CHAPTER 2

Research resources of Japanese Mokkan: Turning information on ancient wooden tablets into research data

Hajime Baba
Nara National Institute

Abstract

This paper focuses on ancient Japanese *Mokkan* (woodcuts) to discuss the background, current status and issues in the digitisation of cultural heritage, including matters of research, protection, and legal rights. As valuable primary sources, *Mokkan* are important in a wide range of academic fields, including history and linguistics. On the other hand, they are extremely fragile and fragmented. Digital technologies can support an improvement in access and study of this material. In this paper, we illustrate the current efforts to improve digitisation workflows and strengthen research capacity: we discuss open data and open access, equal and transparent inter-institutional collaboration, the application of the IIIF standard, the construction of a character database search system, and the application of Deep Learning technology.

The issue of open data and open access is particularly challenging in Japan, where there is still a gap between the legal principles and the actual situation

How to cite this book chapter:
Baba, H. 2023. Research resources of Japanese Mokkan: Turning information on ancient wooden tablets into research data. In: Palladino, C. and Bodard, G. (Eds.), *Can't Touch This: Digital Approaches to Materiality in Cultural Heritage*. Pp. 29–49. London: Ubiquity Press. DOI: https://doi.org/10.5334/bcv.c. License: CC BY 4.0

of cultural heritage institutions. In addition, information needs to be standardized and aggregated, but at the same time the specificity and historical characteristics of the different materials need to be addressed. In this paper, we strongly advocate for the establishment of methods for sharing cultural heritage data. In humanities research, the linkage, open access and standardization of information has become essential, not only to foster data-driven research but also to ensure wide access and dissemination of cultural heritage in Japan. For this reason, we encourage equal and open collaboration not just across cultural heritage stakeholders, but across society as a whole.

チャプターの概要

本稿では、木簡を中心に、デジタル化による文化財の調査・研究、保護・法的権利等について、経緯・現状と課題を説明する。

木簡は、貴重な一次資料として、歴史学・言語学等幅広い学問分野で重要である。一方、木簡は非常に脆弱で断片化が著しい。そのため公開が困難であり、研究には情報の集約が必須である。ここに情報技術への渇望が存在する。

現在は、デジタル化・オープンデータおよびオープンアクセスの強化、平等・継続・透明な機関間連携の実現、デジタル化ワークフローの改善や研究力強化に取り組んでおり、これらを IIIF（International Image Interoperability Framework）規格の利用や、「歴史的漢字データベース検索システム」の枠組構築、Deep Learning　技術の応用などで実現しようとしている。

オープンデータ化の現状を見ると、まだまだ法制度の理念と、実態は乖離している。また、情報は集約・標準化される必要があるが、同時に資料の特異性や歴史的特性にも対応する必要がある。これらの課題を克服し、文化遺産データを共有する方法の確立が必要である。

人文科学研究では、データ駆動型研究の進展など、情報の連携、オープンアクセス、標準化が不可欠となっている。研究機関間だけでなく、社会全体との対等でオープンな連携を求めていきたいと考えている。

1. Introduction

A few decades ago, digital methods and tools were considered just one of many approaches to conducting research and preserving material records, and the only concern about their design was to make them convenient to use.[1] Today,

[1] The present contribution includes research findings pertaining to the JSPS-funded "Development of Integrated knowledge through Establishment of an Interactive Research Scheme based on the Open-Data of Research Resources for Mokkan and Related Topics" (18H05221). The author would

they are the backbone of institutional and research practices, and thinking about digital technologies simply as "tools" is no longer appropriate. Digitization now provides an opportunity to fundamentally review and rebuild traditional workflows, reassess the very objectives of academic research, and rediscover the essential ideas and needs at the origin of traditional analog methods. This reaffirmation extends even more widely to the objectives in handling and managing cultural heritage, from discovery and preservation workflows to the management of legal rights and access (Yamada, Nakamura, Shibuya, Ohmukai & Inoue 2021).

This paper deals with the current project of the digitization of ancient Japanese *Mokkan* at the Nara National Research Institute for Cultural Properties.[2] It will illustrate how the process of digitizing a very specific category of premodern artifact has become, over the years, an opportunity to fundamentally rethink an entire workflow of research, information management, and dissemination to a public audience. This case study is used to further advocate the need for the Japanese government and research institutions to actively pursue open-access policies and practices in the digital publication of cultural heritage material, and to reaffirm some important principles in the translation of analog practices in the realm of digital technologies.

2. The history of Japanese *Mokkan*

In Japanese history and related fields, wooden man-made objects with India ink inscriptions are referred to as *Mokkan*, or "wooden documents." *Mokkan* belongs to the category of excavated artifacts since most of them are recovered from archaeological sites; further, *Mokkan* is classified based on the material of the writing medium, wood, one of the most widely used writing surfaces in Japan after paper (Figure 2.1).

Similar categories of artifacts exist in other areas. There are inscriptions on metal, stone, or earthenware. Similar to wooden documents, bamboo slips with ink inscriptions from China are collectively referred to as *jiǎndú*. Other examples of writing media made of plant materials include the white birch documents of Novgorod or the palm-leaf documents of Southeast Asia. Roman wooden tablets include writing tablets with a wax coating, widely seen on Roman frescos such as the wall painting of Terentius Neo and his wife in

like to express his gratitude for help and advice to Taizo Yamada, University of Tokyo, and Janase Pater, Nara National Research Institute for Cultural Properties. The editors and reviewers, especially Jun Ogawa, are also thanked for their invaluable understanding and assistance.

[2] Nara National Research Institute for Cultural Properties: https://www.nabunken.go.jp/english/.

Figure 2.1: An example of *Mokkan*. Credits: https://colbase.nich.go.jp/?locale=en. Reproduced with permission.

Pompeii, or rougher ones where the writing is performed directly on the tablet with ink, such as the Vindolanda tablets of England.[3]

There are four types of environments in which *Mokkan* are commonly found: arid areas, permafrost, wetlands, and water. In all these environments, the activities of wood-rotting fungi are restricted. Japanese *Mokkan* is mainly excavated from wetlands: in this environment, the supply of oxygen is cut off by moisture; however, the decay of the wood is progressive, albeit slow. *Mokkan* is, therefore, usually decayed and deteriorated considerably at the moment of their finding, which makes them extremely fragile. In addition, damage, breakage, and the fading of the inscriptions can also occur depending on the characteristics of the soil. The influence of metal ions in the groundwater or the presence of wood grain can also make deciphering the inscriptions difficult. In addition, information is usually fragmentary because the writing medium itself is small, and there is a high percentage of shavings generated during the reuse of the wood (up to 80% in the case of the Nara Palace Site). Because of their fragility, it is difficult to put *Mokkan* on public display, or even to conduct regular observations on the artifacts themselves (Figure 2.2).

When referring to "ancient Japanese *Mokkan*" in this paper, we refer to *Mokkan* attested from the end of the seventh century to the end of the eighth century, when *Mokkan* were most widely used in the Japanese archipelago.

[3] Vindolanda Tablets Online: http://vindolanda.csad.ox.ac.uk/.

Figure 2.2: The state of wooden shavings from *Mokkan* immediately after excavation. Credits: https://www.nabunken.go.jp/. Reproduced with permission.

This period overlaps with the formative years of the Japanese state and makes *Mokkan* especially important documents to reconstruct aspects of Japanese premodern history that are not documented otherwise.

The number of *Mokkan* excavated throughout Japan is approximately 470,000: 310,000 exemplars are hosted by the Nara Institute alone. The Nara Institute was established in the mid-1950s with the main purpose of performing comprehensive research on Japanese ancient cultural heritage, with a special focus on the ancient city of Nara and the Nara and Fujiwara Palace site. Before *Mokkan* were recovered from the Nara Palace site in the late 1960s, information about eighth-century Japan was primarily provided by legal documents, such as the Ritsuryo legal code or the Rikkokushi, a historical chronicle compiled by the court. These were official court documents, and as such, they did not provide a complete picture of life in that period. A further complication is that they only exist in copies, and there are no extant originals.

The *Mokkan* found at the Nara Palace site, on the other hand, are primary and original sources about the daily life of the court. They are also free of any intentional editing, like in the case of the *Nihon Shoki*. *Mokkan* represent first-hand accounts: there is less risk of forgery or falsification because they were discovered through archaeological excavations, and they were discarded on the same

site where they were produced and used. Moreover, as their number increases with new excavations, we can expect further developments in the future.

The importance of *Mokkan* lies also in their ability to document the process of creation and establishment of a Japanese writing system from Chinese writing: the kana script and the kanji-kana mixed text. *Mokkan* is located at the periphery of the so-called "Chinese character sphere," the pre-modern East Asian cultural sphere where ideas, religion, administration, and technology were communicated through the use of Chinese characters. Most of the characters inscribed on *Mokkan* are of Chinese origin, and many of the documents written here could be better described as Japanese texts written with Chinese characters. While official documents and chronicles, such as the Rikkokushi, were considered formal documents, and therefore were only written in plain Chinese (the most formal typology of writing), *Mokkan* were considered less formal and therefore could use less formal writing. In everyday writing, various methods were developed to adapt Chinese characters to the Japanese language: while most of the focus was on the shape of the letters, sometimes Chinese characters were given Japanese readings, or the Chinese pronunciation of a character was used to represent something unrelated with a similar sound. This mix-and-match approach to Chinese characters laid the foundation for the current Japanese writing system, and *Mokkan* documented the process through which the Japanese people learned to express their language in writing through the adaptation of Chinese writing (Baba 2022).[4]

Other scripts in the vicinity of the Sinographic cultural sphere, include the Chữ Nôm of Vietnam, which are local adaptations of Chinese characters. We believe that the Sinographic cultural sphere can be seen both as a single coherent entity or a collection of smaller cultural spheres based on local variations of the Chinese writing system. Recent years have seen a rise in interest in approaches to East-Asian tablets that try to analyze several of these local cultural spheres as a whole rather than as isolated entities. Therefore, the characters inscribed on Japanese *Mokkan* are not only important to elucidate the early establishment of the Japanese state, but can also provide hints to reconstruct the broader development of Japanese civilization.

Mokkan are fragile, fragmentary, raw materials, fascinating, and incredibly troublesome. Their peculiarities, due to their status of preservation, their fragility, and the recording of information about them, constitute an interesting case study about current digitization practices of archaeological findings. We believe that the discussion developed around the digitization of *Mokkan* may provide clues to solve problems that are common not only to various inscribed artifacts, especially those made of plant materials but in general to the application of digital technologies to the preservation and management of material cultural heritage in Japan (Yamada, Inoue & Yamaga 2019).

[4] Wooden Tablet Database: https://mokkanko.nabunken.go.jp/en/.

3. Early status of research on ancient Japanese *Mokkan*

As mentioned above, *Mokkan* is incredibly fragile and highly fragmented in both form and content. Because of this characteristic, research on them is conducted necessarily in a comparative fashion: typically, researchers of *Mokkan* have to supplement fragmented information, attempting to combine as many artifacts as possible, cross-referencing fragments, as well as comprehensively examining related materials (such as legislative documents) and conditions of finding (such as the aspect of the site, type of soil, etc.). In other words, to obtain the latent information from *Mokkan*, we have to take a comprehensive and comparative approach. This is made difficult by the very material condition of *Mokkan*, which prohibits extensive analysis.

Extensive digitization and virtual preservation are extremely effective ways to overcome these problems. This was the main driver behind the effort of the Nara National Research Institute for Cultural Properties, in establishing a database of *Mokkan* in the late 1990s, which was later released in 1999 as the Wooden Tablet Database (Baba 2019).

Initially, the database was simply designed to allow researchers to search for inscriptions, including fragmentary ones. The information hosted in the database consisted of data taken from print catalogs and publications, which were more or less systematically turned into metadata. This information included the basic records of print publications from archaeological excavations and the recovery of the materials, as it had been consolidated through practice at the Nara Institute during the excavation of the palace site in the 1960s–1980s: these included the information usually recorded on site, such as the name of the excavation, survey order, excavation grid, excavated remains, stratigraphy and date, and provisional names and identification for the artifacts; later, at the restoration stage, official identification through an R number (a unique number assigned to the wooden batts for storage), storage number, photographs with records, and other information, such as observations on the wood, letters, and handwriting, tree species and woodcutting. This information was transcribed and turned into metadata, and accompanied by tags, such as ancient places and personal names, to assist the exegesis of the material.

This bookkeeping system had been developed gradually, during various stages of the excavation, and it was fundamentally conceived for dissemination through analog methods. Therefore, it did not always have enough structure to be readily digitized. For example, there was a lack of a clear distinction between provenance and artifact information, and interpretive information (such as people or place names) was at the same hierarchical level as material information, such as the legal quantity and storage location. Certain aspects, however, were retained: the combination of excavation grid data and R-number was used as ID for the wooden tablet.

Although at a relatively rudimentary stage of development, the database marked a turning point because it allowed the comprehensive study of fragmented information and its publication on the Internet. Although it was not recognized at the time, the accumulation of digitized metadata on *Mokkan* provided the foundation for all further developments.

A second stage of development was the addition of images, which were connected to the database records. The earlier images consisted of scanned versions of the black-and-white photographs attached to the archaeological inventories. These images were only regarded as supplementary materials, and their quality was low. However, the linking of textual and visual information about each artifact and their publication online as a single entry made it possible to perform cross-searches on fragments, and integrate the photographs of the *Mokkan* with information about their text.

In 2005, we also developed and released an online character dictionary, which was originally designed to allow searches on the characters inscribed on *Mokkan*. This was designed to meet the demand of users who wanted to search for images of specific characters. The database connected a search input based on character codes with images of the characters as they appeared on *Mokkan*: the design was based on how researchers looked up samples of characters when deciphering documents, and therefore it simply translated a traditional paper-based work procedure into the digital world. In essence, it allowed users to upload a small image of a piece of a wood document with an ambiguous character, and show similar characters from the database (Watanabe, Baba & Kurushima 2016; *cf*. Onitsuka, Oyama, Yamada, Inoue & Uchida 2018; Nakamura, Liu, Miyazaki, Inoue, Daisen & Yamada 2022).

The idea that a character search would start from an image, and not from a text, was a very important conceptual and processual innovation. Normally, it is very hard to look up characters that are difficult to decipher, and dictionaries are often ineffective in recognizing hand drawings or require cumbersome scrolling through every character with the same radical or number of strokes (Ly, Nguyen, Nguyen & Nakagawa 2019). The establishment of the images as the foundation for the research work meant that they were no longer regarded as supplementary material to the text and metadata, but they represented effectively an information layer at the same level of importance (Nakamura, Liu & Yamada 2022). The process of digitization offered the opportunity to pile layers of information on top of each other, going from a text-based workflow to one that managed multiple data types. This marked a point in the development where the older black-and-white images were supplemented with additional layers, which included color photographs, infrared photographs, and hand-drawn illustrations, to support the decipherment of the characters as they appeared on the inscriptions.

Another important result of this process was the realization that the database could serve a much wider audience than just researchers of *Mokkan*, and that its generalization could provide the opportunity for full-scale inter-institutional cooperation (see below).

This project has been gradually developed to go beyond the dimension of the character database. In parallel with the image database, other important information about *Mokkan* has been digitized: for example, the information on names of places and people, and literature references, has been made available in a machine-readable format. Location information of *Mokkan* and other artifacts is currently available in a WebGIS interface, where users can search by geographical area, but also object type, name, and prefecture (https://heritagemap.nabunken.go.jp/main).

Moreover, images of *Mokkan* have been updated to versions taken with digital cameras, which allow the automatic conversion of the photographs into data. We also developed a method to acquire multispectral images, which is fundamental when observing wooden tablets, whose ink marks are often obscured.

Today, it is possible to have a system where a single computer connected to the Internet is sufficient for deciphering inscriptions on *Mokkan*. As we move forward, the effort in the digitization of Japanese wooden tablets is increasingly moving in the direction of applying internationally recognized standards for information aggregation, data management, interoperability, and collaboration.

After more than thirty years of collaboration, we are promoting the following fundamental points of development for the current phase of work:

1. Standardization of information and image management through IIIF (International Image Interoperability Framework) standards and Linked Open Data practices; and applications of Deep Learning technologies to improve digitization workflows and to strengthen research capabilities (Clanuwat, Bober-Irizar, Kitamoto, Lamb, Yamamoto & Ha 2018).
2. Implementation of an explicit research and development agenda for Japanese cultural heritage, to empower Open Data and Open Access practices through digitization.
3. Implementation of equal, continuous, and transparent inter-institutional cooperation, within the framework of the Multi-database Search System for Historical Chinese Characters.

4. Current status in the digitization of ancient Japanese *Mokkan*

4.1 Application of IIIF Data Standards

IIIF is becoming a global standard in the field of humanities research, and it enables the creation and publication of high-definition, high-quality metadata, and annotations, at the same time ensuring high versatility and open access (Takada, Fukuyama, Tsutsumi & Kosukegawa 2018).

The IIIF standard includes an Image API, Presentation API, Authentication API, and Content Search API (Liu, Nakamura & Yamada 2022; Baba, Takada & Kuwata 2019). The Image API and Authentication API are standards used for

describing image status and rights, and information is organized accordingly. The Image API is used to display specific data about the image (e.g. area, color, format), but also information such as size and ownership of the material object. The Authentication API typically supports various measures of access restriction for images protected by copyright, and it manages permissions to view comments, annotations, and other types of content.

The Mirador Annotation tool is used to associate various categories of information directly with the images of the artifacts as Web Annotations. The workflow adopted by the Nara Institute is based on the *Mokkan* image, and it reproduces a process already adopted by the Academia Sinica in Taiwan. The character images are annotated, and character codes are superimposed on these annotations. This information is stored as JSON files, but we are also aiming at the creation of a CSV output format. Therefore, after searching for one character, it is possible to cross-reference the entire *Mokkan* where that character appears. Other workflows adopted by other institutes are slightly different: for example, the National Institute of Japanese Literature uses individual character images as the basis, and superimposes metadata, including character codes, as annotation on the images. However, even if the data can be prepared in different ways, as long as it complies with IIIF it can be exchanged across institutions. This will be achieved by improving and applying the 'Repair System' developed and operated by the Historiographical Institute of the University of Tokyo, which consolidates various records of repairs of old documents and observation findings as annotations on historical images. Once this system is in operation, almost all of the information acquired during the process of management of *Mokkan* will be digitized, and it will allow the direct conversion of information originally appearing in print.

The Presentation API enables aggregation of various types of information as cross-referenced annotations based on standardized descriptions, and it is typically used for displaying graphic objects. It is used to describe and display Web Annotations and other metadata related to the *Mokkan* images in the database. This allows storing text images associated with metadata (e.g. character codes, holding institution, etc.) as aggregated information, which can be structured and exported as individual layers as needed. The process of metadata encoding is currently being restructured according to the CIDOC-CRM standard for the best compatibility with Linked Open Data systems.[5] Work is also underway to switch to a system based on ISO 14721 (OAIS reference model) for the creation of information and research resources.

The conventional Content Search API, which is generally used to search within a single manifest, was judged unsuitable because the system needs to search across a large number of manifests, although on a small scale. There-

[5] The CIDOC Conceptual Reference Model (https://www.cidoc-crm.org/) is the most important data model for information integration in the field of cultural heritage.

fore, further implementation is experimented to improve the performance of this API.

The standardization of historical information is a complex process, which requires balancing the recording of detailed individual observations with the requirements of the standard. Sometimes, this means discriminating between types of information regarded as fundamental, and others that end up being ignored, and considered rather as interpretive observations (Unsworth 2000; Shibayama, Morimoto, Tashiro, Kameda, Yamada & Hara 2018). In other words, this process may happen at the expense of certain descriptive aspects of the data for the benefit of exchange and interoperability. This is particularly important in the context of inter-institutional cooperation, where each institution manages different kinds of materials, periods, and records.

A practical example is the process of character decipherment. With traditional analog methods, several types of information are recorded about the characteristics of individual occurrences in context: not just the character code, but also a description of how it appears, writing habits, stroke styles, etc. This information is suppressed when the character is associated with a specific code in a digital database: character codes, most notably Unicode, are a highly versatile way of describing written information, as they ensure seamless exchange of data across different platforms and unique identification. Modern digital research simply cannot do without the association or creation of Unicode characters. However, the external information associated with the specific occurrence of a character in context, such as calligraphic style, writing style, personal habits of calligraphers, or stroke styles, could not be reproduced in Unicode and is therefore suppressed when individual occurrences get associated with a standardized instance.[6]

For this reason, it is also essential to associate object records with high-quality reproductions that can be considered "copies" of the originals. Furthermore, it is essential to allow users to access this material so that they can also record their observations in the form of annotations, and freely accumulate descriptive information. The joint availability of digital copies and standardized metadata is the way to achieve interoperability through versatility. The release of information that is as close as possible to being a "copy" of the original, such as high-resolution images, is fundamental. These reproductions can be used for detailed observation and research, alongside standardized metadata. For this reason, we are preparing born-digital data of visible light and infrared images,

[6] For this point, in Japan, the Center for Open Data in the Humanities (CODH) has recently been promoting the 'curation API' (http://codh.rois.ac.jp/iiif/curation/index.html.en) and made it possible to have a character, its images and other external information together in the common format of IIIF standard. See also Filosa, Gad & Bodard (Chapter 3 in this volume), section 2.2 on the relationship between text transcription, image and explicit encoding.

and in the future, we are considering producing 3D reproductions. Moreover, it is important to release image data in a type and size that is easy to use, even if institutions manage higher-quality and resolution images.

One of the fundamental needs is to strengthen the connection between annotations and related information, and the material object that they describe. Therefore, we are re-building our databases by integrating a production workflow that allows the generation of data (including pre-existing metadata and unstructured observations) in the form of annotations linked to the digital copy of the object.

The annotation workflow, while it allowed user annotation for characters and strokes, was not feasible to generate information about stroke order at a database level. The plan is to combine it with a system of information acquisition about stroke order, developed in the context of the Web App Nazorkun, a citizen science project for acquiring brush stroke order information from general users (Hatano & Baba 2022). Furthermore, a prototype of a wood grain eraser AI system has been completed as an auxiliary tool to sharpen wood grain images and make strokes easier to detect (Terras 2006). The system uses deep learning to extract the outline of the woodblock and ink marks from the woodblock images. The main goal is to facilitate the observation of the materials (e.g. with automatic retrieval of character images) and deepen character observations on them, using the generated output as a starting point to implement the range of information available, especially in the stroke order estimation system (Ohyama, Hatano & Baba 2020).

As a part of this development, we also designed a learning application for children. This application was initially developed as a citizen science project to accumulate records on the stroke order of characters seen on *Mokkan*, to broaden social participation in research. Later, it was developed into a learning platform for elementary and junior high school students, to encourage familiarization and learning with the Japanese culture of writing. This project was well-received by the public and it became a tool for both social participation in research and dissemination for educational use.

Finally, it is important to enable cross-references across different data types, so we are building a Linked Open Data system to appropriately link our data with other institutional repositories.[7] On the one hand, this will lead to a clear distinction and acknowledgment of the role of each piece of information (and managing institution) in the digital ecosystem on *Mokkan* research, while at the same time allowing seamless data aggregation and discovery (Baba 2021).

[7] Japan Search (https://jpsearch.go.jp/) is a good example of Linked Open Data resources about Japanese Cultural Heritage.

4.2 Open Data and Open Access

Open Data needs to be the foundation for this process. The publication and sharing of research data is a method of promoting research in the new era, where large amounts of data can be easily produced, and it is also the responsibility of those involved in research.

While Japan's law for the Protection of Cultural Property clearly states that "cultural properties are the common property of the people", the reality is that the concrete operation of promoting and preserving it is the "burden of the beneficiaries" who are tasked with its curation. In the practical application of Open Data technologies and standards, Japan is considerably lagging behind, especially outside of the research sector. While the government is taking the lead in promoting digitisation and open access,[8] there is still a long way to go to better position the country within a global context.

Globally, Open Data is being promoted at an accelerated pace, with the G8 2013 Summit concluding an agreement on an Open Data Charter. In 2016, the G7 Science and Technology Ministerial Meeting "Tsukuba Communiqué" (Tsukuba, Ibaraki) issued a joint statement that, among other things, emphasized the need of Open Data and sharing policies for Japan.[9] In 2020, the official Japanese translation of the FAIR Data Principles (Wilkinson et al. 2016) was published.[10]

The Law for the Protection of Cultural Properties is the central cultural heritage law in Japan. The chief aim of the law is to "preserve and utilize cultural properties, thereby contributing to the cultural improvement of the people and to the advancement of world culture" (Art. 1). The broad principle is an official acknowledgment that "cultural property is indispensable for a correct understanding of the country's history and culture, and that it forms the basis for cultural improvement and development" (Art. 3). The law also makes various provisions with the direct involvement of the general public, governmental institutions, and individual owners, in the preservation and promotion of Japanese cultural heritage. More specifically, Article 4 stipulates that "the general public shall sincerely cooperate with the measures taken by the Government and the Authorities" in the application of the Law, and that "owners and other persons concerned with cultural property [...] shall preserve it carefully for the public good, and shall endeavor to make cultural use of it,

---

[8] See, for example, the open data portal of the Japanese public administration (https://data.e-gov.go.jp/info/en) and the current discussion on the application of Open Data standards between the USA and Japan (https://liquitous.com/lisearch/journal/2020/09/30/317/).

[9] Tsukuba Communiqué: https://www8.cao.go.jp/cstp/english/others/20160517communique.pdf.

[10] The Fair Data Principles: https://biosciencedbc.jp/about-us/report/fair-data-principle/.

including, as far as possible, *making it available to the public*" (Art. 4.2). At the same time, however, the same article clearly states the preservation of ownership and property rights (Art. 4.3).

In 2016, the Basic Law for the Promotion of the Use of Public and Private Data was enacted to oblige the State and local governments to the use of Open Data. Although this law mainly covers administrative data, its spirit, and direction are considered to expand onto cultural heritage management as well.

In the case of cultural property managed by the government and public institutions, the spirit of this legislation has been actively pursued, and the dissemination of knowledge about the Japanese cultural heritage is actively encouraged. For example, archaeological sites and artifacts and the so-called rescue archaeology are being surveyed and published as part of cultural property protection management. On the other hand, however, the specific application of Open Data and Open Access practices is more complex at various levels.

In terms of the spirit of the law, it seems that the open digitization of cultural heritage is precisely the type of practice that should be promoted and encouraged because it pursues at the same time the goals of preservation and protection through the creation of digital reproductions and the goals of dissemination and promotion through open access publications.

However, on the one hand, the burden of planning and creating digitized resources is entirely placed locally, on the institutions that host cultural heritage collections (see also Takata & Yanase 2021). On the other hand, there is the question of the protection of property rights, which intersects with several other complex legislative questions.

Property rights in the area of cultural heritage include, in addition to ownership of the property, usufruct rights of artifacts, or of the site of excavation, of the surrounding areas, material infrastructures involved, and copyright. These aspects are only minimally covered by the Law on the Protection of Cultural Property, and it is necessary to refer to laws that guarantee separate rights for each area. For example, the ownership of excavated cultural heritage is also governed by the Lost and Found Property Act, but a full explanation of this legislative body would be too extensive for the space of this article.

In other words, it is very difficult to navigate through the various legal implications of digitally publishing cultural heritage data in Japan, partly because there is little direct legal precedent in this area. So, the National Institute for Cultural Heritage, the umbrella organization of the Nara National Research Institute, is taking careful measures and holding study groups with lawyers and lecturers to better understand the implications of copyright legislation.

The current status of *Mokkan* digital data is as follows. First of all, we have to do with artifacts that were created in the 7th century, so there is no author's copyright. Flat photographs of the wood tablets are mainly overhead photographs, and they are regarded as "reproductions" and do not accrue fresh copyright under the current legislation. Therefore, the main aspect to be considered is the ownership rights of the institutions that host and preserve

Mokkan collections. Public institutions in Japan can pursue public data disclosure only after legal consultations and official agreements: for example, the Nara Institute is pursuing this effort with various other entities that host *Mokkan* collections, including the Kyushu Historical Museum, the Kunitachi City Board of Education, Hamamatsu City Board of Education, Higashihiroshima City and the Tohoku Museum of History.[11]

However, the fact that institutions claim ownership rights on cultural heritage collections is a serious hindrance to institutional collaboration in Japan. Research institutions are among the major stakeholders and managers of Japanese cultural heritage: they play a central role not only in the creation of research data, but also in managing rights, ensuring the preservation, and promoting access. Furthermore, they are chiefly responsible for the privately owned materials that they manage and must ensure that the owners' rights are adequately protected according to the law. On the other hand, this exclusivity encouraged a tendency to build extremely specialized resources that emphasized the particularity of individual collections: this approach was seen as a way to provide the public only with the highest quality of information while ensuring that the institution responsible for the management of the collection would retain rights and authority on it. This mechanism created the conditions to seriously hinder cooperation and accessibility of Japanese cultural heritage data: on the one hand, there was the concern that data published openly could be affected by lack of quality control, and on the other, there is still a lack of infrastructure and legislative clarity on ownership and copyright.

For these reasons, efforts in inter-institutional cooperation are exceptional in Japan, and institutions must operate under very specific conditions to ensure a good outcome.

---

[11] A related project is the online publication of the *Journal of Mokkan Studies* (*Mokkan Kenkyu*). This is a research journal published annually by the *Mokkan* Gakkai (Society of *Mokkan* Studies), an academic society that aims to conduct comprehensive research on *Mokkan* and that has strong institutional connections with the Nara institute. It publishes information on the excavation of tablets from all over Japan, as well as articles and book reviews. The journal is already at issue 45, and a vast amount of information has been disseminated through it. The digital publication of the issues was preceded by the obtainment of the consent to publication by the various authors, and for the drawings and photographs the consent of the institution holding the material was obtained (even for drawings, which are normally considered public domain, since they are copies of excavated artifacts, we preferred to err on the side of caution and ask for permission). As a result, the rights to the data published in PDF format are centrally owned by the *Mokkan* Gakkai, but are not currently considered open data.

4.3 Inter-institutional cooperation

The database established by the Nara Institute alone would be sufficient if its only focus were on *Mokkan* themselves. However, when we move towards the digitization of historical characters, we have to acknowledge that images of historical characters are not limited to *Mokkan*, but include many other artifacts—especially documents written on paper. Researchers of *Mokkan* often look up samples of characters used on paper documents. Therefore, it is desirable to have a digitized database that allows comparison and cross-reference of such diverse and rich sources of character images. Therefore, in 2008, we developed and released a term-linked search system in cooperation with the University of Tokyo's Historiographical Institute, which hosts a database of paper-based character images from the thirteenth to sixteenth centuries (Yamada & Inoue 2018).[12]

This project marked the first official collaboration between research institutes in the humanities in Japan, and it was unprecedented in the history of Japanese cultural heritage management: to break the barriers and realize this collaboration, it was necessary to build a shared understanding of the core ideas and goals to pursue, from the conceptualization stage to the concrete implementation, while respecting the uniqueness and particularity of each institution. The engagement in this deeper discourse provided the conditions for an inter-institutional research initiative that went beyond the mere creation of a joint database, and it also led to other organizations taking an interest in the project, placing the conditions for the exploration of multi-organizational collaboration.

On the one hand, it was necessary to manage the concrete differences that marked the creation of very different resources: the *Mokkan* at the Nara Institute on one side, and the paper documents of Tokyo on the other side. These artifacts were very different in material, history, origin, and even period (*Mokkan* being 6th–7th century and paper documents ranging from the 13th to the 17th century): consequently, they had been handled very differently in research and preservation, and the data was created in very different ways. These differences need to be reconciled. An indispensable condition for inter-institutional cooperation was a common commitment in the direction of a more substantial disclosure of the data, through practices of standardization and increased accessibility of the information.

Looking forward, it is desirable to establish a system where all kinds of research resources can be mutually shared, and a relationship of mutual trust among institutions is essential. In 2020, the Nara Institute and various research parties, including the University of Tokyo, the National Institute of Japanese Literature, the National Institute for Japanese Language and Linguistics, the Academia Sinica, and the Kyoto University Research Center for Cultural

[12] Multi-database Search System for Historical Chinese Characters: https://mojiportal.nabunken.go.jp/en/.

Sciences, co-signed and released a Declaration of Cooperation with the common goal of creating common specifications for the digitization and open dissemination of their cultural heritage data (Nara National Research Institute for Cultural Properties, University of Tokyo, Historiographical Literature, National Institute of Japanese Literature, National Institute for Japanese Language and Linguistics, Kyoto University Research Centre for the Cultural Sciences and Institute of History and Philology, Academia Sinica 2020a). This declaration was followed by an Open Data Specification, currently in its first edition, which establishes guidelines for the creation of open data standards, with the explicit intent of following IIIF specifications (Nara National Research Institute for Cultural Properties, University of Tokyo, Historiographical Literature, National Institute of Japanese Literature, National Institute for Japanese Language and Linguistics, Kyoto University Research Centre for the Cultural Sciences and Institute of History and Philology, Academia Sinica 2020b).

The spirit of the declaration and the directions outlined in the guidelines aim at achieving much more openness than in the past, both in terms of collaboration and data exchange. We believe that it is essential in the age of digitization to free the source material from individual stakeholders. Based on this declaration and guidelines, it will be possible to expand and share data through broader collaboration among overseas institutions, domestic institutions, and common citizens, and at the same time maintain the centrality of cultural heritage in society, overcoming the many limitations currently imposed by budget and administration.

5. Conclusions

In conclusion, this examination of the digitization practices for ancient *Mokkan* allows us to expand our considerations into more general issues in the tension between digital technology and the management of material cultural property in Japan.

We want to reaffirm the core ideas and objectives concerning cultural property, in light of the changes initiated by the digital revolution:

1. Cultural property is the common property of the people (of Japan, and even of the world). Therefore, adequate conservation and active public disclosure of the findings are necessary. Institutions and people dealing with cultural heritage should be equal and cooperate. More specifically, open data can support a wide operation of dissemination for a large number of people.
2. Information should be aggregated, standardized, and organized rigorously and according to the specificities of the material. On the one hand, this reaffirms the importance of the practice of cataloging, a method that has a long tradition in Japan since the Heian period. On the other hand,

> digital technology requires that information about artifacts is organized systematically according to specific standards. We need to take care when designing and using standardized models, that the research needs of the project are not sacrificed to the technology and methods of digitization, but that the historical approach is privileged.

The current recommendation is to promote the provision of high-quality and diverse digital resources by first prioritizing cultural heritage that can be easily handled in terms of legal rights, and then to make the significance and benefits of this openness widely known to society so that this becomes accepted as a fact. This will make it easier to expand the benefits of digitization and wide open access in the future, and eventually to address their impact in the area of legislative regulations on the dissemination of cultural heritage data.

Digital technology has made it easier than ever to access cultural properties and to keep costs low. It is now possible, and necessary, to establish a new way of sharing cultural heritage data, overcoming the large concrete discrepancy between the ideal stated by the law and the actual conditions of heritage institutions. Moreover, open access to cultural heritage information will produce richer and better research.

We believe that the role of informatics will only increase in the humanities. Data-driven research will become increasingly important in humanities research, and it is essential to pursue the goals of cooperation, open access, and standardization of information (Yamada & Inoue 2021; Nakamura & Yamada 2021). The attitude of seeking equal and open collaboration, not only across institutions, but also with society at large, will lead to new and important developments in the dissemination of new knowledge, and be beneficial for the community as a whole (*cf*. Yamada 2018).

6. References

馬場　基. Baba, H. (2019). 奈良文化財研究所のICTへの取り組み. Nara National Research Institute for Cultural Properties' ICT Efforts. *Japanese Historical Society 848*(9–13).

馬場　基・高田　祐一・桑田　訓也. Baba, H., Takada, Y., & Kuwata, K. (2019). IIIFの導入による木簡画像データベースの連携強化. Enhancement of linkage of wooden tablet image database by introduction of IIIF. 奈良文化財研究所紀要2019. *Bulletin of the Nara National Research Institute for Cultural Properties* 2019. 18–19.

馬場　基. Baba, H. (2021). 史的文字データベース連携検索システムの理念と未来. Philosophy and future of historical character database linked search system. ふみ *Fumi 16*, 6–7.

馬場 基 Baba, H. (2022). 日本古代漢字漢字運用規範を木簡から探す. Searching for ancient Japanese kanji usage rules from wooden tablets. 日本文学研究ジャーナル. *Journal of Japanese Literature 24*, 25–36.

Clanuwat, T., Bober-Irizar, M., Kitamoto, A., Lamb, A., Yamamoto, K., & Ha, D. (2018). Deep Learning for Classical Japanese Literature. *Neural Information Processing Systems 2018 Workshop on Machine Learning for Creativity and Design*. Retrieved from: arxiv:1812.01718. DOI: https://doi.org/10.20676/00000341.

畑野吉則・馬場基. Hatano, Y., & Baba, M. (2022). 市民参加型筆順情報取得アプリケーション「ナゾルクン」. Nazoru-kun: a Web App for acquiring brush stroke order information. 『奈良文化財研究所紀要. *Bulletin of the Nara National Research Institute for Cultural Properties* 2022, 56–57. DOI: http://doi.org/10.24484/sitereports.129169–119461.

劉冠偉・中村覚・山田太造. Liu, G., Nakamura, S., & Yamada, T. (2022). 研究資源としてのWEB　APIの利用：歴史資料・古典籍の字形を横断的に検索するアプリケーションの開発. The Utilization of WEB API as a Research Resource: Development of an Application for Cross-sectional Search of Glyphs in Historical Materials and Classical Books. じんもんこん2022論文集. *Jinmonkon Proceedings 2022*, 165–170.

Ly, N. T., Nguyen, K. C., Nguyen, C. T., & Nakagawa, M. (2019). Recognition of Anomalously Deformed Kana Sequences in Japanese Historical Documents. *IEICE Transactions on Information and Systems*, Vol. E102-D, No. 8, 1554–1564. DOI: https://doi.org/10.1587/TRANSINF.2018EDP7361.

Nara National Research Institute for Cultural Properties, University of Tokyo, Historiographical Literature, National Institute of Japanese Literature, National Institute for Japanese Language and Linguistics, Kyoto University Research Centre for the Cultural Sciences and Institute of History and Philology, Academia Sinica. (2020a). *Declaration of Cooperation*. Retrieved from https://mojiportal.nabunken.go.jp/files/declaration1en.pdf.

Nara National Research Institute for Cultural Properties, University of Tokyo, Historiographical Literature, National Institute of Japanese Literature, National Institute for Japanese Language and Linguistics, Kyoto University Research Centre for the Cultural Sciences and Institute of History and Philology, Academia Sinica. (2020b). *Open Data Specifications*. First Edition. Retrieved from https://mojiportal.nabunken.go.jp/files/declaration2en.pdf.

Nakamura, S., & Yamada, T. (2021). Development of data-driven historical information research infrastructure at the Historiographical Institute in the University of Tokyo. *Proceedings of JADH2021*, 148–151.

中村覚・劉冠偉・山田太造. Nakamura, S., Liu, G., & Yamada, T. (2022). NDLOCRを用いた東京大学史料編纂所史料集版面画像に対する検

索システムの開発. Development of a Retrieval System for Historiographical Edition Images Using NDLOCR. 研究報告人文科学とコンピュータ（CH) 1月8日. *Research Report Humanities and Computers (CH), 1*(8).

中村覚・劉冠偉・宮﨑・肇・井上聡・大山航・山田太造.Nakamura,S., Liu, G., Miyazaki, H., Inoue, S., Daisen, S., & Yamada, T. (2022). 花押を対象としたデータ駆動型歴史情報学研究の実践. Implementation of data-driven historical informatics research on Kao Signature. じんもんこん2022論文集. *Jinmonkon Proceedings 2022*, 171–178.

大山 航・畑野 吉則・馬場 基. Ohyama, W., Hatano,Y., & Baba, H. (2020, December). 深層学習による木簡実測図の自動作成. Automatic generation of measured drawings for Mokkan using Deep Learning.「人文科学とコンピュータシンポジウム」2020年12月. *Humanities and Computer Symposium*, December 2020, 235–240.

鬼塚　洋輔・大山　航・山田　太造・井上　聡・内田　誠一 Onitsuka, Y., Oyama, W., Yamada, T., Inoue, S., & Uchida, S. (2018). 花押類似検索のための畳み込みオートエンコーダによる画像特徴抽出. Convolutional Feature Extraction for Kaou Image Retrieval. じんもんこん2018論文集. 252–262.

Shibayama, M., Morimoto, S., Tashiro, A., Kameda, A., Yamada, T., & Hara, S. (2018). Building an Ontology-Oriented Archaeological Knowledge-Base "ArcOnBase" in Mainland Southeast Asia. *2018 Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC)*, San Francisco, CA, 1–6. DOI: https://doi.org/10.23919/PNC.2018.8579462.

高田智和・福山雅深・堤智昭・小助川貞次. Takada, T., Fukuyama, M., Tsutsumi, T., & Kosukegawa, S. (2018). 資料画像公開・利用の国際化と高度化の取り組み—「日本語史研究資料［国立国語研究所蔵］の事例—. Implementation of Image Disclosure System Using IIIF. 国立国語研究所論集 第15号. *NINJAL 15*, 10–31. DOI: http://doi.org/10.15084/00001601.

Takata, Y., & Yanase, P. (2021). The Production, Preservation and Dissemination of Archaeological Data in Japan. *Internet Archaeology 58*. DOI: 10.11141/ia.58.11.

Terras, M. (2006). *Image to Interpretation: An Intelligent System to Aid Historians in Reading the Vindolanda Texts*. Oxford: University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199204557.001.0001.

Unsworth, J. (2000, October). *What is Humanities Computing and What is Not?* Paper presented at the Distinguished Speakers Series of the Maryland Institute for Technology in the Humanities at the University of Maryland, College Park MD.

Watanabe, A., Baba, H., & Kurushima, N. (2016). *Mojizo: Image matching search for Mokkan or cursive characters*. Nara National Research Institute for Cultural Properties. Retrieved from https://mojizo.nabunken.go.jp/.

Wilkinson, M. D., Dumontier, M., Aalbersberg, IJ. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Bonino da Silva Santos, L. O., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3, 160018. Retrieved from: https://www.nature.com/articles/sdata201618.

山田　太造. Yamada, T. (2018). 収集史料の体系化と永続的な利用に向けた管理. Organization of Collected Historical Materials and Management for Permanent Utilization. 研究報告人文科学とコンピュータ（CH）. *IPSJ SIG Technical Report*, Vol. 2018-CH-118 1.

Yamada, T., & Inoue, S. (2018). A Common Base of Knowledge for Japanese Historical Materials and its Application. *2018 Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC)*, San Francisco, CA, 1–6, doi: 10.23919/PNC.2018.8579468.

山田 太造・井上 聡・山家 浩樹, Yamada, T. Inoue, S., & Yamaga, H. (2019).　日本史史料データ流通基盤に向けた歴史データリポジトリの整備. Developing a historical data repository for a data distribution platform for Japanese historical materials. じんもんこん2019論文集. *Jinmonkon 2019 Proceedings*, 3–10.

Yamada, T., & Inoue, S. (2021). Personal Name Authority Data Repository for Advancement Data-driven Research in Japanese History. *2021 Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC)*, 13–17. DOI: https://doi.org/10.23919/PNC53575.2021.9672287.

山田太造・中村覚・渋谷綾子・大向一輝・井上聡. Yamada, T., Nakamura, S., Shibuya, A., Ohmukai, I., & Inoue, S. (2021). 日本史史料を対象とした研究データ基盤整備における課題. Issues in developing a research data infrastructure for Japanese historical materials. じんもんこん2021論文集. *Jinmonkon 2021 Proceedings*, 80–87.