First Steps towards Reviving Franke’s ’Chinese Epigraphy in Southeast Asia’: Motivations, Approaches and Data Structures

STREITER, Oliver
oliverstreiter@gmail.com

BINGENHEIMER, Marcus
m.bingenheimer@gmail.com

ZHAN, Hanna Ya-Qing
zhanzhan.ya.qing@gmail.com

TO, Mandy Manwai
tomanwaipopo24@gmail.com

SHIH, Syuan-Fei
syuanfeishih@gmail.com

NUK - National University of Kaohsiung, Kaohsiung, Taiwan
APLAC - Research Center for Asia Pacific Languages and Cultures

Temple University, Pennsylvania, USA
College of Liberal Arts

NTU - National Taiwan University, Taipei, Taiwan
Graduate Institute of Taiwan Literature

National Chenchi University, Taipei, Taiwan
ApSTi - Asia-Pacific SpatioTemporal Institute

Abstract

The goal of this research project is to transform the paper-based and barely accessible Documentation of Chinese Epigraphy produced by Wolfgang Franke and his colleagues (WFCEM) into a permanently archived and freely accessible repository of standardized data that overlays the original documentation with state-of-the-art imagery, enhanced data, geo-localizations and a continuation of the prose describes of sites, object and inscriptions. Doing this we also hope to increase the awareness in various research communities of the
existence of this extraordinarily rich resource, but also raise the awareness of those who 
manage and administer the cultural sites of the treasures history has laid into their hands.

此次研究計畫目標在於：將傅吾康先生與其友人的紙本印刷的東南亞華文碑刻 
銘彙文件，轉化成永久歸檔且容易取得的標準化資料庫，覆蓋了原始文檔，包含 
最先進的圖像、高品質的數據、地理定位，並有註解描述圖片的地點、物品和銘 
文。藉由此研究，希望能增強各種研究團體對於墓誌銘等珍貴資源的意識，也能 
提高歷史性文化遺址管理者的意識。

1 Introduction

Chinese Epigraphy in Southeast Asia counts among the most important sources for the re-
gion’s social and religious history in the last three hundred years. Countless inscriptions in 
temples and graveyards are often the only dated witnesses of local Chinese communities. 
Digitally aggregated, these records could be analyzed on a much wider scale than previ-
ously possible. Based on more than a decade of experience recording tombstone inscrip-
tions on Taiwan, Penghu, and Okinawa,¹ we have begun to work on a ”Digital Database 
of Chinese Inscriptions in SEA”, all the while reshaping our conceptual model to conform 
to the CIDOC Conceptual Reference Model (CRM), for the representation of sites, and to 
TEI, for the representation of inscriptions.

1.1 Epigraphy

An epigraph is an inscription carved into an object composed of hard material, such as 
stone, wood or metal.

The main distinctive features of epigraphs are their uniqueness and their context-dependency. 
Epigraphs, tend to remain on the site or the object on which they have originally been cre-
ated. It is the unique site and the unique object which contribute to the meaning of the 
epigraph. This is particularly the case where deictic expressions are used to refer, e.g., to a 
village as ’本村’ (běncūn, this village), or ’本廟’ (běnmiào, this temple), instead of using 
the name of the village or temple. Likewise, the oldest tombstone, the largest tombstone 
and the only granite tombstone obtain these properties only within a set of tombstones. 
People’s names obtain meaning from nearby shops, factories and temples, and family re-
lations show how within a community families established and maintained their economic 
power.

Second, the object, its size and material, the tools, and the training of the carver influence 
what and how an epigraph ultimately materializes. These factors determine genre, 
style, format, as well as character and semantic variants. Soft materials facilitate the carv-
ing of small characters or characters composed of many strokes, while hard or brittle ma-
terials tend to show large or simplified character variants, e.g. ’顯’ vs. ’顯’ (xiǎn, an 
honorific) or semantic variants, e.g. ’旦’ (dàn) vs. ’穀’ (gǔ) (in the sense of ‘auspicious’). 
Thus, what seems to be a paradigmatic opposition represents only a contextual variant, 
if the conditions of the production are considered. Contextualization thus minimizes the 
danger of overinterpretation.

Beyond the conditions of production, the conditions of the reception of the epigraph too 
are encoded. An epigraph typically assumes a certain posture towards the reader, perhaps 
in front of a tomb, or a temple gate, or at the top of a mountain. Not only the stone, but also 
the assumed position of the readers determine the size and the orientation of the carving.

¹The ThakBong database contains 330.000 images of 682 graveyards in Taiwan and its outlying islands, see 
http://thakbong.dyndns.tv/.
The type of the object, e.g. tombstone or ancestral tablet, will influence the wording, especially in the case of referential linguistic expression on tombs (mù 墓) or ancestral tablets (shénwei 神位).

These features of epigraphs have fundamental implications for research and documentation. Instead of visiting libraries, which collect copies of texts and make them available, researching epigraphs requires the researcher to visit the places where epigraphs were carved or where the objects have been erected in order to document the potential contexts which affect the interpretation of the inscription.

Therefore rubbings or photos of epigraphs, stored in libraries, can only partially replace the object. Beside the properties typical of epigraphs, discussed above, also the spatial properties of the epigraph might have been lost. These include longitude and latitude of the object, the altitude, the bearing of the object in degrees from the north line and the orientation of the object in terms of the spatial configuration within a set of culturally relevant objects of a site. Epigraphs tends to be embedded within a physical space that represents, as symbolic space, social arrangements in a community. The placement and orientation of tombstones, ancestral tablets, donor steles, bells, yunban, censers, couplets or wooden tablets within a temple are thus meaningful since symbolic arrangements and contain clues how an epigraph can to be interpreted.

The value of an epigraph thus derives from its unique position at the intersection of humanities, the science of man-made products, and the social sciences, the science of human behavior. An epigraph combines text, material and object, and has been created by specific agents at specific times, at specific places, in relation to specific events within which they fulfill a specific functions. A wooden tablet attached to the roof of a temple, for instance, is inscribed with the year of its creation, with that space and time, two essential coordinates, are already known. Additional data can be associated, e.g. the content can be transcribed and translated, actors can be identified. Each epigraph is thus a node in a network that can be linked to historical, sociological, anthropological, geographic, linguistic, or economic data, with the potential to change our perception of the past.

1.2 Wolfgang Franke (1912-2007)

Wolfgang Franke was a well-known German Sinologist who taught in Hamburg University. He was the son of Otto Franke (1863-1946), the founder of German Sinology. After his retirement in 1977 Franke moved to Malaysia, where he continued to teach and pursue his research. Originally a specialist of the Ming Dynasty, Franke, perhaps influenced by the destruction of cultural assets during the Cultural Revolution (1966-1975), began to work on a large-scale photographic documentation of Chinese epigraphs in Southeast Asia. We will refer to this multi-volume survey of Chinese epigraphy in Malaysia, Indonesia, and Thailand as the Wolfgang Franke Chinese Epigraphic Materials (henceforth WFCEM).

Franke assembled a large and effective team of collaborators that included renown scholars such as Chen Tieh Fan, Claudine Salmon, Anthony Siu and Porpan Juntarananont, among others. How this team worked is almost undocumented and more research is needed to reveal how this team was able to collaborate successfully for such a long time.
2 Franke’s Documentation of Chinese Epigraphy in Southeast Asia

2.1 Achievements

Building partially on data collected earlier in his career, Wolfgang Franke began to intensify his fieldwork and assemble and publish a multi-volume documentation of Chinese epigraphic materials in Malaysia, Indonesia, and Thailand, covering the time span from the late Ming period to the 1950s (Franke & Fan, 1983), (Franke et al., 1988), (Franke et al., 1997), (Franke & Juntaronanont, 1998a), (Franke & Juntaronanont, 1998b). He and his team were assisted by local scholars who were often crucial in identifying the oldest and most relevant epigraphs in a region.

In two decades, the team around Wolfgang Franke documented and published approximately 10,000 inscriptions, the oldest and most relevant inscriptions they were able to find in the second half of the 20th century. This epigraphy is considered a crucial source for the study of oversea Chinese communities in Southeast Asia. Although in its present form it is difficult to access and cumbersome to use, it is absolutely indispensable for the serious study of Chinese religion and history in Southeast Asia. In total, the team around Wolfgang Franke published eight volumes on South East Asian Chinese epigraphy, each volume containing approximately 1000 epigraphs.

Not only is the content of WFCEM highly significant, also its style and format has become a model for subsequent documentation projects, such as the Chinese Epigraphs in Singapore (Dean & Hue, 2016), in Malaysia (黄文斌, 2013; 张少宽, 2013; 吉隆坡, 2014), in Taiwan (Streiter et al., 2011b; Streiter et al., 2011a; Streiter et al., 2011c; Streiter & Morris, 2016), on Penghu (Streiter & Goudin, 2016; 奥利华, 林莉伦, 陈乃瑜, 2016; 王雅晴, 2016; 奥利华 & 詹雅晴, 2016), in Hong Kong (鄧家宙 & 陳覺聰, 2012), and Chinese Migrants in France (Goudin & Streiter, 2016). Claudine Salmon, having worked with Wolfgang Franke and being listed as editor and collaborator in WFCEM, continued to work on Chinese epigraphies in Vietnam (Salmon & Cac, 1998) and Indonesia (Salmon & Archipel, 2016).

The volumes of WFCEM are organized by country and region. Volume 1 of “Indonesia”, for example, covers Sumatra and Volume 2 covers Java. Some cartographic material

Figure 1 Wolfgang Franke (1912-2007)
has been included, mapping the earliest finds for each administrative region, cf. 3. Sometimes however, additional material from regions documented in earlier volumes has been included in later volumes, making an extensive search in the books more difficult.

WFCEM is indexed, using hierarchic alpha-numeric codes, which will facilitate the conversion of Franke’s material into modern database formats, as Franke’s hierarchical codes match the conceptual distinctions we find today in the CIDOC-CRM Conceptual Reference Model: locations, sites within locations, objects within a site, and inscriptions on objects.  

In Franke’s coding system, K, stands for central Java, K1 for Semarang, with K1.13 for

---

2 The CIDOC-CRM Conceptual Reference Model, designed for implementation in database formats, describes the implicit and explicit concepts and relationships used in cultural heritage documentation. The CIDOC-CRM standard is one of the formal standards into which we will transform WFCEM.
a specific graveyard and K1.13.1 for the first tombstone in this graveyard. Likewise K1.3 represents a temple and K1.3.1 an inscription in that temple. Generally, the first character represents a larger administrative region, a two-character-code a town or village, a three-character-code a site and and four-character-code an objects that bears and inscription. Sites can be graveyards, temples, schools or residences. Items can be tombstones, wooden tablets, steles, censers, couplets, sculptures etc.

Using this simple coding system, Franke had no means to describe complex relations among objects, such as the inclusion of one object in another. E.g. the relation between an altar, a couplet at the altar and each part of the couplet could not be formally expressed in his notation.

Generally, WFCEM provides one photo for each epigraph, a normalized transcription and a relatively literal and uncritical translation. In the normalized transcription, character variants are normalized throughout. The transcriptions are Üncritical” in the sense that the content of the inscription is presented as fact. A placename, e.g. Tong’an is usually translated as from Tong’ an, no matter whether the person migrated from that place, whether the ancestors migrated from that place or the ancestors are believed to have migrated from that place. Likewise the sons and daughters are always translated to have erected the tombstone, although they might also just have sent money or the tombstone might have been erected without their knowledge or consent. he translations should ideally matched with family genealogies, where available. Non-factual inscriptions expressing attitudes or identities, such as 皇上 (huángqīng, of the August Qing Dynasty), are not translated and not explained. In total, the team around Wolfgang Franke published eight volumes on South East Asian Chinese epigraphies, each volume containing an estimate of 1.000 epigraphs.3

The way Franke and his team selected and presented the material reflect a deep under-

---

3The exact number of epigraphs will only be known after the project has been completed, as Franke frequently grouped related epigraphs under one number.
standing and appreciation for local cultures, histories and languages. Temple names, for example, are romanized in Southern Min where that was the local language. Inscriptions in Bahasa Malay or Peh-oe-ji are equally documented along with inscriptions in Chinese, Japanese, Arabic and Thai. Where Mandarin is romanized, the earlier volumes on Malaysia use Wade-Giles, later volumes use Hanyu Pinyin.
Figure 5  Pilot Study: The spatial distribution of some sites extracted from WFCEM. Once completed these maps will guide our fieldwork.
Figure 6  Pilot Study: The spatial distribution of some sites extracted from WFCEM, split by the type of site
By today’s standards, the black-and-white photography seems rather primitive. In order to improve the readability of carved characters, Franke’s teams used chalk that was rubbed into the inscriptions.\footnote{This method, effective as it is, is today depreciated, not only for cultural reasons. Chalk is very abrasive and will provide food for lichen and fungus which will further destroy the stone. In addition, it contain a cocktail of chemical substances that might react with the stone and accelerate the degradation of the material}. On the images the material of the chalked object becomes almost impossible to assess.

Although most photos in WFCEM have been taken in situ, objects are depicted in isolation from their site. E.g. we find photos of tombstones, without photos of the tomb, of wooden tables without photos of the temple.

Except for one overview map per book (e.g. Figure 3, WFCEM lacks, an analytic component that visualizes and interprets the bulk of collected data. Timelines, maps, social networks, and linguistic analysis are only implicitly encoded in the transcriptions and the descriptions that WFCEM provides for each object. Once the information is digitized, it becomes computationally accessible and can be analyzed and visualized in various ways.
A wooden tablet in the Jiulongtang Lin family temple in Batu Pahat: Left side photo in MYS-I, right side photo taken on in August 2018. On the recent photo, the plaque has been renovated and the original date ‘民國廿一年’ has been replaced by ‘歲次壬申’ and disambiguated on the left side with ‘西元一九三二年立’.

The full potential of the WFCEM could be enjoyed by the growing research community that studies Chinese communities in SE Asia.

3 Research Goals and Objectives

3.1 Revisiting and Re-Documenting WFCEM Sites - Result of a pilot study

Our aim is to re-visit as many sites as possible, to re-document the epigraphs once documented in WFCEM, and to identify and document new material. We intend to get better photos and more data, such as: information on symbols and motifs, cf. Figure 2, GPS-localization, orientation, localization within the site, material, size, etc. We will also document the transformation of the epigraphs. Some were change on purpose, e.g. in Figure 4, some change is due to environmental reasons of the site, as shown in Figure 3. Others are due to degradation or abandonment of the site, or to an inappropriate storage.

In the pilot study, we aimed to get a general view on what happened in the last thirty years in the regions we visited. Of the 49 sites we visited in Malaysia, 34 had been visited before by Wolfgang Franke and his team. Of them, 33 were accessible and in the same place as 30 years ago. One temple has moved, one has completely burned down and has been, like most others reconstructed. This reconstruction work took a took a heavy toll on the objects and epigraphs once documented in these temples. Destroyed by the community, destroyed during the reconstruction, moved into sheds or simply unfindable or stolen, about 50% of the objects and epigraphs documented by Franke seen to have been lost. Of those epigraphs still available, many are being stored under inadequate conditions, among tools, paint cans, roof tiles and scaffolds. Rummaging in sheds, however, revealed epigraphs that Wolfgang Franke either did not have found or not documented, adding an extra value to our pilot study.

3.2 Processing Fieldwork Material and Data

Each digital file, each photo, video or audio, is annotated with meta-data that refer to the field trip (date, technique, researcher), the site, the object and all the particularities of the item. We also classify the data files as to which photo is the best photo showing the epigraph, the best photo showing the environment of the epigraph, a photo showing mea-
measurements, a photos showing parts of the epigraph, photos documenting our fieldwork in relation to that epigraph, etc. Pieces of information that cannot be transcribed, such as motifs and symbols are mentioned in the meta-data as well.

3.3 Scanning and OCRing of WFCEM

All books of WFCEM are now scanned in the highest possible resolution at the National University of Singapore, by a team under the leadership of Kenneth Dean. Dean and his associates have also kindly shared a georeferenced dataset listing the locations of the original Franke sites.

On these these scans we have applied bilingual Chinese and English OCR (optical character recognition) to produce a bilingual docx-document on the basis of ABBYY FineReader. Besides recognizing the Chinese and English texts, the OCR attempted to identify and preserve the images in bilingual docx output. The preserved images can be extracted automatically with Pandoc. Where the OCR did not recognize an image, the image has to be cut manually from the scans. Likewise, OCR mistakes must be corrected manually in the docx output. This can be done semi-automatically for the English text, using a spelling checker. Chinese OCR output, e.g. the names of temples and sites, and the transcriptions of the epigraphs, must be proofed manually.

The irregular layout of the pages in WFCEM is a major challenge for OCR, c.f. Figure 4. In the printed volumes, English paragraphs, Chinese paragraphs and photos are mixed in creative ways in order to allow for the maximum of photos and texts on one page. Frequently, the OCR software thus attempts character recognition on images (especially as many depict visible characters), turning the photo incorrectly into text. The photo thus disappears from the docx output-file and has to be manually cut from the scan.

3.3.1 Transforming OCRed Scans into a Intermediate Annotation Format (IAF)

After the final corrections of OCR mistakes, we transform the output of the OCR into a format which is easy to understand and easy to manipulate, and to compute with. Scripts running over this intermediate representation can add redundant information, insert GPS-data, or correct obvious mistakes.

The IAF (Intermediate Annotation Format) follows closely the structure that underlies the hierarchically organized entries in WFCEM. An example of this format for the first entries in IDN-IIa is given in Appendix A.

3.3.2 Entity Recognition (ER)

Entity recognition refers to the recognition, extraction and storage of references to people, places and temporal units and their relations, such as events, rituals, locationzation etc. Sources from which entities are recognized are the metadata of epigraphs, the transcriptions of epigraphies, the translation of epigraphies and the prose description of sites and objects provided in WFCEM.

The IAF contains already in isolated forms entities and of entity relations, e.g. in the table of content, in the names of temples, and in the transcribed date. These entities can be searched and indexed in the textual elements. From the recognized entities, templates can be compiled which interactively allow to extract more entities. Entities are searched extensively before their relations are identified. The steps we will followed are:

- the extraction of entities
- the storage of entities in an intermediate entity database
the extraction of relations
the storage of relations in an intermediate entity relation database
the transfer of the intermediate entity and entity relation database into the XML TEI markup and into CIDOC-CRM.

Examples of the Intermediate Entity Database (IEDB) and the Intermediate Entity Relation Database (IERDB) are reproduced in Appendix B and Appendix C. In all these processing steps we follow a three-step procedure.

Fully automatic processing with a high degree of confidence, e.g. a full match of unambiguous entities.

Semi-automatic processing with a lower degree of confidence, e.g. the system ask an expert for confirmation when different entities match onto one string.

Fully manual processing for cases that are too complex or too ambiguous to be recognized or handled automatically. E.g. the foundation date cannot be extracted automatically from IDN-IIa C5.1 ("The only, rather new Chinese temple of the small town.")

3.3.3 Conceptualizing Models in TEI and CIDOC-CRM

One of the main purposes of our research is to represent WFCEM in standardized digital formats, to stimulate related research in various research communities. The two standard formats we will use in the creation of WFCEM in digital formats are TEI and CIDOC.

TEI (Text Encoding Initiative) is the main standard for text-based research in the humanities. Linguists, historians, anthropologists, literary scientists and information scientist collaboratively created this standard and share their annotated data for interdisciplinary research. With TEI we will create a digital representation of the volumes that Wolfgang Franke has created. The TEI representation can be digitally archived by librarians and serve as a master format of other output formats, e.g. a website or a PDF file. Sites documented during our fieldwork, which have not already been documented in WFCEM, will in a first step not be elaborated in TEI.

In contrast to TEI, which is used by textual studies and historians, CIDOC-CRM is used mainly by anthropologists, archaeologists, museum scientists and professionals involved in the management of archaeological or heritage sites. With CIDOC-CRM the focus of description shifts from the text to the site and its objects, their relations to other objects and the transformation they underwent. We record all sites, even those not previously documented by Franke, in this model. The CIDOC-CRM standard has been developed from 1996 under the auspices of the ICOM-CIDOC-CRM (International Council for Museums – International Committee on Documentation) Documentation Standards Working Group. The CIDOC-CRM model is an object-oriented ontology and can be implemented in various formats, of which the most commonly used is RDF (Resource Description Framework). In 2006, the International Organization for Standardization (ISO) adopted CIDOC as standard ISO 21127:2014, called "Information and Documentation: A Reference Ontology for the Interchange of Cultural Heritage Information".

The focus of CIDOC-CRM lies on the formal description of cultural heritate objects, their locations, periodization, their connections etc..

CIDOC-CRM is not desigend as an all-comprising standard. Instead, it is one standard in the CIDOC-CRM family of models, which can be combined according to the needs of a project. Additional complementary standards are CRMgeo (a spatiotemporal model), CRMinf (an argumentation model), CRMsci (a scientific observation model), CRMdig
(a model for provenance metadata), CRMba (a model for archaeological buildings), CR-
Marcheo (an excavation model), FRBRoo (a model of intellectual processes) and PRES-
Soo (a model of underlying semantics of bibliographic information). However, none of
these standards is nearly complete to formally represent the inscriptions on the object and
their linguistic features. TEI will thus not only be used to represent the documentation
published in WFCEM, but also to represent the inscriptions further referenced by CIDOC-
CRM.

The example of a TEI annotation is shown in Appendix D for the objects shown in
Figure 4. Entities collected in the IEDB have already been annotated in that example.

The TEI files produced this way can serve many purposes. They can be accessed by
visualization tools, e.g. for the creation of timelines, maps and social networks, or by statistics programs for theory testing. Alternatively the TEI files can be presented as hypertexts in HTML or PDF. The trans formation from a TEI format into another human-readable format is performed through XSLT style-sheets. One example for the result of such a trans formation, a simple visualization as HTML hypertext is shown in Appendix E.

CIDOC data, which by definition follow an object-oriented model, is created by adding
relations to the set of relations that the object entertains. The resulting data can be part of
a database system or used (and distributed) in XML format.

3.4 Analyzing the Data Using Digital Techniques: GIS, Time-
lines and Social Network Analysis

The data, when summarized in generalizing views, reveal common patterns that tell us
how communities usually developed. The development of a community that deviates from
these common patterns may have its cause in a particular history that either can be directly
pinpointed at by the data or that an extended research might reveal.

Wolfgang Franke, for example reasons that the onset of epigraphy in a community
depends on a certain level of technical, literary and cultural development. From there he
assumed that epigraphy, once started, developed simultaneous across media there should
be a common and general onset of epigraphs, and that, for instance, the earliest tombstones
and the earliest wood tablets had been carved at about the same time.

Pursuing this idea, we have digitized a few pages of IDN-IIa. Then we have plotted
the year of the carving of the extracted epigraphs, their object types and the name of their
respective community. As can be seen in Figure 7 and Figure 8, in some communities
the onset of tombstone and tablet carvings did indeed occur, as Franke assumed, in the
same period. Inscriptions on altars and censers, however, tend to occur later, and mostly
together, while bells, usually imported from China, show no regularity with respect to when
they have been ordered and installed. We might therefore argue that bells depend less on
the cultural than on the economic development, in as far as its wealth allowed to purchase
the expensive bronze bells. The appearance of temple bells, imported of locally produced,
might thus serve as a marker for an economic development that can be used to compare
communities, regions and countries.

Where the first gravestones appear long after wooden tablets, such as in Pematang
Siantar, or vice versa, such as in Bengkalis, researchers would have to find a reason in the
local history of these communities to explain these particular distributions. For Pematang
Siantar, Wolfgang Franke himself, apparently surprised, noticed that “no tombstones of the
late Qing or early Republican area were found”. This illustrates how ditial approaches can

\footnote{Introduction to MYS-I, pg. 6: “From overall experience it can be established as a rule that at a certain town or settlement the earliest dates of inscribe objects in a temple of other building correspond roughly to the earliest tombstone dates.”}
Figure 7 Pilot Study: The first occurrences of epigraphs on objects of different types, split by the location in Indonesia.
detected deviations from regular patterns, prompting researchers to look deeper into a region of site. Without data visualization, nobody would assume that there might be a story.

4 Conclusion

Digitizing and formalizing WFCEM is a daunting task. Yet, filling gaps in Franke’s documentation and following the traces of more than thousand sites and roughly ten thousand epigraphs is an even greater challenge. While few researchers might start with this task, its completion will have to rest on many shoulders.

As the expected outcomes of this project can be predicted to be highly significant for many researchers and research communities, we hope that more and more researchers might join this effort. It thus will become necessary to find a social, organizational and computational structures that would support this long-term collaborative effort, similar to the collaboration that formed the bases for the creation of WFCEM. As this cultural heritage is eroding day by day, no time has been left to wait.
<table>
<thead>
<tr>
<th>Location</th>
<th>1850</th>
<th>1950</th>
</tr>
</thead>
<tbody>
<tr>
<td>Padang</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tebing Tinggi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telok Betong</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanjung Pura</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanjung Balai</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanjung Morawa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanjung Pandan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanjung Pinang</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pangkal Pinang</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sabang</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seruwai</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sibolga</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sungai Rampah</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palembang</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pangkal Pinang</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pem. Siantari</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Padang</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tebing Tinggi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telok Betong</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanjung Pura</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanjung Balai</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanjung Morawa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanjung Pandan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanjung Pinang</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pangkal Pinang</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pem. Siantari</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 8  Pilot Study: The first occurrences of epigraphs on objects of different types, split by the location in Indonesia (continued)
References


Streiter, Oliver, & Goudin, Yoann. 2016 (March 30 – April 1). Epigraphic practices on Taiwan and Penghu as expressions of power, resistance and mediation. Prague, Czechia.

Streiter, Oliver, & Morris, James X. 2016 (October 17–21). Between documentation and sample: Creating a digital cultural heritage archive of gravesites, tombs and tombstones. Kuala Lumpur, Malaysia.

Streiter, Oliver, Lin, Ann Mei-fang, & Goudin, Yoann. 2011a. Close to heart, close to home: Gravesites in Taiwan.

Streiter, Oliver, Goudin, Yoann, & Huang, Chun (Jimmy). 2011b. ThakBong, digitizing Taiwan’s tombstones for teaching, research and documentation.

Streiter, Oliver, Goudin, Yoann, Huang, Chun (Jimmy), & Lin, Ann Mei-fang. 2011c. 紀錄與追跡台灣墓碑地名：以製圖表現其地理分佈及文化意義. National Taiwan University.


奧利華，林莉倫，陳乃瑜，莫詹姆，詹雅晴. 2016. 論澎湖西嶼新發現之「皇明洪門楊氏」墓. 台灣學誌, 14 期, 49–74.

張少寬. 2013. 槟榔嶼華人寺廟碑銘集錄. 南洋田野研究室.

A An Example of the Intermediate Annotation Format (IAF)

typ:loc
loc:region
tic:印度尼西亚; 亞齊
ide:A
tie:Indonesia; Aceh
nor:4.695135
eas:96.7493993
complete:y

===:
typ:loc
loc:village
ide:A
tic:印度尼西亚; 亞齊; 班達亞齊, 大亞齊
tie:Indonesia; Aceh; Banda Aceh
inf:Former name Kota Raja 古打拉惹, is the capital of Aceh Province.
eas:95.3237559
nor:5.5482904
complete:y

===:
typ:sit
ide:A.1
sit:museum
tic:印度尼西亚; 亞齊; 班達亞齊, 大亞齊; 博物館
tie:Indonesia; Aceh; Banda Aceh; The Museum in the centre of the town
eas:95.3149826
nor:5.5479392
complete:y

===:
typ:obj
ide:A.1.1
obj:bell
mat:bronze
tic:印度尼西亚; 亞齊; 班達亞齊, 大亞齊; 博物館; 鐘銘
tie:Indonesia; Aceh; Banda Aceh; The Museum in the centre of the town; \\Bronze bell
dat:成化五年冬月吉日造
yea:1469
img:2.jpeg
eas:95.3149826
nor:5.5479392
complete:y
inf:A large bronze bell, 1.25 m high and 0.75 m in diameter, suspended ...
The codes used in this representation are:

typ: type, possible values loc, site or obj for locations, sites and objects respectively,

loc: the kind of locality, possible values region or village for larger administrative units or cities, towns or villages,

sit: the kind of site, possible values museum, house, temple, school, memorial site or gravesite

obj: the kind of object, possible values altar, bamboo tube, bell, censer, couplet, gate couplets, horizontal wooden tablets, lion, memorial stone, niche, pillar, stele, tablet, tombstone and yunban,

tic: the Chinese name of the location, site or object,

tic: the English name of the location, site or object,

ide: the identifier assigned in WFCEM

nor: the latitude of the location, site or object, not provided in WFCEM

eas: the longitude of the location, site or object, not provided in WFCEM

inf: textual explanations provided in WFCEM,

yea: the year of the creation, erection or establishment of the object, site or location

img: the image associated with the site or object
B  An Example of the Intermediate Entity Database (IEDB)

Each entity is stored with its type, e.g. 'person', 'date', 'event' etc, all expression that refer to that entity, e.g. '黃華生' and a unique identifier for each entity, e.g. '1'. For each entity the source of its identification is retained.

<table>
<thead>
<tr>
<th>id</th>
<th>entity.id</th>
<th>entity.type</th>
<th>entity.reference</th>
<th>source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>tombstone</td>
<td>黃華生之墓</td>
<td>WFCEM entry: 黃華生之墓</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>inscription</td>
<td>黃華生之墓</td>
<td>WFCEM entry: 黃華生之墓</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>person</td>
<td>黃華生</td>
<td>WFCEM entry: 黃華生之墓</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>person</td>
<td>Huang Huashen</td>
<td>WFCEM entry: 黃華生之墓</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>date</td>
<td>民國三四年</td>
<td>WFCEM entry: 黃華生之墓</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>date</td>
<td>公元一九四五年</td>
<td>WFCEM entry: 林媽之墓</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>location</td>
<td>粵邑</td>
<td>WFCEM entry: 黃華生之墓</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>material</td>
<td>concrete</td>
<td>WFCEM entry: 黃華生之墓</td>
</tr>
</tbody>
</table>

C  An Example of the Intermediate Entity Relation Database (IERDB)

The relations are stored temporarily in triples, the subject, the predicate and the object of the relation. In addition, for each entity the semantic role in which it stands to the relation is stored. Again for each relation, we retain its source.

<table>
<thead>
<tr>
<th>id</th>
<th>entity1.id</th>
<th>entity1.role</th>
<th>relation</th>
<th>entity2.id</th>
<th>entity2.role</th>
<th>source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>theme</td>
<td>contain</td>
<td>1</td>
<td>location</td>
<td>entry: 黃華生之墓</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>theme</td>
<td>create</td>
<td>4</td>
<td>date</td>
<td>entry: 黃華生之墓</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>theme</td>
<td>create</td>
<td>4</td>
<td>date</td>
<td>entry: 黃華生之墓</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>theme</td>
<td>create</td>
<td>4</td>
<td>date</td>
<td>entry: 黃華生之墓</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>theme</td>
<td>jiguang</td>
<td>5</td>
<td>location</td>
<td>entry: 黃華生之墓</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>theme</td>
<td>create</td>
<td>6</td>
<td>material</td>
<td>entry: 黃華生之墓</td>
</tr>
</tbody>
</table>
D An Example of a TEI Encoding

<?xml version="1.0" encoding="UTF-8"?>
<xml-model
href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_allPlus.rng" type="application/xml"
schematypens="http://relaxng.org/ns/structure/1.0"/>
<xml-model href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_allPlus.rng" type="application/xml"
schematypens="http://purl.oclc.org/dsdl/schematron"/>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
<header/>
<fileDesc>
<titleStmt><title>Chinese Epigraphy in SEA - Ind 2 - Î.1 - Fengshan miao 凤山廟</title></titleStmt>
<pubstmt><pubstmt>
<p>Part of this data is digitized from W.Franke (s.b.), other parts were collected by the authors in the course of their fieldwork.</p>
<p>The data that was collected as part of the project "Chinese Epigraphy in Southeast Asia" 1998-2019, is published under a Creative Commons License (CC BY-SA 4.0). </p>
</pubstmt>
<sourceDesc>
</sourceDesc>
</fileDesc>
</header>
</TEI>
The Fengshan miao (鳳山廟) is located at Jalan Kemenangan Tiga no. 48, not far from the Jinde Temple. Its present Indonesian name is Wihara Dharma Wijaya (湧泉寺) or “Temple of the Victorious Law”. According to the stone inscription of 1839, the temple was founded in 1755 by a group of merchants from Changtai in the prefecture of Zhangzhou, Fujian (see ref target="#ind2I-1-2"> I.2</ref>). In honour of Dashi gong (大使公) or Qingyuan zhenjun (清元真君), a deity especially worshipped by the people of Changtai (see also K 19.1). According to one of the oldest inscriptions (ref target="#ind2I-1-2"> I.1.2</ref>), it seems that the deity was first worshipped on the premises of Changtai huiguan (長泰會館), the association founded by the merchants native to that county. It would thus be the oldest association of that kind in Jakarta. Changtai merchants in Singapore proceeded in the same way when, about a century later (in 1849), they founded their own association. 


A wooden incense burner donated by Wang Tianci (王天慈, native of Xiayu (霞浦)), Quanzhou prefecture, Fujian. Dated 1751. The first line of the inscription alludes to the
fact that the <persName key="#deity0002"> Immortal Wuan zhenjun 武安真人 </persName>. (Wuan is a former name of Changtai) has transmitted incense to <placeName>Balang</placeName> (</placeName> Batavia</placeName>). The last line is quoted from an inscription on an incense burner written by emperor <persName type="DDMperson" key="#A002426"> Yuandi 元帝</persName> <!-- this is a name already contained in the DDM authority db, we could also use CBDB -->. Apart from an inscribed wooden altar dated <date>1724</date> (<ref target="#ind2I-1-2-1">ī.2.1</ref>), this is the earliest known inscription in Jakarta.<p>For a translation in French see <bibl>Salmon & Lombard, 1980, p. 140</bibl>.</p>
E  An Example of a TEI visualization

An example of the objects of Ind-IIa shown in Figure 4 and encoded in TEI as shown in Appendix D, transformed into an HTML hypertext. This HTML visualization is one of many possible visualizations that can be created and tuned through style-sheets.
The Mapping of WFCEM Entity Types onto CIDOC-CRM Entities

CIDOC-CRM makes a principal distinction between entities and properties, where properties describe the relations between the entities. Entities include objects, their parts and their surrounding. The basic entity types used in WFCEM, mapped onto CIDOC-CRM Entities.

<table>
<thead>
<tr>
<th>WFCEM entity type</th>
<th>CIDOC-CRM Entity</th>
<th>Example</th>
</tr>
</thead>
</table>
| objects            | E19 Physical Object | 11.1.1 木香爐乾隆拾六
年 (1751)                                                   |
| persons            | E21 Person        | Wang Tianci                                                 |
| sites              | E27 Site          | The Fengshan miao 鳳山廟                                    |
| photo              | E31 Document      | photos                                                      |
| title              | E35 Title         | Indonesia Sibolga Guany-inting                             |
| foundry            | E40 Legal Body    | Dingxin Co., Hongkong                                       |
| location name      | E41 Appellation   | “Batavia”                                                  |
| date               | E49 Time Appellation | 大清乾隆                                                   |
| date               | E50 Date          | 大清乾隆拾陸年歲在辛末余月                                  |
| location           | E52 Time-Span     | between 1905 and 1908                                       |
| location           | E53 Place         | Batavia                                                     |
| material           | E55 Material      | wood                                                        |
| material           | E61 Time Primitive | 1905                                                        |
| community          | E74 Group         | Chinese Community of Lhokseumawe                            |
| person’s name      | E82 Actor Appellation | 武安真君傳香吧浪                                      |
| inscriptions       | E90 Symbolic Object | 武安真君傳香吧浪                                         |
| identifiable inscriptions | E73 Information Object | 武安真君傳香吧浪                                      |
| space-time         | E92 Spacetime Volume | Located at Jalan Jen. Gatot Subroto, but no more extant in the 1980s. |

Some entity types are left unexpressed in WFCEM, or are expressed only implicitly or vaguely. Yet, with the help of external knowledge resources, reasoning algorithms or new field data, these missing data can be recovered and represented as CIDOC-CRM Entities.

<table>
<thead>
<tr>
<th>CIDOC-CRM Entity</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>E54 Dimension</td>
<td>size, weight, diameter, density</td>
</tr>
<tr>
<td>E94 Space Object</td>
<td>6.1427,106.8119</td>
</tr>
</tbody>
</table>

Different sub-types of the WFCEM entity types, e.g. the sub-types of the type ‘site’ are distinguished by the CIDOC-CRM property P101 has_as_general_use. Likewise objects will be specified using the property P101 has_as_general_use. How these objects and sites are distinguished is not specified within CIDOC-CRM.
<table>
<thead>
<tr>
<th>WFCEM entity type</th>
<th>CIDOC-CRM Entity</th>
<th>CIDOC-CRM Property</th>
</tr>
</thead>
<tbody>
<tr>
<td>graveyard</td>
<td>E27</td>
<td>P101=community burial</td>
</tr>
<tr>
<td>tomb</td>
<td>E27</td>
<td>P101=personal burial</td>
</tr>
<tr>
<td>school</td>
<td>E27</td>
<td>P101=community teaching</td>
</tr>
<tr>
<td>monument</td>
<td>E27</td>
<td>P101=memorial</td>
</tr>
<tr>
<td>museum</td>
<td>E27</td>
<td>P101=exposition</td>
</tr>
<tr>
<td>temple</td>
<td>E27</td>
<td>P101=community worship</td>
</tr>
<tr>
<td>side-temple</td>
<td>E27</td>
<td>P101=community worship</td>
</tr>
<tr>
<td>tombstone</td>
<td>E19</td>
<td>P101=memorial of deceased</td>
</tr>
<tr>
<td>stone stele</td>
<td>E19</td>
<td>P101=memorial of temple renovation</td>
</tr>
<tr>
<td>wooden tablet</td>
<td>E19</td>
<td>P101=expression of blessing</td>
</tr>
<tr>
<td>incense burner</td>
<td>E19</td>
<td>P101=burning incense</td>
</tr>
<tr>
<td>altar</td>
<td>E19</td>
<td>P101=collecting offerings to venerate a deity</td>
</tr>
</tbody>
</table>

Many entities in CIDOC-CRM can be further specified by part-of relations, specifying, e.g. that the site tomb can be part of the site graveyard, or that the object tombstone is part of the tomb. Objects not included into a larger object are double-encoded as site and object, each with its specific properties. For example, an isolated removed tombstone is in the site of this isolated unused tombstone.