

データを記録・保存する

中村 覚

1. はじめに

歴史資料に対する科学研究は、分析手法の高度化により、この 10 年間で大幅に発展している。一方、これらの分析に使用されたデータは個々の研究者によって保存・管理されることが多く、分析方法および結果を共有・再現することが一般的には困難である。データの共有と再利用を支援することにより、データの出典の明確化と、研究結果の再現性の確保につなげることができる。またオープンデータ¹⁾の増加により、科学研究のさらなる発展に寄与することができる。

このような背景にもとづき、東京大学史料編纂所では caid（ケイド、classification and annotation for image data）というデータ管理ツールを開発している。画像およびそれに関するデータを簡易に、効率的に、柔軟に管理可能な機能を提供することにより、研究データの蓄積と共有、歴史学におけるオープンサイエンスの推進を目指す。本章では本ツールについて紹介する。

2. caid について

図 1 に caid の概要を示す。caid はオンライン / オフラインで動作可能なデスクトップアプリケーションであり、インターネット接続が困難な調査先などでも利用することができる。PC 内の特定のフォルダを監視し、当該フォルダに新しい画像が追加されると、caid にも自動登録される。この機能を用いることにより、図 1 左に示すように、調査先での顕微鏡を用いた画像撮

影などと並行したデータ蓄積を行うことができる。caid に登録されたデータは、後述する機能を用いることで、画像データの分類や、画像内の情報に対するアノテーション付与[★]などを行うことができる。このようにして蓄積されたデータを分析し、また外部のシステムと連携させることで、料紙研究におけるデータ共有を支援する。

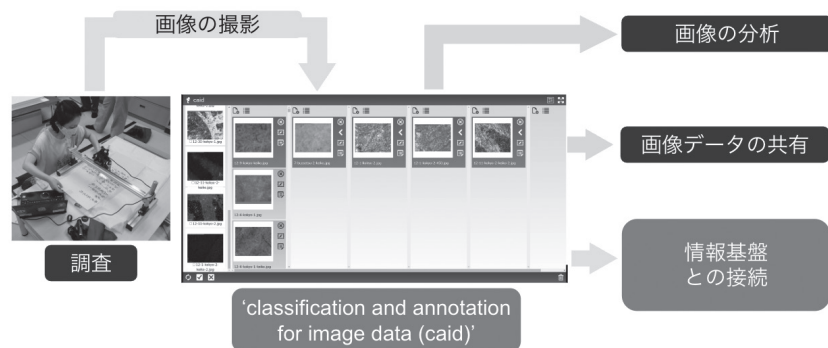


図1 caidの概要

以下、caid の使用方法の詳細について説明する。

3. 監視フォルダの設定

上述した通り、本アプリケーションでは PC 内の監視対象のフォルダを設定することで、各種機材で撮影した画像の自動登録を行う。複数のフォルダを設定できるようにしており、調査目的や機材の違いに応じて、それらを切り替えられるようにしている。

4. アプリケーションの操作

本アプリケーションの利用には、ウェブブラウザを用いる。なお、一般的にはウェブブラウザはインターネット上のコンテンツの閲覧に使用されるが、今回は PC 上で起動したアプリケーションの閲覧・操作のために使用する。

そのため、上述した通り、オフラインで使用可能なアプリケーションとなっている。

caid におけるデータ管理の全体像を図2に示す。「メイン画面」では資料の重層的な情報を階層構造によって管理する。各階層では、データの一覧・編集機能（「画像データ一覧」）、編集対象とするメタデータの設定画面（設定フォーム）[★]などを提供する。これらの機能を用いることにより、調査・分析対象に応じた多様なデータの分類と蓄積を支援する。

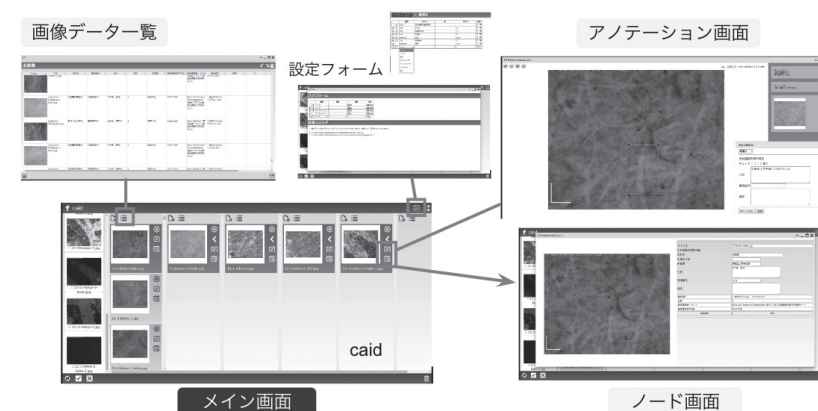


図2 caid におけるデータ管理の例

以下、各機能について説明する。

5. 画像とノード

本アプリケーションでは、監視フォルダに保存されたファイルを画像、それらを分類するための枠をノードと呼ぶ。

図3に示すように、📁を押して、空のノードを作成する。ノードの画像部分に画像をドロップすると画像が登録される。枠の部分にドロップすると、ノードの関連画像として登録され、一つ下の階層にノードが追加される。

このように画像どうしを階層的に保存することにより、資料の重層的な情報を管理することができる。



図3 画像とノード

6. メイン画面

メイン画面を図4に示す。画面左部には未分類画像（画像監視フォルダに追加された画像ファイル）が登録された日時の降順で表示される。画面右部にはノードが階層ごとに表示される。

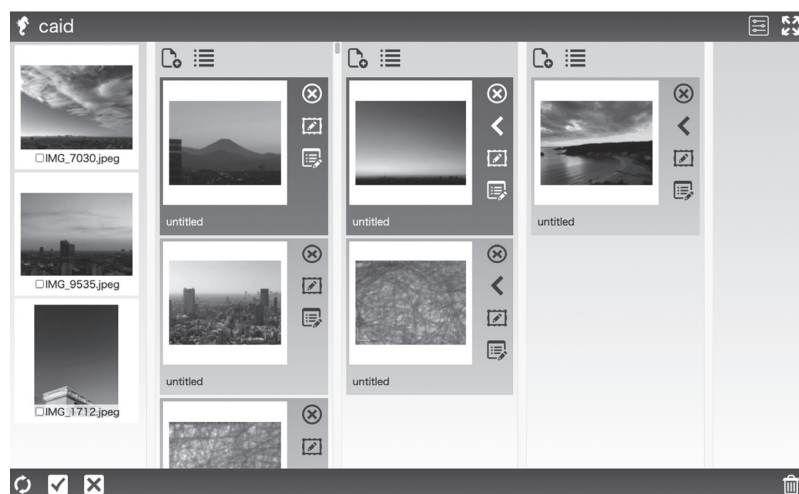


図4 メイン画面

7. データの編集

caidは図5に示す、画像に対するデータ編集機能を提供する。図5左に示すメイン画面から特定の画像データを選択し、図右に示すアノテーション付与画面に遷移する。以下、アノテーションの編集画面、ノード情報の編集画面について説明する。

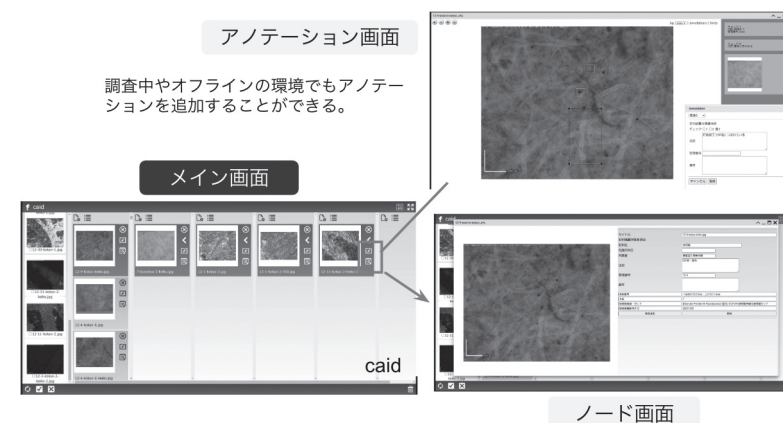


図5 データの編集

8. アノテーション編集画面

アノテーション編集画面を図6に示す。画像の任意の矩形箇所に対して、アノテーションを付与することができる。これにより、調査中および調査後のメモなど、テキストデータによるアノテーションを付与することができる。アノテーション付与の項目については、分析者・利用者がカスタマイズ可能な機能を提供することで、さまざまなニーズに応じたデータ蓄積を支援する。この設定方法について後述する。



図6 アノテーション編集画面

また本アプリケーションの特徴として、関連する画像もアノテーションとして付与することができる。これにより、史料画像の特定の箇所に対して、顕微鏡で撮影した画像を紐づける、といった画像管理を実現する。

9. ノード情報編集画面

各ノードについては、図7に示すノード情報編集画面から、そのメタデータを編集することができる。この画面で編集可能なメタデータ項目についても、後述する設定画面において、階層ごとにカスタマイズすることができる。

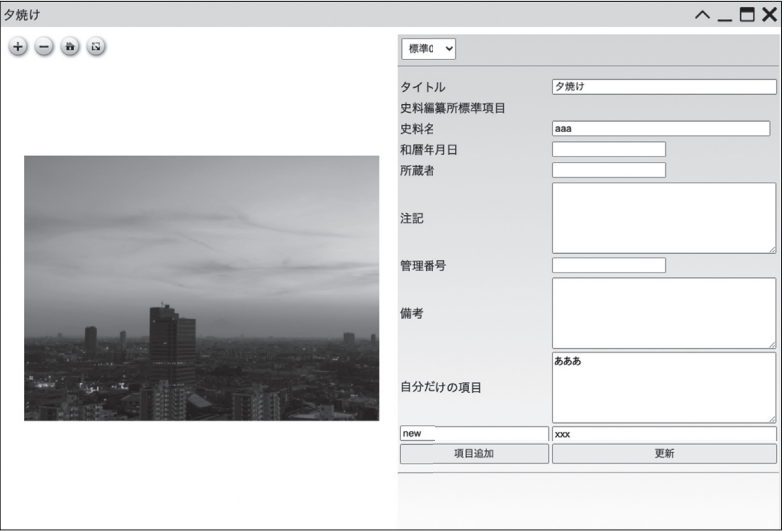


図7 ノード情報編集画面

10. 画像一覧画面

各階層における画像およびそのメタデータは、図8のように一覧表示することもできる。列の項目をクリックすると、昇順・降順で並び替えを行うことができる。また各項目の右側の▼をクリックすることで、フィルタリングを行う。

さらに表中のデータをダブルクリックすることにより、データの編集を行うことができる。本画面を用いることにより、複数の画像に対するデータの

一括修正などを容易に行うことができる。



図8 画像一覧画面

11. 設定画面

図9に示すように、入力フォームを柔軟に設定することができる。フォームを追加し、種別・階層・名称を設定する。フォームの種別として、アノテーション編集に用いるものと、ノード情報の編集に用いるものをそれぞれ設定できるようにしている。また左側の矢印をクリックし、表示される順番を変



※ 監視フォルダはタスクトレイのアイコンから「Set Watch Folders」を選択して、変更することができます。

1. /Volumes/SSD 1TB/Dropbox/Machines/macmini8/Development/workspace.201909/seahorse/./data/app/watch
2. /Volumes/SSD 1TB/Dropbox/Machines/macmini8/Development/tmp/agent/build/classes/init_data

図9 設定画面

更することができる。さらに、名称をクリックすると、後述するフォーム内容の編集画面に遷移する。

12. フォーム編集画面

入力フォームの内容についても、図 10 に示すように、柔軟に設定することができる。フォーム部品を追加し、種別・ラベル・値・サイズを設定する。また左側の矢印をクリックし、表示される順番を変更することができる。

	種別	ラベル	値	サイズ	操作
	label	史料編纂所標準項目			✕
↑ ↓	text	史料名		40	✕
↑ ↓	text	和暦年月日		20	✕
↑ ↓	text	所蔵者		20	✕
↑ ↓	textarea	注記		40x5	✕
↑ ↓	text	管理番号		20	✕
↑ ↓	textarea	備考		40x5	✕
↑	textarea	自分だけの項目			✕

クリア ラベル ▼

図 10 フォーム編集画面

フォーム部品の種別として、以下のものを用意している。

- ・ラベル：そのまま表示される。値・サイズは使用しない。
- ・テキスト：値は初期値となる。サイズは入力フィールドの大きさとなる。
- ・数値：テキストと同様だが、数値のみが入力できるフィールドとなる。
- ・テキストエリア：テキストと同様だが、サイズは横 x 縦（例：「40x5」とすると 40 桁が 5 行）となる。
- ・チェックボックス：設定する値を半角カンマで区切って指定する。サイズは使用しない。入力時には複数の項目を選択することができる。
- ・ラジオボタン：設定する値を半角カンマで区切って指定する。サイズは使用しない。入力時には一つだけ項目を選択することができる。

- ・選択：ラジオボタンと同じだが、表示形式がプルダウンとなる。

13. データのエクスポート

caid を用いて入力されたすべてのデータは、ユーザの PC 内に保存される。このデータはインポートおよびエクスポートが可能である。この機能により、ある調査データとほかの歴史資料の分析結果の比較などを行うことができる。同じデータに対する複数ユーザーによる解釈や注釈の違いなどに対して、caid を用いた比較・分析を支援する。また USB メモリなどにコピーすることで、データを保全することができる。

さらに図 11 に示すように、PC に蓄積されたデータを外部の情報システムと連携させることにより、より発展的なデータ共有および分析が可能となる。特に AI を活用した画像解析技術の発展は著しく、蓄積および共有されたデータに対する応用が期待できる。東京大学史料編纂所では、多様なデータを管理・蓄積する「史料情報統合管理システム」および AI によるデータ活用などを目的とした「データ駆動型歴史情報研究基盤の構築 [2]」プロジェクトを進めている。今後、これらのシステムおよびプロジェクトとの連携を図り、料紙研究のさらなる発展を目指す。

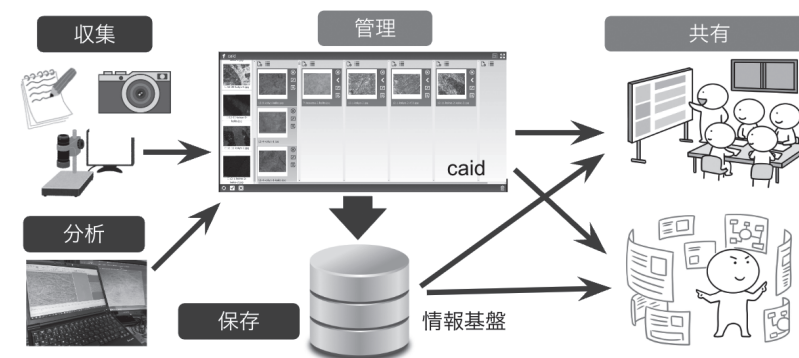


図 11 caid を用いた研究データの共有と活用

14. まとめ

2022年9月時点において、caidは限られた関係者のみで利用可能なツールであるが、将来的にはオープンソースソフトウェアとしての公開も見据えて現在開発を進めている。汎用的に利用可能なツールを目指すことにより、本研究成果の社会還元も視野に入れて活動を継続する。これにより、国内外の料紙研究の発展にも寄与していきたい。

謝辞

本研究はJSPS科研費20H00010の助成を受けたものである。また、開発にご協力いただいた株式会社フォーイチ様に深く感謝いたします。

1 Open Knowledge International [1] による定義は以下である。

オープンデータとは、自由に使えて再利用もでき、かつ誰でも再配布できるようなデータのことだ。従うべき決まりは、せいぜい「作者のクレジットを残す」あるいは「同じ条件で配布する」程度である。

参考文献

- » [1] Open Knowledge International, Open Data Handbook, <http://opendatahandbook.org/guide/ja/what-is-open-data/>, (参照 2022-11-15).
- » [2] データ駆動型歴史情報研究基盤の構築 | 東京大学, https://www.u-tokyo.ac.jp/adm/fsi/ja/projects/sdgs/project_00214.html, (参照 2022-11-15).

史料の形態データと 内容データを関連付ける

——複合的史料研究推進に向けた史料情報統合——

山田太造

1. はじめに

東京大学史料編纂所（以下、史料編纂所）は、日本史史料の調査・研究を行うことで、『大日本史料』『大日本古文書』『大日本古記録』といった日本史研究に欠かせない基幹的史料集の編纂を行い、冊子体としての出版やデータベースシステムにより、その成果を公開・共有してきた。史料の調査は発足当初から継続して行っており、調査では史料収集も行ってきた。以前は写本作成やマイクロカメラ撮影による収集だったが、いまではデジタルカメラによる収集を行っており、130年を超えた収集により、膨大な日本史史料コレクションとして形成するに至った。

複製史料を用いて、史料に記述されている内容をもとに、史料の様式論・機能論・伝来論、さらには、登場する人物・地名・時間などを軸として史料の分類を行うなど、批判的な史料分析を行ってきた。いわば、史料研究の基本である。1997年、史料編纂所附置センターとして画像史料解析センターが発足した。これを契機に、絵図・肖像画・錦絵・古写真といった画像史料を対象とし、さらには、古文書にある花押や文字の形状そのものをも画像史料としてとらえ、文字データのみならず画像データも対象として史料データの解析を行うことで、史料研究の進展を行ってきた。