

## Übersicht von Projekten zur Schaffung von historischen Datenbanken

### 1. Historisches Datenzentrum Sachsen-Anhalt

Katrin Moeller (Halle)

### 2. DeHiStaEd - An Edited Database of German Historical Statistics

Davide Cantoni (München) / Erik Hornung (Köln)

### 3. Historische Statistik von Deutschland

Sibylle Lehmann (Hohenheim) / Mark Spoerer (Regensburg) / Jochen Streb (Mannheim)

### 4. Datenbankmanagement SPP 1859/SBB

SPP 1859 (Ingo Köhler) / Staatsbibliothek zu Berlin (Christian Mathieu)

### 5. Historical Statistics of Germany

Universitätsbibliothek Mannheim (Irene Schumm) / Leibniz-Informationszentrum Wirtschaft (ZBW) / Ulrich Pfister (Münster) / Mark Spoerer (Regensburg)

### 6. Erschließung historischer Datenbestände deutscher Finanz- und Wirtschaftsunternehmen (1871 - heute)

Leibniz-Institut für Finanzmarktforschung SAFE (Uwe Walz) / Universitätsbibliothek Mannheim (Irene Schumm) / Institut für Bank- und Finanzgeschichte

### 7. Forschungsdaten im Journal Data Archive der ZBW

Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte / Leibniz-Informationszentrum Wirtschaft (Sven Vlaeminck)

### 8. Digitalisierung und semi-automatisierte Erschließung der Handelskammerberichte bis 1914

Staatsbibliothek Berlin (Christian Mathieu) / Manuel Burghardt (Leipzig) / Mark Spoerer (Regensburg)

### 9. Aufbau und Analyse eines Datenkorpus zur demographischen Entwicklung Deutschlands vom 16. zum 19. Jahrhundert auf der Basis einer umfassenden Aufnahme von Kirchenbüchern

Georg Fertig (Halle) / Ulrich Pfister (Münster)

## 1. Historisches Datenzentrum Sachsen-Anhalt

Katrin Moeller (Halle)

Präsentation: <https://cloud.uni-halle.de/s/V8lcoy3OvFhWoaQ>

Das historische Datenzentrum Sachsen-Anhalt versteht sich als fachliches Datenzentrum, das unter Einhaltung der FAIR-Prinzipien alle Etappen des Data-Lifecycle berät.

- Beratungen und Services zum Forschungsdatenmanagement (Planung, Erfassung, Modellierung, Analyse, Rechtsberatung, Lizenzierung).
- Kuration, Dokumentation und Publikation von Forschungsdaten (Verlag von Forschungsdaten) auf qualitätsgesicherten Repositorien (z.B. <https://opendata.uni-halle.de/handle/1981185920/66>).
- Aufbereitung, Erschließung und Publikation von Altdaten, z.B. Tonarchiv der Bonner Längsschnittstudie des Alterns (<https://bolsa.uni-halle.de/>).
- Werkzeuge und Services zur Datenanreicherung und Entitätsextraktion von Daten (Geocodierung, Berufsklassifikation, Normdatenanreicherung, Eigennamenerkennung, Kuration)
- Herausgabe der „Ontologie der historischen, deutschsprachigen Amts- und Berufsbezeichnungen (OhdAB)“ zur Klassifikation historischer Berufe mit verschiedenen internationalen Standards.
- Massendatenaufnahmen und Betreuung von Citizen Science-Projekte (z. B. <http://wiki-de.genealogy.net/Halle/Heiratskirchenb%C3%BCcher/Projektbeschreibung>).

## **2. DeHiStaEd - An Edited Database of German Historical Statistics**

Davide Cantoni (München) / Erik Hornung (Köln)

Eine zeitgemäße Aufarbeitung der historischen Statistik Deutschlands muss grundlegend als digitale Edition und Datenbank konzipiert werden. Wir planen, historische Statistiken aus folgenden Korpora digital zu erfassen, und als hyperdigitale Edition zu veröffentlichen:

- alle wesentlichen Merkmale aus der amtlichen Statistik des deutschen Reiches (ab 1870) und der Länder im Deutschen Bund/Reich (ab 1815), die in einer Aufschlüsselung im Querschnitt (nach Landkreisen oder Gemeinden) verfügbar sind;
- die wesentlichen, quantitativ erfassbaren Dimensionen der städtischen Entwicklung – in wirtschaftlicher, sozialer und institutioneller Perspektive – für alle knapp 2300 Städte mit Stadtrechten im deutschsprachigen Raum, ab dem Zeitpunkt ihrer Gründung.

Diese digitale Edition soll folgende technische Desiderata erfüllen:

- Geokodierung aller Statistiken (Kennzeichnung von Datenpunkten durch tags, Rekonstruktion von amtlichen Grenzen);
- Möglichkeit zum Download der Daten (open access);
- dynamisch generierte Visualisierungen (Karten/Tabellen).

Stand: beantragt über Bayerische Akademie der Wissenschaften

## DeHiStaEd - An Edited Database of German Historical Statistics

Davide Cantoni (München) / Erik Hornung (Köln)

Eine zeitgemäße Aufarbeitung der historischen Statistik Deutschlands muss grundlegend als digitale Edition und Datenbank konzipiert werden. Wir planen, historische Statistiken aus folgenden Korpora digital zu erfassen, und als hyperdigitale Edition zu veröffentlichen:

- alle wesentlichen Merkmale aus der **amtlichen Statistik** des deutschen Reiches (ab 1870) und der Länder im Deutschen Bund/Reich (ab 1815), die in einer Aufschlüsselung im **Querschnitt** (nach Landkreisen oder Gemeinden) verfügbar sind;
- die wesentlichen, quantitativ erfassbaren Dimensionen der **städtischen Entwicklung** – in wirtschaftlicher, sozialer und institutioneller Perspektive – für alle knapp 2300 Städte mit Stadtrechten im deutschsprachigen Raum, ab dem Zeitpunkt ihrer Gründung.

Diese Statistiken liefern Einblicke in die gesamte Bandbreite des wirtschaftlichen und sozialen Lebens der Vergangenheit:

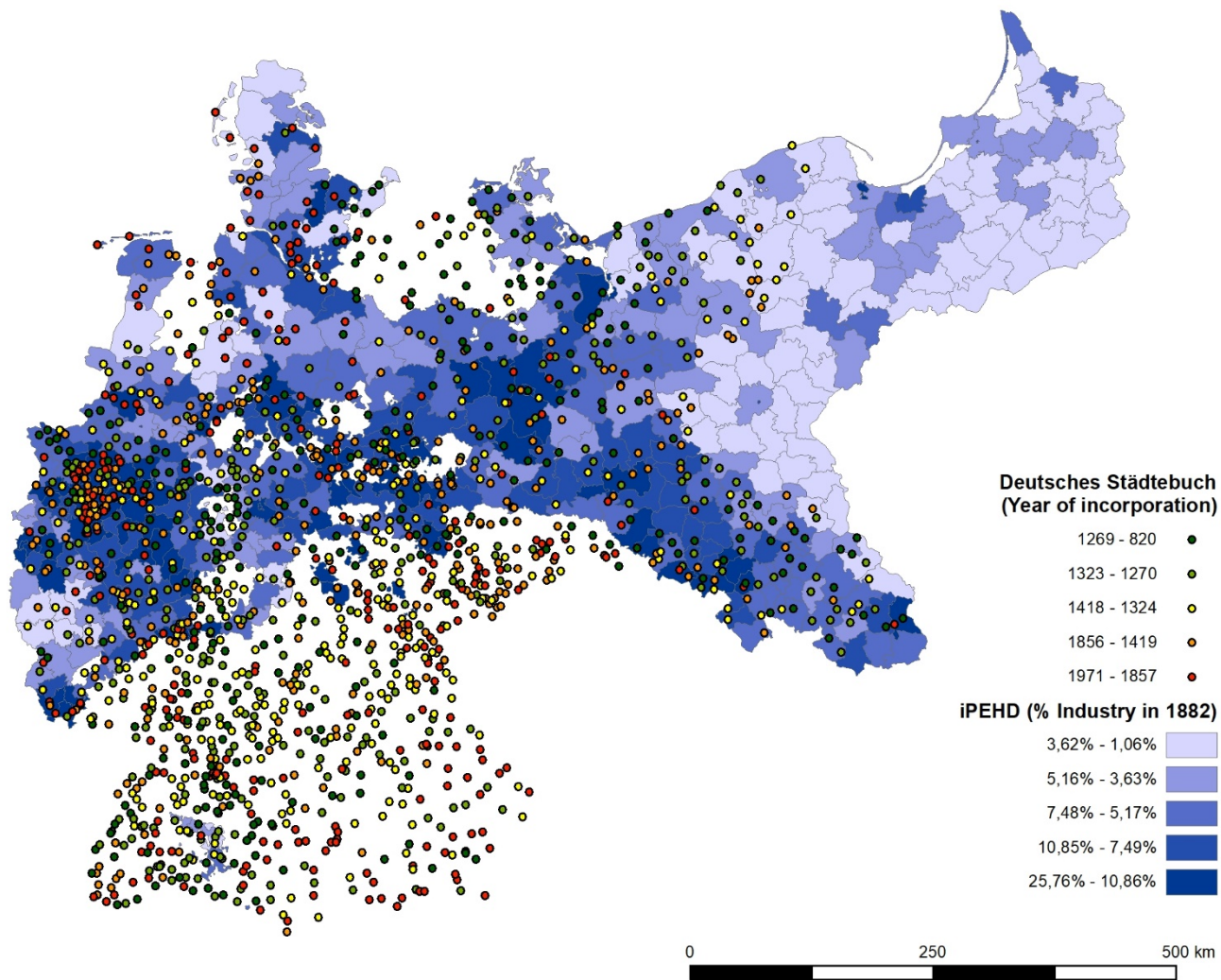
- Bevölkerung und Demographie (Bevölkerungsbewegung)
- Landwirtschaftliche Erträge, Viehzählungen, Meteorologische Daten
- Handwerkliche und industrielle Produktion (Berufs- und Gewerbezahlungen)
- Bildung, Gesundheitswesen (Medizinalstatistik) und Kirche
- Öffentliche Finanzen, Steuern
- Handel, Schiffsverkehr, Warenwirtschaft
- Kriminalität, Arbeitskämpfe
- Politische Entwicklung, Wahlergebnisse

Diese digitale Edition soll folgende technische Desiderata erfüllen:

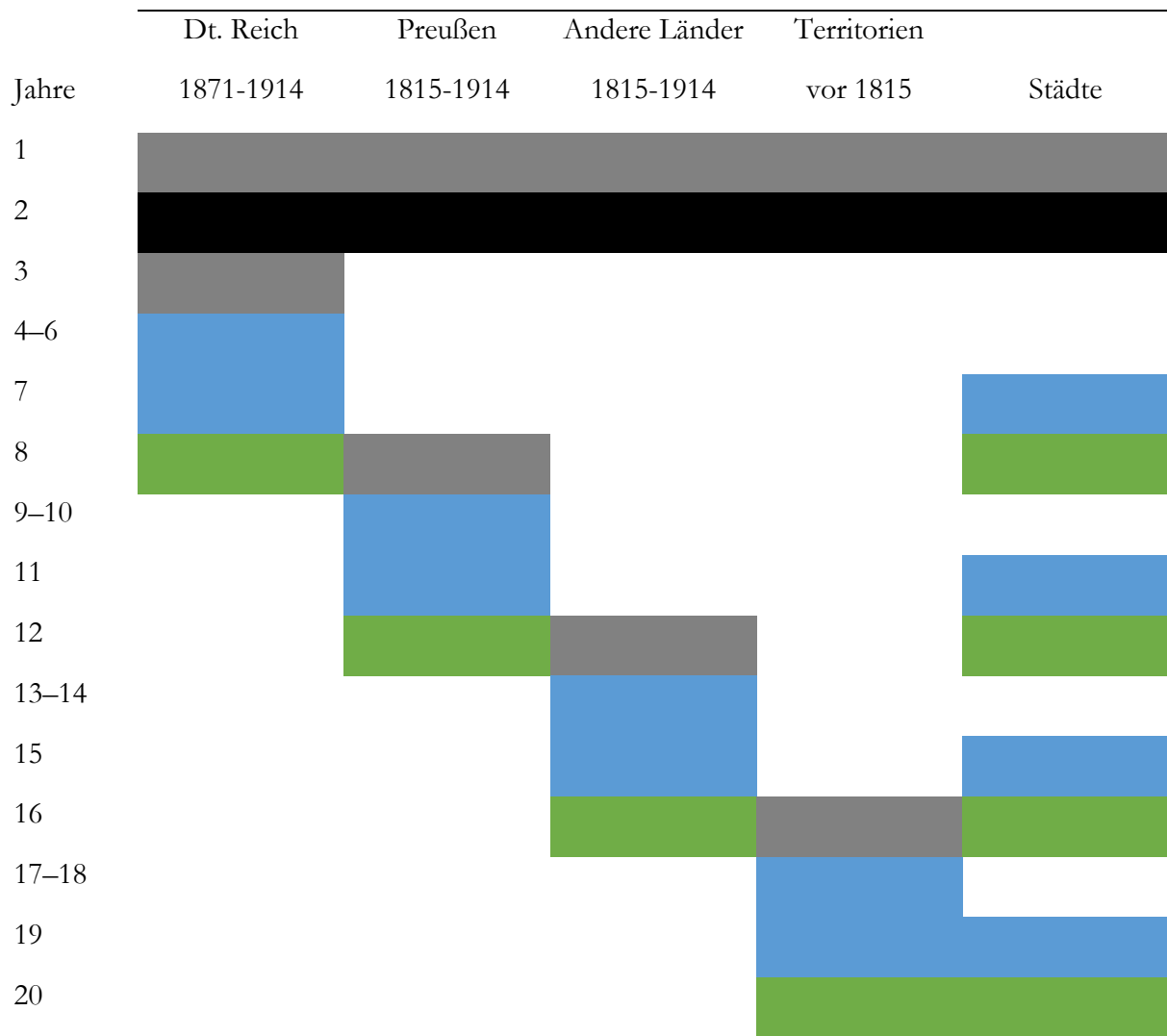
- Geokodierung aller Statistiken (Kennzeichnung von Datenpunkten durch tags, Rekonstruktion von amtlichen Grenzen);
- Möglichkeit zum Download der Daten (open access);
- dynamisch generierte Visualisierungen (Karten/Tabellen).

Das Ergebnis dieses Projekts wird eine Webseite sein, die für Wissenschaftler:innen wie für das interessierte Laienpublikum (Lokalhistoriker:innen, Journalist:innen, etc.) folgende Informationen zur Verfügung stellt:

- edierte, auf Konsistenz geprüfte, und geokodierte Statistiken zu wirtschaftlich und sozial relevanten Variablen, in der Zeitreihe und im Querschnitt;
- Möglichkeit zum Download der Daten für Forschungszwecke in allen gängigen Formaten;
- dynamisch generierte Visualisierungen (als geographische Karten oder Tabellen) der Statistiken in benutzerdefinierten Ausschnitten.



Gebiete



### **3. Historische Statistik von Deutschland**

Sibylle Lehmann (Hohenheim) / Mark Spoerer Regensburg / Jochen Streb (Mannheim)

Empirisch arbeitende Wissenschaften greifen zunehmend auch weit zurückreichende Daten, insbesondere Zeitreihen auf, um über die Beschreibung von wirtschaftlichen, sozialen, gesellschaftlichen und kulturellen Veränderungen hinaus die Ursachen und Folgen dieser Entwicklungen zu untersuchen und mögliche Kausalitäten herauszuarbeiten. An dieser Schnittstelle zwischen den historischen und den systematischen Wissenschaften wird die Einrichtung eines Online-Angebots beantragt, das neben umfangreichen Datensätzen auch wissenschaftliche Kommentierungen und ausführliche thematische „Essays“ im Open Access zur freien Nachnutzung zur Verfügung stellen wird. Darüber hinaus wird es möglich sein, die historischen Daten in Form von Diagrammen und in regional tiefer Gliederung in Form von Karten georeferenziert zu visualisieren. Im Rahmen des Projekts sollen auch neue Daten in quellenorientierter Grundlagenforschung erschlossen werden. Alle Daten, insbesondere die Zeitreihen, bleiben gegenüber Erweiterungen und Aktualisierungen offen.

Stand: beantragt über Bayerische Akademie der Wissenschaften

#### **4. Datenbankmanagement SPP 1859/SBB**

SPP 1859 (Ingo Köhler) / Staatsbibliothek zu Berlin (Christian Mathieu)

In Kooperation mit der SBB entwickelt das SPP 1859 „Erfahrungen und Erwartungen“ ein Repository, um die Speicherung, Vernetzung und Zugänglichkeit von Forschungsdaten aus ca. 30 wirtschaftshistorischen Teilprojekten zu ermöglichen. Die Bandbreite der Datenformate reicht von historischen Statistiken und Paneldaten bis hin zu georeferenzierten Vektordaten und Textmining-Analysen. Genutzt wird eine MyCoRe-Anwendung, die die notwendige Multifunktionalität von Datenarchivierung, Metadatenverwaltung sowie Datensuche und -austausch über Standard-Schnittstellen gewährleistet. Das Projekt ist als „bottom-up“ konzeptioniert, indem die SPP-Daten genutzt werden, um langfristig eine auch breiter einsetzbare Datenplattform entstehen zu lassen. Derzeit ist eine Testinstanz etabliert, mit deren Hilfe die projektspezifischen Anforderungen an das Datenmanagement (u.a. Interoperabilität, Normdaten) geklärt sowie entsprechende Umsetzungsszenarien modelliert werden.

Stand: Pilotgruppe integriert erste Daten



## 5. Historical Statistics of Germany

Universitätsbibliothek Mannheim (Irene Schumm) / Leibniz-Informationszentrum Wirtschaft (ZBW) / Ulrich Pfister (Münster) / Mark Spoerer (Regensburg)

Zusammenführung, Bereitstellung und Pflege *qualitätsgesicherter* Forschungsdaten für ein breiteres Publikum (Geistes- und Sozialwissenschaften, Organisationen, Lehrer\*innen, Journalist\*innen usw.) bezüglich langer Datenreihen, die sich auf die (bundes-) staatliche Ebene beziehen und sich zeitlich vom frühen 19. Jahrhundert bis zur Gegenwart erstrecken.

1. Aufbau einer Plattform zur Datenbereitstellung auf der Basis von CKAN.
2. Erarbeiten von fachlichen und formalen Qualitätsstandards sowie – soweit erforderlich – Erweiterung von Normdaten im Hinblick auf eine historische Statistik Deutschlands.
3. Übernahme existierender Datenreihen, wobei dies bei Bedarf eine Bearbeitung nach Maßgabe der erarbeiteten Qualitätsstandards beinhaltet. Dies ist der bei weitem umfangreichste Arbeitsschritt.
4. Einrichtung eines Zertifizierungsmechanismus zur Integration neuer Datenreihen.

Stand: Antrag kurz vor Einreichung.

## **6. Erschließung historischer Datenbestände deutscher Finanz- und Wirtschaftsunternehmen (1871 - heute)**

Leibniz-Institut für Finanzmarktforschung SAFE (Uwe Walz) / Universitätsbibliothek Mannheim / Institut für Bank- und Finanzgeschichte

Wir planen mit unserem Vorhaben die empirische wirtschaftswissenschaftliche Forschung im deutschen Kontext durch die Erschließung von historischen Datenquellen wesentlich voranzutreiben. Erreichen möchten wir dies durch die Erstellung einer Forschungsdatenbank mit Firmendaten und Finanzkennzahlen deutscher real- und finanzwirtschaftlicher Unternehmen sowie zusätzlicher interpretativer Dokumentation über einen Zeitraum von über 150 Jahren (1871 - heute). Die historischen Daten werden digitalisiert, strukturiert, verknüpft, nachhaltig gespeichert und für die Wissenschaft zugänglich gemacht.

Bedingt durch die Besonderheiten der deutschen Entwicklung im Betrachtungszeitraum (Hyperinflation, Bankencrash, etc.) eignen sich die Daten hervorragend für Forschungsarbeiten zu den großen Herausforderungen für das Finanzsystem bzw. zu den damit verbundenen wirtschaftspolitischen Maßnahmen. Ein auf dieser umfassenden Sammlung aufbauendes Forschungsprogramm, das sich auf eine große Bandbreite an finanzhistorischen bzw. aus der Langfristperspektive betrachteten Themen erstreckt, wird den enormen Nutzen der Datenbank aufzeigen.

## **7. Forschungsdaten im Journal Data Archive der ZBW**

Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte / Leibniz-Informationszentrum Wirtschaft  
(Sven Vlaeminck)

Forscher\*innen stellen in der mit Heft 2019/1 eingeführten neuen Rubrik „Forschungsdaten“ in einem etwa 10-15seitigen Aufsatz historische Daten vor, die sie für ein Projekt erhoben haben und der Forschung zur Weiternutzung zur Verfügung stellen wollen. Dabei erläutern sie die Quellen, das Vorgehen und die (Grenzen der) Interpretation der Daten.

Die Daten werden durch das Journal Data Archive des Leibniz-Informationszentrums Wirtschaft (ZBW) bereitgestellt. Die Daten sind mit einem DOI versehen, sind zitierfähig und können open access heruntergeladen werden. (<https://journaldata.zbw.eu/journals/vswg>).

Stand: Bislang sind fünf Datensätze vom Hochmittelalter bis Anfang des 21. Jahrhunderts eingestellt, zwei weitere sind in Vorbereitung. Bis Ende März 3.936 Downloads.

## **8. Digitalisierung und semi-automatisierte Erschließung der Handelskammerberichte bis 1914**

Staatsbibliothek Berlin (Christian Mathieu) / Manuel Burghardt (Leipzig) / Mark Spoerer (Regensburg)

In dem Projekt sollen (fast) alle Jahresberichte der etwa 200 Handelskammern im deutschen Raum sowie den Kolonien digitalisiert und die darin enthaltenen statistischen Übersichten, die nach einem weitgehend einheitlichen Muster aufgebaut sind, semi-automatisiert in Tabellen übertragen werden. Die Vorlagen entstammen überwiegend den sehr umfangreichen Beständen der Staatsbibliothek zu Berlin; zur Füllung der verbleibenden Lücken ist eine Kooperation mit den Universitätsbibliotheken Köln und Berlin (HU) sowie regionalen Wirtschaftsarchiven beabsichtigt.

Das Projekt hat somit zwei Komponenten:

1. Massendigitalisierungsprojekt im DFG-Programm „Digitalisierung und Erschließung“
2. Exploratives Vorhaben zur automatischen Extraktion von Tabelleninhalten im DFG-Programm „e-Research-Technologien“

Stand: Antrag kurz vor Einreichung.

## **9. Aufbau und Analyse eines Datenkorpus zur demographischen Entwicklung Deutschlands vom 16. zum 19. Jahrhundert auf der Basis einer umfassenden Aufnahme von Kirchenbüchern**

Georg Fertig (Halle) und Ulrich Pfister (Münster)

Das Vorhaben zielt erstens auf eine systematische Erschließung der Quellengattung Kirchenbuch sowie auf die digitale Erfassung einer großen Stichprobe in Zusammenarbeit mit Archiven und Bürgerwissenschaft. Darauf aufbauend soll zweitens ein Korpus von Grundlagendaten zur Bevölkerung Deutschlands vom 16. zum 19. Jahrhundert entwickelt werden. Es beinhaltet desaggregierte Informationen zu individuellen Vitalereignissen, Personennamen sowie – soweit möglich – zu Berufen und Todesursachen. Diese Informationen werden mit georeferenzierten Daten geowissenschaftlicher und historischer Art verknüpft. Auf der Basis des Korpus sollen datenintensive Forschungen zu langfristigen Wandlungsprozessen betrieben werden. Zu diesen gehören die Überwindung der sog. „malthusianischen Falle“, das Einsetzen des modernen Wirtschaftswachstums und der Fruchtbarkeits- bzw. epidemiologische Übergang.