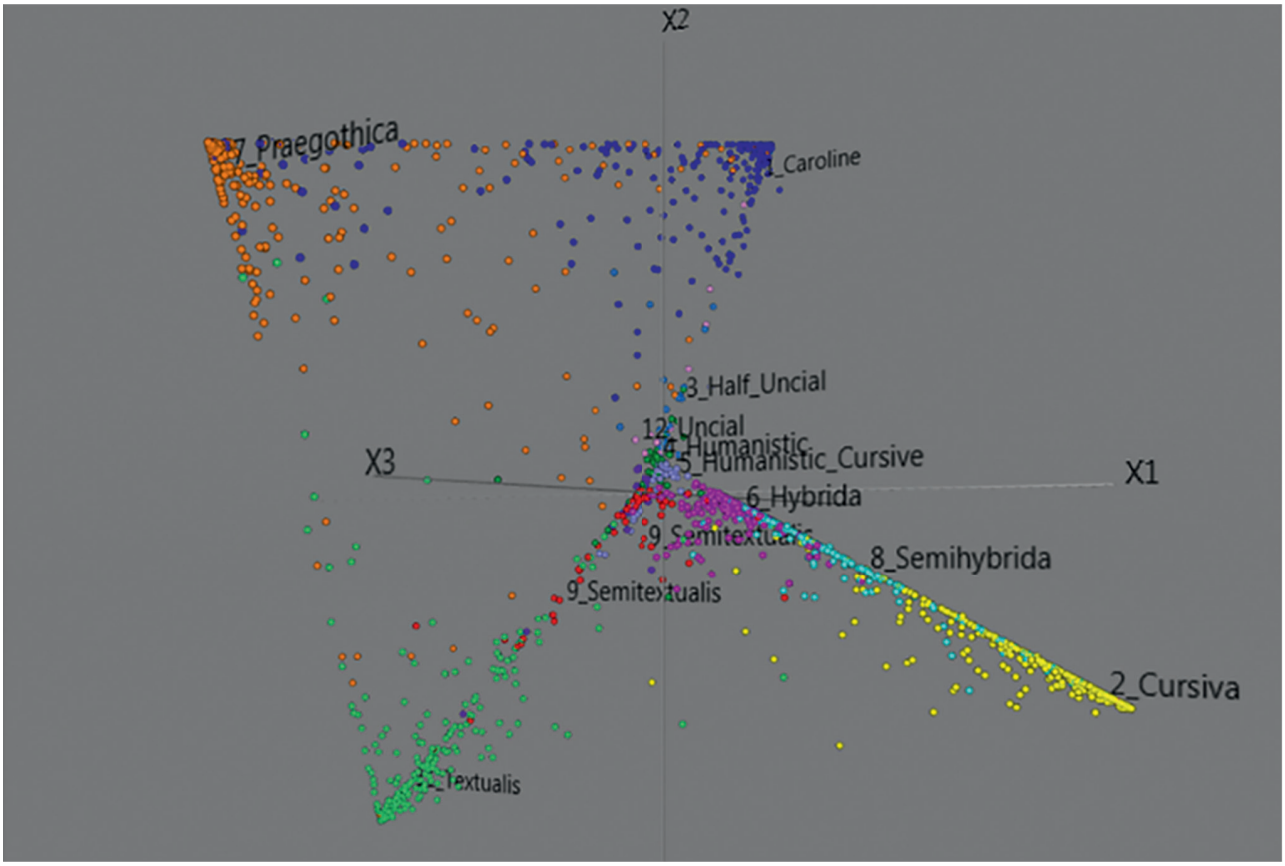
mc Nº 15 2020

manuscript cultures

Hamburg | Centre for the Study of Manuscript Cultures ISSN 1867–9617



Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG
CENTRE FOR THE STUDY OF MANUSCRIPT CULTURES
DFG

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.

Editors
Prof. Dr Michael Friedrich
Universität Hamburg
Asien-Afrika-Institut
Edmund-Siemers-Allee 1/ Flügel Ost
D-20146 Hamburg

Tel. No.: +49 (0)40 42838 7127
Fax No.: +49 (0)40 42838 4899
michael.friedrich@uni-hamburg.de

Prof Dr Jörg Quenzer
Universität Hamburg
Asien-Afrika-Institut
Edmund-Siemers-Allee 1/ Flügel Ost
D-20146 Hamburg
Tell. No.: +49 40 42838 - 7203
Fax No.: +49 40 42838 - 6200
joerg.quenzer@uni-hamburg.de

www.csmc.uni-hamburg.de

Editorial Office
Dr Irina Wandrey
Universität Hamburg
Centre for the Study of Manuscript Cultures
Warburgstraße 26
D-20354 Hamburg
Tel. No.: +49 (0)40 42838 9420
Fax No.: +49 (0)40 42838 4899
irina.wandrey@uni-hamburg.de

Layout
Miriam Gerdes

Cover
Image of ICDAR2017 Tensmeyer's distance matrix (axes 2 and (1 and 3)), see article by Dominique Stutzmann, Christopher Tensmeyer and Vincent Christlein in this volume.

Translation and Copy-editing
Amper Translation Service, Carl Carter, Fürstenfeldbrück

Print
AZ Druck und Datentechnik GmbH, Kempten
Printed in Germany

ISSN 1867–9617

CONTENTS

ARTICLES

Article

Building an Evaluation Framework Researchers Will (Want to) Use

Joseph Chazalon | Le Kremlin-Bicêtre

Abstract

Researchers often face difficulties regarding the availability and the quality of public evaluation frameworks when attempting to compare individual contributions with the state of the art. Firstly, they are often limited to a poorly documented raw dataset that is hard to obtain. Secondly, they seldom provide consistent specifications of the tasks they are designed to evaluate – the dataset and the evaluation methods. Lastly, they often are complex to use, require the use of centralized platforms or may not be available in the long term. Based on the experience we have in building and using evaluation frameworks, we present here a proof of concept for a modified version of the smartDoc competition (challenge 1). This prototype takes the form of a fully open Python package that offers great ease of use: it can be installed with a single command; loading the dataset and its associated ground truth can be reduced to one line of code; and benchmarking a given method takes only 8 lines of code.

1. Introduction

To cope with the ever-increasing pace of research work, researchers need to be able to re-use research products that can be trusted and for which results can be easily reproduced. Open initiatives about methods, tools, services, datasets etc. lay the foundations we can build on. We believe *Evaluation Frameworks* play a particularly critical role here, as the comparison between methods is of paramount importance to allow researchers to make the right choice at the right moment.

By *Evaluation Frameworks (EFs)*, we mean a set composed of:

1. the clear definition of a processing task (input and outputs, goal) — see Fig. 1 for an example;
2. a set of representative input and expected output data;
3. an evaluation protocol based on proven procedures (along with tools implementing them) that make it possible to measure how well predicted results match expected ones.

Input image

Ideal output segmentation

Fig. 1: Overview of the segmentation task ('Task 1' – Original task of the SmartDoc competition – challenge 1): from a raw video frame, locate the coordinates of the 4 corners of the document outline.

Building such evaluation frameworks is often considered a complex task, and few researchers invest time in this direction. However, we believe that using the right approach makes this goal easier to achieve than expected and rewards the creators of *EFs* with many benefits.

What we present here is a proof of concept for an evaluation framework focused on a particular set of tasks for Document Analysis and Recognition (DAR). One of these tasks (a segmentation task) is illustrated in Fig. 1. We have

tried to make these tools as simple as possible to re-use for anyone with basic experience in the DAR field. They take the form of a Python package, which can be installed with one single command:

```
$ pip install smartdoc15_ch1
```

Given a function *you defined* named detect_object(image) → contour that takes an image as input and returns the outline of the document found in the image, the *complete* evaluation code of *your* method can be reduced to:

```
from smartdoc15_ch1 import (
  Dataset,
  evaluate_segmentation)
from my_module import detect_object
d = Dataset()
pred_seg = []
for frame in d:
  seg = detect_object(frame.read_image())
  pred_seg.append(s)
evaluate_segmentation(
  pred_seg,
  d.segmentation_targets,
  d.model_shapes,
  print_summary=True)
```

The source code for all the material presented here is available at the following URLs:
<https://github.com/jchazalon/smartdoc15-ch1-dataset>
<https://github.com/jchazalon/smartdoc15-ch1-pywrapper>.

After a brief review of a selection of notable initiatives that support open or reproducible research, with a focus on DAR (Sec. 2), we present our contribution (Sec. 3): a proof-of-concept evaluation framework based on a task-oriented definition of the SmartDoc 2015 dataset (Sec. 4.1), a new distribution scheme for the raw dataset (Sec. 4.2) and an easy-to-use Python wrapper for data loading and method evaluation (Sec. 4.3).

2. Related work

As mentioned in the introduction, we take into consideration three aspects of evaluation that constitute a consistent evaluation framework: task definitions, datasets and evaluation methods.

Within the Document Analysis and Recognition (DAR) community, the Technical Committees 10 and 11 maintain a list of publicly available datasets for research use.[1] While the datasets listed by the curators are generally free to obtain and have research-friendly licenses, it is not uncommon that the download links get broken, causing long-term availability issues because of unreliable hosting solutions. Furthermore, there are even fewer options for evaluation tools and it is very rare that free and open tools are released. A notable exception is the excellent UNLV-ISRI set of OCR evaluation tools,[2] which is still used and maintained, even 20 years after its release.

To our knowledge, the first initiative that really embraced the three aspects of an evaluation framework is the Robust Reading Competition (RRC) series.[3] RRC provides a clear definition of the different problems (or tasks) that researchers can evaluate their methods against. Datasets are hosted on the platform; upon submission of method results, an evaluation and a ranking of the methods are automatically performed. This approach now faces the issue of rising hosting and maintenance costs. In response, the platform is now being progressively opened to avoid interrupting the service. This solution could also solve the issue of having closed-source evaluation procedures and secret ground truth for some datasets.

The Document Analysis and Evaluation (DAE) platform[4] was another important initiative. It was more general in the sense that it aimed at offering an orchestration platform that could be used to compose data processing units (exposed as web services) and data sources. The complexity of the packaging of processing units made the dataset distribution aspect much more successful than the processing one.[5] To avoid availability issues for datasets, a new version of the platform (DAE-NG) now focuses on building a federation of synchronized dataset repositories. However, this still requires specific deployment and maintenance skills.

[1] Available at http://iapr-tc10.univ-lr.fr and http://tc11.cvc.uab.es.

[2] Rice and Nartker 1996.

[3] Karatzas et al. 2011.

[4] Lamiroy and Lopresti 2012.

[5] Lamiroy 2017.

A last initiative from the DAR community we would like to mention is the DIVAServices platform.[6] It enables the easy packaging of document processing tools using Docker containers and their testing or combination using a REST API. Hosting evaluation methods on this platform is promising, but it requires hosting datasets on a separate platform and can lower the consistency between data and evaluation. Finally, while DIVAServices creators build the platform with openness in mind, deploying and maintaining a complex platform is necessary to run experiments, penalizing long-term availability unless a consortium eventually takes over the project.

It is interesting to note that the DAE, DAE-NG and DIVA platforms were inspired by other communities like bioinformatics, which successfully built massive platforms[7] where researchers can download and upload massive datasets (genomics for instance) and even launch computations on public clouds against standard datasets. We fully support the initiatives of our colleagues from the DAR community, as these efforts are federating the community around important problems and encourage collaboration, and we do not want to minimize their impact. Our goal here, however, is slightly different and fully compatible: we believe datasets and associated evaluation frameworks should be as open as possible and available over the long term. Such evaluation frameworks should then be hosted on those platforms, but we believe that if users can easily reproduce results by themselves, this will make things a bit more durable. Furthermore, even in communities like bioinformatics or medical imaging, evaluation methods and tools are often used for competition, but not always released in an open way. This reduces the impact of such work, as other researchers may not be able to reproduce the results by themselves.

An approach that reconciles long-term availability, ease of use and a strong consistency between dataset content and evaluation procedures is the Scikit-learn Dataset API.[8] This free and open Python library can be installed with a single command, and Scientific Python is a new standard for computer vision research. It often makes evaluation very easy to implement, thanks to the many standard functions provided. The Dataset API makes it possible to load many datasets in a computable format with a single line of code. We believe such an approach is an interesting alternative to centralized evaluation platforms.

3. Building an open evaluation framework

Based on the original version of challenge 1 of the smartDoc 2015 competition, we revised its task definition to enable a wider use of the dataset and the evaluation tools we created (Sec. 3.1). To facilitate the dissemination and evolution of this new evaluation framework, we separated the implementation of the raw dataset distribution (Sec. 3.2) from the Python library ('wrapper'), which enables the manipulation of its content as *computable* objects, as well as the straightforward evaluation of any method that complies with the previous task definition (Sec. 3.3).

As previously mentioned, all the sources and products of this proof of concept are freely available online. Our main goals were to build an evaluation framework that would be:

1. as easy to use as possible;
2. available in the long term;
3. reliable and trustworthy.

3.1 New task definitions

As previously mentioned, the original dataset we used as the basis of this proof of concept is the SmartDoc 2015 database for document capture (challenge 1).[9] This dataset was initially created to evaluate the performance of smartphone applications for document image acquisition, focusing on one of the first stages of the pipeline: the segmentation of the document outline in video frames or pictures, in order to allow the correction of perspective distortion.

The dataset was built by capturing 30 document models (5 for each of the 6 different types as shown in Fig. 2) under 5 different background scenarios (as visible in Fig. 3). Some small noise and margins from the original document images were removed and finally the images were rescaled to have the same size and fit an A4 paper format, resulting in several variants of the 30 model images. In addition to the video clips, a picture of each of the documents was captured to be used as another set of models variants.

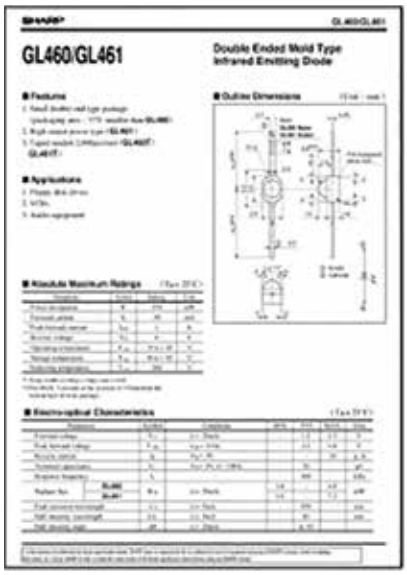
Each of these documents was printed using a colour laser jet and captured using a Google Nexus 7 tablet. The dataset consists of 150 video clips, comprising nearly 25,000 frames, captured by hand while holding and moving the tablet. The

[6] Würsch, Ingold, and Liwicki 2016.

[7] For an example, see <https://www.genouest.org/hosted-resources-and-tools/>.

[8] Buitinck et al. 2013.

[9] Burie et al. 2015.



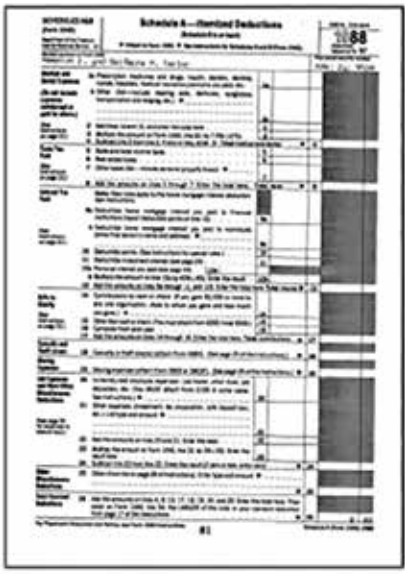
Datasheet 2

Letter 4

Magazine 2

Paper 5

Patent 4



Tax 3

Fig. 2: Sample documents used in our dataset.

video frames present realistic distortions, such as focus and motion blur, perspective, change of illumination and even partial occlusions of the document pages. The ground truth of segmentation data was created by semi-automatically annotating the quadrilateral coordinates of the document position for each frame in the collection.

The new version of the dataset makes use of the model images we created and captured. Each of the new tasks we are about to introduce can have two variants: 1) a *model-agnostic* variant with no knowledge of the original document models; 2) a *model-aware* variant based on the knowledge of the complete set of document model images. This second type of scenario allows researchers to test applications like augmented reality or form digitization.

Researchers can test their methods against three tasks using this new dataset. For each of these tasks, the *model-aware* variant is obtained by adding one extra input: the set of model images (or the result of the indexation of the latter).

3.1.1 Task 1: Segmentation (original task)

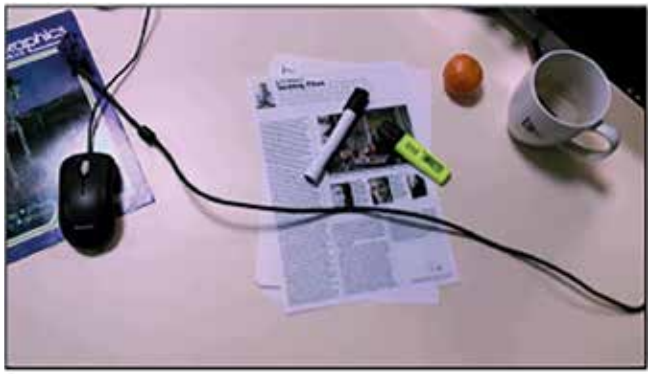
Inputs are video frames and expected output comprises the coordinates of the four corners of the document image in each frame (top left, bottom left, bottom right and top right). The evaluation is performed by computing the intersection

Background 1

Background 2

Background 3

Background 4



Background 5

Fig. 3: Sample backgrounds used in our dataset.

over union ('IoU' or also 'Jaccard index') of the expected document region and the detected region. The frame coordinates are projected onto the document referential to allow comparisons between different frames and different document models. The original evaluation code is available online,[10] and the Python wrapper also contains an implementation using the new data format.

3.1.2 Task 2: Model classification (new task)
Inputs are video frames and expected output is the identifier of the document model represented in each frame. There are 30 models named 'datasheet001' to 'tax005'. The evaluation is performed using usual multi-class classification metrics: mean accuracy, precision, recall etc.

3.1.3 Task 3: Model type classification (new task)
Inputs are video frames and expected output is the identifier of the document model *type* represented in each frame. There are 6 model types, each having 5 members, named 'datasheet', 'letter', 'magazine', 'paper', 'patent' and 'tax'. The evaluation is performed using usual multi-class classification metrics: mean accuracy, precision, recall etc.

3.2 Raw dataset distribution
We changed the dataset format to adapt it to the new task specifications. We did so with the objective of supporting long-term availability, which also has consequences for the hosting strategy we followed. The resulting product has a dedicated GitHub project available online.[11] GitHub hosting

[10] <https://github.com/jchazalon/smartdoc15-ch1-eval>.

[11] <https://github.com/jchazalon/smartdoc15-ch1-dataset>.

has the immediate benefit of making the project easy to find and self-documented, thanks to the embedded documentation viewer.

To design a dataset still usable in ten years, we reviewed and changed the data format used for the original version of the dataset. The original distribution was based on a set of video files, along with XML files in a custom format for the ground truth. The files were available at some secure file server using a procedure sent by email to users after they registered and accepted the dataset license on the main website. This process is sustainable, thanks to email automation, but it does not support automated download from a Python script. Furthermore, the explicit license agreement is an unnecessary burden for a standard Creative Commons license.

We started by producing a format that makes reading data as simple as possible. We first extracted all the frames from the video files and saved them as JPEG images to minimize decoding issues. We then created a simple CSV file for storing all metadata about each video frame: each line represents a frame observation, the columns store either information about file location or the ground truth for each task. The format is documented to remove any ambiguity about content types. The resulting files (images, metadata, documentation, license etc.) are packaged into a gzipped TAR archive for maximal compatibility. The model images were packaged using the same process.

Regarding data hosting and distribution, we considered several options with the constraint of being widely accessible and durably available. At the time we started working on this project, we chose to use GitHub[12] *releases*, as they feature very simple HTTPS upload and download while supporting version numbers. Since then, we started using Zenodo,[13] an EU-funded platform that specifically targets the archiving and distribution of research datasets. Zenodo and GitHub work well together, as Zenodo can automatically archive GitHub files (sources and binaries in *releases*). We now strongly encourage researchers to consider Zenodo, as it is entirely dedicated to the archiving and dissemination of research products and has many interesting features.

The resulting solution (either using GitHub or Zenodo) makes data archives directly downloadable (and verifiable using SHA256 checksums) with an implicit license agreement. Users will be able to download specific versions of the dataset based on the version number they carry. We hope versioning will encourage collaboration and the improvement of datasets, while allowing researchers to keep track of which sets of version numbers produce comparable evaluation results.

3.3 Python wrapper for data loading and evaluation

Using this reliable dataset distribution, we built a Python wrapper with three goals in mind: 1) making the dataset '*computable*' thanks to the automated loading of the raw archive as ready-to-compute objects; 2) making evaluation so easy that researchers will want to work with it; and 3) leveraging openness to maximize trust in the correctness of the method, as well as enabling long-term availability.

The resulting code has a very simple API organized around two kinds of functions: loading functions (one for the frames, one for the models), which have options to load images, pre-process them and load the associated ground truth for each task; evaluation functions (one for each of the three tasks). We tried to comply as much as possible with Python and SciPy philosophies in order to provide researchers with a plug-and-play library, as illustrated in the listings of the first page. In particular, every data series is a Numpy array that supports all the useful operations researchers are familiar with. For instance, is it possible to directly pass the result of the dataset loading to the automated function of Scikit-learn to separate training and test sets to perform cross-validation. The evaluation therefore complies with the usual evaluation pattern for estimators, comparing the target results with actual ones.

To maximize the usability of this proposed evaluation framework, we packaged this Python wrapper as a PIP package listed on the central Python Package Index (PyPI) and installable with a single `pip install` command. Finally, the source of these tools can be inspected, forked, maintained, improved and contributed to at: <https://github.com/jchazalon/smartdoc15-ch1-pywrapper>. Thanks to the online documentation viewer, this address also features a simple and nice entry point for any researcher willing to experiment with this prototype.

4. Conclusion

We presented a proof of concept for the distribution of evaluation frameworks within the document analysis and recognition community. Greatly inspired by initiatives

[12] <https://github.com>.

[13] <https://zenodo.org>.

from other communities like bioinformatics, we proposed leveraging the Python Package Index to distribute code, as well as public storage offered by GitHub or Zenodo to distribute data and archive code. This facilitates the distribution of complete evaluation frameworks to researchers. We believe that distributing data, documentation and code all together is a key to engaging researchers in using such evaluation frameworks and to encouraging reproducible research.

Of course, we still rely on external platforms like GitHub (now owned by Microsoft) and its release feature (a core element of their product), Zenodo (backed by the CERN and part of the infrastructure that stores data for the Large Hadron Collider) or software frameworks like SciPy and Python (which receive massive attention and funding these days). These choices are, in our opinion, the most reliable options today, but this will obviously change. However, we believe that providing researchers with tools they can use easily, with few dependencies and great benefits, are a key enabler for long-term support of software (hence the implemented methods) and massive dissemination of datasets. We hope our simple proof of concept will encourage researchers to look for ready-to-use *evaluation platforms* and maybe even start distributing some.

REFERENCES

Buitinck, Lars, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, and Jaques Grobler (2013), 'API Design for Machine Learning Software: Experiences from the Scikit-Learn Project', *arXiv.org*: <arXiv: 1309.0238>.

Burie, Jean-Christophe, Joseph Chazalon, Mickaël Coustaty, Sébastien Eskenazi, Muhammad Muzzamil Luqman, Maroua Mehri, Nibal Nayef, Jean-Marc Ogier, Sophea Prum, and Marçal Rusiñol (2015), 'SmartDoc: Smartphone Document Capture and OCR Competition', *13th International Conference on Document Analysis and Recognition*, 1161–65. doi: <10.1109/ICDAR.2015.7333943>.

Karatzas, Dimosthenis, S. Robles Mestre, Joan Mas, Farshad Nourbakhsh, and P. Pratim Roy (2011), 'Robust Reading Competition-Challenge 1: Reading Text in Born-Digital Images (Web and Email)', *International Conference On Document Analysis and Recognition (ICDAR)*, 1485–1490. doi: <10.1109/ICDAR.2011.295>.

Lamiroy, Bart (2017), 'DAE-NG: A Shareable and Open Document Image Annotation Data Framework', *14th IAPR International Conference On Document Analysis and Recognition (ICDAR)*, vol. 4, 31–34. doi: <10.1109/ICDAR.2017.309>.

Lamiroy, Bart, and Daniel Lopresti (2012), 'The Non-Geek's Guide to the DAE Platform', *10th IAPR International Workshop on Document Analysis Systems*, 27–32. doi: <10.1109/DAS.2012.87>.

Rice, Stephen V. and Thomas A. Nartker (1996), *The ISRI Analytic Tools for OCR Evaluation* (Las Vegas: University of Nevada, Information Science Research Institute; Technical Report TR-96-02). <https://isri-ocr-evaluation-tools.googlecode.com/svn-history/r2/trunk/analytic-tools/user-guide.pdf.>.

Würsch, Marcel, Rolf Ingold, and Marcus Liwicki (2016), 'SDK Reinvented: Document Image Analysis Methods as RESTful Web Services', *12th IAPR Workshop on Document Analysis Systems (DAS)*, 90–95. doi: <10.1109/DAS.2016.56>.

Written Artefacts as Cultural Heritage

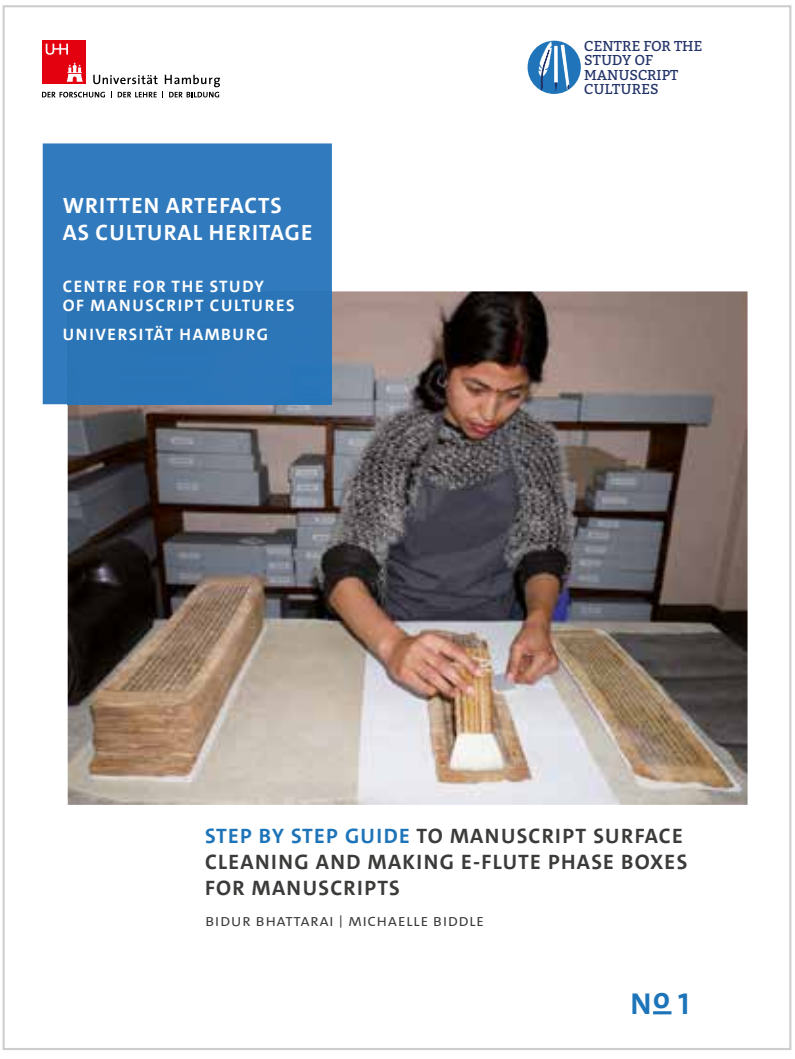
Ed. by Michael Friedrich and Doreen Schröter

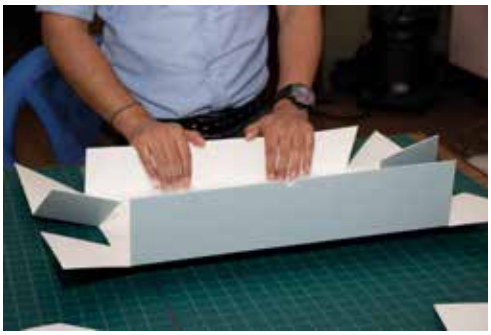
Written Artefacts as Cultural Heritage was established in 2020. The series is dedicated to the double role of written artefacts as representations and generators of humankind's cultural heritage. Its thematic scope embraces aspects of preservation, the identity-defining role of artefacts as well as ethical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





manuscript cultures (mc)

Editors: Michael Friedrich and Jörg B. Quenzer
Editorial office: Irina Wandrey

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

manuscript cultures 14

manuscript cultures 13

manuscript cultures 11

manuscript cultures 10

manuscript cultures 9

manuscript cultures 8

manuscript cultures 7

manuscript cultures 6

manuscript cultures 5

manuscript cultures 4

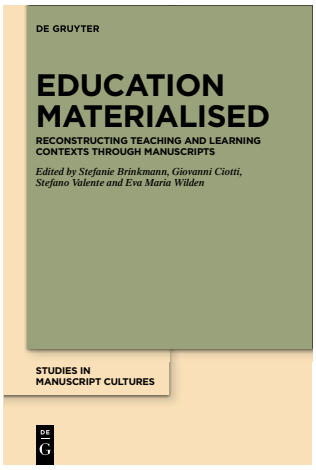
Studies in Manuscript Cultures (SMC)

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/

DE GRUYTER

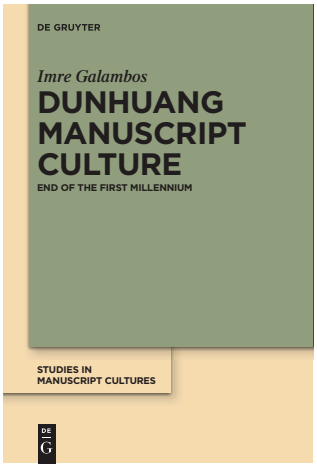
Forthcoming



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.

New release



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.

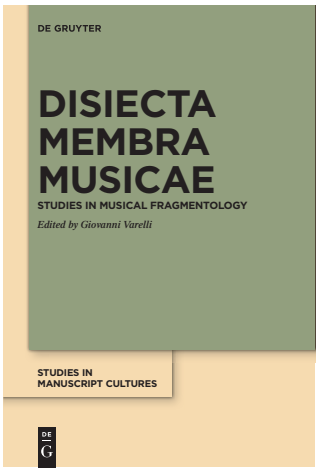
Studies in Manuscript Cultures (SMC)

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/

DE GRUYTER

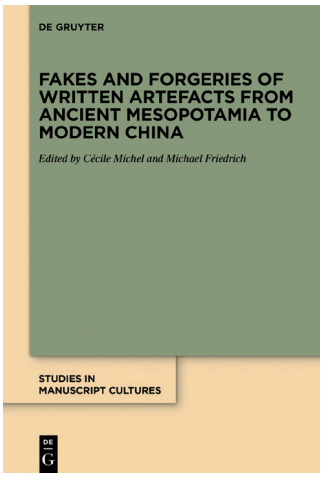
New release



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 serie of thirteen articles devoted to fakes and forgeries of written artefacts from th beginnings of writing in Mesopotamia to modern China. The studies empha sise the subtle distinctions conveyed by an established vocabulary relating t the reproduction of ancient artefacts and production of artefacts claiming t be ancient: from copies, replicas and imitations to fakes and forgeries. Fake are often a response to a demand from the public or scholarly milieu, or ever both. The motives behind their production may be economic, political, reli gious or personal – aspiring to fame or simply playing a joke. Fakes may be revealed by combining the study of their contents, codicological, epigraphi and palaeographic analyses, and scientific investigations. However, certain fa mous unsolved cases still continue to defy technology today, no matter hov advanced it is. Nowadays, one can find fakes in museums and private collec tions alike; they abound on the antique market, mixed with real artefacts tha have often been looted. The scientific community's attitude to such objects calls for ethical reflection.

ISSN 1867–9617

Centre for the Study of Manuscript Cultures
Universität Hamburg
Warburgstraße 26
D-20354 Hamburg

www.csmc.uni-hamburg.de