The database of the ISMI project

Opportunities and challenges of research-driven development in the humanities

Robert Casties, Dirk Wintergrün

The Islamic Scientific Manuscript Initiative (ISMI)

The project's mission is to collect and make accessible information on all Islamic manuscripts in the exact sciences (astronomy, mathematics, optics, mathematical geography, music, mechanics, and related disciplines), whether in Arabic, Persian, Turkish, or other languages from the 8th to the 19th century.

The collected data contains basic bibliographic information but also paleographic and codicological information, and information about the content of the texts and the attested uses of and the persons using the manuscripts.
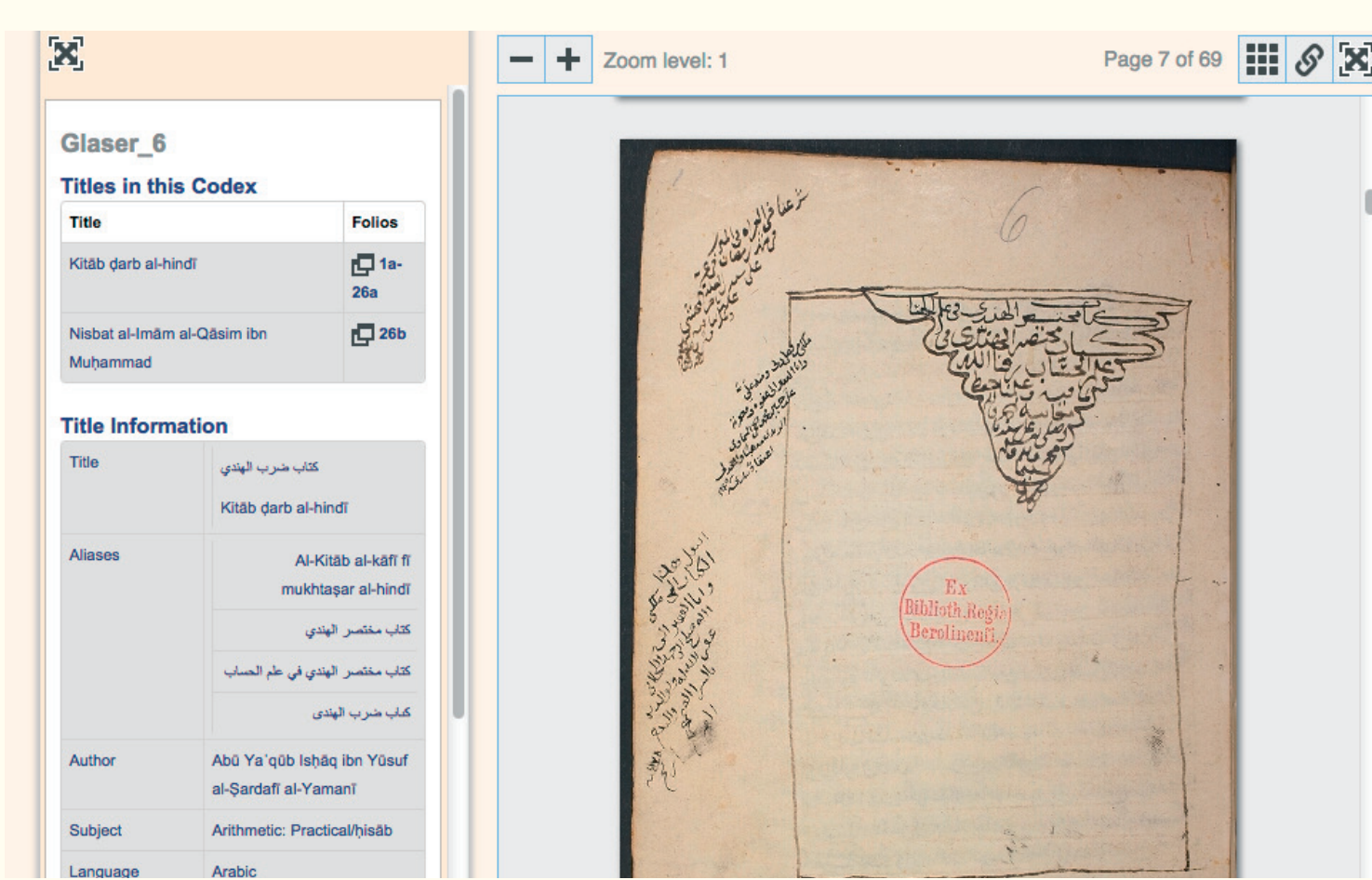

Web display of manuscript *Kitāb ḍarb al-hindī* in codex *Glaser 6* (Staatsbibiliothek Berlin)

Since 2007 the project has collected information on more than 4000 texts existing in 14500 witnesses in 7500 codices, as well as information on over 2000 persons. In a cooperation with the MPIWG library and the Staatsbibliothek Berlin over 1000 codices have been scanned in high resolution and made available to the public.

The database as a network of data objects

Cataloguing old manuscripts quickly overwhelms standard bibliographical databases. Among the problems are unknown titles and anonymous authors, a proliferation of authors with the same name and variations of different names for the same author and generations of copies of the same text in different versions.

The problem was re-examined from the ground up together with the scholars in 2006 and a new data model was born based on objects representing conceptual entities and flexible relations creating together a network of data objects.
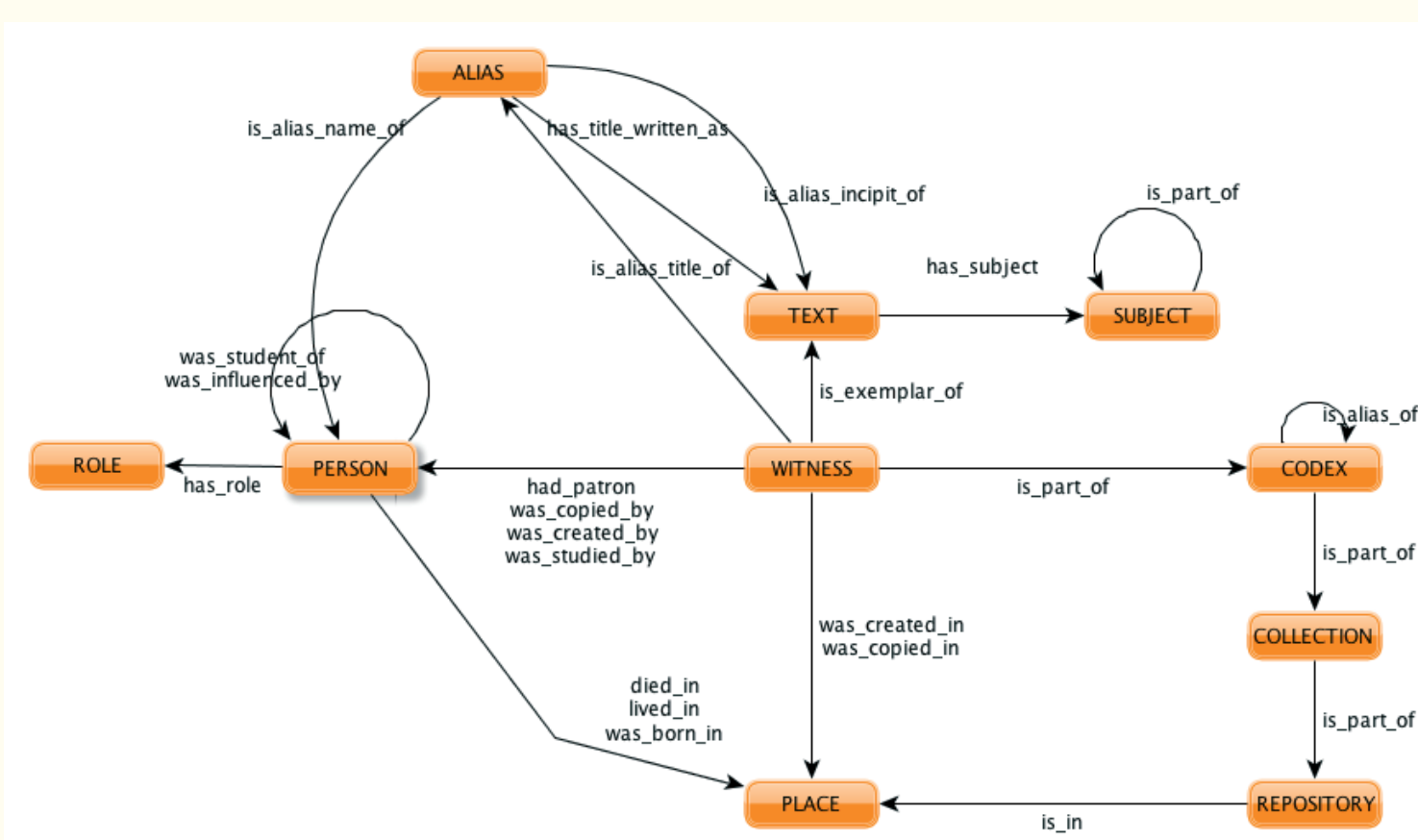
Data model of the ISMI database showing conceptual object types and their relations.

Thinking about the data in this way led to many elegant solutions. In the new model a title is an abstract text that is connected to all the witnesses/manuscripts that are exemplars which in turn are parts of codices in collections. Persons can be related to texts and witnesses not only as authors but also as copyists, readers, annotators or sponsors.

The ability to constantly evolve the data model and establish new relation types has broadened the scholar's view of the historical material and its representation and the research questions the data will be able to answer.
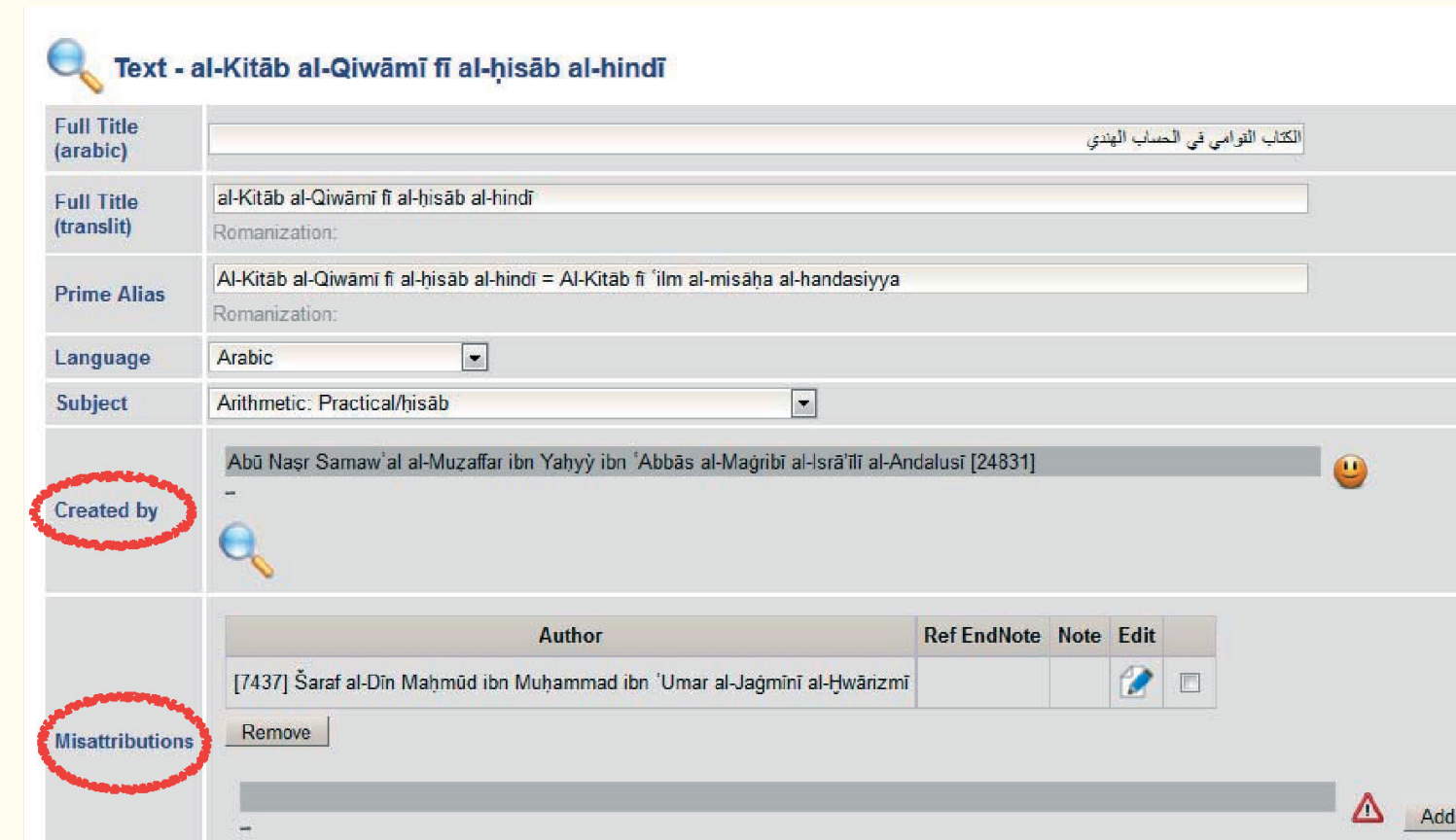

Example of authorship misattribution for a text.

Representing wrong information

A requirements that arose in the project early is the need to record and present outdated information. In many cases authorship of texts has been misattributed widely and for long times so that scholars coming to the database looking for specific information may search under a wrong name or assume the database to be in error unless the misattribution is also presented with arguments why the old information is superseded.

Going beyond catalogue information

In the cases where a manuscript can be examined at first hand or as a digital facsimile a lot of additional information can be learned: there are notes and seals of ownership, notes documenting the transfer of ownership or the study of the manuscript and additional annotations to the text. All this information gives important insights into the social and historical context of the knowledge and the historical actors.
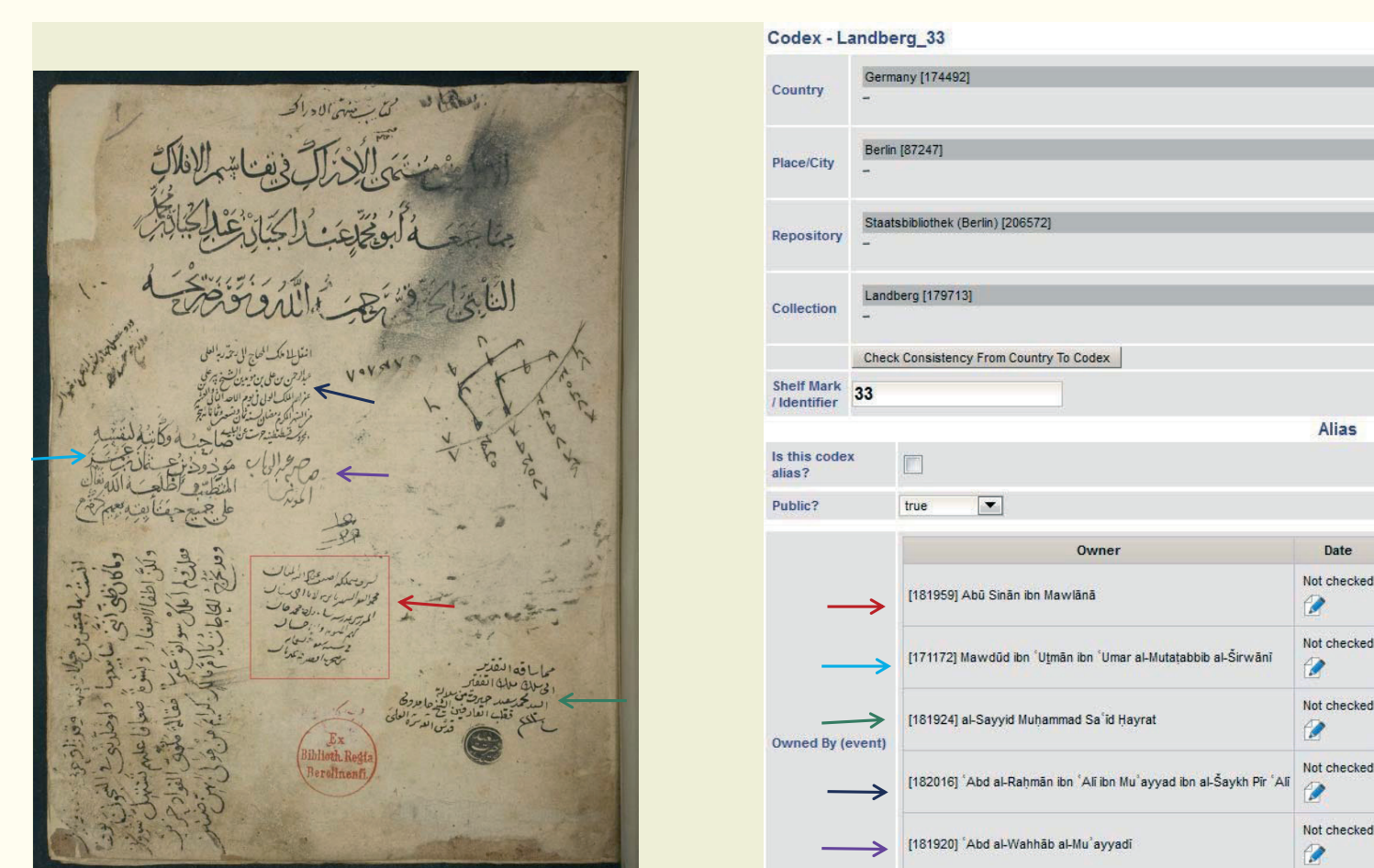

Ownership notes in codex *Landberg 33 (Staatsbibliothek Berlin)* and their representation in the database.

Ownership notes without additional information are represented as *was_owned_by* relations connecting the codex object with the person objects representing its owners.

Adding events to the database

The flexible nature of the relations made it easy to introduce new objects and relation types as it became necessary in the research process, for example to record the documented transfer of ownership or the study of a manuscript.
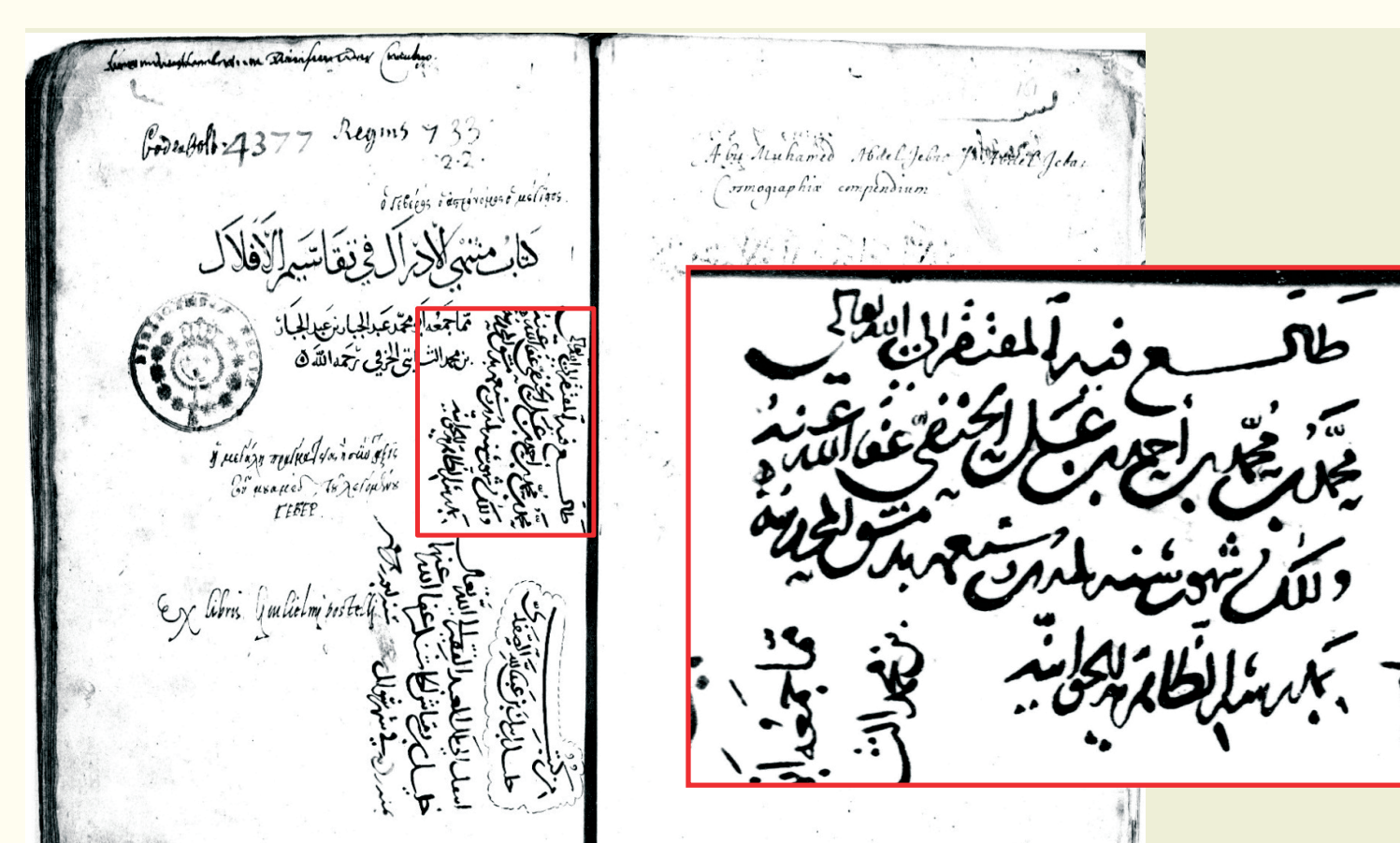
Study event in *BNF Arabe MS2499: Muḥammad b. Muḥammad b. Aḥmad b. ʿAlī al-Ḥanafī studied this at the Ẓāhiriyya Juwāniyya madrasa 730 H. [1329-30] in Damascus*

A study event object connects the witness that was studied with the person studying and the place and institution where this took place as well as the date of the event.

Challenges of software development

The concept of a network of objects with flexible relations is also called a *directed attribute-graph* and it exists in widely used database products like *Neo4J* today but those were not available in 2006 which led to the development of the custom database software *OpenMind*. The software is Open Source, written in Java, and uses a conventional SQL database backend and a Web-based frontend.

OpenMind is in use today for data entry and display together with the *Diva.js* viewer for manuscript display and *Drupal* for managing the bibliography of secondary literature and public information pages.

The continuing maintenance and development of the custom software pieces places a growing burden on the project's limited development resources.

Challenges of networked data

The network-like structure of data in the database makes it easy to add new relation types or new attributes to objects. At the same time it is more difficult to create simple forms for entering data for things that are composed of multiple objects in the data model.
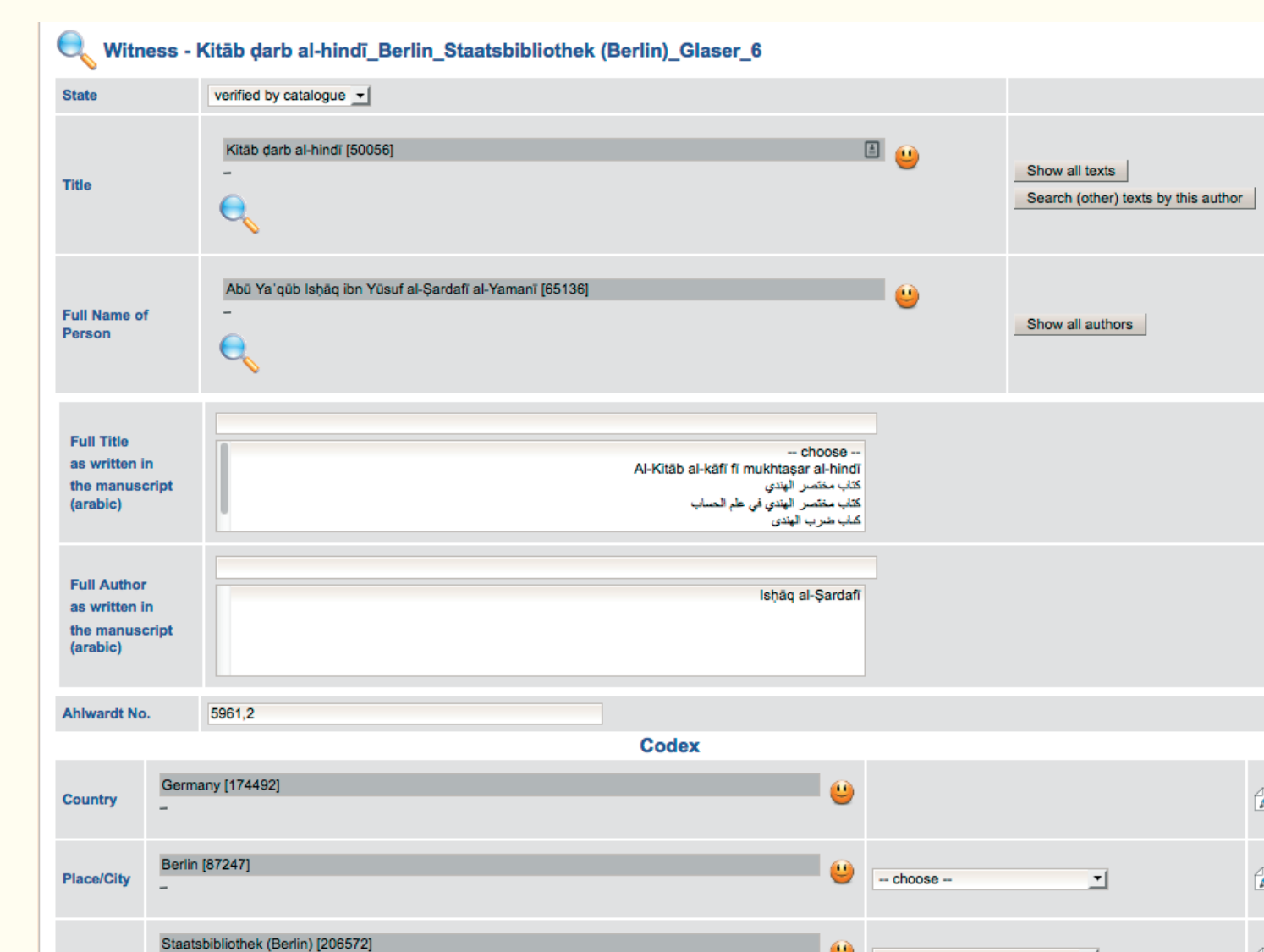
Part of witness entry form, editing attributes of the witness object like *Full title as written in manuscript* and relations like *Title (is_exemplar_of)* and *Full Name of Person (was_created_by)*.

The form for entering the information for a new manuscript not only creates a witness object and fills its attributes but it also creates relations to a text object, multiple person objects, a codex object, a collection object, a library object and a place object, creating those objects if they do not exist.

Working with networks of data objects

Visualising and querying this kind of networks of data objects is a new area where few established tools and concepts exist.

The data has been exported from *OpenMind* into *Neo4J* to use the more readily available support for this platform and some experimental visualisations and query tools have been developed.

The applicability of established network-analytical methods is evaluated in cooperation with research in Department I of the MPIWG.

Expert scholars in the field are invited in workshops to explore new tool prototypes and give feedback for the ongoing development of new software and new research questions.
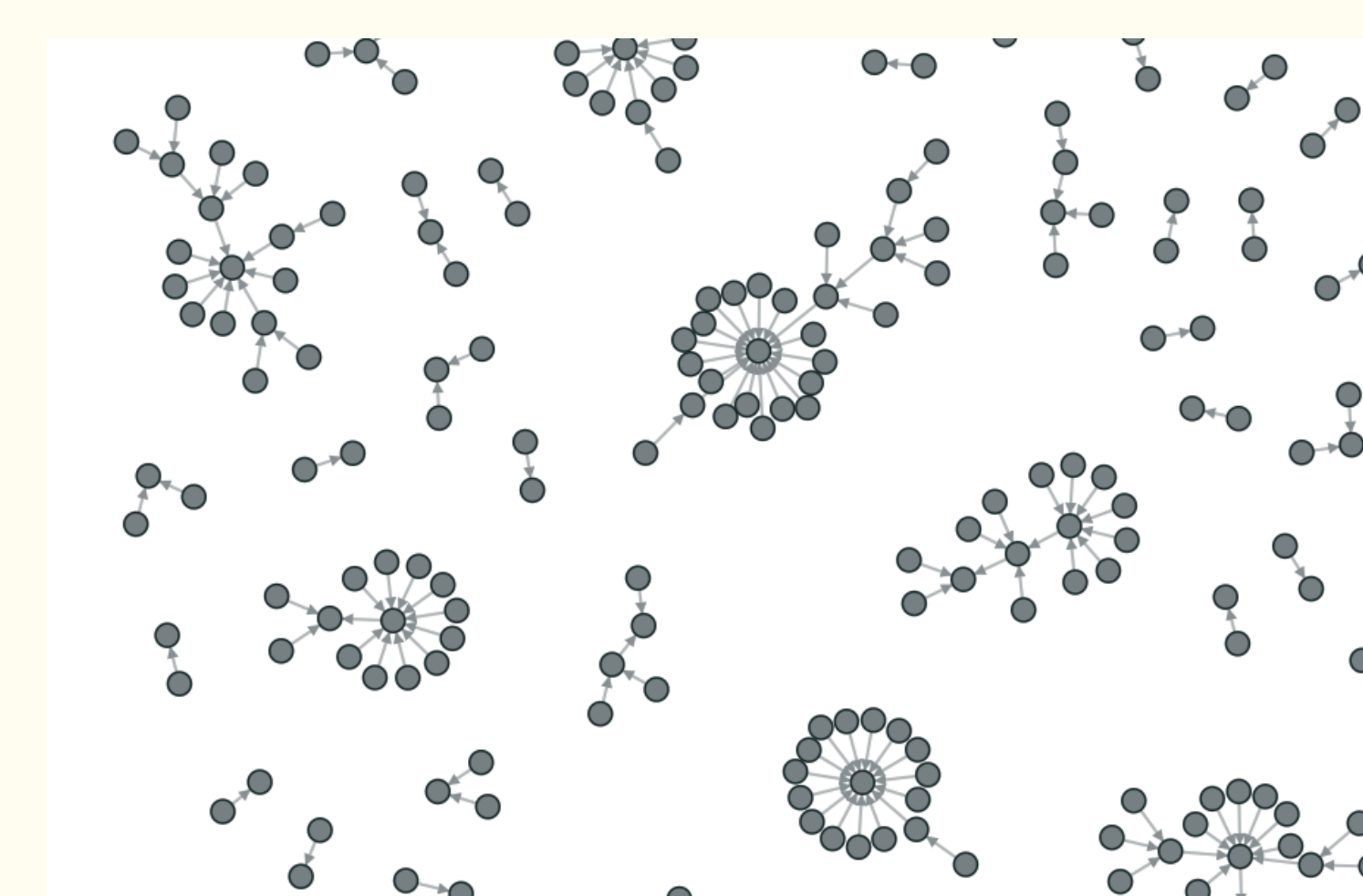
Partial graph of commentary relations between texts in the ISMI database.

Cooperation Partners
F. Jamil Ragep, Sally P. Ragep,
Institute of Islamic Studies, McGill University

The **Islamic Scientific Manuscript Initiative** is a cooperation project of the Department of Lorraine Daston at the Max Planck Institute for the History of Science (MPIWG) and the Institute of Islamic Studies (IIS) at McGill University in Montreal, Canada.