel enger gefasst. Dabei fehlt häufig einen saubere Trennung von Daten die DSGVO-relevant sind, von solchen, die in keinem Bezug zu lebenden Personen stehen.⁶

Kleinere, regionale Forschungsverbünde arbeiten inzwischen mit digitalen Repositorien oder Portalen zur Bereitstellung von Quellenmaterial speziell für die Provenienzrecherche, doch stehen diese in der Regel nur einem eingeschränkten Kreis an Nutzer_innen zur Verfügung, da neben der rechtlichen Lage auch viele moralisch-ethische Fragen im Umgang mit den erarbeiteten Informationen offen bleiben, so etwa privaten Informationen zu den Geschädigten (Familiendokumente, etc.). So erfüllen die existierenden Lösungsansätze bis heute kaum die Anforderungen an umfassende Transparenz im Umgang mit (Meta-)Daten zur Herkunft der in deutschen Einrichtungen verwahrten kulturellen Objekte, ein Problem, das auch auf die in jüngerer Zeit begründeten Netzwerke zum Umgang mit Objekten aus kolonialen Kontexten - hier sei exemplarisch auf das künftige CCC-Portal der Deutschen Digitalen Bibliothek verwiesen - übertragen werden kann."

Allerdings gibt es aktuell verschiedene Ansätze zur Aufarbeitung von Archivbeständen stärker maschinell gestützten Verfahren einzusetzen, die über Volltexterschließung, Natural Language Processing (NLP) oder der Named Entity Recognition (NER) Dokumente zugänglich machen.8 Doch auch hier bliebt fraglich wie z.B. ein bundesweit angelegtes Projekt zur Digitalisierung der Wiedergutmachungsakten die zahlreichen archiv- und personenschutzrechtlichen Beschränkungen und Fristen umgehen wird bzw. wie transparent die Ergebnisse schließlich publiziert werden können, womit alle Digital-Projekte in diesem sensiblen Bereich in rechtlichen Grauzonen agieren. Bedarf es nicht sogar gerade in diesem Fall eines speziellen Schutzes von Daten zu Opfern der NS-Verfolgung? Die Akten, mit denen die Provenienzforschung arbeitet, sind Zeugnisse eines totalitären Unrechtsregimes. Neben den Akten der Finanzämter, die Auskunft über Vermögenswerte und fiskalische Verfolgung geben, erlauben Auflistungen des vor der Emigration in den Speditionen deponierten Hausrats unmittelbare Einblicke in Hausstand und Familienleben. Schließlich finden sich in Entschädigungsverfahren nicht selten Zeugnisse, die Foltermethoden benennen, die die Geschädigten über sich ergehen lassen mussten, ebenso sowie die davongetragenen medizinischen Spätfolgen. Dürfen wir heute über dieses Wissen, diese Daten frei verfügen?

Doch gerade für Forschende ist der Ansatz z.B. über die maschinelle Erschließung von möglichst vielen Dokumententexten die für die Provenienzrecherche wichtige objekt- oder personenbezogenen Daten schnell herauszufiltern natürlich essentiell, da auf Basis der unüberblickbaren Menge der europaweit verstreuten Quellen zum Kunst- und Kulturgutraub der Nationalsozialisten einzelfallbezogene Prüfungen und Sondierungen oft nicht effizient und nachhaltig bearbeitet werden können. Zwar ist es denkbar, maschinelle Verfahren der Texterschließung anzuwenden, aber es braucht zusätzliche Methoden, sensible Inhalte zu filtern und Teile von Beständen – wie bislang auch analog gehandhabt – nur auf begründeten Antrag zur Verfügung zu stellen.

Gleichzeitig steht die digitale Provenienzforschung vor der Herausforderung eine Basis für die Etablierung effizienter Modelle zur standardisierten Erfassung von eindeutigen bzw. uneindeutigen Provenienzdaten zu schaffen und Datenkompetenzen auszubilden. Denn neben der Recherche von Objektbiografien, Eigentumsübergängen und -verlusten,

hat Provenienzforschung auch das Ziel den heutigen Anspruchsberechtigten und der interessierten Öffentlichkeit gerecht zu werden - ergo zu dokumentieren, aufzuklären, zu informieren und damit auch Daten weltweit - auffindbar zu machen. Bereits im Sommersemester 2020 sind Studierende am Fachgebiet Digitale Provenienzforschung der TU Berlin in einem Seminar der Frage nachgegangen, ob und inwiefern sich Nachfahren der im NS-Regime rassisch oder anderweitig Verfolgten über laufende Projekte und bestehende Online-Datenangebote im Bereich der Provenienzforschung in Deutschland informieren können. Im Zentrum stand die Frage, ob die bereits bestehenden Datenbanken/-tools auch für ein nicht spezialisiertes Publikum zugänglich, ob sie auffindbar, transparent und verständlich sind, ob es Sprachhürden gibt und wie viel Vorwissen erforderlich ist, um die in ihnen enthaltenen Informationen richtig zu interpretieren. In Gesprächen mit heute international ansässigen Nachfahren, kristallisierte sich heraus, dass das Online-Angebot nicht nur unübersichtlich ist, sondern wesentlichen Adressat*innen bzw. Interessentengruppen weitestgehend verschlossen bleibt. Die aktuellen Angebote produzieren folglich nicht nur Expertenwissen, sondern auch Ausschlüsse der eigentlich Betroffenen.

Vor diesem Hintergrund haben wir uns seit Herbst 2020 intensiv mit Optimierungsprozessen im Umgang und bei der Erschließung von öffentlich verfügbaren Quellenbeständen zur Provenienzforschung befasst. Unsere Verfahren zielten dabei unter anderem auf den heute im Bundesarchiv Koblenz befindlichen, komplett digitalisierten aber bisher nicht effizient zu durchsuchenden Bestand B323 der in den Nachkriegsjahren agierenden ehem. Treuhandverwaltung von Kulturgut beim Oberfinanzpräsidium in München ab, der wesentliche Original-Quellen zum NS-Kunstraub sowie zu den alliierten Rückführungs-Bemühungen enthält und der in seiner sehr gemischten Zusammensetzung einen optimalen Testbestand lieferte. Ausgangspunkt des Pilotprojekts waren zunächst Metadaten aus der Archiverfassung,9 die wir als Graph aufbereitet haben (Bussche, Hopp 2022a und 2022b). Hierbei orientieren wir uns an bisherigen Bemühungen, den Nutzen von Linked-Data- und Semantic-Web-Technologien für archivarische Sammlungen zu untersuchen (Ferris 2014, Gracy 2015, Niu 2016). Um diese zu einer verlässlichen Ressource für die Provenienzforschung zu entwickeln, wurden alle über die Rechercheplattform des Bundesarchivs zur Verfügung stehenden Digitalisate mittlels OCR erfasst und die Daten in einer Suchmaschine zur Verfügung gestellt. Bereits auf dieser ersten Grundlage zeichnete sich ab, dass die Erschließung über Volltexte die Anforderungen der Provenienzforschung wesentlich besser abbilden kann, als das bisherige Online-Angebot mit Erschließungen zu Einzelakten über die Bestandbeschreibungen, denn der Praxis geht es häufig darum die Dokumente mit Erwähnungen bestimmter Personen, Werke oder Institutionen erst einmal aufzufinden.

Die bisher in unserem experimentellen Projekt verwendeten Verfahren machine learning basierter Clouddienste, wie etwa azure OCR oder Google NER ermöglichten es uns auch mit geringen personellen Kapazitäten durchaus umfangreiche Aktenbestände zu verarbeiten, wenngleich die Nutzung der genannten Dienste in öffentlichen Einrichtungen problematisch wäre. Wesentliche Hindernisse für eine Zugänglichmachung von Provenienzdaten beginnen also nicht erst bei der Frage der Online-Stellung, sondern schon sehr viel früher bei der Verarbeitung der Daten. Die von verschiedenen Institutionen unterschiedlich strikt gehandhabten Regeln hierzu blieben ebenso uneinheitlich wie die Vorgaben zum Datenschutz selbst, wobei Empfehlungen für "rechtskonforme" Ersatzprodukte fehlen.

Die Möglichkeiten der maschinellen Verarbeitung der bisher erzeugten Daten zum Bestand B323 gehen allerdings weiter (Stork 2021, Moss et al. 2018, Krenn 2019). Zum gegenwärtigen Zeitpunkt geht es uns darum, die Qualität der bislang eingesetzten Verfahren zu evaluieren und an einer künftigen verbesserten Verarbeitungspipeline zu arbeiten.

Welche qualitativen Vorteile bringt eine Vorverarbeitung der Scans (z.B. Ränder entfernen)?

Wo müssen Elemente eines Digitalisats segmentiert werden um die Auslesung der Daten bzw. Texte zu optimieren? Das betrifft vor allem kleine Notizen und Belege oder ausgeschnittene Bilder aus Mikrofilmen, die auf DinA4 Seiten montiert wurden.

Wo kann hingegen Layouterkennung eingesetzt werden (z.B. bei Listen, Karteikarten, Korrespondenzen)?

Welche Möglichkeiten haben wir Dokumente auch inhaltlich erkennen zu lassen, um z.B. eine Filterung nach Rechnungen oder Transportlisten vorzunehmen? Welche Algorithmen stehen für buchhalterische Dokumente wie Quittungen oder Rechnungen zur Verfügung, um deren Inhalte strukturiert erkennen zu lassen?

Wie können wir Texte über Entitäten erschließen?

Für die Erschließung von Dokumentenbeständen über Entitäten gibt es bereits Vorbilder: so wurden Dokumente aus der Zeit der deutschen Besatzung in den Niederlanden beispielsweise im Projekt Oorlogsbronnen aufbereitet (Borggräfe et al. 2020). ¹⁰ Ein weiteres bemerkenswertes Beispiel aus dem Bereich der Provenienzforschung ist der Archivführer zur deutschen Kolonialgeschichte, der archivische Sammlungen katalogisiert und mit Wikidata-Elementen verknüpft und somit neben dem niedrigschwelligen Einstieg in die Recherchen auch die weitere Bearbeitung der Entitäten und den Einbezug von Exper-

t_innenwissen erlaubt, um die Qualität der Erkennung und Disambiguierung zu verbessern (Jung 2019). Die beiden Beispiele liefern damit Entitäten, über die Texte mit der entsprechenden Qualität maschinell erschlossen werden können.

Das bedeutet im Folgeschluss, dass es immer wichtiger wird, dass strukturierte Daten auch als offene Daten zur Verfügung stehen. Plattformen wie PROVEANA haben die dazu notwendige Struktur und Qualität, bieten aber über Einzelrecherchen hinaus keinen Zugang zu den Rohdaten oder API an.

Die Anwendung von Semantic-Web-Technologien in der Provenienzforschung benötigt unserer Meinung nach: einen offenen Diskurs zum standardisiertn und FAIRen Umgang mit den Digitalisaten und Forschungsdaten an allen Provenienzforschung betreibenden Einrichtungen, Verfahren zur effizienteren Aufbereitung von Dokumenten, um die nötige Qualität der Dokumentenverarbeitung (OCR, Layouterkennung usw.) sicherstellen und Methoden um Unterscheidungen zwischen den für die Forschung offenzulegenden Daten von den sensiblen Informationen vorzunehmen.

Fußnoten

- 1. Vgl. https://www.proveana.de (aufgerufen am 12.12.2022).
- 2. Vgl. https://www.dhm.de/datenbank/ccp/dhm_ccp.php?seite=9 (aufgerufen am 12.12.2022).
 3. Vgl. https://www.errproject.org (aufgerufen am 12.12.2022).
- 4. Forschungsschwerpunkte zum Kunstmarkt gibt unter anderem am Zentralinstitut für Kunstgeschichte (ZI) München, hier sogar mit ausgewiesenem Schwerpunkt im Bereich der Provenienzforschung: https://www.zik-g.eu/forschung/provenienzforschung-werte-von-kulturguetern, am Zentralarchiv für deutsche und internationale Kunstmarktforschung an der Uni Köln: https://khi.phil-fak.uni-koeln.de/fachgebiete/kunstmarkt/forschungsschwerpunkte oder am Forum Kunst und Markt der TU Berlin https://www.kuk.tu-berlin.de/menue/forum_kunst_und_markt/ (alle aufgerufen am 12.12.2022).
- 5. Siehe: https://dsgvo-gesetz.de/erwaegungsgruende/nr-158/ (aufgerufen am 12.12.2022).
- 6. Der rechtliche Rahmen findet sich skizziert in Schlagk 2019; Eisenberger et al. 2018.
- 7. Vgl. https://ccc.deutsche-digitale-bibliothek.de/ (aufgerufen am 12.12.2022).
- 8. Wesentliche Projekte sind hier die Akten des Oberfinanzpräsidenten Brandenburg, s. https://blha.brandenburg.de/index.php/projekte/ofp-projekt/ und das Projekt zu den Wiedergutmachungsakten, s. https://www.bundesarchiv.de/DE/Content/Pressemitteilungen/publikation-portal-wiedergutmachung.html und https://www.fiz-karlsruhe.de/de/forschung/wiedergutmachung (alle aufgerufen am 12.12.22).
- 9. Das Bundesarchiv stellt die Daten aus der Bestandserfassung als offene Daten zur Verfügung, vgl. https://www.bundesarchiv.de/DE/Content/Artikel/Ueber-uns/Aus-unserer-Arbeit/open-data.html .

10. Vgl. https://www.oorlogsbronnen.nl (aufgerufen am 12.12.2022).

11. Vgl. https://archivfuehrer-kolonialzeit.de (aufgerufen am 12.12.2022).

Bibliographie

Borggräfe, Henning, et al., editors. "Linking and Enriching Archival Collections in the Digital Age: The Dutch War Collections Network." Tracing and Documenting Nazi Victims Past and Present, De Gruyter Oldenbourg, 2020, pp. 315–38, https://www.degruyter.com/document/doi/10.1515/9783110665376-018/html.

"Digitale Provenienzforschung" Provenienz & Forschung, edited by Deutsches Zentrum Kulturgutverluste, vol. 1, 2020

"EAC-CPF." Encoded Archival Context for Corporate Bodies, Persons, and Families, https://eac.staatsbibliothek-berlin.de.

Eisenberg, Iris et al. "Zeitgeschichtliche Forschung im Spannungsfeld von Archiv- Datenschutz- und Urheberrecht." Wien, 2018.

Ferris, Virginia L. Beyond "Showing What We Have": Exploring Linked Data for Archival Description. University of North Carolina at Chapel Hill, 2014, https://doi.org/10.17615/6n9k-q582.

Gracy, Karen F. "Archival Description and Linked Data: A Preliminary Study of Opportunities and Implementation Challenges." Archival Science, vol. 15, 2015, pp. 239–94, https://doi.org/10/gdcck6.

Greenberg, Jane. "The Applicability of Natural Language Processing (NLP) to Archival Properties and Objectives." The American Archivist, vol. 61, 1998, pp. 400-25, https://meridian.allenpress.com/american-archivist/article/61/2/400/23942/The-Applicability-of-Natural-Language-Processing.

Hopp, Meike. "Provenienzrecherche und digitale Forschungsinfrastrukturen in Deutschland: Tendenzen, Desiderate, Bedürfnisse" ...(k)ein Ende in Sicht. 20 Jahre Kunstrückgabegesetz in Österreich, edited by Eva Blimlinger and Heinz Schödl, Wien, 2018, pp. 37-61.

Jung, Uwe. "Archivfüher zur deutschen Kolonialgeschichte" Archivar, no. 4, 2019, pp. 325–327.

Krenn, Brigitte. "Methoden der künstlichen Intelligenz und ihre Anwendung in der Erschlieβung von Textinhalten." Die Zukunft der Vergangenheit in der Gegenwart, Archive als Leuchtfeuer im Informtionszeitalter, edited by Elisabeth Schöggel-Ernst, Thomas Stockinger and Jakob Wührer, Wien, 2019, pp. 169-184.

Marciano, R., et al. "Archival Records and Training in the Age of Big Data." Re-Envisioning the MLS: Perspectives on the Future of Library and Information Science Education, vol. 44B, Emerald Publishing Limited, Bingley, 2018, pp. 179–99, https://doi.org/10/gndpbh.

Moss, Michael, et al. "The Reconfiguration of the Archive as Data to Be Mined." Archivaria, vol. 86, 2018, pp. 118–51, https://archivaria.ca/index.php/archivaria/article/view/13646.

Niu, Jinfang. "Linked Data for Archives" ARCHIVARIA, The Journal of the Association of Canadian Archivists, vol. 82, 2016, pp.83–110, https://archivaria.ca/index.php/archivaria/article/view/13582.

Nadeau, David, and Satoshi Sekine. "Linked Data for Archives." ARCHIVARIA, The Journal of the Association of Canadian Archivists, vol. 82, 2016, pp. 83–0, https://archivaria.ca/index.php/archivaria/article/view/13582.

Schlagk, Particia. "Die datenschutzrechtlichen Privilegien von im öffentlichen Interesse liegenden Archivzwecken." Bachelorarbeit, Fachhochschule Potsdam 2019, urn:nbn:de:kobv:525-24311.

Stork, L. Knowledge Extraction from Archives of Natural History Collections. Leiden, July 2021, https://scholarlypublications.universiteitleiden.nl/handle/1887/3192382?solr_nav%5Bid %5D=dc5e9a7ed703338e5b78&solr_nav%5Bpage %5D=0&solr_nav%5Boffset%5D=0

von dem Bussche, Ruth, and Meike Hopp. "Der .Bestand B323' als Knowledgegraph Provenienzforschung. Methodische Überlegundie Verarbeitung Archivdaten gen zur von Linked Open Data." Archivar, no. 2022. 60-63, https://www.archive.nrw.de/sites/default/files/ media/files/

Archivar_2022-1_Internet-NEU-28032022_Mod.pdf.

von dem Bussche, Ruth, and Meike Hopp.
"The Archive as a Graph - Provenance Research on Bundesarchiv B323" Graphs and Networks in the Humanities. Technologies, Models, Analyses, and Visualizations, 6th International Conference, 2022, https://graphentechnologien.hypotheses.org/files/2022/01/

The_Archive_as_a_Graph_Provenance_Research_on_etc-Hopp.pdf