

Hieratische Paläographie DB の構築

中村 覚・永井 正勝（東京大学）

和氣 愛仁（筑波大学）・高橋 洋成（東京外国語大学）

本研究では、Möller のヒエラティック字典を対象とし、その利活用を目的として開発した Hieratische Paläographie DB について述べる。具体的には、IIIF Curation API を用いてページ画像内の部分領域を管理し、文字番号や音価などのメタデータも合わせて RDF データとして記述することで、それらの部分画像に基づく検索や、ページおよび巻を横断した画像比較を可能とした。さらに、研究利用シナリオに基づく画像比較などを通じ、開発した DB の有用性を検証した。

Development of Hieratische Paläographie DB

Satoru Nakamura / Masakatsu Nagai (The University of Tokyo)

Toshihito Waki (Tsukuba University) / Yona Takahashi (Tokyo University of Foreign Studies)

In this study, we describe the Hieratische Paläographie DB, which was developed for the purpose of supporting the utilization of the Möller's Hieratic dictionary. Specifically, the IIIF Curation API is used to manage the partial area in the page image, and metadata such as character numbers and pronunciation are also described with RDF format. This enables to search items based on those partial images and compare images across pages and volumes. Furthermore, we verified the usefulness of the developed DB through examples such as image comparison based on research use scenarios.

1. はじめに

東京大学附属図書館アジア研究図書館上廣倫理財団寄付研究部門（U-PARL）ではアジア研究図書館の構築支援の一環として蔵書構築を行っている。その蔵書の一つとしてヒエラティック字典の George Möller, Hieratische Paläographie, 4 vols, 1909-36（初版本セット。以下、Möller の字典）がある。

また、U-PARL では研究図書館の機能開拓研究も実施している。その成果の一つとして資料のデジタル公開を行っており、それが「東京大学アジア研究図書館デジタルコレクション[1]」である。そして、「東京大学アジア研究図書館デジタルコレクション」に設けられた「Digital Resources for Egyptian Studies」にて、先に紹介した Möller の字典の画像を IIIF 形式で公開している。

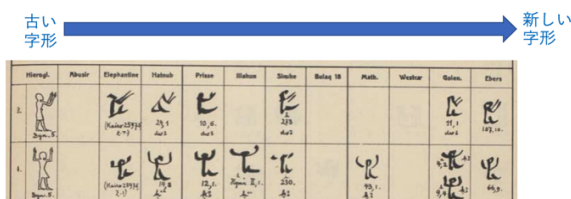


図 1 Möller の字典のページ画像例
Figure 1 Example of page images.

Möller の字典のページ画像例を図 1 に示す。ページ内の各行ごとにヒエラティックの字形が列

挙されており、列を右に進むごとに新しい時代の字形となっている。

この Möller の字典は第 1 巻の刊行から既に一世紀以上も経っている書籍であるが、ヒエラティックを学習・研究する際には現在でも必要不可欠な字典である。一方、古い書籍であることと、紙媒体で分冊されていることにより、以下に示すような使用方法に関する課題がある。

- Möller（ヒエラティック）の文字番号はヒエログリフの文字番号とは異なるシステムであり、対応表が別途必要となる。
- 文字の音価や語の子音転写から文字を調べることができない。
- ある 1 つの文字の解釈候補を探す際に複数のページを確認しなければならない。

そこで筆者らは、上述した課題を解決するシステムとして、Hieratische Paläographie DB[2]を構築している。本論文では、このシステム構築におけるデータ構築やアプリケーション開発について述べる。さらに、研究利用シナリオに基づく画像比較の例や、他の DB との連携事例などを通じ、本 DB の有用性を検証する。

2. データ作成

先述した通り、Möller の字典のページでは行ごとにヒエラティックが分けられている。本研究では行ごとの画像およびメタデータの管理を行う

ことにより、行ごとに与えたメタデータ（文字番号や音価など）に基づく検索や、ページおよび巻を横断した画像比較を可能とし、1. で挙げた課題を解決する。

2.1. IIF Curation API を用いた画像部分領域の管理

行ごとの画像およびメタデータの管理にあたり、本研究では画像が IIF 形式で公開されている利点を生かし、IIF Curation API を採用した。IIF Curation API では、画像の部分領域を一つのデータ単位として管理し、各々のメタデータも合わせ

て記述することができる。IIF Curation API を活用した研究例として、顔貌コレクション[3]などがある。

Möller の字典の各ページの行（画像の部分領域）の取得にあたっては、すでに人手によって切り出し済みのサンプル画像が存在したため、ページ画像と切り出し済み画像の特徴量マッチングを行った。図 2 に示すように、A-KAZE と KNN (K-Nearest Neighbor algorithm) を用いて、ページ画像中の行の座標情報を取得し、それを IIF キュレーションリストの形式に変換した。精度は 90%程度であり、最終的には人手による目視で誤り箇所を修正した。

①切り出し済み画像を用いた特徴量マッチング (A-KAZE, KNN)



②座標情報を使って、IIFキュレーションリストを作成

図 2 特徴量マッチングによる座標情報の取得

Figure 2 Acquisition of coordinate information by feature matching.

2.2. メタデータ設計

各々の行（画像領域）に対して、Möller の字典では提供されていない情報を含め、RDF を用いたメタデータ記述を行った。このメタデータの記述にあたり、以下の 3 つのクラスを定義した。

- Item : Möller の字典のページ画像内の行領域
- HieroglyphNo : ヒエログリフ番号
- HieraticNo : ヒエラティック番号

以下では、それぞれのクラスのリソースに対して定義したプロパティを示す。なお、接頭名詞空間「hpdb」は「<https://w3id.org/hpdb/api/>」を示す。

2.2.1. Möller の字典のページ画像内の行領域

本リソースが持つプロパティを表 1 に示す。当該領域に関する「ヒエラティック番号

(`hpdb:properties/hieraticNo`)」「ヒエログリフ番号 (`hpdb:properties/hieroglyphNo`)」については、後述するクラスを値域として与えた。また、Möller の字典の巻数 (`hpdb:properties/vol`) や、備考 (`hpdb:properties/notes`) を記述するプロパティを定義した。さらに、サムネイル画像へのリンク (`schema:image`) や IIF Curation Viewer へのリンク (`schema:relatedLink`)、ライセンス (`schema:license`) のプロパティも用意した。

2.2.2. ヒエログリフ番号

本リソースが持つプロパティを表 2 に示す。これらのプロパティ選定については、`enwp>List_of_Egyptian_hieroglyphs`[4]を参照した。

2.2.3. ヒエラティック番号

本リソースが持つプロパティを表 3 に示す。これらは音価/語を記述するプロパティである。

表 1 「Möller の辞書のページ画像内の行領域」リソースが持つプロパティ
Table 1 List of properties in “Item” resources.

プロパティ名	説明 ↑	プロパティ値の例	必須・省略の別	繰り返しの有無	値域
rdfs:label	名称	1(A26)	必須	なし	xsd:string
hpdb:properties/hieraticNo	ヒエラティック番号	hpdb.term/hieraticNo/1	必須	なし	hpdb:classes/HieraticNo
hpdb:properties/hieroglyphNo	ヒエログリフ番号	hpdb.term/hieroglyphNo/A26	必須	なし	hpdb:classes/HieroglyphNo
hpdb:properties/vol	巻数	3	必須	なし	xsd:integer
hpdb:properties/notes	備考	See 39(A4/A5).	省略可能	なし	xsd:string
dct:identifier	識別子	0001	必須	なし	xsd:string
schema:image	サムネイル画像	https://iif.dl.itc.u-tokyo.ac.jp/iif/asia/hp/2/2_023.tif/9...	必須	なし	rdfs:Resource
schema:relatedLink	IIIF Curation Viewerへのリンク	http://codh.rois.ac.jp/software/iif-curation-viewer/demo/?m...	必須	なし	rdfs:Resource
schema:url	URL	http://w3id.org/hpdb/item/0001	必須	なし	rdfs:Resource
schema:license	ライセンス	http://creativecommons.org/licenses/by/4.0/	必須	なし	rdfs:Resource

表 2 「ヒエログリフ番号」リソースが持つプロパティ
Table 2 List of properties in “HieroglyphNo” resources.

プロパティ名	説明	プロパティ値の例 ↑	必須・省略の別	繰り返しの有無	値域
rdfs:label	名称	A19	必須	なし	xsd:string
hpdb:properties/unicode	ユニコード	U+13017	必須	なし	xsd:string
rdfs:description	文字の説明	bent man leaning on staff	省略可能	なし	xsd:string
hpdb:properties/transliteration	翻字	old, to be fragile (jrw) Elder (smsw) bigger (wr) to learn...	省略可能	なし	xsd:string
hpdb:properties/notes	備考	perhaps also wise and sage	省略可能	なし	xsd:string

表 3 「ヒエラティック番号」リソースが持つプロパティ
Table 3 List of properties in “HieraticNo” resources.

プロパティ名	説明	プロパティ値の例	必須・省略の別	繰り返しの有無	値域
rdfs:label	名称	1	必須	なし	xsd:string
hpdb:properties/phonetic	音価/語	nis, aA	省略可能	あり	xsd:string

2.3. データ作成

2.3.1. IIIF キュレーションリストの作成

2.1 で作成した IIIF キュレーションリストに対して、2.2 で作成したメタデータをキュレーションリスト内のメンバーに追加した。IIIF Presentation API 2.1 及び Curation API 1.0 の Curation におけるメタデータは人間可読性を重視した仕様となっているため、2.2.1 で定めた [rdfs:label](#) の値を `metadata` のラベル名として与えた。また、RDF データへのリンクを `seeAlso` 要素に与えることで、機械可読性を担保した。作成した IIIF キュレーションリストには以下からアクセスすることができる。

<https://moeller.jinsha.tsukuba.ac.jp/data/curation.json>

2.3.2. RDF データの作成

RDF データの作成にあたっては、江草[5]らの手法を参考として、所定のフォーマットで記述した Excel ファイルのデータを、SHACL データ、および 2.2 で述べた RDF データに変換し、それぞれ JSON-LD ファイルとして出力した。SHACL (Shape Constraint Language) は、検証可能な RDF データモデルの形式記述であり、リソースが持つ構造を説明するための文書としても利用することができる。

3. アプリケーション開発

開発したアプリケーションは、Dydra[6]を用いた SPARQL エンドポイントと、Nuxt.js を用いた検索アプリケーションの2つに分けられる。以下、それぞれについて述べる。

3.1. SPARQL エンドポイントの構築

2. で作成した RDF データへのアクセスを容易とするため、Dydra を用いた SPARQL エンドポイントを作成した。さらにユーザインタフェースの改良を目的として、トリプルストアを探索するための ajaxy フロントエンドである Snorql[7]を用いたインタフェースを提供している(図3)。このインタフェース提供に関わる静的ファイルは、後述する GitHub Pages 上で公開している。

Description of <https://w3id.org/hpdb/api/items/0001>	
rdf:type	hpdb:classes/Item #
rdfs:label	1 (A26)
dct:identifier	'0001'
schema:image	<https://iif.dl.itc.u-tokyo.ac.jp/iif/asia/hp/2/2_023.tif/983,2131.3913,476/200/0/default.jpg> #
schema:license	<http://creativecommons.org/licenses/by/4.0/> #
schema:relatedLink	<http://codh.rois.ac.jp/software/iif-curation-viewer/demo/?man...s/p23&xywh=983,2131.3913,476&xywh_highlight=border> #
schema:url	<http://w3id.org/hpdb/item/0001> #
hpdb:properties/hieroglyphNo	hpdb:term/hieroglyphNo/A26 #
hpdb:properties/möllerNo	hpdb:term/möllerNo/1 #
hpdb:properties/phonetic	'aA'
hpdb:properties/vol	'nis'
hpdb:properties/vol	2

図3 SPARQL エンドポイントの画面例
Figure 3 Interface of SPARQL Endpoint.

3.2. 検索アプリケーションの開発

3.2.1. 提供環境

先に作成した IIF キュレーションリストを読み込み、メタデータに基づく絞り込みや画像比較を可能とするアプリケーションを開発した。

アプリケーションを開発するにあたり、以下の工夫を実施した。

- 静的サイトとしての構築
- 永続的 URL の設定

まず、総データ数が 1,626 件と比較的小規模な点¹、および運用コストの軽減を目的として、静的サイトとして構築した。静的サイトの利点として、表示の高速化、セキュリティリスクの低減、運用コストの低減、などが挙げられる。具体的には、Nuxt.js の SSG (Static Site Generator) 機能を用いたサイトとして構築し、GitHub Pages を用いた配信を行っている。サイトビルド時に 2.3.2 で作成

¹ 現在のところ、収容データは I-III 巻の基本字のみであり、1-3 巻に掲載されている数字、尺度、続け字、

した RDF データを読み込み、静的な HTML ファイルを生成している。

2 点目は、永続的 URL の設定である。手段としては、江草ら[5]の手法を参考として、w3id.org[8]を利用した。この時、GitHub Pages へのリダイレクト設定において、特定のディレクトリ以下では、拡張子を無視したアクセスを可能とした。これにより、JSON-LD 形式で記述した RDF データに対して、拡張子のない URI でのアクセスを可能とした(表4)。

表4 w3id.org を用いた URL の永続化

Table 4 Permanent URL with w3id.org.

リダイレクト元	リダイレクト先
https://w3id.org/hpdb/api/items/0001	https://moeller.jinsha.tsukuba.ac.jp/api/items/0001.json

3.2.2. 提供機能

検索画面の例を図4に示す。画面上部の検索フォームや、画面左部のファセット検索を利用して、ヒエログリフ番号や音価に基づく行の絞り込みが可能である。また、検索結果の各アイテムにはチェックボックスが用意されており、ここで選択した画像群を Mirador で比較することができる。

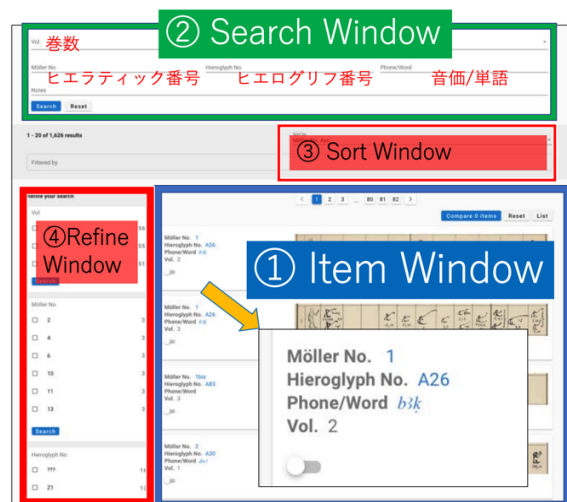


図4 検索画面例
Figure 4 Search interface.

また各アイテムから、個別の詳細画面に遷移する。図5に示す詳細画面では、IIF Curation Viewer Embedded[9]を通じて当該箇所の画像が表示される他、そのメタデータが合わせて表示される。さらに、IIF Curation Viewer や、上述した SPARQL エンドポイント、各種 SNS のリンクも提供する。

そして 1-3 巻の補遺としての 4 巻のデータは未収録である。

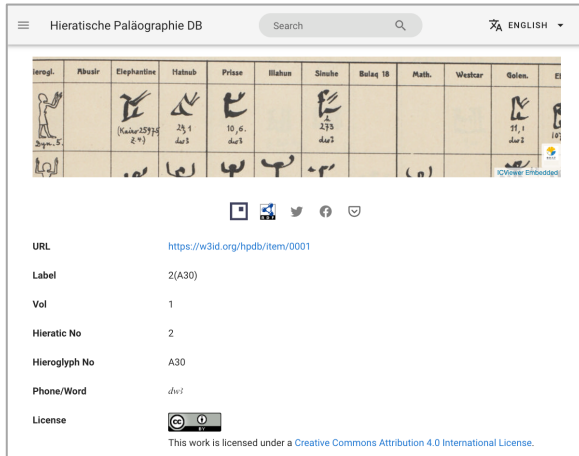


図 5 アイテム詳細画面例
Figure 5 Item detail interface.

さらに高久[10]による手法を参考として、2.3.2で述べた SHACL データから自動変換して作成したプロパティの説明ページも提供している。表 1 から表 3 は、本機能を用いた作成した画面である。

3. 2. 3. 利用例

次に、エジプト語研究の立場から本システムの利用例を述べる。ヒエログリフの E23 [文字の象形対象=ライオン；音価= rw, r] と U13/U14 [文字の象形対象=鋤；音価= sn] は、文字の象形対象、音価、用法の異なる別の文字である(図 6 左部)。しかしながら、これらの文字が互いに交替することが知られている。このような、関係のないヒエログリフが交替する現象の背後に、ヒエラティックの字形の類似性が隠されていることがある。そこで、3.2.2 で述べた機能を利用して E23 と U13/U14 に対応するヒエラティックの 125 と 600 を比較する。まず、ヒエラティック番号で 125 (あるいはヒエログリフ番号で E23) を検索した後、2 巻の 125 を選択する。その後、同様にヒエラティック番号の 600 (あるいはヒエログリフ番号 U13 あるいは U14) を検索した後、2 巻の 600 を選択する。この状態で「Compare 2 items」ボタンをクリックすると、Mirador 画面に遷移し、選択した文字が 1 つの画面に上下に配置されて示される。その結果、文字の対比が容易となり、両者のヒエラティック字形が類似していることを確認することができる(図 6 右部)。

ヒエラティックの 125 と 600 は Möller の字典で別ページにあるため、紙媒体の書籍を使用して相互比較を行おうとした場合、たとえばページをコピーして両者を比較するなどのことをしない限り、参照者の記憶を頼りに別ページの字形同士を比較することとなる。しかし、本システムを使用することにより、物理的に離れたページの文字の比較が 1 画面で達成されることとなるため、参照者の記憶に頼ることなく、文字の比較を行うことができる。

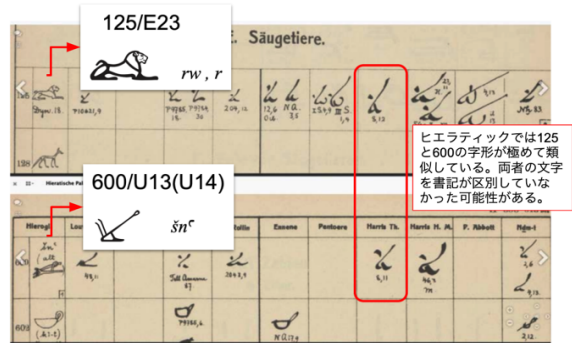


図 6 Mirador を用いた選択行の比較画面例
Figure 6 Image comparison with Mirador.

3. 2. 4. 外部システムとの連携例

さらに、永井や和氣らが別途開発している Hieratic Database Project[11][12]の DB と IIF を介して連携することにより、本 Hieratische Paläographie DB の字典ツールとしての相互運用性の高さを確認した。具体的には、図 7 に示すように、Hieratic Database Project の DB が提供するアノテーション表示箇所に本 DB へのリンクを与え、字典として適宜参照可能な環境を構築した。IIF Image API を用いて切り出したサムネイル画像を表示するとともに、当該文字をクエリした検索結果画面に遷移する。



図 7 Hieratic Database Project との連携
Figure 7 Connection with Hieratic Database Project.

4. 考察

4.1. IIF Curation API によるデータ記述

本研究では、IIF Curation API を用いたデータ記述を行った。この利点として、相互運用性の高さ、管理の容易さが挙げられる。

IIF Curation API に準拠したデータを用いることにより、切り出し画像の表示部分を IIF Curation Viewer などに任せることができ、本アプリケーションで肝となる検索部分の開発に注力することができた。また、本検索アプリケーション以外でも、IIF Curation Platform が提供する各種ツール (ex. ICEditor) を利用できた。これらの相互運用性の高さから、効率的なデータ作成とアプリケーション開発を行うことができた。

また、複数の資料(本研究では 3 巻からなる字典)からそれぞれ切り出した画像領域を、それぞ

れのメタデータとともに管理するという要件に対して、IIIF キュレーションリストの1ファイルのみで対応することができた。IIIF のアノテーションリストやその他のファイル形式でも実現可能であるが、IIIF キュレーションリストはその中でも特に簡易な記述手法の一つである。このデータ管理の容易さは、データの長期保存にもつながるという点で有用である。

4.2. 静的サイトジェネレータを用いたアプリケーション開発

上述した通り、IIIF や RDF を用いたデータ作成により、利活用性・相互運用性の高いシステム構築を志向したが、さらに本研究では持続可能性も考慮したアプリケーション開発を行った。具体的には、IIIF 画像サーバや Dydra などの外部システムに依存している部分もあるものの、静的サイトジェネレータを用いたアプリケーションを構築し、システムの運用コストを軽減した。データ規模や機能要件に依存するが、データが比較的小規模であり、個々のアプリケーションの独自性が高く、また運用コストの低減と長期利用保証が求められる人文系の研究成果の発信手法として、本手法は有効な手段の一つだと考える。

4.3. エジプト語研究からの考察

1. で示したように、Möller の辞書はヒエラティック番号に従って文字が配列されているため、Möller の辞書の後に整備されたヒエログリフの文字番号に慣れているエジプト学者は、Möller の辞書の文字配列に戸惑うことがある。また、Möller の辞書では、一部の文字に音価が付されているものの、すべての文字に音価や語の子音転写が付されているわけではなく、音価の索引も用意されていない。それゆえ、現行のヒエログリフ番号、文字の音価や語の子音転写からヒエラティックを検索することができなかった。このような課題を克服した世界初のシステムとして開発されたものが Hieratische Paläographie DB である。ヒエラティック番号での検索は言うまでもなく、ヒエログリフ番号、音価/語からヒエラティックを検索する仕組みは、エジプト学におけるヒエラティックの学習や研究に大いに役立つことが期待される。特に、3.2.3. で述べたように、ページや巻の異なるデータを一覧表示させる仕組みは、翻刻の判断を支援する画期的なシステムであると高く評価される。

5. 結論

本研究では、Möller の辞書の利活用を目的として構築した Hieratische Paläographie DB について述べた。IIIF を活用することで、IIIF Curation API によるデータ記述構造の標準化や、Mirador や他

DB との接続といった高い相互運用性を実現することができ、情報学分野への貢献がなされた。また文字比較に関する研究利用シナリオを通じて、人文学に依拠したエジプト学の研究に有効なシステムが構築されたことが確認された。

謝辞

本研究は JSPS 科研費 18K00525 の助成を受けたものである。

参考文献

- [1] “東京大学アジア研究図書館デジタルコレクション”. <https://iiif.dl.itc.u-tokyo.ac.jp/repo/s/asia/page/home>, (参照 2020-08-26)
- [2] “Hieratische Paläographie DB”. <https://moeller.jinsha.tsukuba.ac.jp/>, (参照 2020-08-26)
- [3] 鈴木親彦, 高岸輝, 北本朝展. 顔貌コレクション (顔コレ): 精読と遠読を併用した美術史の様式研究に向けて, じんもんこん 2018 論文集, Vol.2018, pp.249 - 256, 2018.
- [4] “List of Egyptian hieroglyphs - Wikipedia”. https://en.wikipedia.org/wiki/List_of_Egyptian_hieroglyphs, (参照 2020-10-29)
- [5] 江草由佳, 高久雅生. 教科書 Linked Open Data (LOD) の構築と公開, 情報の科学と技術, Vol.68, No.7, pp.361-367, 2018.
- [6] “Dydra”, <https://dydra.com/>, (参照 2020-10-29)
- [7] “kurtjx/SNORQL: a ajaxy front-end for exploring triple stores”, <https://github.com/kurtjx/SNORQL>, (参照 2020-10-29)
- [8] “w3id.org - Permanent Identifiers for the Web”, <https://w3id.org/>, (参照 2020-10-29)
- [9] “IIIF Curation Viewer Embedded | ROIS-DS 人文学オープンデータ共同利用センター (CODH)”, <http://codh.rois.ac.jp/software/iiif-curation-viewer-embedded/>, (参照 2020-10-27)
- [10] “masao/ttl2html: Static site generator for Linked Data”, <https://github.com/masao/ttl2html>, (参照 2020-10-27)
- [11] 永井正勝, 和氣愛仁. 古代エジプト神官文字写本を対象とした言語情報表示システムの試作, じんもんこん 2012 論文集, Vol.2012, No.7, pp.225-230, 2012.
- [12] 和氣愛仁, 永井正勝, 高橋洋成. アノテーション付与型画像データベースプラットフォームの IIIF 対応, 人文科学とコンピュータ研究会報告, Vol.2019-CH-119, No.15, pp.1-6, 2019.