



Open Library of Humanities

Mapping Late Hokusai Research: Digitizing and Publishing Bilingual Research Data

Stephanie Santschi, University of Zurich, CH, stephanie.santschi@uzh.ch

The initiative “Late Hokusai: Thought, Technique, Society” took place at the British Museum (BM) and SOAS, University of London (2016–2019). As part of its activities, it built a linked-data platform prototype on ResearchSpace. The prototype offers a redesigned process for how museum researchers and users find, research with, discuss and expand bilingual data about early modern Japanese artist Katsushika Hokusai (1760–1849) and instigated a discussion about what a collaborative research platform for the Hokusai research community could look like. While Japanese resource specialists have long recognized the complexity of Japanese script as a challenge for multilingual research and collection platforms, the processes for and results of integrating Japanese source data into bi- or multilingual museum databases remained unsatisfactory.

This paper revisits the challenges posed by “non-Latin script” (NLS) in museum databases in the case of the Hokusai research platform at the British Museum, which integrated Japanese and English languages. It localizes the issues arising from working with Japanese source data in the Latin script project environment and accompanies the museum researchers’ tasks regarding the correct input, rendering and display of the source script at each step: 1) object analysis, 2) registering NLS metadata, 3) processing NLS information and 4) visualizing LS and NLS information for general and specialist audiences. After assessing these practices, the paper critically reflects on selected approaches, successes, and shortcomings experienced while creating such a prototype. By sharing its experiences, the project hopes to aid prospective research projects on a similar path regarding project setup and documentation. Furthermore, it advocates the sustainability of research practices according to data reusability parameters.

L’initiative « Late Hokusai : Thought Technique and Society » (Hokusai tardif : Pensées techniques et société) a eu lieu au British Museum (BM) et SOAS, l’Université de Londres (2016–2019). Dans le cadre des activités, cette initiative a produit une plateforme prototype de Web des données sur ResearchSpace. Le prototype offre un processus redessiné aidant les chercheurs de musée et les usagers à trouver, à faire de la recherche, à discuter et à étoffer les données bilingues concernant l’artiste Katsushika Hokusai (1760–1849), du début de l’ère moderne japonaise. Cela a déclenché



une discussion sur l'apparence possible d'une plateforme de recherche collaborative dédiée à la communauté de recherche sur Hokusai. Tandis que les spécialistes de ressources japonaises reconnaissent depuis longtemps la complexité de l'écriture japonaise comme un défi pour la recherche multilingue et pour les plateformes de collection, les processus et les résultats de l'intégration des données sources japonaises dans des bases de données de musées bi- ou plurilingues demeurent insatisfaisants.

Cet article réexamine les défis liés à des « écritures non-latines » (NLS, non-Latin script) dans des bases de données de musée dans le cas de la plateforme de recherche sur Hokusai au British Museum, ce qui a intégré les langues japonaise et anglaise. L'article localise les questions qui se posent durant le travail avec les données sources japonaises dans un environnement de projet en écriture latine et accompagne les tâches des chercheurs de musée concernant l'entrée correcte, le rendu et l'affichage de l'écriture source à chaque étape : 1) les analyses d'objet, 2) les enregistrements de métadonnées NLS, 3) le traitement de l'information NLS et 4) la visualisation de l'information LS (écriture latine, Latin script) et NLS pour des audiences générales et spécialistes. Cet article présentera une évaluation de ces pratiques et, ensuite, considérera de façon critique les approches sélectionnées, les succès et les défauts rencontrés pendant la création d'un tel prototype. En partageant ces expériences, ce projet vise à aider des projets de recherche prospectifs qui se trouvent dans un cas similaire, considérant la configuration de projets et la documentation. En outre, ce projet promeut la viabilité de pratiques de recherche conformément à des paramètres de réutilisation de données.

1 Introduction

NLS resource specialists long recognized the lack of literature about “research into languages that do not utilize the Roman-based alphabet” (Goggin and McLelland 2009, 5). With characters representing both word and image, Japanese script, however, has long challenged the notion of how writing can manifest and has fascinated Japanese artwork viewers from within and outside the Japanese sociolinguistic community with its logographic nature. When larger quantities of Japanese artworks arrived in Europe in the late 19th and early 20th centuries, Japanese scripts arrived with the objects. Japanese prints were a significant vehicle that introduced the qualities of Japanese script to the Latin script (LS) environment. These prints fascinated viewers with their visual qualities and a sense of the “exotic” (for a discussion of orientalism regarding this term, see Saïd 2014). Artists like Vincent van Gogh or Paul Klee, to name just two, are examples of this Japonism. The former framed his “Flowering Plum Tree (after Hiroshige)” in 1887, with symbolic renderings of Japanese characters, and the latter published on and integrated the concept of “text as an image” (“Schriftbild” in Spiller 1956) in his various paintings. In the age of digital processes used to research and reproduce these works, museum researchers, when engaging with scripts as features of artworks and a multilingual public, remain confronted with altered reading and conceptualization process requirements. For example, Elkins (2001, 238) argues that highly pictorial scripts like Chinese alter reading processes by deviating attention from the textual content to the image. The fluid boundary between art and text makes for various ways these characters communicate meaning. Workflows in digital humanities need to find appropriate methods for reacting to semiotic and semantic processes. The Late Hokusai prototype platform on ResearchSpace documents engaging with such conceptual and procedural challenges of NLS in a museum database.

2 The Late Hokusai Research Project

The Arts and Humanities Research Council (AHRC)-funded research project “Late Hokusai: Thought, Technique, Society” took place at the British Museum and SOAS, University of London, between April 2016 and March 2019. The project focused on investigating Japanese artist Katsushika Hokusai (1760–1849) from the perspective of his works during his last three decades. Its core research questions were how Hokusai’s context influenced his artworks and how digital methodology can support recognizing implicit connections between the artist and his society, faith, and technological setting over time. Inter-institutional collaboration with partners in the US, Europe, and Japan

resulted in researchers employing multiple languages and scripts, outside and within the digital realm. Hereafter, the paper focuses on the latter.

In autumn 2015, researchers from the fields of Japanese art history, Japanese collections, and museum information infrastructures met at the British Museum to discuss the creation of a digital platform for Hokusai research. Their expectations influenced the specification of a “pilot online interface.” The proposal initially envisioned the online resource to provide “a new model for the online study of cultural materials, provide open access to research findings, bring together material from multiple collections, and enable innovative, flexible searching” (AHRC 2019). Their key area of interest was how the life of and works by Hokusai relate. When Hokusai scholars refer to Hokusai’s life, they usually refer to the posthumous *Katsushika Hokusai den*, written 40 years after Hokusai’s passing (current edition: Iijima and Suzuki 1999). However, studying works created with Hokusai’s involvement, such as paintings, prints, letters, and illustrated books, is a method to complement the otherwise scarce historical documentation.

To make such information accessible to researchers and general users alike, the Late Hokusai Project planned a research platform in the knowledge infrastructure of ResearchSpace. The platform would contain two spheres. One of those would be for an interested public, and the other for specialized Hokusai researchers. Both in audiences unable to read Japanese script or understand Japanese languages, and in Japanese-only audiences, the ability to understand more than one language cannot be assumed. Therefore, an ideal platform should be customizable to work in Japanese and English and invite visitors and specialized users to view and enrich new information in the language they feel most comfortable.

3 Where does the NLS script appear?

Generally, script destined for visitor consumption appears in different environments within the museum. Japanese art-related script at the BM includes special exhibitions, galleries displaying permanent collections, catalogues, and digital media. Interestingly, language-specific catalogues tend to get published as separate publications because bilingual exhibition catalogues sell comparatively worse. So, for example, the catalogue of the 2017 Hokusai exhibition (Clark 2017) in London, which later moved to Osaka, Japan, was translated into separate publications in Japanese (Asano and Clark 2017), Italian (Clark 2018a), and Chinese (Clark 2018b). Linguistic segregation appears common. The priority of English results from the primary language present in the sphere wherein the BM is located, English, and the lingua franca used among

museum visitors and researchers. As for the reasons for including Japanese scripts in exhibition environments, the decorative aspect of Japanese scripts is a major factor. For example, the BM Mitsubishi Corporation Japanese Galleries, which reopened in 2018, prioritizes the display of the English language over the Japanese but does include Japanese characters in gallery title, section headings, and title and artist name on each object label.

The *linguae francae* among project researchers were Japanese and English, especially in oral exchanges. Primary object data existed in Japanese and secondary sources in languages including English, Japanese, French, and others. Whereas the European language use LS, Japanese is an example of a tripartite NLS. The characteristic qualities of Japanese as NLS, in a digital framework, challenged the established procedures of project researchers working on the digital platform creation and implementation in several ways. Below, the paper looks at the challenges posed by Japanese NLS in a procedural framework. It describes the problem at the level of source data, data input, processing and presenting, and reflects on the chosen path among potential solutions in retrospect.

3.1 NLS in source data

Initially, Japanese scripts are present at the level of the objects that the Late Hokusai Project researched. In paintings, prints, books, and letters, among other objects, the text, besides images, is a primary mode of communicating meaning. As a result, the works scrutinized by the Late Hokusai project as historical objects present two kinds of challenges. On the one hand, the works encapsulate a distinct historic moment with their own writing conventions. On the other hand, the Japanese script inscriptions bridge the dichotomy between text and images.

For example, **Figure 1** shows a page from the first volume of *Ehon hayabiki* 絵本早引. Here, readers can see the three types of script used in early modern Japan: *kanji*, *hiragana*, and *katakana*. The captions of the whimsical illustrations on the left primarily consist of a Chinese character script derivative. Kanji, as they are called in Japan, are derived from a set of Chinese characters developed before and during the Han dynasty (206 BC–220 AD) and accompanied Buddhism's arrival in Japan around the 6th century (Seeley 1991). Literally translated as "Chinese characters," they are composites that convey meaning through form. As writers join brush strokes to radicals, characters, and character compounds to a "system of script" (Ledderrose 2000, 8) of Chinese characters, they create a complex modular system with phonetic, semantic, and semiotic functions.



Figure 1: Page from: Hokusai, K. 1817. *Ehon hayabiki* vol. 1. Edo: publisher n.n. Object no. JH.431, vol. 1, image 4. The Trustees of the British Museum. <https://www.britishmuseum.org/collection/image/1521623001>.

The katakana used to transcribe the captions on the left page in Figure 1, and the hiragana, on the top right, are distinct to Japan. Both developed as a shorthand from the *manyōgana*, a Chinese syllabary script used in Japan. On the left page of Figure 1, some katakana accompany the kanji, and on the top right of the same page, the “i-ro-ha” pangram, which contains all hiragana syllables exactly once, appears as an index. Both writing systems are phonetic syllabaries of, in their modern iteration, 46 standard syllables. However, here on the top-right, the syllabary displays the historical 47 characters. This element exemplifies one of the historical irregularities relevant when working with classical versions of Japanese scripts in a digital environment. Kana can accompany or replace the kanji. The cursive hiragana tend to appear in prose, and the sharp-edged and straight-cornered katakana denote official documents (Frellesvig 2010). Each “mora,” phonological unit in either kana system, is represented by a syllable. Today, katakana mainly mark exclamations, foreign origin words, and proper names, and hiragana denote particles and grammatical modifiers (compare with Miller 2015, 5–45).

Together with those three scripts, Japanese text employs reading aids. One of them is *furigana*, minuscule syllabic characters printed next to kanji that render their

pronunciation in kana script. For an example of ruby annotations, see the left page of **Figure 1**, which shows katakana characters used as furigana for rendering the pronunciation of the kanji. In addition, *kunten*, explanatory marks, introduce Japanese grammatical structures to reading Japanese approximations to classical Chinese texts, *kanbun*. The right page of **Figure 2** shows an example of *kanbun*, instantly recognizable by the minor characters to the bottom right and left of the characters. These are diacritical and syntactic markers such as the *kaeriten*, which mark the order in which sentence fragments should be read. Seals use a stylized kanji version, the seal script *tenshotai*. For example, it is visible in the red seals on the right page of **Figure 2**. Further, punctuation marks (see Seeley 1991, 183–185) and Arab numbers may accompany Japanese characters, even though this is rare for the early modern script. Finally, not all characters remain the same over time. Unusual characters not part of the modern sets are called *hentaigana*, “variant characters.”

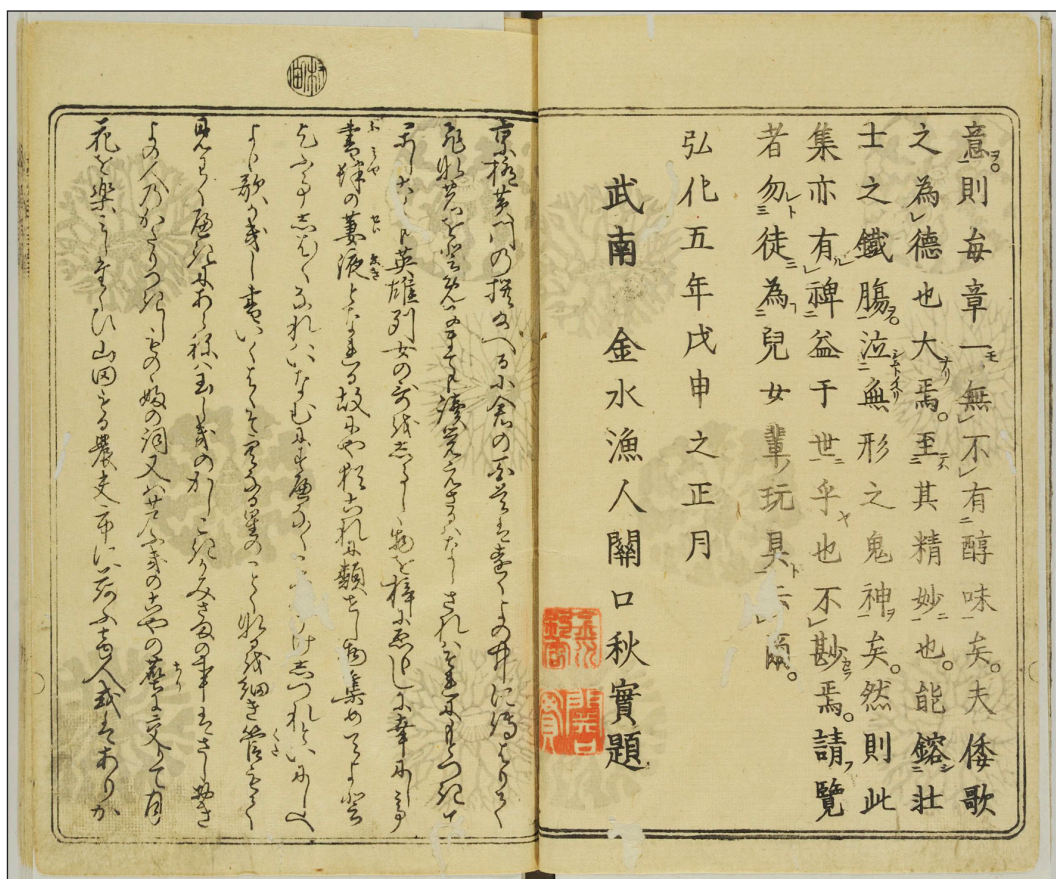


Figure 2: Page from Hokusai, K. 1848. *Shuga hyakunin isshu*. Publisher info n.n. Object no. JH.466, image 3. The Trustees of the British Museum. <https://www.britishmuseum.org/collection/image/1542977001>.

In some cases, the lines between script and image become blurry. Hokusai used to play with the effects of this hybrid device in early manuals such as the *Ehon hayabiki* vol. 2. The example of the title page (left page of **Figure 3**) shows characters assembled from the shapes of auspicious objects. Also, Jippensha Ikku's rake-like doodle on the bottom left of the right page, next to his signature of the preface, displays calligraphic qualities that resist classification according to a pictorial-literary dichotomy. Another visual feature of the script is its writing direction. All examples demonstrate vertical script, running from top right to bottom left. There are few exceptions to this practice in Hokusai's works, such as a small-format print of the *Oshiokuri hatō tsūsen no zu* ("Express delivery boats rowing through waves"), ca. 1800–1805, titled and signed horizontally to evoke a European style. Rarely do works carry handwritten notes in LS. The frontispiece in **Figure 6** is such an example. Those function as indexes to the object's title or content for anyone familiar with LS but unable to read Japanese scripts.

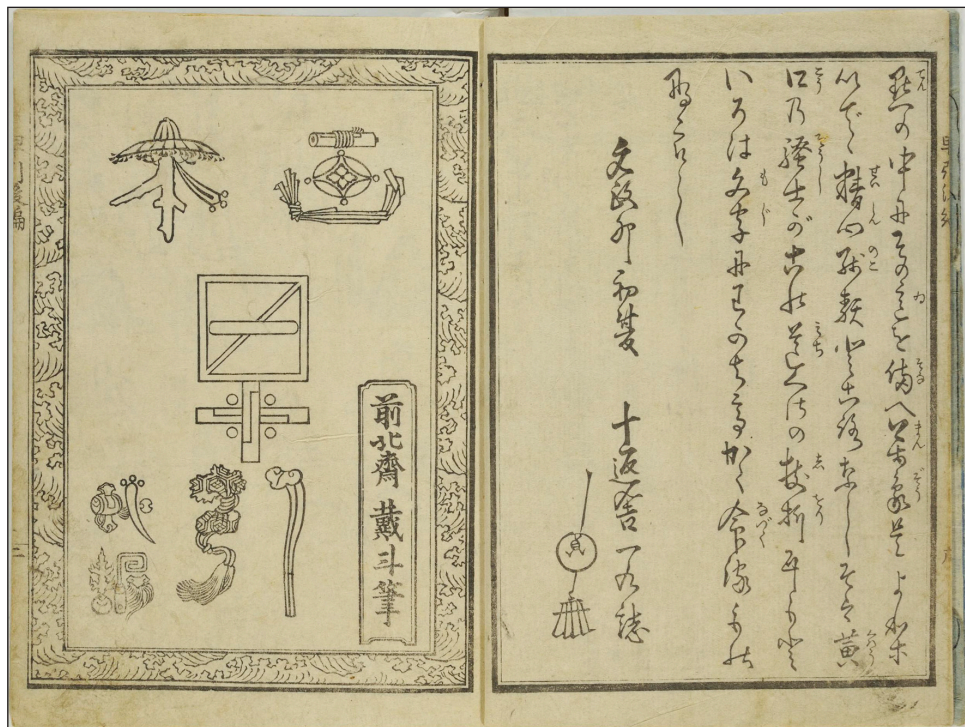


Figure 3: Page from: Hokusai, K. 1819. *Ehon hayabiki* vol. 2, Edo: publisher n.n. Object no. JH.431, vol. 2, image 3. The Trustees of the British Museum. <https://www.britishmuseum.org/collection/image/1521657001>.

In summary, the main challenges regarding the script's presentation are its changes over time and textual-visual features. Furthermore, while the object is historical documentation, its virtual representation in digital collections is a modern-day manifestation whose registration is bound to the technology used. The problems

arising from these historicity and digital representation challenges resurface in the next three steps: data input, transformation, and display.

3.2 NLS during data input

The British Museum researchers working on the Late Hokusai Research Project were bilingual to a professional level in the target languages (English and Japanese). Ideally, such tacit knowledge should find application when registering object information. However, neither institutional priorities nor the collection management system (Merlin) in place at the start of the project in 2016 was optimized for multiple languages and scripts. As a relational database for the museum, Merlin supported collection management and publishing selected information to its platform, Collections Online. In the application interface, museum researchers entered information into fixed fields. While the British Museum upgraded its collection management system to MuseumIndex+ during the research project (see MuseumIndex+ 2022), a large part of Hokusai-related information was digitized even before the project started. Therefore, it was only partially affected by this change in the first place. While Merlin allowed curators to add native language source data and a transliteration or translation for inscriptions, it did not record the hierarchy of terms with the information. In other places, such as for the object title or descriptive fields, adding information in Japanese was only possible by overfilling the target field with several values (e.g., English title / Japanese kanji title / transliterated title). While this is sufficient for virtual display, such data is not usable for digital processes. Further data harvested from the online collections of institutions like the Metropolitan Museum of Art, New York, or the Freer-Sackler Gallery, Smithsonian Institution, Washington DC, two research partners of the Hokusai project, contains little to no Japanese script source texts.

Regarding the general encoding of Japanese scripts for inputting, processing, and displaying, the Unicode consortium, particularly its Ideographic Research Group (IRG), developed and encoded the most comprehensive character sets, known as ISO/IEC Standard 10646 or Unicode. The two scripts distinct to Japan, hiragana and katakana, are each encoded in one code block. However, kanji characters across CSVK (Chinese, Japanese, Korean and Vietnamese scripts—in the case of the latter two, mainly historically used) are part of the same code block that encodes Han script. To make them look regionally distinct, users can change the appearance of characters encoded in Unicode by applying local fonts. Non-standard characters can be submitted to IRG (compare with Unicode Inc. 2020).

When museum researchers enter Japanese into databases, they do so via their computer terminals. On the most basic level, a selective writing system is in place. Morita (1989) documented the first approaches to writing Japanese characters and forecasted

the popularity of the syllabic input process, where users select a character after inputting its pronunciation in syllables. Then, by writing the syllable or word phonetically, users prompt a browsable kanji dictionary to select the right one. Such automatic conversion proved the key to selecting the correct characters from thousands of possible ones. Today, users can install Japanese keyboard applications such as Microsoft-IME (“input method editor”) or Apple’s KOTOERI on any computer. Once activated, users can type syllables and select appropriate characters such as kanji and kana.

While the achievements and documentations by institutions like the Unicode consortium or library services organization OCLC provide invaluable guidance on Japanese as NLS (OCLC 2022), museum researchers remain challenged in their daily tasks by the specificities of Japanese scripts. Those challenges are of a conceptual and technical nature. On a technical level, the challenge originates in non-standard elements such as classical characters not yet part of the Unicode character set and ruby annotations (furigana, kunten, kaeriten). Ruby annotations can only be formatted appropriately with the character or character compound they explain if the digital presentation allows rendering small-scale characters above, to the side, or below full-size characters. While HTML supports ruby markup, Merlin did not, and therefore source data did not contain the appropriate ruby information. (W3C 2016). Finally, the unique kanji-based seal script is not yet part of Unicode; however, its inclusion is under discussion (Everson et al. 2021).

Regarding transcription of terms into LS, Hepburn and *Kunrei-shiki* co-exist as the main romanization systems. However, transcriptions can vary depending on the referred period and target language that determine how a term is pronounced. For example, a transliteration for a French reader can appear different from that for an English reader: for example, the ubiquitous English “Hokusai” and its French historical variation “Hokousai,” first noted in Goncourt (1896), would not automatically be picked up as the same entity by a computational system. Moreover, culturally distinct entities, like local date expressions (duration expressed in a single format), get vague if not explicitly defined. For example, a well-established tacit practice in the field of early modern Japanese (art)history is noting Japanese dates as “the 3rd day of the 4th month.” Such a date refers to a date of the lunisolar calendar unlikely explicitly documented for or known by researchers external to the discipline. If external researchers understood the date as April 3rd in the Gregorian calendar, they introduced errors into the data collection.

In the cases of historical characters that were not part of the Unicode character set at the time of information registration, museum researchers developed piecemeal solutions for encoding them into the database. In the example of “今様櫛☆雛形 (*Imayō sekkin hinagata*),” translated as “Modern Designs for Combs and Tobacco Pipes” (BM,

object no. 1915,0823,0.111), a star was employed in the Japanese title to indicate the presence of the non-standard character. In addition, the key to understanding was added in the object description field: “☆ = 竹 + 搥.” Later, the star appeared changed to the phonetic rendering of the character as “今様櫛きん雛形.” Unfortunately, the semantic relation between symbol and character gets severed in either case. Moreover, the split notation in title and description fields complicates automated data searching and visualization processes.

Figure 4 shows an example from the relational version of the British Museum database, where the Japanese script title accompanies its transcription. No translation is visible at first glance. Besides translations, transcriptions are the prime method for individuals outside the sociolinguistic Japanese sphere to form knowledge about Japan. For example, many British Museum Japanese book collection illustrated books carry title transcriptions on the inside covers. Transcription and translation are shortcuts to understanding the language without fully grasping its script (Miller 2015). However, they do not form part of the data directly associated with the object, which, in this case, would mean the title in Japanese script. This distinction should be respected when registering transcriptions and translations with source data.

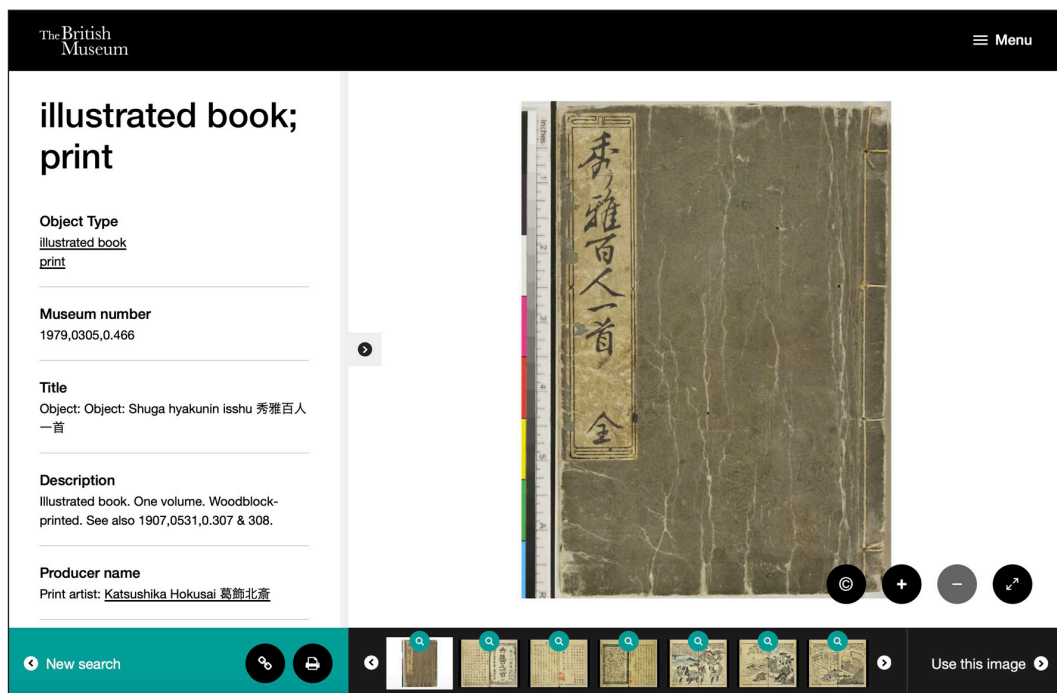


Figure 4: Screenshot of a Japanese illustrated book in the Collection Online of the British Museum. https://www.britishmuseum.org/collection/object/A_1979-0305-0-466 (retrieved 27.1.2022).

Besides relying on existing institutional data, the Late Hokusai Research Project also transcribed, translated, and digitized analogue information. Among original texts were Hokusai's letters, the posthumous biography *Katsushika Hokusai den* (Iijima and Suzuki 1999), the *Catalogue Raisonné of the Surviving Single Sheet Woodblock Prints of Katsushika Hokusai* by Keyes and Morse (1972–2007), and the various prefaces to Hokusai's books. Pictorial and textual digitization tasks involved Japanese scripts to different degrees and depended on the amount of script in the source document. The illustrated books in this paper's illustrations show how much text Hokusai's works can contain. Notably, not all objects contain this much text. Nevertheless, suppose a solution worked for the complex object of the multi-page book in which Japanese script information is present at various levels. In that case, processes are likely to work for bibliographically simpler objects such as hanging scrolls or single-sheet prints.

Overall, regarding data input applications into collections management systems (relational or linked data), pre-formatted input fields modelled on the ontological reality of the object and respecting the conceptual differences between primary and secondary information (in inscriptions or dates) would ideally minimize the risk of accidental mistakes. Also, additions such as language tags could be requested from users automatically at the input level.

3.3 NLS during information processing

Research data processing presented the third area where Japanese as an NLS challenged established processes and applications. The procedural goal of the Late Hokusai Research Project was to investigate means for linking structured data from existing databases, structured analogue information digitized during the project, and unstructured data such as orally expressed opinions and written publications. To weave such a linked-data “mind-map of information,” it tested the semantic web development and humanities data visualization tools provided by the knowledge representation platform ResearchSpace (for description, see Oldman and Tanase 2018). The argumentation model extension to CIDOC CRM, CRM_{inf}, promised to enable a semantically associated, side-to-side display of primary information and secondary argumentation (CIDOC 2022). With ResearchSpace's support, the Late Hokusai project envisioned, the Late Hokusai knowledge graph would become an interactive digital workspace.

Most challenges encountered on this level were of procedural rather than technical nature. Not all museum colleagues who helped process Japanese script data were versed in Japanese, which caused a communicational hurdle to overcome. However, the insufficient quality of existing data for automated processing resulted in a substantial need for data curation even earlier. The source data in its original format was not usable

for digital processes, particularly multilingual embedding. The project had to spend a significant amount of time on data cleaning, including manual and half-automated processing and enriching. A suggestion to facilitate similar projects in the future could be sharing script libraries, for example, via GitHub, to make the overall field benefit from grassroots solutions.

Aiming to create a network of information, the project team extended the conceptual discussion around mapping Hokusai data from paintings and single sheet prints to complex objects such as illustrated books. Because the institutional records document objects and not contextual research, their description fields contained only minor references to research done on and with them. For example, information on artists' networks, colourants, or ways of rebinding books, which is documented external to the platforms in books, journals, and exhibition catalogues, were only occasionally inserted as references for database viewers to consult if they wanted to learn more about the object. Therefore, in parallel to data curation, the Late Hokusai project created models of conceptual relationships between objects, people, events, and spatiotemporal data present in the included repositories. Their goal was to build starting points onto which researchers of specific objects, concepts or phenomena could tie their argumentation. The project used the CIDOC CRM ontology v.6.2.1 (CIDOC 2020) to describe central entities and their connections. These models were then used to add structure and unique resource identifiers to institutional participants' XML-derived mono- and bilingual source records. Finally, the entire dataset was uploaded into a Hokusai knowledge graph on ResearchSpace, forming a network of Hokusai-related information.

Unique resource identifiers are a crucial component of linked data applications. Research projects across linguistic divides can establish common grounds for discussion by using common denominators for entities. Such denominators emerge from reference data repositories like Wikidata, where entities like Hokusai are noted via unique "item identifiers." The person of Hokusai, for example, is linked to the identifier Q5586. Such modules create an ontological representation of real-world relationships independent of encryption and language. An example is Hokusai's name. "Hokusai" appears differently transliterated in the different databases. Any collection might hold a variation of his name, which the artist had changed over 30 times throughout his life. Ideally, resource specialists employ standard identifiers from global repositories (for example, Wikidata, GND or Getty AAT); however, an imbalance towards Latin script terms and non-Japanese entities (particularly regarding lesser-known individuals like rarely mentioned block cutters) was evident in these repositories during the project duration.

Regarding modelling NLS and LS script information, the project focused on distinguishing source data and child-level information. A model is visible in **Figure 6**, which originates from the exhibition interface created for *Hokusai: The Great Picture Book of Everything* (Hokusai: The Great Picture Book of Everything 2021), launched a bit more than two years after the culmination of the Late Hokusai project at the British Museum (30 September 2021 – 30 January 2022, see Clark 2021). This platform builds on the work done for Late Hokusai ResearchSpace. It visualizes the conceptual mapping of terms and transcriptions/translations in a display of the frontispiece of the drawing collection *Banmotsu ehon daizen zu* 万物絵本大全図. The frontispiece carries the title 西戎中華. Researching using the ResearchSpace prototype, researcher Timothy Clark had marked the area in the image that contains these characters as feature regions using an application of the Mirador viewer added into the project interface of Late Hokusai ResearchSpace. This image region is encoded as an informational object “Inscription” and conceptually linked to the page that carries this inscription. The Inscription itself is the centre of a conceptual map onto which transcriptions and translations can be tied. Such a model could be expanded further to distinguish, for example, historical script from modern-day Japanese transcription or accommodate different translations by different researchers. The distinction between evidence and argument is crucial and the semantic model’s main benefit.

3.4 NLS as part of visualizations for specialized and general users

The fourth and last level on which Japanese as NLS poses distinct challenges is the interface level. On the interface, users see digitally encoded information. As users, the Late Hokusai ResearchSpace envisioned two types: specialized researchers and a loosely defined public. What distinguishes specialized from the general user is which understanding they bring to the platform. While Hokusai researchers, regardless of their primary language, are qualified to work with Japanese scripts and language, general users might not read and understand Japanese, not even in LS transcriptions. For the user interface, this means catering to both, including, on the one hand, the Hokusai researchers interested in Japanese scripts’ historical variability, and, on the other hand, monolingual users that selected the language according to their preference in a low-threshold user interface.

The qualities of text as image and image as text are intimately connected to the properties of Japanese script. Regarding the digital treatment of word-image interplay, the Late Hokusai project wanted to allow users to link observations about images to other pieces of information (text or image) stored in the knowledge graph. Referring

to the various semantic and semiotic encoding of meaning regarding textual content seems particularly relevant when looking at the pictorial potency of the classical Japanese language, where the allusive use of characters invites readers to think creatively beyond simple literal translation (Miller 2015, 179). ResearchSpace offered the possibility of linking image content and textual information and converting visual information into conceptual items. Those conceptual items could then present the evidence for consecutive argumentation with image annotators. Annotated items build a thesaurus of terms users refer to in ongoing image annotation work. Users can then access visual and textual evidence by search tools and integrate them into their argumentation (compare with Oldman and Tanase 2018). Visual annotations are a means to encode visual inscriptions, such as in **Figure 3**. In that example, the scroll forming the top horizontal stroke of the character 画, *ga*, in the title, could simultaneously be encoded as a conceptual entity “scroll” (which, in a formalized thesaurus, could be translated into Japanese or other languages or even linked to repositories such as Wikidata as Q720106). At the same time, the full title could be registered as an inscription with the content 画本早引, from which its transcription as “*Ehon hayabiki*” and respective translations could be mapped as informational objects.

While the Late Hokusai project envisioned the user interface to have a fully bilingual search and input functionality, the issue of bilingual data quality and the lack of a coherent updated translation of the semantic properties prevented ultimately realizing this vision within the scope of the project (2016–19). User interfaces would need to integrate an updated version of the CIDOC CRM reference model (in Japanese and English) to display the properties and entities information in the Japanese language. The newest translation available at the termination of the project in 2019 was version 3.4, edited by Kujirai (2003). However, new efforts seem to be underway (Rekihaku 2021) in connection to the *Khirin* project (Rekihaku 2022). In a fully bilingual digital workspace, the users would benefit from understanding the intricate details of semantic properties on either side of the linguistic divide. A complete translation of the ontology and its explanation likely prevents mapping, labelling, and visualizing mistakes. Finally, a last challenge for the interface conceptualization was the simultaneous presence of English and Japanese scripts. Font selection and design had to react to the different needs regarding space and directionality. As the design included informational cards for each entity accompanying LS scripts of varying length, the team chose a horizontal left-to-right display of the scripts. **Figures 5 and 6** show examples of this.

Late Hokusai The project ▾ Example Resources ▾ Dashboard Thinking Frames 🔍 ?

Clipboard

Late Hokusai 北斎の晩年

Katsushika Hokusai (1760-1849) is a Japanese artist who has created art with a truly global reach. During his exceptionally long life, Hokusai created brush paintings, drawings, woodblock prints, and illustrated books.

In the Late Hokusai research project, we use these various works to understand how the environment in which he created them influenced Hokusai's individual artistic style, technique and choice of subject matter.

Project Details > Project Partners > Search All >

The research for this project is conducted on [ResearchSpace](#).

Discover existing works and the work of researchers companions to the artworks themselves (see example resources). Looking beyond the works themselves, our online semantic research space will encourage more scholars and the public to discover the many strands of meaning that connected Hokusai to the world around him.

Figure 5: Screenshot of the Start Page of Late Hokusai ResearchSpace at the British Museum. <https://latehokusai.researchspace.org/resource/rsp:Start> (retrieved 28.1.2022).

4 Evaluation

In the virtual space, the Japanese language and script exist in two conceptual spheres, firstly, as a reference to the object that carries them as primary source information, secondly, as a de-contextualized language with which information can be accessed. Therefore, using language to document meaning in a museum environment requires awareness of the language's regionally specific properties and finding global solutions. This awareness is apparent at the level of museum researchers who deal with the challenges of bilingual, multiscript data every day, but seems of lesser priority to institutions of the magnitude of the British Museum who would need to drive appropriate changes at levels of data collection, processing, and visualization. In the case of the British Museum, its sheer size, and the precedent that decisions for one linguistic area would present for another, might slow this development. Accordingly, despite the broad knowledge of their staff, museums are losing values that would be available to them at a little extra investment. In essence, the multidimensional layers

The screenshot shows a digital interface for an exhibition. At the top, it reads "Hokusai: The Great Picture Book of Everything" and "The exhibition online Browse 103 drawings". Below this, the title "Display Case 2, exhibition section 2: Hokusai and the picture encyclopedia - 北斎と絵入り百科事典" is displayed. There are three buttons: "All drawings", "Related subjects", and "Clear highlight".

The main content is a knowledge map centered on a drawing titled "EXHIBITION DRAWING - N.2020.3015.2" with the title "Frontispiece: 'India, China'". This central card is connected to three other cards:

- Left card:** "Visual Item: Chinese boy". It features a drawing of a boy and an inscription: "INSCRIPTION 中華" (China) with translations and transliterations.
- Right card:** "Visual Item: Indian man". It features a drawing of a man and an inscription: "INSCRIPTION 西戎" (India) with translations and transliterations.
- Bottom-left card:** "Hokusai 103 dessins, y compris le titre" (Hokusai 103 drawings including the title) in French, with a translation and transliteration.
- Bottom-right card:** "MARK" showing a red seal with the text "anri veveeru" and a note: "Collector's seal of Henri Veber (1854-1942)".

Relationships are indicated by arrows: "has feature" connects the central drawing to the Chinese and Indian boy items, and "has inscription" and "has mark" connect it to the bottom cards. A search and zoom interface is visible in the bottom right corner.

Figure 6: Screenshot of the Great Picture Book ResearchSpace exhibition portal, showing a map of inscriptions, transcriptions, and translations on the example of the Frontispiece of *Banmotsu Daisen zu*. Display Case 2, Knowledge map of the Frontispiece, https://hokusai-great-picture-book-everything.researchspace.org/resource/rsp:Exhibition_greatPictureBook (retrieved 26.1.2022).

of tacit knowledge and implicit understanding among resource specialists get lost if not expressed explicitly. However, by not challenging existing collection management systems with multilingual and multi-script information, creating appropriate input fields remains a low priority. Therefore, the author urges developing content and appropriate documentation processes in synergy and active communication with resource specialists and governing bodies.

Automatization could support museum researchers on data understanding, input, and processing levels. In recent years, tools based on machine learning have shown promise. For example, accessing Japanese script content includes applications that combine optical character recognition with machine learning (for example, Kuronet for cursive script, see Lamb, Clanuwat, and Kitamoto 2020). On the level of translation,

these include automated translators such as Google Translate, Microsoft Translator, and DeepL Translator (for an evaluation of these translators for ukiyo-e titles, see Song, Batjargal, and Maeda 2020). While none of these tools is an end-all solution to the challenges of Japanese language and script in museum environments, they are tools that, when used correctly, can substantially facilitate researchers' digitalization tasks. Further automation at the input level could include the compulsory and automated addition of language tags for source data and its transliterations (into kana and Latin script characters), according to IANA's current recommendations (IANA 2020). Regarding the issue of date reference systems, widgets such as "NengoCalc" (Schemm 2006) are helpful tools for transforming individual dates. Unfortunately, no widget was available at open-source licenses during the project duration, and contacting its creator was unsuccessful. Therefore, the process could not be automated within the Hokusai platform.

Working more intuitively at the intersection of image and text was one goal of the Hokusai project. Bridging the gap between semantics and semiotics is significant for the field of Japanese art history that looks at illustrated books, single sheet impressions bearing titles or signatures, *surimono* that illustrate poems, or seals that themselves represent text in a stylized, pictorial format. To comprehensively understand those materials' textual and visual information, researchers need to study, document, and access those factors together. Conceptual models for the data structure allow encoding images into processable code with relationships. Of course, conceptual entities would still need to be translated for selected (or ideally all) cases. However, how information connects could be tied to the models and linguistic encoding.

5 Conclusion

Initially, the Late Hokusai Project planned the platform to enable researchers to enrich the knowledge graph with information on provenance, authenticity, style, influence, or artistic networks in multiple scripts and languages. The basis for this would be the real-world conceptual relationships modelled with computer ontologies. In addition, building an informational graph would ensure that human researchers and computer processes can perform searches, analyses, and visualizations with the data. Employing ResearchSpace's semantic web toolset in a multilingual environment brought the status quo of noting and processing multilingual information in museum repositories into question. The prototype achieved by the end of the project, and its expansion with the Great Picture book exhibition website, demonstrate that working with a linked-data knowledge graph benefits researching Hokusai. The shortcomings of the platform regarding a fully bilingual operability also show that to ultimately realize a

sustainable research platform, developing and subscribing to standards with long-term institutional support across the linguistic spheres remains necessary.

The paper sees digital applications as applicable to museum researchers who process Japanese language and script in two ways. On the one hand, researchers can use digital transcription and translation applications to process Japanese as NLS. As the applications evolve with machine learning algorithms, they are even more promising regarding advancing research processes and subject knowledge collaboratively. On the other hand, artworks carry unique linguistic, scriptural, and visual features that demand specialized local treatment and for which semantic modelling offers a conceptual solution that bridges individual and global requirements. Unfortunately, largely dependent on legacy data, the online portals of museum collection databases that hold Japanese works outside Japan still contain only small amounts of Japanese script. The reason for this scarcity of Japanese as NLS in research and application lies in three areas, as the paper has shown:

1. The connection between institutional linguistic embedding and the selection of scripts and languages appears politically motivated. Institutions' linguistic environment and institutional priorities influence how their researchers deal with additional languages and scripts.
2. The space required to display several languages makes demands on exhibition layouts, which want to display just the right amount of information in an easy-to-read language to the user. Therefore, in LS spaces, the Japanese script has predominantly decorative functions.
3. The simultaneous display of several languages conceptually challenges research.

The third area, workflows challenged by NLS scripts, is the most interesting for future developments. If museums remain interested in offering their data for computational processing, they need to find ways to improve their data quality on a conceptual and formal level. Temporary solutions of data registration, processing, and display, to which the Hokusai project had to resort to lack other possibilities, do not suffice to ensure full interoperability between languages from a data processing and access perspective.

The exhibition platform for *Hokusai: The Great Picture Book of Everything*, which section 3.3. mentions and **Figure 6** illustrates, demonstrates that the initial Late Hokusai prototype on the ResearchSpace platform finds application in goal-oriented use for exhibition research and information display. It also shows that some of the issues present in the conceptualization of the bilingual and multiscript environment of

the Late Hokusai project could be resolved at least on the level of linking data. Having learned by experience, the Hokusai Research Project encourages future projects to devise a concise data management plan that considers multilingual and -scriptural points from the start. Best practice recommendations and (linked) open data standards help achieve global solutions that allow more effective communication and long-term sustainability of project outcomes. They will need to be respected at all stages of the data collection and transformation process if museums aim to contribute to FAIR data (see Wilkinson et al. 2016) that can be used in digital humanities analyses. Valuable resources for museum researchers are the community resources of consortiums such as OCLC, Unicode, and Wikidata. Particularly resource specialists among librarians are key information points that can help ensure the reusability and reliability of data. Finally, communicating standards that ensure data quality within the domain of museum data management could help achieve this goal and support users in conducting research more effectively.

Competing Interests

The author has no competing interests to declare.

Contributions

Editorial contributions

Section Editor

Shahina Parvin, The Journal Incubator, Brandon University, Canada

Copy and Layout editor

Christa Avram, The Journal Incubator, University of Lethbridge, Canada

References

- AHRC (Arts and Humanities Research Council). 2019. "Late Hokusai: Thought, Technique, Society." UK Research and Innovation. Accessed May 7, 2022. <https://gtr.ukri.org/projects?ref=AH%2FN00440X%2F1>.
- Asano, Shūgō, and Timothy Clark. 2017. *Hokusai: Fuji o koete*. Osaka: Abeno Harukas.
- CIDOC (International Committee for Documentation). 2020. "CIDOC Conceptual Reference Model (CRM)." ICOM (International Council of Museums) and CIDOC. Accessed May 7, 2022. <http://www.cidoc-crm.org/>.
- . 2022. "CRMinf Argumentation Model." ICOM (International Council of Museums) and CIDOC. Accessed May 16, 2022. <https://cidoc-crm.org/crminf/>.
- Clark, Timothy, ed. 2017. *Hokusai: Beyond the Great Wave*. London: Thames & Hudson.
- , ed. 2018a. *Hokusai: Oltre la grande onda*. Torino: Einaudi.
- , ed. 2018b. *Geshi beizhai chaoyue ju lang*. Wuhan: Huazhong University of Science and Technology Press.
- . 2021. *Hokusai: The Great Picture Book of Everything*. London: The British Museum Press.
- Elkins, James. 2001. *The Domain of Images*. Ithaca: Cornell University Press. DOI: <https://doi.org/10.7591/9781501723902>.
- Everson, Michael, Rick McGowan, Ken Whistler, and V.S. Umamaheswaran. 2021. "Roadmap to the TIP." *Roadmaps to Unicode. Unicode*. Accessed May 16, 2022. <https://www.unicode.org/roadmaps/tip/tip-14-0-0.html>.
- Frellesvig, Bjarke. 2010. *A History of the Japanese Language*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511778322>.
- Goggin, Gerard, and Mark McLelland. 2009. "Introduction: Internationalizing Internet Studies." In *Internationalizing Internet Studies: Beyond Anglophone Paradigms*, edited by Gerard Goggin and Mark McLelland, 3–17. New York: Routledge. DOI: <https://doi.org/10.4324/9780203891421>.
- Goncourt, Edmond de. 1896. *Hokousai: L'art japonais au XVIIIe siècle*, edited by E. Flammarion and E. Fasquelle. Paris.

- Hokusai: The Great Picture Book of Everything. 2021. "From the Rediscovery to the Exhibition." Semantic Exhibition Platform on ResearchSpace. Accessed May 16, 2022. <https://hokusai-great-picture-book-everything.researchspace.org/resource/rsp:Start>.
- IANA (Internet Assigned Numbers Authority). 2020. "Language Subtag Registry." Internet Assigned Numbers Authority. Accessed May 16, 2022. <https://www.iana.org/assignments/language-subtag-registry/language-subtag-registry>.
- Iijima, Kyoshin, and Jūzō Suzuki. 1999. *Katsushika Hokusai den*. Tokyo: Iwanami Shoten.
- Keyes, Roger S., and Peter Morse. 1972–2007. *Catalogue Raisonné of the Surviving Single Sheet Woodblock Prints of Katsushika Hokusai*. 90 vols. Unpublished manuscript deposited at the Department of Asia, British Museum, London.
- Kujirai, Hidenobu, ed. 2003. *Bunka isan jouhou no data moderu to CRM*. Tokyo: Bensei Shuppan.
- Lamb, Alex, Tarin Clanuwat, and Asanobu Kitamoto. 2020. "KuroNet: Regularized Residual U-Nets for End-to-End Kuzushiji Character Recognition." *SN Computer Science*, 1:177. DOI: <https://doi.org/10.1007/s42979-020-00186-z>.
- Ledderose, Lothar. 2000. *Ten Thousand Things: Module and Mass Production in Chinese Art*. Princeton: Princeton University Press.
- Miller, Roy Andrew. 2015. *Nihongo: In Defence of Japanese*. London: Bloomsbury.
- Morita, Ichiko T. 1989. "Japanese Characters, Computer Input Of." *Encyclopedia of Computer Science and Technology*, edited by Allen Kent and James G. Williams, 20:309–325. New York: Marcel Dekker, Inc.
- MuseumIndex+. 2022. "MuseumIndex+." System Simulation. Accessed May 27, 2022. <https://www.ssl.co.uk/museumindex>.
- OCLC (Online Computer Library Center). 2022. "Discover How to Catalog Non-Latin Records in Connexion Client." *International Cataloging*. OCLC. Accessed May 27, 2022. https://help.oclc.org/Metadata_Services/Connexion/Connexion_client_3_0/International_cataloging.
- Oldman, Dominic, and Diana Tanase. 2018. "Reshaping the Knowledge Graph by Connecting Researchers, Data and Practices in ResearchSpace." In *Lecture Notes in Computer Science: International Semantic Web Conference 2018*, edited by Denny Vrandečić, Kalina Bontcheva, Mari Carmen Suárez-Figueroa, Valentina Presutti, Irene Celino, Marta Sabou, Lucie-Amiée Kaffee, and Elena Simperl, 11137:325–340. DOI: https://doi.org/10.1007/978-3-030-00668-6_20.
- Rekihaku (National Museum of Japanese History). 2021. "CIDOC CRM nihongo-yaku purojekuto no shidō." *Metaresource* (blog). May 26. Integrated Studies of Cultural and Research Resources. Accessed May 27, 2022. <https://www.metaresource.jp/cidoc-crm>.
- . 2022. *Khirin: Knowledgebase of Historical Resources in Institutes*. Accessed May 17, 2022. <https://khirin-ld.rekihaku.ac.jp>.
- Saïd, Edward W. 2014. *Orientalism*. New York: Random House.
- Schemm, Matthias. 2006. "NengoCalc, v4." Application for the conversion of Japanese dates into their Western equivalents. Accessed June 22, 2022. <http://bibliothek.kyoto.uni-tuebingen.de/NengoCalc/index.htm>.

Seeley, Christopher. 1991. *A History of Writing in Japan*. Leiden: Brill.

Song, Yuting, Biligsaikhan Batjargal, and Akira Maeda. 2020. "A Preliminary Attempt to Evaluate Machine Translations of Ukiyo-e Metadata Records." In *Lecture Notes in Computer Science: International Conference on Asian Digital Libraries (ICADL): Digital Libraries at Times of Massive Societal Transition*, edited by Emi Ishita, Natalie Lee San Pang, and Lihong Zhou, 12504:262–268. DOI: https://doi.org/10.1007/978-3-030-64452-9_24.

Spiller, Jürg. 1956. *Paul Klee: Das Bildnerische Denken*. Basel: Benno Schwabe.

Unicode, Inc. 2020. "Chinese and Japanese." *Frequently Asked Questions*. Unicode. Accessed May 16, 2022. https://www.unicode.org/faq/han_cjk.html.

W3C (World Wide Web Consortium). 2016. "What Is Ruby?" W3C Internationalization. Accessed May 16, 2022. <https://www.w3.org/International/questions/qa-ruby>.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, *et al.* 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data*, 3:160018. DOI: <https://doi.org/10.1038/sdata.2016.18>.

