

»Data Adaptation« als Analysemethode für geisteswissenschaftliche Forschung

Ergebnisse einer Untersuchung zur wissenschaftlichen Kommunikation in der modernen Physik

Martin Fechner, MPI für Wissenschaftsgeschichte

Die Lücke zwischen Forschung und Daten

Probleme

Die Digital Humanities unterstützen die Forschung durch Digitalisierung, Softwareentwicklung, Datenanalyse und Visualisierung. Datenbasierte Forschung kann zu vielfältigen Ergebnissen führen, wie man an den Resultaten der Klimaforschung sieht. Auch bietet das Web viele Möglichkeiten Ergebnisse zu präsentieren und zu teilen. Heute stehen die Geisteswissenschaften vor der Frage, wie sie diese Möglichkeiten nutzen können, ohne die eigenen Forschungsziele aus dem Auge zu verlieren. Die unreflektierte Nutzung der neuen Methoden oder von automatisch generierter »Big Data« läuft Gefahr die Forschung auf Fragen zu reduzieren, die von den Daten bestimmt werden. Es gibt also eine Lücke zwischen den Daten und der Forschung.

Lösungsansatz

Die Methode der »Data Adaptation« startet mit der Forschungsfrage und kombiniert in sechs Schritten Daten und Forschung. So wird eine Datensammlung auf transparente und systematische Weise erzeugt. Die »angepassten Daten« sind konsistent mit den Forschungsfragen und machen durch ihre Heterogenität und Wiederbenutzbarkeit verschiedene Auswertungen möglich.

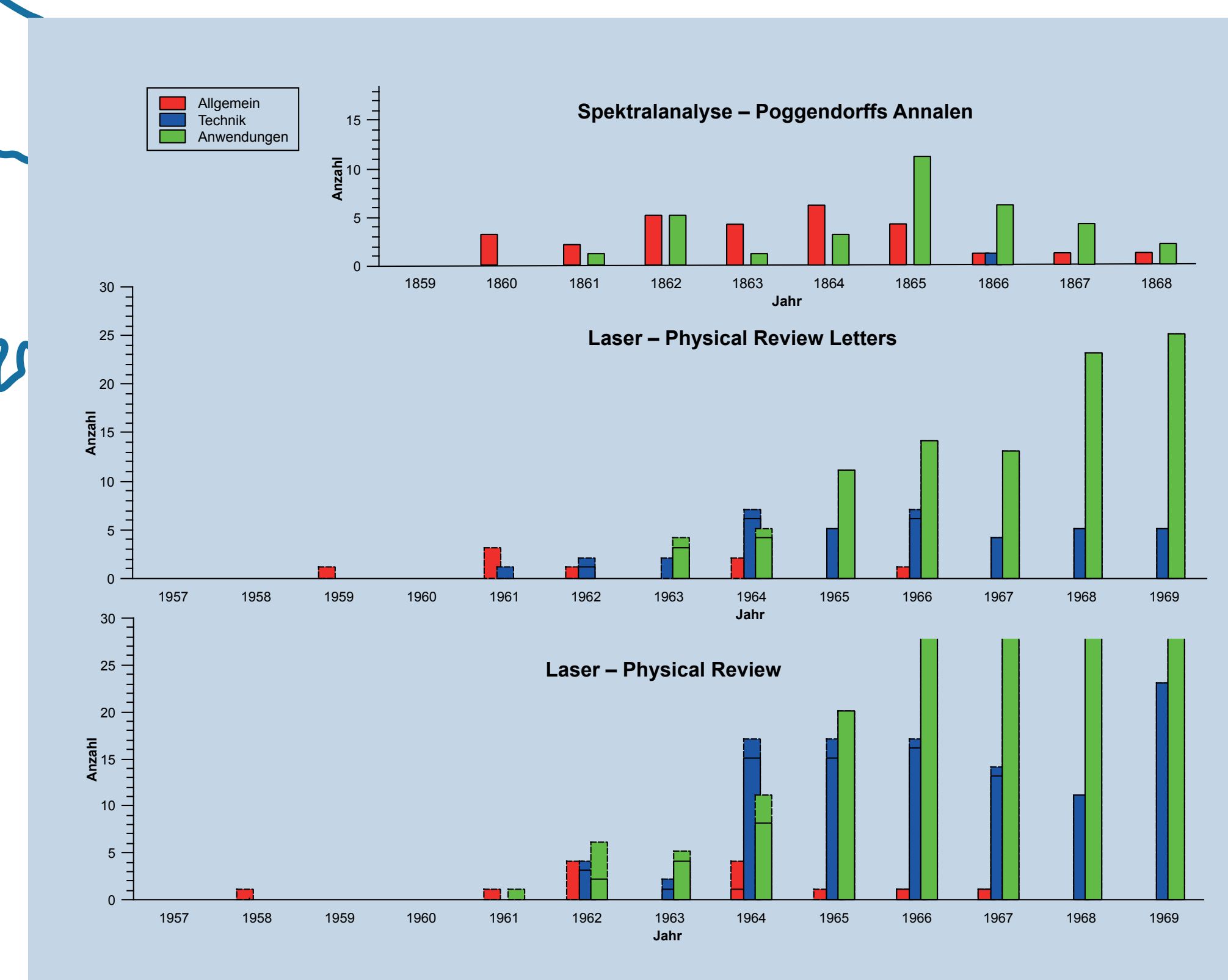
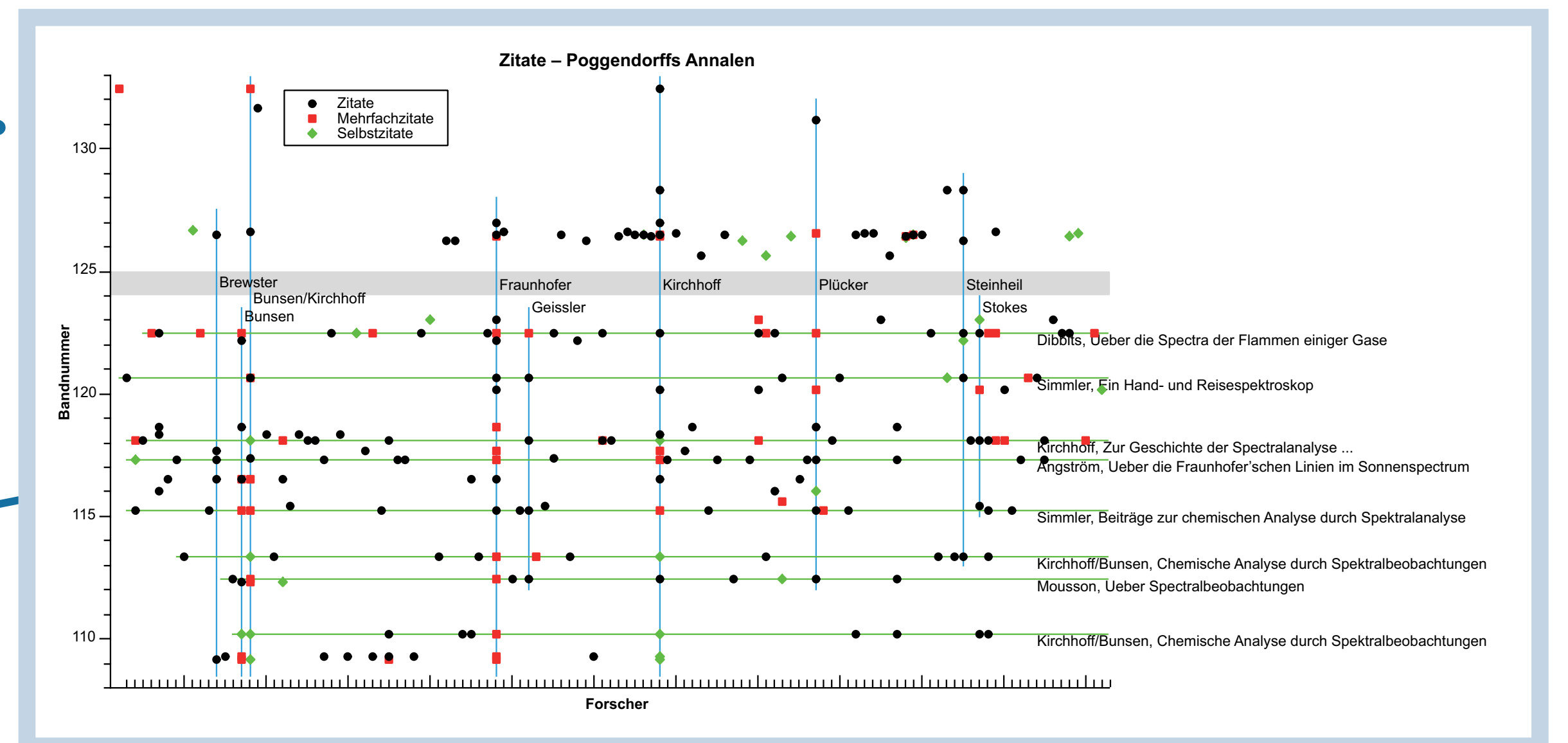
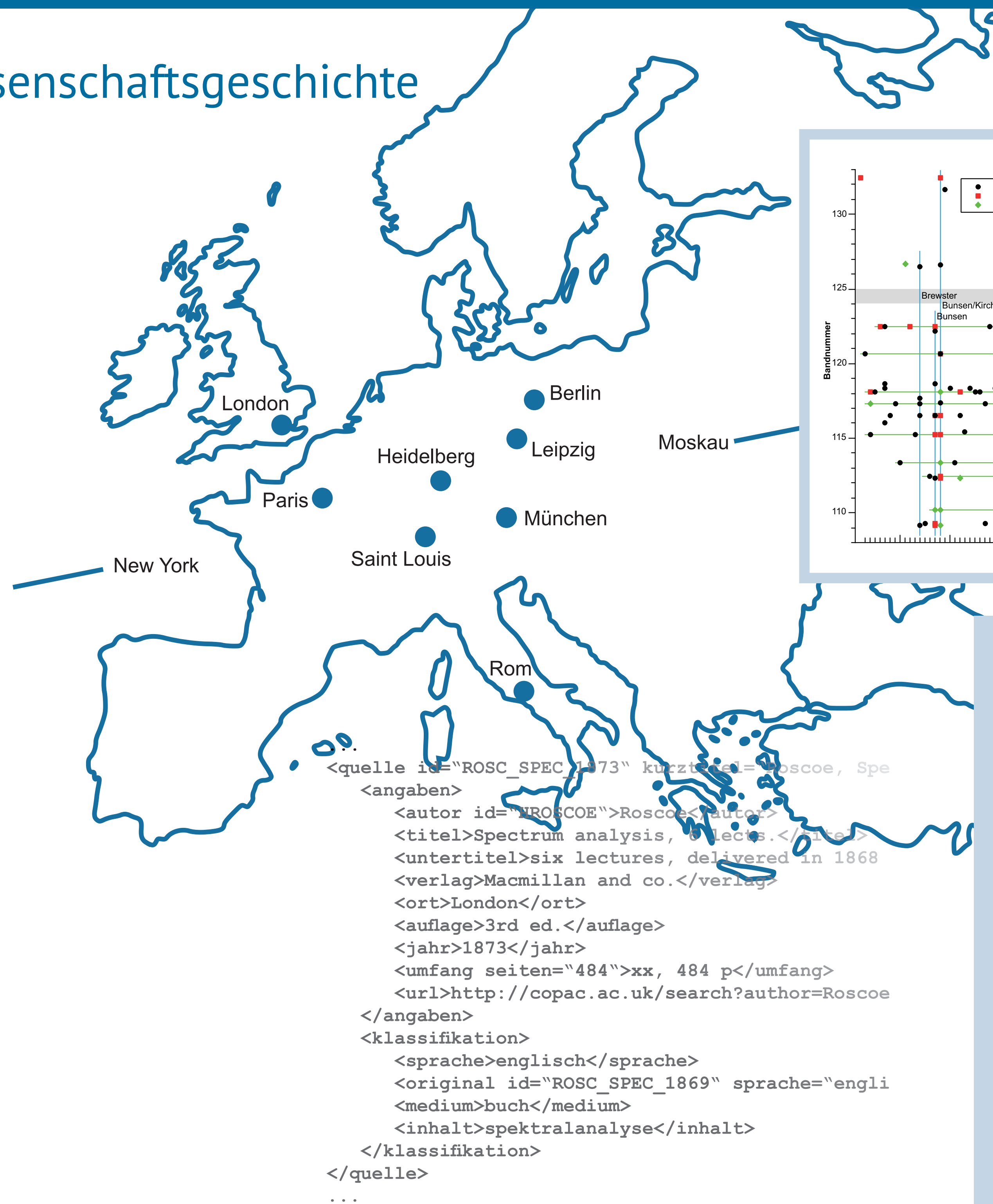
Fazit

Probleme

Schlechte Kategorisierungen oder eine unzureichende Quellenbasis können auftreten, lassen sich aber beheben, indem das Datenmodell weiter verfeinert und angepasst wird und indem weitere Quellenrecherchen angestellt werden.

Erweiterbarkeit

Die Methode der »Data Adaptation« zeigt, wie sich neue Erkenntnisse durch Datenanalyse auch im Rahmen einer hermeneutischen und geisteswissenschaftlichen Arbeit gewinnen lassen. Das mehrstufige Verfahren gibt klare Schritte zur Anwendung in anderen auf Forschungsobjekten basierten Forschungsfragen vor.



Data Adaptation

Schritt 1: Forschungsziele und Quellen

Die Forschungsfragen und die Forschungsgegenstände werden einander gegenübergestellt: Wie funktioniert wissenschaftliche Kommunikation? Wie breiten sich wissenschaftliche Nachrichten aus? Wie hängen Themen und Medienformen voneinander ab? Mögliche Quellen sind wissenschaftliche Artikel, wissenschaftliche Publikationen, Briefe, Zeitungen, Biographien. Als Fallbeispiele werden die Erfindung der Spektralanalyse im 19. Jahrhundert und des Lasers im 20. Jahrhundert gewählt.

Schritt 2: Beschreibungsmodelle

Ein theoretisches Modell muss definiert werden, mit dem sich die Forschungsfragen beantworten und in das sich die Quellen anhand ihrer Qualitäten einordnen lassen: Ein Kommunikationsraum besitzt mehrere Eigenschaften, die sich separat betrachten lassen. Jede Publikation kann hinsichtlich dieser Eigenschaften untersucht werden. Inhalt, Darstellung und Form separieren verschiedene Kommunikationsräume, während Zitationen Verbindungen aufzeigen. Eine zeitliche, geographische, personelle und institutionelle Verortung zeigt Strukturen und Prozesse auf.

Schritt 3: Datenmodell

Ein auf den Beschreibungsmodellen basiertes Datenmodell wird entwickelt. Es enthält automatisch die vorgegebene Forschungsperspektive auf die Forschungsobjekte: In einem XML-Schema können einerseits die bibliographische Angaben zu Publikationen festgehalten werden. Weiterhin können im Detail die behandelten Wissensgebiete, Darstellungsformen, Zitationen und besondere Merkmale auf festgelegte Weise näher beschrieben werden, um eine systematische Weiterverarbeitung zu erlauben.

Schritt 4: Datensammlung und -anpassung

Die Daten werden aus verschiedenen Quellen gesammelt, bereinigt und an das Datenmodell angepasst. Weiterhin werden basierend auf einem intensiven Studium der Forschungsobjekte neue Daten generiert: Angaben zu mehr als 1.500 Publikationen zur Spektralanalyse und zum Laser wurden aus biographischen Datenbanken gesammelt und thematisch klassifiziert. Mehr als 600 Zeitschriftenartikel aus drei Zeitschriften wurden im Detail analysiert und klassifiziert. Es wurden Angaben zu Zitationen, Themen und Darstellungsformen notiert.

Schritt 5: Datenexploration

Der Datenkorpus lässt sich unter Beachtung des Beschreibungsmodells im Einklang mit den anfangs formulierten Forschungsfragen explorativ erforschen. Unbekannte Zusammenhänge zwischen den Daten können sichtbar gemacht werden: Die statistische Auswertung verdeutlicht die Unterschiede zwischen Artikeln, kurzen und langen Publikationen. Die medienabhängige Ausbreitung, Spezialisierungs- und Popularisierungsprozesse werden sichtbar. Die geographische Räume bilden die historischen Strukturen und Veränderungen ab. Implizite Verbindungen zwischen Publikationen werden sichtbar und weisen auf unterschiedliche Rollen von Wissenschaftlern hin.

Schritt 6: Detailanalyse

Die explorative Analyse zeigt Stellen auf, wo ein weiteres Detailstudium der Forschungsobjekte lohnenswert sein kann: Die Publikationen, die die wissenschaftlichen Diskurse zeitlich und geographisch verbinden oder trennen, werden sichtbar und können separat studiert werden. Den Popularisierungsprozessen und den Hinweisen auf die verschiedenen Rollen von Wissenschaftlern, Autoren und Forschern kann nachgegangen werden.