Publishing Information

Natural Sciences, Technology and Informatics in Manuscript Analysis
Edited by Oliver Hahn, Volker Märgner, Ira Rabin, and H. Siegfried Stiehl

Proceedings of the third International Conference on Natural Sciences and Technology in Manuscript Analysis and the workshop OpenX for Interdisciplinary Computational Manuscript Research that took place at the University of Hamburg, Centre for the Study of Manuscript Cultures, on 12–14 June 2018.
CONTENTS

3 | Editorial
  by Oliver Hahn, Volker Märgner, Ira Rabin, and H. Siegfried Stiehl

ARTICLES

5 | On Avoiding Segmentation in Handwritten Keyword Spotting: Overview and Perspectives
  Marçal Rusiñol*

11 | Writer Identification and Script Classification: Two Tasks for a Common Understanding of Cultural Heritage
  Dominique Stutzmann, Christopher Tensmeyer, and Vincent Christlein*

25 | Z-Profile: Holistic Preprocessing Applied to Hebrew Manuscripts for HTR with Ocropy and Kraken
  Daniel Stökl Ben Ezra and Hayim Lapin*

37 | On Digital and Computational Approaches to Palaeography: Where Have we Been, Where Are we Going?
  Peter A. Stokes*

47 | Creating Workflows with a Human in the Loop for Document Image Analysis
  Marcel Gygli (Würsch), Mathias Seuret, Lukas Imstepf, Andreas Fischer, Rolf Ingold*

53 | Building an Evaluation Framework Researchers Will (Want to) Use
  Joseph Chazalon*

61 | Turning Black into White through Visual Programming: Peeking into the Black Box of Computational Manuscript Analysis
  Vinodh Rajan Sampath and H. Siegfried Stiehl*

73 | Legally Open: Copyright, Licensing, and Data Privacy Issues
  Vanessa Hannesschläger*

77 | A Comparison of Arabic Handwriting-Style Analysis Using Conventional and Computational Methods
  Hussein Mohammed, Volker Märgner, and Tilman Seidensticker

87 | Illuminating Techniques from the Sinai Desert
  Damianos Kasotakis, Michael Phelps, and Ken Boydston

91 | Image Quality in Cultural Heritage
  Tyler R. Peery, Roger L. Easton Jr., Rolando Raqueno, Michael Gartley, and David Messinger

105 | When Erased Iron Gall Characters Misbehave
  Keith T. Knox

115 | ‘Dürer’s Young Hare’ in Weimar – A Pilot Study
  Oliver Hahn, Uwe Golle, Carsten Wintermann, and Ira Rabin
123 | Material-Techical Details on Papyrus as Writing Support
    Myriam Krutzsch

133 | The Techniques and Materials Used in Making Lao and Tai Paper Manuscripts
    Agnieszka Helman-Ważny, Volker Grabowsky, Direk Injan and Khamvone Boulyaphonh

163 | Inks Used to Write the Divine Name in a Thirteenth-Century Ashkenazic Torah Scroll: Erfurt 7 (Staatsbibliothek zu Berlin, Or. fol. 1216)
    Nehemia Gordon, Olivier Bonnerot, and Ira Rabin

185 | Contributors
Legally Open: Copyright, Licensing, and Data Privacy Issues

Vanessa Hannesschläger | Vienna

Abstract

In the field of digital science, the methods and approaches of open science are gaining momentum. A vital precondition for applying these methods is knowledge of various aspects of the legal landscape, which this paper aims to address. Specifically, it will discuss the topics of copyright/intellectual property rights (IPR) in an international context, the possibilities and pitfalls of open licensing and the legal restrictions brought about by the European Union’s new data privacy legislation (GDPR).

1. Preliminary remark: disclaimer

Please note that this contribution is not professional legal advice, but a mere collection of thoughts and information compiled from legal sources.

Readers might be wondering why this is explicitly pointed out here, and they would be right to do so: in fact, it is not advisable to include a disclaimer in an online presence (or any sort of publication, really). The reason is that if rights are infringed, the existence of the disclaimer will not affect the judgement of or sentence resulting from the infringement. However, the disclaimer might inspire imposters or predatory companies that specialize in (justified or unjustified) intellectual property right claims to contact the disclaimant and lodge excessive claims. The best way to avoid IPR claims is to avoid using material whose rights holders are undetermined and to make sure that appropriate credit is always given (e.g. when using a licensed image, its license and all other information it requires the user to include should be explicitly stated).

2. Copyright/intellectual property rights (IPR)

The legal frameworks we are embedded in define if, how and how long texts, material, meta-/data and software can be made (and kept) available. The most relevant area of legislation affecting digital research, and especially computational manuscript research, is copyright. In this context, researchers are always in a Janus-faced position: as creators of content, on the one hand, and as (re-)users of content, on the other. The two sides of the copyright coin are expressed by the Universal Declaration of Human Rights (UDHR), which states, ‘Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits,’ but goes on to say that everyone also ‘has the right to the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he [sic] is the author’.¹ It has even been argued that, due to ‘the requirement for researchers to make their publicly funded work available to the public’,² copyright is an unsuitable legal structure for scientific works. Scientific norms guide scientists to reproduce and build on others’ research, and default copyright law by its very purpose runs counter to these goals.³ Still, copyright law is a reality that contemporary research communities have to face. Especially for researchers working with computational methods, the ‘increasingly rapid development of new media continuously leads to new and unanticipated ways of distributing copyrighted works’⁴ — which researchers both as creators and as users. Knowledge of the basic aspects of national copyright law in the country where the research is carried out is vital in both roles, as it affects the possibilities of working with existing data, on the one hand, and the rights to distribute and be compensated for one’s own work, on the other.

¹ UDHR, Art. 27 no. 1 and 2.  
² Stodden 2009, 40.  
³ Stodden 2009, 35.  
⁴ Darling 2012, 485.
A second topic of interest in this context is the main principles of national and international copyright legislation. In a European context, employment of the term ‘copyright’ itself is already problematic, as it refers to a concept from the Anglo-American legal tradition: copyright primarily aims at regulating the right to replicate and reproduce. However, in most European countries, the Germanic legal tradition, which puts a stronger focus on the persona of the creator (‘Urheberrecht’, ‘droit d’auteur’), has shaped ‘copyright’ legislations. Within (most of) Europe, it is therefore more accurate to speak of ‘intellectual property rights’ (IPR) rather than ‘copyright’ law. However: ‘As IP law in the European Union is merely harmonized and not unified, the exact scope of copyright and similar rights may differ between Member States (e.g. some Member States recognize an exclusive right for ‘scientific and critical editions’, while others don’t).’

It is crucial for researchers to be aware of this fact. However, it is also crucial to know that the EU has taken a first step in the direction of IP law harmonization by accepting the Copyright in a Digital Single Market Directive (COD) proposal, by means of which IPR is supposed to be modernized and adapted to the realities of the digital world throughout the EU. From the point of view of the research community, this proposal has its upsides and downsides; while it does aim to implement a general permission of text and data mining in a scientific context, the articles on ‘Protection of press publications concerning digital uses’ (‘link tax article’) and ‘Use of protected content by information society service providers storing and giving access to large amounts of works and other subject-matter uploaded by their users’ (‘upload filter article’) are being discussed controversially. As the directive proposal was approved by the EU Parliament in September 2018, but has not yet been formalized, the effects on the legal conditions of research in Europe are not yet clear.

3. Open licensing
Due to the territorial limitations of copyright, the digital space that transcends national borders calls for new legal arrangements that are able to protect the researchers’ rights, on the one hand, and ensure the reusability of their work, on the other. Open licensing models enable long-term preservation of and international research on data collected in local research projects, thus greatly supporting emerging open approaches in computational manuscript research. However, scholars often lack an overview of the various possibilities to license their findings. The most established model, which has gained great popularity for creative content and is increasingly also applied to research data, is Creative Commons licensing (CC). The use of widely known licenses such as the ones provided by CC is advisable (and preferable to writing one’s own individual licenses) because it will enable others to understand the conditions under which material is available immediately (without having to read a complicated legal text). Although CC has become a de facto standard for licensing research data, scholars are often unaware of the details of the different CC modules and their consequences; choosing appropriate licenses for software is an even more complex task. Therefore, awareness of the available options of ready-made open licenses and their benefits and potential pitfalls and of license selection (such as license compatibility issues, copyright preconditions and other legal commitments such as work contracts) is crucial.

3.1 The Public License Selector
Creative Commons offers a basic license selection tool that is helpful for researchers who are already sure that a) their content is licensable under CC and b) they have made a conscious decision to use CC. However, in some cases, CC licenses might not be the best choice, for example in the case of code. A very nifty tool that helps select appropriate open licenses for both data and/or code is the Public License Selector developed by the European research infrastructure consortium CLARIN ERIC. Users start with a total selection of 22 open, publicly available, ready-made licenses and have to answer a sequence of questions. Each answer narrows down the licenses compatible with the respective preconditions, leaving the user with a final choice of open licenses suited to their specific situation (as well as further information about the individual qualities of all available licenses) at the end of the process.

4. Data privacy and the General Data Protection Regulation (GDPR)
As a third main topic, this paper addresses the EU’s General Data Protection Regulation (GDPR), which came into effect on May 25, 2018. As the GDPR is a regulation, it took legal

---

5 Kamocki, Stranák, and Sedlák 2016, 2534.
6 COD, Art. 3.
7 COD, Art. 11.
8 COD, Art. 13.
effect in all EU member states immediately on the day of implementation (in contrast to a mere directive such as the COD, which has to be submitted to and approved by national legislatures before going into effect). The GDPR replaces the EU’s Personal Data Directive (1995). Although the GDPR does not differ from the Personal Data Directive in terms of fundamental concepts, it does establish a few new requirements, as well as tangible punishments (penalties) in case of infringement. While the main aim of the GDPR is to protect citizens and individuals from abuse of their personal information by international corporations, it affects everyone working in digital space (despite several ‘research exceptions’ such as archiving in the public interest).

Hence, the main concepts of the GDPR are outlined below, explaining the most vital points to be considered in the context of computational manuscript research.

4.1 ‘Data subjects’

Although it might seem obvious to many, not everyone is aware of this basic, but crucial fact: the GDPR applies only to natural persons. This means that the GDPR does not apply to legal bodies (i.e. institutions), but only to real people. It also means that it applies only to living people, not historical ones. As manuscript researchers work with data about dead people more often than with data about living people, this might be a relief for some. However, while the GDPR does not protect data about dead people, other (national) laws might still restrict the collection and publication of data about dead people (this is especially true for relatively recently deceased persons).

The GDPR defines the rights that data subjects have with regard to their personal data. These include the right to information (e.g. about the data themselves, their processing and its purpose, their storage and its duration, their accessibility and their protection), the right to access the data (regardless of whether the data subject was the provider of the data or they were taken from elsewhere) and to rectify them if necessary, to restrict their processing or object to it, and, most importantly, the right to erasure. These rights of the data subject can be restricted to a certain extent by the research exceptions defined by the GDPR. However, as this regulation is so relatively new and no court rulings about uncertainties in the GDPR are yet available, it is not advisable to build a research project that relies on these exceptions.

4.2 ‘Personal data’

According to the GDPR, ‘personal data’ is ‘any information relating to an identified or identifiable natural person’. This means any information about a person that might enable someone to identify that person. Examples include name, date of birth, age, sex, gender, address, pictures of the person, audio recordings of the person’s voice etc. Be aware that a combination of data might make a person identifiable even if the person’s name is not included in the data collection; e.g. the information that one of the data subjects is male, 80 years old and lives in a village with 300 inhabitants will likely make the data subject identifiable. Also be aware that the form or format of the data is irrelevant. All personal data are protected by the GDPR, even if it is only stored in handwriting on a slip of paper in a desk drawer.

Pseudonymised data (i.e. data that cannot be directly associated with an identifiable individual in their current state, but that could still be transformed into a state that makes it possible to reconstruct whom the data describes) is considered personal data by the GDPR. Anonymised data (i.e. data that cannot be associated with an identifiable individual in their current state, which must be irreversible) is not personal data according to the GDPR.

4.3 ‘Processing’

According to the GDPR, ‘data processing’ is ‘any operation or set of operations which is performed on personal data’. This means that any action done with data is part of the processing – even collecting, storing and deleting data. This is also true if the data is only in analogue form.

4.4 What does the GDPR require a researcher to do?

Researchers are likely to find themselves in the position of a ‘data controller’ (a person who ‘determines the purposes and means of the processing of personal data’) or a ‘processor’ (a person who ‘processes personal data on behalf of the controller’).

9 See GDPR, Art. 6, Art. 89, and Rec. 50, Rec. 47, Rec. 113, Rec. 157, Rec. 159.

10 GDPR, Art. 4 no. 1.

11 GDPR, Art. 4 no. 2.

12 GDPR, Art. 4 no. 7.

13 GDPR, Art. 4 no. 8.
If work with personal data is carried out in a research project, researchers have to make sure to follow the data protection principles the GDPR defines. These include lawfulness (data subjects must give explicit consent to the collection and processing), fairness, transparency (data subjects must be informed about the ways and purposes of data collection and processing), purpose limitation (data can be collected and subsequently used only for purposes specified in advance), data minimization (only the data that are necessary for the defined purpose may be collected), storage limitation (data may be stored only as long as necessary for the defined purpose), accuracy, integrity, confidentiality and accountability (the data controller has to be able to demonstrate that all these principles are met). In addition, the data controller is obliged to protect the data and keep records of all processing activities.

5. Conclusion
Understanding the legal frameworks within which research operates is a vital skill for researchers in the digital age. Not only is it important to understand in what ways one is allowed to work with the findings and material provided by others: researchers also need to understand their own rights to their research in order to open it up for others. A good way to do that is to provide open licenses. In addition to considering intellectual property laws, awareness of data privacy legislation is vital in order to ensure that operations with information about users, data contributors and research subjects are carried out in a legal manner. While the EU has already provided a uniform framework for the latter, the development of the former will have to be observed on a European level over the next years.

REFERENCES

‘Creative Commons’ (CC) <https://creativecommons.org/> (accessed September 26, 2018).


Written Artefacts as Cultural Heritage

Ed. by Michael Friedrich and Doreen Schröter

Written Artefacts as Cultural Heritage was established in 2020. The series is dedicated to the double role of written artefacts as representations and generators of humankind’s cultural heritage. Its thematic scope embraces aspects of preservation, the identity-defining role of artefacts as well as ethical questions.

The mix of practical guides, colloquium papers and project reports is specifically intended for staff at libraries and archives, curators at museums and art galleries, and scholars working in the fields of manuscript cultures and heritage studies.

Every volume of Written Artefacts as Cultural Heritage has been peer-reviewed and is openly accessible. There is an online and a printed version.

If you wish to receive a copy or to present your research, please contact the editorial office: https://www.csmc.uni-hamburg.de/publications/cultural-heritage.html
CSMC’s academic journal was established as newsletter of the research unit ‘Manuscript Cultures in Asia and Africa’ in 2008 and transformed into a scholarly journal with the appearance of volume 4 in 2011. *manuscript cultures* publishes exhibition catalogues and articles contributing to the study of written artefacts. This field of study embraces disciplines such as art history, codicology, epigraphy, history, material analysis, palaeography and philology, informatics and multispectral imaging.

*manuscript cultures* encourages comparative approaches, without regional, linguistic, temporal or other limitations on the objects studied; it contributes to a larger historical and systematic survey of the role of written artefacts in ancient and modern cultures, and in so doing provides a new foundation for ongoing discussions in cultural studies.

Every volume of *manuscript cultures* has been peer-reviewed and is openly accessible: https://www.csmc.uni-hamburg.de/publications/mc.html

If you wish to receive a copy or to present your research in our journal, please contact the editorial office: irina.wandrey@uni-hamburg.de
Studies in Manuscript Cultures (SMC)

Ed. by Michael Friedrich, Harunaga Isaacson, and Jörg B. Quenzer

From volume 4 onwards all volumes are available as open access books on the De Gruyter website:
https://www.degruyter.com/view/serial/43546
https://www.csmc.uni-hamburg.de/

23 – Education Materialised: Reconstructing Teaching and Learning Contexts through Manuscripts, edited by Stefanie Brinkmann, Giovanni Ciotti, Stefano Valente and Eva Maria Wilden

Manuscripts have played a crucial role in the educational practices of virtually all cultures that have a history of using them. As learning and teaching tools, manuscripts become primary witnesses for reconstructing and studying didactic and research activities and methodologies from elementary levels to the most advanced.

The present volume investigates the relation between manuscripts and educational practices focusing on four particular research topics: educational settings; teachers, students and their manuscripts; organising knowledge: syllabi; exegetical practices: annotations; modifying tradition: adaptations.

The volume offers a number of case studies stretching across geophysical boundaries from Western Europe to South-East Asia, with a time span ranging from the second millennium BCE to the twentieth century CE.

22 – Dunhuang Manuscript Culture: End of the First Millennium, by Imre Galambos

Dunhuang Manuscript Culture explores the world of Chinese manuscripts from ninth–tenth century Dunhuang, an oasis city along the network of pre-modern routes known today collectively as the Silk Roads. The manuscripts have been discovered in 1900 in a sealed-off side-chamber of a Buddhist cave temple, where they had lain undisturbed for for almost nine hundred years. The discovery comprised tens of thousands of texts, written in over twenty different languages and scripts, including Chinese, Tibetan, Old Uighur, Khotanese, Sogdian and Sanskrit. This study centres around four groups of manuscripts from the mid-ninth to the late tenth centuries, a period when the region was an independent kingdom ruled by local families. The central argument is that the manuscripts attest to the unique cultural diversity of the region during this period, exhibiting – alongside obvious Chinese elements – the heavy influence of Central Asian cultures. As a result, it was much less ‘Chinese’ than commonly portrayed in modern scholarship. The book makes a contribution to the study of cultural and linguistic interaction along the Silk Roads.
Studies in Manuscript Cultures (SMC)

Ed. by Michael Friedrich, Harunaga Isaacson, and Jörg B. Quenzer

From volume 4 onwards all volumes are available as open access books on the De Gruyter website:
https://www.degruyter.com/view/serial/43546
https://www.csmc.uni-hamburg.de/

New release

21 – *Disiecta Membra Musicae: Studies in Musical Fragmentology*, edited by Giovanni Varelli

Although fragments from music manuscripts have occupied a place of considerable importance since the very early days of modern musicology, a collective, up-to-date, and comprehensive discussion of the various techniques and approaches for their study was lacking. On-line resources have also become increasingly crucial for the identification, study, and textual/musical reconstruction of fragmentary sources. *Disiecta Membra Musicae. Studies in Musical Fragmentology* aims at reviewing the state of the art in the study of medieval music fragments in Europe, the variety of methodologies for studying the repertory and its transmission, musical palaeography, codicology, liturgy, historical and cultural contexts, etc. This collection of essays provides an opportunity to reflect also on broader issues, such as the role of fragments in last century’s musicology, how fragmentary material shaped our conception of the written transmission of early European music, and how new fragments are being discovered in the digital age. Known fragments and new technology, new discoveries and traditional methodology alternate in this collection of essays, whose topics range from plainchant to *ars nova* and fifteenth- to sixteenth-century polyphony.

New release

20 – *Fakes and Forgeries of Written Artefacts from Ancient Mesopotamia to Modern China*, edited by Cécile Michel and Michael Friedrich

Fakes and forgeries are objects of fascination. This volume contains a series of thirteen articles devoted to fakes and forgeries of written artefacts from the beginnings of writing in Mesopotamia to modern China. The studies emphasise the subtle distinctions conveyed by an established vocabulary relating to the reproduction of ancient artefacts and production of artefacts claiming to be ancient: from copies, replicas and imitations to fakes and forgeries. Fakes are often a response to a demand from the public or scholarly milieu, or even both. The motives behind their production may be economic, political, religious or personal – aspiring to fame or simply playing a joke. Fakes may be revealed by combining the study of their contents, codicological, epigraphical and palaeographic analyses, and scientific investigations. However, certain famous unsolved cases still continue to defy technology today, no matter how advanced it is. Nowadays, one can find fakes in museums and private collections alike; they abound on the antique market, mixed with real artefacts that have often been looted. The scientific community’s attitude to such objects calls for ethical reflection.