

# Die Ontologie historischer deutschsprachiger Berufs- und Amtsbezeichnungen. Interoperationalität und Berufsklassifizierung durch semantisches Topic Modeling

**Nasarek, Robert**

robert.nasarek@geschichte.uni-halle.de  
Martin-Luther-Universität Halle-Wittenberg, Deutschland

**Moeller, Katrin**

katrin.moeller@geschichte.uni-halle.de  
Martin-Luther-Universität Halle-Wittenberg, Deutschland

## Forschungsstand:

Berufsbezeichnungen sind eine der häufigsten Angaben von individualspezifischen Quellen. Besonders in den Sozial- und Politikwissenschaften, den Geisteswissenschaften und einigen naturwissenschaftlichen Disziplinen (Sozialtopografie, Medizin, Arbeitsmedizin, Epidemiologie etc.) bieten Berufsbezeichnungen einen wichtigen Bezugspunkt sozialstruktureller Analysen. Dazu kommen verschiedene Formen von Berufsklassifikationen zum Einsatz. Während für die Berufswelten des 20./21. Jahrhunderts verschiedene normierte Klassifikationsmodelle als Standard sowohl auf nationaler wie internationaler Ebene existieren, kann die interdisziplinäre Forschung für deutschsprachige, historische Berufe nicht auf ein solches Normsystem zurückgreifen. Insgesamt gibt es bisher keinen gültigen Standard. Zu den renommiertesten und qualitativ hochwertigsten historischen Klassifikationsmodellen gehören bisher die Historical International Standard Classification of Occupations (HISCO) (van Leeuwen et al. 2002), das PST-System der Cambridge Group um Wrigley und Davies (Wrigley 2010). Sie zeichnen sich durch eine theoretisch nachvollziehbare Konzeption der Tätigkeit aus, beinhalten bisher aber kaum deutschsprachige Berufsamen. Berufsklassifikationssysteme wie zu Altona 1803 (Brandenburg et al. 1991) und das Berufsklassifikationsmodell zur Analyse von Bürgerlichkeit von Schüren und Hettling (Schüren 1989, Hettling 1999) liefern Ansätze für deutschsprachige Berufe. Daneben existieren etliche andere, im Zuge der Städteforschung entstandene Systeme, die jedoch eher als

Teilaspekt einer Arbeit entstanden sind (bspw. François 1982; Kill 2001; Rödel 1985; Sachse 1987).

Aufgrund des hohen personellen Aufwands werden Systematiken häufig induktiv entwickelt und die Einordnung der Berufe (Abklären des genauen historischen Tätigkeitsprofils; Zuordnung zur richtigen Beschreibungsform der Berufsklassifizierung) gar nicht durchgeführt bzw. nicht dokumentiert. Daher entwickeln Forschungsprojekte jeweils neue Klassifizierungsmodelle, was eine Vielzahl nicht vergleichbarer Ergebnisse und für das einzelne Projekt ein enorm zeitaufwendiges Verfahren produziert. Dies schmälert den wissenschaftlichen Mehrwert, da die Ergebnisse von sozialstrukturellen Aussagen und Analysen letztlich nicht reproduzierbar sind. Zudem unterbleibt bei solchen Vorhaben sowohl eine konzeptionelle als auch eine direkte Anbindung an moderne Klassifikationsmodelle, da diese keine unmittelbare „Schnittstelle“ zu historischer Beruflichkeit bieten.

Mit der Ontologie deutschsprachiger Berufs- und Amtsbezeichnungen möchten wir diese Lücke schließen und eine Berufssystematik für historische, deutschsprachige Berufe entwickeln, die sowohl an die (modernen) Klassifikationssysteme anknüpfen (Klassifikation der Berufe 2010: Wiemer et al. 2011) als auch die deutschen Berufsbezeichnungen für internationale, historische Klassifikationssysteme (HISCO, PST) anschlussfähig macht und die sowohl als maschinenlesbare Ontologie als auch mittels eines Webservice für einen individuellen Zugriff genutzt werden kann.

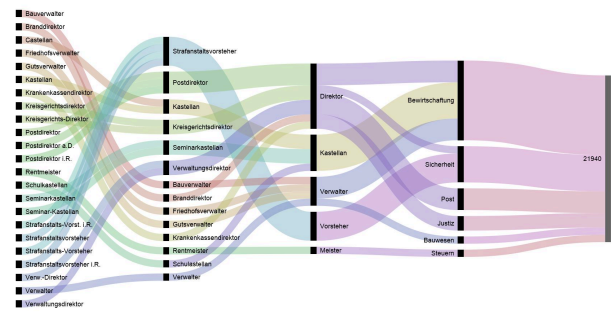
## Ziel des Projekts

Ziel des Projektes ist es einen nationalen Standard zur Normierung und multiperspektivischen Klassifizierung von deutschsprachigen, historischen Berufsbezeichnungen zu entwickeln und als webbasierte, offene Ressource für die Nachnutzung in Forschungsprojekten und für Infrastrukturen des Forschungsdatenmanagements zur Verfügung stellen. Der hohe Mehrwert für die wissenschaftliche Analyse und die enorme Arbeitserleichterung sollte erheblich dazu beitragen, auf positive Weise tatsächlich einen anreizbildenden Standard zu etablieren, zumal dieser dann auch international anschlussfähig sein wird. Da sich die meisten Klassifizierungsmodelle auf Berufssysteme des 19./20. Jahrhundert konzentrieren, möchten wir diese Systematiken um Beruflichkeit der Frühen Neuzeit und des Spätmittelalters erweitern. Im Einzelnen umfasst dieser Service:

- Einen Standard zur Normierung von deutschsprachigen Berufsbezeichnungen. Dazu stehen ca. 200.000 Varianten von historischen Berufsschreibungen zur Verfügung, die sich in ca. 20.000 normierte Einheitsbezeichnungen zusammenfassen lassen.

Dieser enorme Korpus ist Ergebnis jahrelanger eigener Datenproduktionen und Kooperationen mit qualitativ hochrangigen Projekten. Im Gegensatz zu vielen Handbüchern, Lexika und anderen gedruckten Kompendien (Ebener 2015; Gerholz 2005; Haemmerle 1933; Molle 1975; Palla 2014; Puchner / Stadler 1935; Reith 1991; Ulm-Sanford 1975) werden die Originalschreibungen der Varianten erhalten. Dies ermöglicht einen hohen Mehrwert für interdisziplinäre Zugriffe z.B. linguistische und etymologische Analysen, aber auch für den technischen Vorgang der Normierung und Zuweisung der Kodierung in einer Vielzahl von Erschließungsprojekten. Damit bietet das Werkzeug eine außerordentliche gute Grundlage, zur Erschließung (Lemmatisierung und Kodierung) von textuellen Quellen jeder Art.

- Um die Klassifizierung zu erleichtern, wie auch einen Informationsverlust durch die Klassifizierung zu verhindern, werden alle Berufsbezeichnungen zunächst in kategorial geordnete semantische Einheiten aufgetrennt. Dies erleichtert erheblich die Nutzbarkeit des Angebots für unterschiedlichste Anforderungen. Diese semantische Trennung und Normierung der Namenseinheiten ermöglicht später eine individuell flexible inhaltliche Auswahl verschiedener logischer Einheiten (Topic Modeling) nach eigenen Prinzipien. In unserem Vortrag möchten wir die Logik dieses Modells erläutern und dem Fachpublikum zur Diskussion stellen, um Anregungen für eine interdisziplinär optimal zu nutzende Methode zu erhalten. Diese Vorgehensweise unterscheidet das Projekt von allen anderen rein auf ID-Identifikatoren zugeschnittenen Verfahren und macht es wesentlich flexibler und breiter für verschiedenste Bedürfnisse geistes- und sozialwissenschaftlicher Forschung anwendbar. Bei herkömmlichen Verfahren kann lediglich über eine fest zugewiesene ID-Nummer auf den verschiedenen Ebenen des Kodierungssystems klassifiziert und sortiert werden. In dem hier vorgeschlagenen Modell erfolgt zusätzlich auch auf der semantischen Ebene die Möglichkeit zum intuitiven Systematisieren, Anordnen und Auswählen. Dies gilt auch für alle weiteren Informationen (Geschlecht, Herrschaft, Rechtsbeziehungen, Ruhestand, Karriere etc.), die mit typisch frühneuzeitlichen, quellenbasierten Berufsbezeichnungen einhergehen und in moderne Berufssystematiken meist keinen Eingang finden. Moderne Klassifikationssysteme für frühneuzeitliche Berufsamen sind daher tendenziell mit einem sehr hohen Informationsdefizit belastet und können unter Umständen in ahistorischen Analysen enden.



- Zur Klassifizierung von Berufen werden aus zeittypischen Lexika des 17./18. Jahrhunderts Tätigkeitsbeschreibungen abstrahiert, welche die Anforderungsprofile und Kompetenzen einzelner Berufe definieren. Sie dienen dem Abgleich mit modernen Tätigkeitsprofilen und erlauben erstmals überhaupt eine tätigkeitsgenaue Zuordnung historischer Berufe. Im Vortrag möchten wir zeigen, welche Möglichkeiten zur Systematisierung und Kategorisierung Lexika der frühen Neuzeit bieten und welche inhaltlichen Anknüpfungspunkte sie für moderne Klassifikationssysteme liefern. Dieser Aspekt bietet zudem hervorragende Möglichkeiten zur fachwissenschaftlichen Analyse, werden auf diese Weise doch Definitionskriterien frühneuzeitlicher Beruflichkeit überhaupt ermittelt und mittels einer logischen Auszeichnung auffindbar und auswertbar (xml/TEI). Wir möchten zeigen, welche Probleme dabei auftreten und welche Lösungsansätze wir hierfür entwickelt haben.



Weitere Ziel des Projektes (aber nicht unmittelbar unseres Vortrages, auch wenn wir auf die einzelnen Punkte sicherlich hinweisen) sind zudem folgende Punkte:

- Die Tätigkeitsprofile ermöglichen die nachvollziehbare, transparente Einordnung von Berufsbezeichnungen zum bereits existierenden englischsprachigen Berufsklassifizierungssystems HISCO. Momentan gibt es in HISCO lediglich 1.306 deutschsprachige Berufsbezeichnungen. Die Daten wurden nicht

von einem Muttersprachler kodiert, weshalb viele Zuordnungen korrigiert werden müssten. Die Kodierung von HISCO ermöglicht das Mapping weiterer historischer Klassifikationsmodelle wie HISCLASS oder PST. Da HISCO moderne Beruflichkeit misst, ist eine Erweiterung und Präzisierung für die Berufe der Frühen Neuzeit unerlässlich (van Leeuwen et al. 2002; HISCO Tree Of Occupational Groups [<http://historyofwork.iisg.nl/major.php>]).

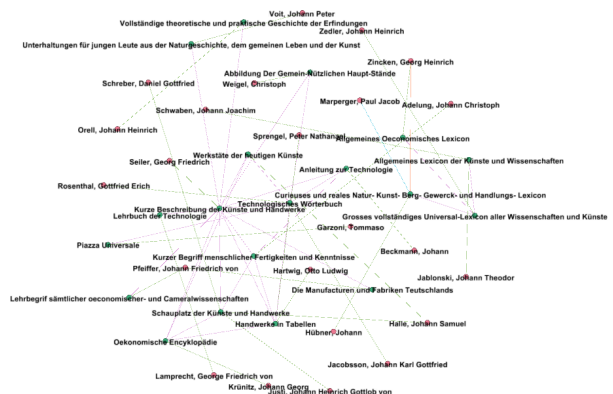
- Zudem werden die Einheitsbezeichnungen der KldB 2010 zugeordnet. Sie bieten nicht nur einen Zugriff auf (weitere) moderne Berufsklassifikationsmodelle, sondern durch den theoretischen Perspektivwechsel auch eine Systematisierung nach Anforderungsprofilen (Wiemer 2011).
- Die offenen Daten ermöglichen es Nutzern, transkribierte, quellenbasierte Originalberufsbezeichnungen zu normalisieren und automatisiert in ein ausgewähltes Kodierungssystem zu überführen. Ein Usecase wäre bspw. ein Forschungsprojekt, welches eine Netzwerkanalyse durch Heiratsverbindungen vornimmt. Der Beruf bietet hier häufig den einzigen Hinweis auf die sozialstrukturelle Dimension von Gruppen. Über das Werkzeug können die Daten in einem gewünschten Klassifizierungsmodell ausgegeben werden, indem die Originalschreibungen der Berufe in das Tool geladen und automatisiert kodiert werden.
- Varianten die keine automatisierte Kodierung beinhalten, können anschließend recherchiert und kodiert und in das Kompendium übertragen werden. Das Werkzeug ist damit beliebig erweiterbar. Das Projekt entwickelt ein Geschäftsmodell, wie solche Daten als Teil eines Serviceangebots nachkodiert und integriert werden können. Als Teil der langfristigen Infrastruktureinrichtung DARIAH soll die Ontologie der Berufe durch die Niedersächsische Staats- und Universitätsbibliothek auf Dauer angeboten werden.

## Zwischenergebnisse

- Im Bereich der Berufssegmentierung wurde ein Prototyp eines Datenschemas entwickelt, um die Berufe in adäquate Informationseinheiten aufzutrennen.
- Der Informationsgehalt der Lexika bezogen auf berufskundliche Informationen wurde tiefergehend untersucht und erste Aussagen können darüber getroffen werden.

[illegible]

Des Weiteren ist ein erstes Netzwerk von Autorenschaften und Informationsflüssen nachweisbar.



- Die Taxonomie der Klassifikationssysteme von HISCO und der KldB 2010 wurden ausführlich untersucht und erste Ansätze einer eigenen Systematik können vorgestellt werden.
- Mehrere Workflowalternativen von der Quelle bis zum Ergebnis der Inhaltsanalyse wurden erprobt, bewertet und können mit ihren Vor- und Nachteilen präsentiert werden. An dieser Stelle platziert sich am deutlichsten die „Kritik an der digitalen Vernunft“, vor allem in Bezug auf die
  - Arbeit mit OCR-Erkennung (am Beispiel der OCR-Programmsammlung „ocropy“)
  - dem Pre- und Postprocessing von Quellendigitalisaten (im Sinne von Ordnerstrukturen und Bildaufarbeitung)
  - XML vs. QDA-gestützten hermeneutischen Verfahren der Inhaltsanalyse (zur Implementierung und Weiterverarbeitung der Ergebnisse und Daten) und
  - der Nutzung von Objekt- oder relationalen Datenbanken zur Datenverwaltung und Verarbeitung.

Hierbei wird über die Vermeidung von epistemischen Fallstricken durch informationstechnische Automatisierung reflektiert und die Effizienzsteigerung digitaler Werkzeuge kritisch betrachtet, aber auch über die neuen Möglichkeiten zur Bewältigung von Big Data und dem dazugehörigen Erkenntnisgewinn referiert.

## Bibliographie

**Brandenburg, Hajo / Gehrmann, Rolf / Krüger, Kersten / Künn, Andreas / Rüffer, Jörn** (1991): *Berufe in Altona. Berufssystematik für eine präindustrielle Stadtgesellschaft anhand der Volkszählung*. Kiel: Arbeitskreis für Wirtschafts- und Sozialgeschichte Schleswig-Holsteins.

**Brückner, Carola / Möhle, Sylvia / Pröve, Ralf / Roschmann, Joachim** (1988): *Vom Fremden zum Bürger: Zuwanderer in Göttingen 1700-1755*. In: Hermann Wellenreuther (Hg.): *Göttingen 1690-1755. Studien zur Sozialgeschichte einer Stadt*. Göttingen (Göttinger Universitätschriften. Serie A, Schriften, Bd. 9), S. 88–174.

**Bundesanstalt für Arbeit** (1988): *Klassifizierung der Berufe. Systematisches und alphabetisches Verzeichnis der Berufsbenennungen*. Nürnberg. Online verfügbar unter <http://statistik.arbeitsagentur.de/Statistischer-Content/Grundlagen/Klassifikation-der-Berufe/KldB1975-1992/Generische-Publikationen/KldB1988-Systematischer-Teil.pdf>, zuletzt geprüft am 07.06.2016.

**Bundesanstalt für Arbeit** (2011): *Klassifikation der Berufe 2010. Systematischer und alphabetischer Teil mit Erläuterungen*. 2 Bände. Nürnberg (1). Online verfügbar unter <https://statistik.arbeitsagentur.de/Statistischer-Content/Grundlagen/Klassifikation-der-Berufe/KldB2010/Printausgabe-KldB-2010/Generische-Publikationen/KldB2010-Printversion-Band1.pdf>, zuletzt geprüft am 07.06.2016.

**Ebeling, Dietrich** (1987): *Bürgertum und Pöbel. Wirtschaft und Gesellschaft Kölns im 18. Jahrhundert*. Köln (Städteforschung. Reihe A, Darstellungen, Bd. 26).

**Ebner, Jakob** (2015) *Wörterbuch historischer Berufsbezeichnungen*. Berlin u.a.

**Fischer, Volker** (2000): *Stadt und Bürgertum in Kurhessen. Kommunalreform und Wandel der städtischen Gesellschaft 1814-1848*. Kassel (Hessische Forschungen zur geschichtlichen Landes- und Volkskunde, 35).

**François, Etienne** (1982): *Koblenz im 18. Jahrhundert. Zur Sozial- und Bevölkerungsstruktur einer deutschen Residenzstadt*. Göttingen (Veröffentlichungen des Max-Planck-Instituts für Geschichte, Bd. 72).

**Gerber, Roland** (2001): *Gott ist Bürger zu Bern. Eine spätmittelalterliche Stadtgesellschaft zwischen Herrschaftsbildung und sozialem Ausgleich*. Weimar (Forschungen zur mittelalterlichen Geschichte, 39).

**Gerholz, Heinrich** (2005): *Gerholz-Kartei. Eine Sammlung alter Berufsbezeichnungen*. Lübeck.

**Haemmerle, Albert** (1966): *Alphabetisches Verzeichnis der Berufs- und Standesbezeichnungen vom ausgehenden Mittelalter bis zur neueren Zeit*, (Reprografischer Nachdruck der Ausgabe München 1933), Hildesheim.

**Hahn, Hans-Werner** (1991): *Altständisches Bürgertum zwischen Beharrung und Wandel*. Wetzlar, 1689-1870. München (Stadt und Bürgertum, Bd. 2).

**Hettling, Manfred** (1999): *Politische Bürgerlichkeit. Der Bürger zwischen Individualität und Vergesellschaftung*

in Deutschland und der Schweiz von 1860 bis 1918. Göttingen.

**ILO** (1958): *International Standard Classification of Occupations*. Geneva. Online verfügbar unter [http://www.ilo.org/public/libdoc/ilo/1958/58B09\\_81\\_engl.pdf](http://www.ilo.org/public/libdoc/ilo/1958/58B09_81_engl.pdf), zuletzt geprüft am 31.05.2016.

**ILO** (1969): *International Standard Classification of Occupations. Revised Edition 1968*. Geneva.

**ILO** (2004): *ISCO-88. Main Objectives*. Online verfügbar unter <http://www.ilo.org/public/english/bureau/stat/isco/isco88/pub11.htm>, zuletzt geprüft am 06.07.2016.

**ILO** (2008): *Resolution Concerning Updating ISCO 1 Resolution Concerning Updating the International Standard Classification of Occupations*. Online verfügbar unter <http://www.ilo.org/public/english/bureau/stat/isco/docs/resol08.pdf>, zuletzt geprüft am 31.05.2016.

**Jägers, Regine** (2001): *Duisburg im 18. Jahrhundert. Sozialstruktur und Bevölkerungsbewegung einer niederrheinischen Kleinstadt im Ancien Regime (1713-1814)*. Köln (Rheinisches Archiv, 143).

**Kill, Susanne** (2001): *Das Bürgertum in Münster 1770-1870. bürgerliche Selbstbestimmung im Spannungsfeld von Kirche und Staat*. München.

**Kroll, Stefan** (1997): *Stadtgesellschaft und Krieg. Sozialstruktur, Bevölkerung und Wirtschaft in Stralsund und Stade 1700 bis 1715*. Göttingen (Göttinger Beiträge zur Wirtschafts- und Sozialgeschichte, Bd. 18).

**Krüger, Kersten** (1986): *Sozialstruktur der Stadt Oldenburg 1630 und 1678. Analysen in historischer Finanzsoziologie anhand staatlicher Steuerregister*. Oldenburg.

**Laufer, Wolfgang** (1973): *Die Sozialstruktur der Stadt Trier in der frühen Neuzeit*. Bonn (Rheinisches Archiv, 86).

**Lundgreen, Margret Kraul; Ditt, Karl** (1988): *Bildungschancen und soziale Mobilität in der städtischen Gesellschaft des 19. Jahrhunderts*. Göttingen.

**Manke, Matthias** (2000): *Rostock zwischen Revolution und Biedermeier. Alltag und Sozialstruktur*. Rostock (Rostocker Studien zur Regionalgeschichte, Bd. 1).

**Molle, Fritz** (1975): *Wörterbuch der Berufs- und Berufstätigkeitsbezeichnungen*. Wolfenbüttel.

**Müller, Christina** (1992): *Karlsruhe im 18. Jahrhundert. Zur Genese und zur sozialen Schichtung einer residenzstädtischen Bevölkerung*. Karlsruhe (Forschungen und Quellen zur Stadtgeschichte, 1).

**Palla, Rudi** (2014): *Verschwundene Arbeit. Das Buch der untergegangenen Berufe*. Wien.

**Paulus, Wiebke / Matthes, Britta** (2013): *Klassifikation der Berufe. Struktur, Codierung und Umsteigeschlüssel*. Online verfügbar unter [http://doku.iab.de/fdz/reporte/2013/MR\\_08-13.pdf](http://doku.iab.de/fdz/reporte/2013/MR_08-13.pdf), zuletzt geprüft am 11.06.2016.

**Puchner, Karl / Stadler, Josef Klemens** (1935): *Lateinische Berufsbezeichnungen in Pfarrmatrikeln und sonstigen orts- und familiengeschichtlichen Quellen*. Hirschenhausen (Obby).

**Raschke, Helga** (2001): *Bevölkerung und Handwerk einer thüringischen Residenzstadt. Gotha zwischen*

1640 und 1740. 1. Aufl. Jena (Palmbaum Texte. Kulturgeschichte, Bd. 9).

**Reith, Reinhold** (1991): Lexikon des alten Handwerks. Vom späten Mittelalter bis ins 20. Jahrhundert. München.

**Rödel, Walter Gerd** (1985): Mainz und seine Bevölkerung im 17. und 18. Jahrhundert. Demographische Entwicklung, Lebensverhältnisse und soziale Strukturen in einer geistlichen Residenzstadt. Stuttgart (Geschichtliche Landeskunde, Bd. 28).

**Sachse, Wieland** (1987): Göttingen im 18. und 19. Jahrhundert. Zur Bevölkerungs- und Sozialstruktur einer deutschen Universitätsstadt. Göttingen (Studien zur Geschichte der Stadt Göttingen, Bd. 15).

**Schüren, Reinhard** (1989): Soziale Mobilität. Muster, Veränderungen und Bedingungen im 19. und 20. Jahrhundert. St. Katharinen.

**Schüren, Reinhard** (1989): Soziale Mobilität: Muster, Veränderungen und Bedingungen im 19. und 20. Jahrhundert. St. Katharinen.

**Statistisches Bundesamt** (1992): Klassifikation der Berufe 1992 (KldB 92). Gliederungsstruktur bis zur 4 Steller-Ebene. Stuttgart: Metzler-Poeschel. Online verfügbar unter [https://www.destatis.de/DE/Methoden/Klassifikationen/Berufe/klassifikationkldb92\\_4st.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Methoden/Klassifikationen/Berufe/klassifikationkldb92_4st.pdf?__blob=publicationFile), zuletzt geprüft am 06.06.2016.

**Straubel, Rolf** (1995): Frankfurt (Oder) und Potsdam am Ende des Alten Reiches. Studien zur städtischen Wirtschafts- und Sozialstruktur. 1. Aufl. Potsdam (Quellen und Studien zur Geschichte und Kultur Brandenburg--Preussens und des Alten Reiches, Bd. 2).

**Ulm-Sanford, Gerlinde** (1975): Wörterbuch von Berufsbezeichnungen aus dem siebzehnten Jahrhundert, gesammelt aus den Wiener Totenprotokollen der Jahre 1648 - 1668 und einigen weiteren Quellen, Bern 1975.

**van Leeuwen, Marco H.D. / Maas, Ineke / Miles, Andrew** (2002): Historical international standard classification of occupations. Leuven.

**Weichel, Thomas** (1993a): Die Berufsstruktur der Städte - erste Ergebnisse und Vergleiche. In: Lothar Gall (Hg.): Stadt und Bürgertum im Übergang von der traditionellen zur modernen Gesellschaft. München (Historische Zeitschrift. Beihefte, n.F., Bd. 16), S. 51–74.

**Weichel, Thomas** (1993b): Die Bürger in ihrer beruflichen und sozialen Stellung. In: Lothar Gall (Hg.): Stadt und Bürgertum im Übergang von der traditionellen zur modernen Gesellschaft. München (Historische Zeitschrift. Beihefte, n.F., Bd. 16), S. 93–103.

**Weichel, Thomas** (1997): Die Bürger von Wiesbaden. von der Landstadt zur "Weltkurstadt", 1780 1914. München.