

テクニカルサポート通信

**Technical
support
Newsletter**

〈ARC のデータベースの基本解説 その1〉

◇プロジェクト別の専門データベースを運用してみよう◇

- ・ポータルデータベースと個別プロジェクト型データベース
- ・リソースデータベースとレファレンスデータベースによるリサーチ空間
- ・多言語インターフェイスと個別専門 DB の専門用語対応

1, 個別プロジェクト型 DB を実現するポータルデータベース

アート・リサーチセンター（ARC）のデータベース（DB）システムは、2014年頃から「ポータルデータベース」という考え方を導入し、運用されています。研究プロジェクト活動の中で生まれてくる、あるいは必要とされるDBを、その内容の類似性で統合し、できるだけ一つのDBシステムにまとめて利用してもらい、しかし、プロジェクト単位、所蔵者単位でも独立して運用できる仕組みです。一つの巨大なDBシステムの中にデータが搭載されているながら、データ1件単位でどのプロジェクトが使えるデータなのかを管理しているため、不特定多数の利用者はもちろん、他のプロジェクトからも閲覧・活用できないデータを登録でき、また、そのデータを一般公開するのか、限定利用するのかの管理もできます。

さらに、ARCのDBは、開発当初からマルチメディア型で、コンテンツ閲覧・鑑賞を容易にすることを特徴としていたので、このコンテンツの閲覧のセキュリティにも配慮されており、DBのセキュリティ・レイヤーとコンテンツセキュリティを組み合わせると、重層型で情報管理ができます。

このように研究プロジェクトにとっては、利用者をきめ細かくコントロールできるので、蓄積段階・研究段階・公開段階など、状況に応じた様々な用途に使えるようになっているわけです。文系の研究プロジェクトにとっては、対象となる研究資源・素材の蓄積と管理（まさにアーカイブ）は、必須の機能ですから、本拠点を活用する研究プロジェクト、メンバーの皆さん、是非ともARCポータルデータベースを使って見てください。

2, リソースデータベースとレファレンスデータベースによるリサーチ空間

また、ARCのポータルDBは、デジタル化された文化資源そのものを活用するリソースデータベースと、研究者らが作成する解説や考証データを蓄積するレファレンスデータベースという2本立てのカテゴリでシステムが組まれています。

現在、ARCのHP上にある、Database List (<https://www.arc.ritsumei.ac.jp/j/database/>) には、リソースデータベースとしては、その形態別に、

- ①浮世絵（版画・絵画を含む）、②古典籍、③近代書籍、④番付、⑤板木、
- ⑥催事型写真、⑦景観型写真、⑧日本関係地図、⑨文化資源（音声・動画を含む）

の9件のDBが稼働しており、⑨が、①～⑧に当てはまらない文化資源諸々を担うことで、どのような種類の文化資源が提示されても、DBに登録できると考えています。

レファレンスDBとしては、

①年表記事、②文化人・芸能人人物名、③演劇・映画興行年表をリストに掲示しているが、近々、④人物典拠DBも運用を開始する予定である。ここに空間（地理）情報が加わると、時空間情報のレファレンスが可能となる。時空間の中に置かれるのが「コト」であり、人物名や人物典拠DBで「ヒト」の情報がDB化されます。ARCでは、「モノ」を文化資源と捉えており、リソースデータベースは、基本的に「モノ」を対象としてDB化したものです。こうして、「モノ」「コト」「ヒト」の文化をかたちづくる基本要素を構造化できるわけ

です。そして、この三つの要素の情報が連携することで、「リサーチ空間」を構築できるのです。

ARCのデータベース群は、どのような対象・素材にでも対応できるように、ARCはシステム開発を続けており、想定外の新しい対象が出現しても、テクニカル・サポートボードが出来る限り迅速に新しいシステムを開発していくのを使命としています。さらには、このリサーチ空間上で、このそれぞれのDBが持つシナジーを最大限に生かすべく、DB間を連携させるシステム改良にも力を入れています。シナジーは、データが蓄積されてから始めて気づくこともあるため、DBを使いながら「このDBのこの部分とこちらのDBのこの部分を直接繋げると、有用性が抜群に高くなるはずだ」と気づいたなら、すぐにテクニカル・サポートボードに相談してもらえれば、と思います。

3. 多言語インターフェイスと個別専門DBの専門用語対応

ARCのリソースDBのメタデータは、日本語のみのものが多いのですが、基本的に研究対象は日本の文化資源であり、まずは、日本語で専門性の高いデータを蓄積していくのが目標となると考えます。そして、余力があればバイリンガルでデータが蓄積していく、これも当然の方向性です。

一方、日本の文化に興味を持ちながら、日本語が堪能ではないユーザーが世界には意外と多く存在するという指摘も以前から受けてきました。しかし、一つのデータベースに対し、言語単位でインターフェイスを用意すると、システムの改良・更新時には、すべての言語用のシステムファイルを変更しなくてはならないため、システム担当者の負担が大きくなります。

そこで、DBのインターフェイスに使われる日本語の項目名や使用方法などのフレーズと、表示したい言語とを対応させた「言語ファイル」を用意し、それをシステムに組み込むことで、日本語を基本言語として、インターフェイス上は、別言語で表示されるような仕組みを組み込んであります。たとえば、

“作品名” => “Title”, “サムネイルのみを表示” => “Thumbnail Only”,

のような対応表を用意すればよいのです。言語ファイルは、いくつも登録できるので、言語ファイルさえ用意すれば、Web上で扱えるすべての言語で表示が可能ということになります。

さらに、この仕組みを応用し、ポータルデータベースの中に構築できるプロジェクトごとの個別データベースのフォーマットファイルに、言語ファイルと同様の用語対応リストを書込むことで、表示する用語を変更して表示できます。

たとえば、浮世絵（版画・絵画）データベースには、近代書籍の口絵も登録されており、独立した「口絵DB」が稼働しています。口絵を研究する場合、浮世絵で「版元」と呼ぶ概念は、明治以降の書籍ですので「出版社」と呼びたい。このように概念の内容は同じでも、分野によって用語に違いがあるのです。この場合は、

“版元” = “出版者”

という対応リストをフォーマットファイルに記述します。また、口絵にはほとんど題名が記載されていないのですが、それを収録している書籍には書名があります。場合によっては、副題が付いていることもあります。一方、浮世絵にはシリーズ名と個別名がある場合があり、描かれている画題も特定できることがあります。そこで、メタデータ項目については、

“シリーズ名（作品名1）” = “書名（書名1）”

“個別名（作品名2）” = “副題（書名2）”

“画題” = “口絵題名”

という項目のマッチングをフォーマットファイルに加えることで、システム全体を書換えることなく、口絵 DB だけの用語を表示することができ、個別の専門 DB として違和感なく運用できるようになります。

以上の機能をご理解の上、より多くの研究プロジェクトに使ってもらえれば幸いです。

(以上、文責 赤間亮)

< ARC Database Basics: Part 1 >

◇ How to Operate Specialized Databases by Project ◇

1. Portal Databases and Individual Project-Type Databases
2. Research Space with Resource and Reference Databases
3. Multilingual Interface and Terminology Support for Individual Specialized DBs

1. Portal Databases that Realize Individual Project-Type Databases

The database (DB) system of the Art Research Center (ARC) introduced the concept of a “portal database” around 2014, which was to integrate DBs that were created or required in the course of research project activities, based on the similarity of their contents. While we ask the research projects to combine their DBs as much as possible into a single DB system, each project or collector can also operate the system independently. While the data is stored in one large DB system, we manage which projects can use each piece of data, which makes it possible to register data that cannot be viewed or used by other projects, as well as unspecified users. We can also control whether the data shall be open to the public or for limited use.

Furthermore, characterized as a multimedia type since the beginning of their development, the ARC’s DBs have been facilitating the viewing and appreciation of contents. That means that, we have given much thought to the security of viewing the contents, as well. Combining the security layers of the DB and content security, we have been taking a multi-layered approach to manage information.

Since research projects can manage their users precisely, they can use the system for a variety of purposes, depending on their situation, such as the accumulation, research, and publication phases. As the accumulation and management of target research resources and materials, i.e., the archive, is an essential function for research projects in the humanities, we encourage all research projects and members to make use of the center’s portal DBs.

2. Research Space with Resource and Reference Databases

As a system, the ARC’s Portal DBs are organized under two major categories: a resource database that makes use of the digitized cultural resources themselves, and a reference database that accumulates commentary and reference data created by researchers.

Currently, the following, 9 resource databases are in the Database List on the ARC website (<https://www.arc.ritsumei.ac.jp/e/database/>), according to the forms of resources:

- (1) ukiyo-e (including prints and paintings) ; (2) early Japanese books; (3) modern Japanese books; (4) banzuke (playbills); (5) woodblocks; (6) photographs of events; (7) landscape photographs; (8) maps related to Japan; and (9) cultural resources (including audio and videos). (9) is used for various cultural resources that may not fit into the categories (1) to (8), which, we believe, can accommodate any kinds of cultural resources presented.

As for the reference DB, its list includes the following:

- (1) chronological articles; (2) names of people in culture and entertainment; and (3) chronology of performing arts and cinema. Besides these, we are planning to start operating: (4) the personal name authority

DB soon. When spatial (geographic) information is added to this, it will be possible to refer to spatiotemporal information. It is an “event” that shall be placed in time and space, and the information of “people” is made into a DB with the personal name authority DB. The ARC regards “objects” as cultural resources, and the resource database is basically a database for “objects”. In this way, we can structure the basic elements that shape the culture, such as “objects”, “events” and “people”. By connecting the information from these three elements with each other, we can construct a “Research Space”.

The ARC continues its system development so that its group of databases can handle any subjects or materials, and when new and unexpected subjects emerge, the Technical Support Board carries out its mission, developing new systems as quickly as possible. Furthermore, to maximize the synergies of each DB in this research space, we are also focusing on improving the system that links the DBs with one another. Synergies may only become apparent after the data has been accumulated. Thus, should you realize while using a DB that “Part A of this DB should be much more useful if it were directly connected with Part B of that DB”, please feel free to consult the Technical Support Board immediately.

3. Multilingual Interface and Terminology Support for Individual Specialized DBs

Most of the metadata in the ARC’s resource DBs is only available in Japanese, but since the basic research target is Japanese cultural resources, we believe that our primary goal is to accumulate highly specialized data in Japanese. Then, should there be spare capacity, the data shall be accumulated bilingually. This sounds a natural course of action we shall take.

On the other hand, it has long been brought to our attention that there are a surprisingly large number of users in the world who are interested in Japanese culture but are not fluent in Japanese. What we would like you to know is that, preparing a different interface in each language for a single database would mean an extra burden on the system staff, when the system should be improved or updated, as they would have to change all of the system files for each language.

In order to deal with this, therefore, we have prepared “language files” which correspond to the phrases used in the DB interface, such as Japanese item names, usage, and the language to be displayed. By incorporating these into the system, we have created a mechanism where Japanese is the basic language while the interface can be displayed in different languages. For example, we prepare a table of correspondence as follows:

“作品名” => “Title”, “サムネイルのみを表示” => “Thumbnail Only”

Since you can register as many language files as you like, any language available on the web can be displayed as long as you have the language file.

Furthermore, this mechanism can be applied to the format file of the individual database for each project that can be built in the portal database, by writing the same list of terms as in the language file, so that the terms to be displayed can be changed.

For example, *kuchi-e* prints (woodblock printed frontispieces) in modern books are registered in the Ukiyo-e (prints and paintings) Database while an independent “Kuchi-e Portal DB” is also in operation. When researching *kuchi-e*, the concept of “publisher”, which is referred to as “版元 (*hanmoto*)” in the ukiyo-e research, shall be called “出版社 (*shuppansha*)”, since *kuchi-e* can be found in books published after the Meiji Period. Thus, while the concept means the same, the terms might be different, depending on research subjects. In this case, the format file should contain a corresponding list of:

“版元” = “出版者”.

Furthermore, although titles are rarely given to *kuchi-e*, the books that contain them do have titles. In some cases, the books also have subtitles. Ukiyo-e, on the other hand, may have both series and individual names, and their themes may be identifiable, as well. Therefore, by adding the corresponding items for the metadata items to the format file as follows, it is possible to display the terms only for the Kuchi-e Portal DB without rewriting the entire system, and you will be able to operate it as a separate specialized DB without any sense of discomfort.

“Series name (title 1)” = “Book title (title 1)”

“Individual name (title 2)” = “Subtitle (title 2)”

“Picture theme” = “Title of Kuchi-e”

We sincerely hope that you will find the above-mentioned features useful, and that you will use them in more of your research projects.

(Responsibility for this article: Ryo Akama; and translated by Emily Li, and Keiko Suzuki)

【システム開発から】

ARC ポータルデータベースシステム上での多言語化と専門用語への対応

山路 正憲 (立命館大学衣笠総合研究機構 研究員)

E-mail m-yamaji@fc.ritsumeit.ac.jp

赤間 亮 (立命館大学文学部 教授)

E-mail rat03102@lt.ritsumeit.ac.jp

「多言語化」「個別専門用語」対応システムの開発経緯と、システムの概要について、具体的な事例を交えて詳しく述べておく。

1. DB のサステナビリティ

アート・リサーチセンター (ARC) は、その設立以来、デジタル環境下での文理融合型研究拠点となることを目標に掲げ活動してきた。その手法を、デジタル・ヒューマニティーズと呼ぶようになって以降は、とりわけ文系・理系の双方からのコラボレーションによる研究が活発に展開する研究拠点として、注目され評価を得てきた。

文理融合研究活動を実現するために最も重要なキーワードは、「デジタル」であり、その可能性を蓄える倉庫が、デジタル・アーカイブ (DA) である。そして、それは保存・蓄積され、それをエネルギーとして継続的に活用されなければアーカイブではない。そして、そのアーカイブを活用させるエンジンがデータベース (DB) ということになる。

世にさまざまなデジタル・アーカイブをテーマにした研究活動があるが、実際には、それが単なるデジタル化であってアーカイブ (蓄積) されない事例、あるいは、折角アーカイブされても継承されずに短期間で消滅してしまう事例が何と多いことだろうか。ARC は、そうした状況を踏まえて、国際共同利用・共同研究拠点としての重要な課題として、「データベースのサステナビリティ」を大きな柱としている。

日本の研究組織では、自組織の所有する研究資源を蔵品目録として公開するDBと、その組織が共

同利用型であれば、研究成果としてのDBを公開するというパターンでDB群を運用している場合が多い。問題は、後者のタイプで、例えば、国際日本文化研究センターでは、41件のデータベースがリスト化されているが、所蔵図書以外のOPAC以外は、プロジェクト単位で作成されているため、たとえば、「風俗図会」には、「絵双紙 (絵草子・絵草子) の高精度画像データベース」としながら、浮世絵を含むさまざまな形態の画像付データ458件のアイテムが登録されている。一方、内容・資料形態で重複するDBとしては、たとえば「鯰絵コレクション」(87件)や「米国議会図書館所蔵浮世絵」(2741件)「在外日本美術」(8673件)など浮世絵を対象とするDBが、独立して稼働している。

ARCの場合も、20年以上に亙り国内外の文化資源を対象にDA活動を活発に行い、また研究者自身による効率的なDB構築手法・ARCモデルの確立により、100件以上にもなる多数のDBを構築・運営することとなった。プロジェクトによっては、景観型写真に写り込んでいる内容 (オブジェクトやテーマ) ごとに、別々のDBを用意するように求めて来られる場合もある。しかし、システム開発の担当は、他の組織も同様であろうが、人員が限られており、開発の遅延や、サステナビリティへの危機感が極限状態に達していた。

2. ポータルDB化と専門DB対応問題

その後、ポータルDB化構想¹⁾のもと、DB群の改修を行い、類似の性質を持ったDBや、同一の

資料を異なるテーマで扱っているDBなどを統合し、柱となるいくつかのポータルDBへと集約する動きを進めてきた。このことにより、データ管理の一元化・簡便化が行われ、より効率の良いデータ収集や蓄積、横断検索などが可能となった。

しかし一方では、この統合によって元の個別DBそれぞれに存在した特徴的な項目を汎用的な名称に変更せざるを得ず、個別DBの利用者にとっては、各分野の事情に沿った項目名とは違う用語に馴染めないまま、利用して、フラストレーションが誘発するという問題を引き起こすことになった。

ARCは、2019年度に文部科学省による国際共同利用・共同研究拠点に採択され、様々な国籍・多言語による研究者チームの研究活動を、これまで以上に積極的に招き入れることがミッションとなる。そして、その研究活動や研究成果として用いられるDBについても、効率的な国際化・多言語化が急務とされた。

こうして、各言語別にその分野特有の特徴的な項目名表記が存在するものに対しては、それぞれのデータベース項目に対して「言語別表記」「専門分野別表記」を実現することが大きな課題となったのである。

この両方の課題に柔軟に対応するべく、言語・用語とDB項目の表記についての関係性をコントロールする機構の実現を目指して、システムの改修を行うこととなった。

DBシステムの多言語化については、項目名や操作説明文などのユーザーインターフェースを自動翻訳するものなどを既にサービスとして提供している企業もある。また、DB内のメタデータを自動翻訳するサービスや研究などもあるが、今回の内容はユーザーインターフェースについての多言語化・専門用語対応となるため、これも目的が異なる。

今回の目的は自動化ではなく、自動処理では対応不可能な、学術的に専門性の高い分野で、DBそれぞれの内容に応じた、DB制作者・データ提供者の意図通りのデータ項目名やインターフェイス上の用語を表示して公開する手法の確立が主たる課題である。そのため、新たなシステムを独自に開発することとした。

まず、具体的な事例によって説明しよう。ARCの

浮世絵ポータルDBでは、浮世絵や絵画など、一枚の画像で構成される資源形態のDBを統合したものである。

元々の個別DBでは、それぞれの資料種別、研究テーマごとの項目呼称を各々自由に設定していた。一例を挙げると、ダブリンコアでいう「title」となる項目について、DBによっては「作品名」「書名」など様々な名称がつけられていた。しかも、個別DBは研究者の個人管理であるので、同一の資料が複数の個人DBに別個に登録され、メタデータの項目内容、項目名称も統一されていない状況があった。これでは、システム管理は容易ではない。

これに対し、資源1件と画像1枚が対応するタイプの資源DBを統合し、一元管理することのできるポータルDBを構築し、これを「浮世絵ポータルDB」²⁾とした。

一方で、大規模なポータルDBのみでは、従来個別DBが持っていた利便性が失われてしまう。たとえば、範囲への限定（ノイズが少なくなる）や、閲覧資料の個別の認証などの機能も維持すべく、資料の所蔵先や利用する研究チームなどの実態に即して、検索ターゲットとする資料を個別に指定して制限することのできる「入口別分割」システムを設けた。これにより、巨大なポータルDBの一部を切り取り、各所蔵機関、各研究プロジェクトグループの個別DBとして活用することも可能となった。

データの統一的な管理や隣接分野での横断検索は容易となったが、一方で個々のデータベース特有の項目名や指示フレーズなどに関しては、絶対に代替できない項目は、システム改修を行うことで対応するとして、軽度の汎用的な項目名・用語については、利用者に我慢を強いていたという状況であった。

共同利用・共同研究拠点としての活動が継続し、海外の博物館・美術館・個人コレクションなどの所蔵先との連携、銅版画、地図、書簡、屏風などが登載資源の分野として加わると、以前からの浮世絵型項目のメタデータとはそぐわない事例が増加し、抜本的な対策が必要となった。そこで着手したのが今回の「多言語化」「専門用語対応」カスタマイズシステムの開発となる。

3. システムの構成

本システムの構造は、ユーザーインターフェース部・汎用言語設定・個別言語設定の三層からなる（図1）。まずポータル DB システム全体からメタデータ項目名や操作上のメッセージなども含めた、すべての用語を登録した辞書ファイルを作成する。同時に、そのすべての用語に対して、多言語化したい言語へ翻訳する内容を定義した言語ファイルを作成し、それぞれの言語について、各用語をどのように表記するかを設定する。

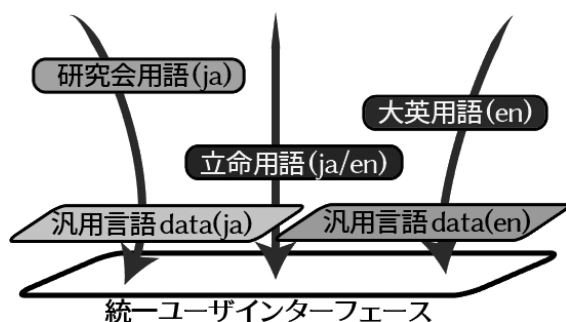


図1 ジャンル・言語別データ切替

これで多言語化は完了するが、「入口別分割」システムを利用し、各個別の所蔵先・研究プロジェクトグループごとに特有の用語があれば、入口別の設定ファイル内において「言語別」「専門別（入口別）」固有のカスタマイズを柔軟に設定することが可能となる。検索システムのユーザーインターフェース側では、まずすべての用語を登録した「汎用言語設定」ファイルのうち、DB アクセス URL のパラメータ「lang」により設定された言語ファイルを読み込み、指定された言語で画面表示を記述していく。その後、全データを表示するポータル入口以外の場合には、更に「入口別分割」設定により、各所蔵機関別・専門別入口の事情に応じた言語設定が上書きされ、個別言語設定が優先されることとなる。個別設定がない用語については汎用言語設定が優先されるため、個別の入口言語設定に関するデータベース作成者の作業量は必要最低限となる。

実際の活用例では、ARC 浮世絵ポータル DB システムにおいて「作品名」となっている項目は、英語表記では「Title」となっているが、「入口別分割」によって作成された「口絵ポータル DB」³⁾では、そ

のジャンル特有の用語として、この項目は「書名」と表示される。また、浮世絵では「配役」となる項目は口絵においては「被写人物」となる。これらジャンル入口別の用語指定がなければ、汎用言語設定の項目名がそのまま表示される（図2）。

フィールド定義名	field83	field83	field6	field1
汎用言語 data	(ja) 作品名	(en) Title	(ja) 配役	(en) AcNo.
口絵 (ジャンル)	(ja) 書名	-	(ja) 被写人物	(en) ARCNo.
画面表記	書名	Title	被写人物	ARCNo.

図表2 口絵データベースでの対応例

この例では日本語ファイル（language_jp）、英語ファイル（language_en）の2言語の例ではあるが、同様に language_cn, language_kr など、2文字の国コードを末尾につけた言語ファイルを作成することで、新しい切り替え言語を増やすことができる。

4. システムの展望

本システムは拡張性を重視して開発しており、設計上はすべての言語に対応可能である。また、この汎用言語設定やジャンル分野別言語設定ファイルは平易なテキストファイルで構成されており、特殊なソフトウェアやプログラミングの知識は不要である。また、今後更なるメンテナンス性の向上を目指し、検索システム上で使われている単語の一覧と、その翻訳後の用語を一覧表示し、既存の翻訳リストの編集、新しい言語の追加を行い、言語ファイルを出力できるツールの開発を計画中である。

このことにより、デジタル・アーカイブを活用する各プロジェクトチームは、システム側の都合で実際のデータ内容に適した項目名を使えないという状況に悩まされることなく、各研究分野の事情に細やかに対応させた検索システム上の項目名や指示フレーズなどのカスタマイズを行うことができる。

さらに、本システムを用いれば、このような調整を、各研究者の専門分野に関する知識をシステム管理者に理解させる手間をかけることなく、文系研

究者自身が独自に各々の専門知識を反映させ、設定を行うことが可能となる。

この方式をすべてのDA作成者が利用できるようなことになって、専門性の高い個別DBが多数作成され、今後、分野別用語のサンプルを多数収集することができれば、ダブリンコアやIIIFのようなフォーマットを利用してメタデータのやり取りを行う際に、各言語・各分野でそれぞれの項目をどのような用語に置き換えればよいかという候補を提供することができる。

さらには、分野別・言語別用語DB、およびその情報を提供するAPIサービスとして公開することもできよう。

新規DBの開発時や、IIIFなどを連携的に利用する時、メタデータ項目を実際のユーザーインターフェースに当てはめる際にどのような表示名称が適切なのかを示唆することで、広くDAの普及に貢献できるのではないかと考えている。

〔注釈〕

- 1) 赤間亮「専門分野別研究資源ポータルデータベースと相互リンクによるユーザビリティ」(アートドキュメンテーション学会216年次大会予稿集, 2016.6, pp. 38-43.
- 2) 収録する資料の90%以上が浮世絵であったので、現在もそのままの名称となっているが、浮世絵以外のジャンルの資源も含まれるようになってきたため、名称の変更が必要である。
ARC 浮世絵ポータルデータベース, https://www.dh-jac.net/db/nishikie/search_portal.php, (Accessed 2021-02-10)
- 3) 口絵ポータルデータベース, https://www.dh-jac.net/db/nishikie/search_kuchie.php, (Accessed 2021-02-10)