

オープンサイエンスと 手書きプロセスデータ流通基盤の構築

梶田 将司^{1,2} 青木 学聡^{1,2} 喜多 一^{2,3}

概要: 我々は、人が文字や図を手で書く動作は人の思考過程に最も近く、かつ複雑な情報の表現を可能にする行為であるとの立場から、手書きプロセスデータを収集・蓄積・活用するための手書きプロセスデータ流通基盤とそのアプリケーションに関する研究開発を行っている。本報告では、カリフォルニアデジタルライブラリが開発した研究データマネジメントに関する成熟度モデルを参照しながら、京都大学における研究データマネジメント支援環境の利用を前提とした手書きプロセスデータ流通基盤の構築について議論する。これにより、オープンサイエンスを推進するためのユーザ側・サービス提供側に求められる要件について検討する。

キーワード: オープンサイエンス, 研究データマネジメント, 学習支援, 手書き, オープンスタンダード, オープンソース, 次世代デジタル学習環境

Open Science and Development of Handwriting Process Data Sharing Platform

SHOJI KAJITA^{1,2} TAKAAKI AOKI^{1,2} HAJIME KITA^{2,3}

Abstract: We have been studying about a widely sharable infrastructure for handwriting process data and its applications, with consideration that we think that handwriting actions for writing letters and drawing figures reflect much of the intellectual activity in the human brain. In this report, we discuss the current project status in terms of Research Data Management (RDM) through the evaluation using California Digital Library RDM Maturity Model, in order to envision what we should do in the context of Open Science as a researcher and the RDM service-provider at Kyoto University.

Keywords: Open Science, Research Data Management, Learning Management, Handwriting, Open Standard, Open Source, Next-generation of Digital Learning Environment

1. はじめに

高等教育機関では、「学生がどう学んでいるか」「教員がどう教えているか」が、データとして大量に日々蓄積されるようになり、「ラーニングアナリティクス」という言葉

で語られるデータに基づく教育学習活動の改善が可能になってきている。この流れの中で、欧米の大学では、次世代デジタル学習環境 (Next-generation of Digital Learning Environment, NGDLE) の議論が行われながら、研究レベルから実践レベルまでラーニングアナリティクスに関する様々な取り組みが幅広く行われている。

NGDLE の検討においては、学生が扱う情報端末として、タッチ入力、ペン入力、音声入力等が標準的に使えるデバイスが普及してきていることから、それらで取得可能な手書きプロセスデータとそのコンテキスト情報を「人間の思考過程を把握する有用なデータ」として教育学習活動で利活

¹ 京都大学情報環境機構 IT 企画室
IT Planning Office, Institute for Information Management and Communication, Kyoto University

² 京都大学学術情報メディアセンター
Academic Center for Computing and Media Studies, Kyoto University

³ 京都大学国際高等教育院
Institute for Liberal Arts and Sciences, Kyoto University

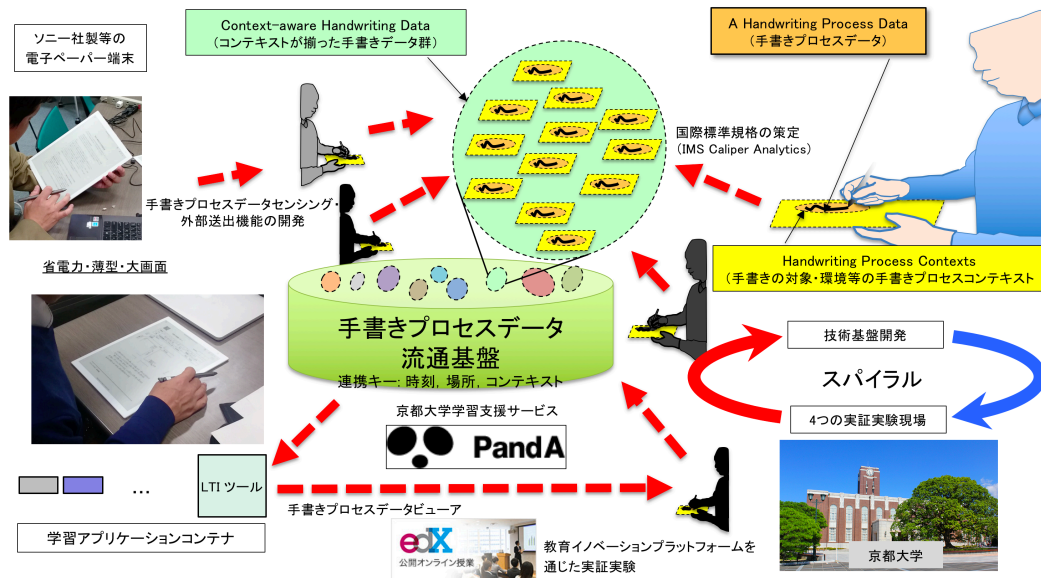


図 1 手書きプロセスデータ流通基盤の構築プロジェクトの概要 [4].

用することが期待されている [1]. しかしながら、手書きプロセスデータをセンシングし、外部システムに送出・蓄積・共有するための技術基盤がない。そこで我々は、ラーニングアナリティクスの標準規格を策定している IMS Global Learning Consortium の Caliper Analytics Workgroup[2] において、手書き入力デバイスを有するベンダや LMS ベンダとともに、「手書きプロセスデータ」およびそれが生成された状況についての「手書きプロセスコンテキスト」を記録するための国際標準規格の策定を行うことで、システムや大学の枠を超えて、コンテキスト手書きプロセスデータを蓄積し、大規模な研究データとしてビッグデータ化することにより、手書きプロセスデータ流通基盤の構築を目指している [3], [4].

一方、このような研究データに関する新たな潮流として、研究公正のための研究データの長期保管・検証（論文公開後の 10 年間）や、OECD・G8 での各国政府によるオープンサイエンス推進（国内は内閣府）、国際潮流になりつつある研究助成機関によるデータマネジメント計画（Data Management Plan; DMP）の提示要求への国内への波及の可能性等、学術分野に関係なく研究環境の情報化に対する研究リテラシ能力の強化が求められている。

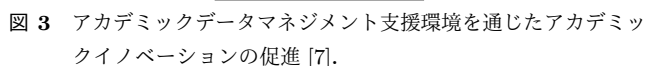
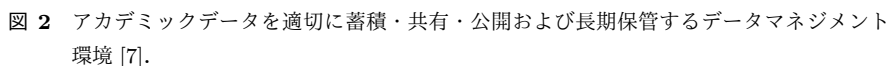
しかしながら、蛸壺化した研究室における情報環境の急速な変化への対応能力の低下や、定常的な研究予算の激減等、研究者や研究グループが上述の新たな潮流に個別に対応することはできなくなっているため、各大学は、全学的観点から研究基盤の抜本的な見直しと再構築が必要になっている。例えば、欧米の大学では、米国国立科学財団・国立衛生研究所等の研究資金配分機関に研究費を申請する際、配分機関が求める研究データマネジメントに関する要件を満たす計画が容易に作成できるよう、各大学の支援環

境が整備・強化されている（例えば、Purdue University の PURR[5]）。我が国でも、オープンサイエンスが内閣府主導で進められる中、競争的資金における研究データマネジメントに関する方針策定の必要性が議論されており、近い将来、科研費等の競争的資金申請時に DMP の策定が求められる、研究データマネジメントに関する研究者の興味関心が急速に高まることが予想される。

そこで、本報告では、California Digital Library が開発した研究データマネジメントに関する成熟度モデル [6] を参照しながら、京都大学における研究データマネジメントに関係する支援環境の利用を前提とした手書きプロセスデータ流通基盤の構築について議論する。これにより、オープンサイエンスを推進するための研究者側・大学側に求められる要件について検討する。

2. 京都大学における研究データマネジメント環境整備に関する取り組み

京都大学では、理学研究科附属地磁気世界資料解析センター主催の「オープンサイエンスデータ推進ワークショップ」が 2015 年度より 6 回開催され、これを契機に、オープンサイエンスや研究公正のためのデータマネジメントに興味関心を持つ有志が集い、総長裁量経費や全学経費の支援を受けながら、京都大学における研究データに関わる広範な議論や学内外連携が進み始めた。その成果として、研究データに係る状況をボトムアップでかつ全学的に調査研究するため、京都大学学際融合教育研究センターにアカデミックデータ・イノベーションユニット（通称「葛ユニット」）。この名前には研究データマネジメントを通じて研究データを屑（くず）としてゴミ箱に破棄するのではなく、葛（くず）まんじゅうのように見栄えも味も確かなものに



まず、(1)については、国内は国立情報学研究所オープンサイエンス基盤研究センターや大学 ICT 推進協議会研究データマネジメント部会*1等、海外は California Digital Library (CDL)*2等の関係者と連携しながら、国内外の既存あるいは計画中的の支援サービスとの連携を前提とした全学的な研究データマネジメント環境の整備のあり方を模索

*2 UC バークレイ校, UC ロサンゼルス校, UC サンディエゴ校等,
10 のキャンパスからなる University of California System の
本部組織

一方, (2) については, 葛ユニットが所属する京都大学学際融合教育研究センターの支援 (2018 年度分野横断プラットフォーム構築事業「研究大学強化促進事業『百家争鳴』プログラム」) を受けながら, ハンズオンワークショップを開催 (2018 年 10 月 6 日, 2019 年 2 月 28 日) し, 「京都大学研究データマップ」と「京都大学研究データマネジメントルブリック」の開発を CDL 研究データマネジメントルブリック (表 1 参照) をベースに進めている. これにより, 研究者それぞれの研究データマネジメント状況が向上するための方策や分野融合・連携の可能性を明らかにすることを目標にしている [8].

手書きプロセスデータ流通基盤の構築あたって、我々が想定しているユースケースは、手書きプロセスデータの取得・蓄積・流通の観点から想定したもの(図4参照)と手書きプロセスデータの利活用の観点から想定したもの(図5参照)からなる[3]:

UC1-1 学習者によるノートテイキング、課題への回答、教授者による添削等により生成された手書きプロセスデータを、手書きストロークデータとそのコンテキストデータ（手書き対象物や手書きに関するメタ情報）としてセンシングし、それぞれを独立に、あるいは、何らかの方法でパッケージ化した上で LRS に蓄積する。

3

表 1 CDL 研究データマネジメントルーブリック (日本語訳)[8].

	その都度	1度のみ	常に見直す	再利用のために最適化
データマネージメント計画	データを入手した段階で、自分なりに考える。標準化や文書化はしていない。	最初に計画を作成するが、通常は、途中で見直すことはしない。	詳細な計画を作成し、プロジェクト期間を通じ参照と見直しを行う。	自分あるいは他人が将来スムーズにデータを利用できるように計画している。
データの体系的整理	一貫した手法に従っていない。そのため、しばしば探し出すのに時間を要する。	手法はあるが、それを適用するのは、プロジェクト終了時のみ。	先を見越した手法を導入しているが、必ずしも標準化されてはいない。	他人が自分でデータを検索、内容を理解し、利用できる。
データの保存とバックアップ	作業中にどのデータが重要であるかを決め、通常はこれを特定の1か所に保存する。	どのデータを保存すべきかを理解している。損失のリスクを緩和するため、作業後にバックアップをとる。	作業中でも定期的にデータを保存するシステムを用いている。複数のバックアップを保持している。	再利用する機会を最大限生かせるよう考慮した方法と場所にデータを保存している。
データを解析可能な状態にする	標準化もしくは十分に文書化された手続きはない。	解析に適した形式を検討し整備するが、事例ごとに異なる方法をとっている。	データ処理方法は標準化され、かつ十分に文書化されている。	将来も自分自身及び第3者が利用しやすい方法で、データ処理を行っている。
データ解析と成果の取扱い	解析や実験がどのような手順やパラメータにより実施されたか確認するために、これらをやり直すことが多々ある。	解析作業終了後、パラメータの詳細や解析手順、プロトコルを文書化している。	データ解析の際、決められた方法で解析のワークフローや判断のプロセスの詳細をいつも記録している。	自身が行った解析のワークフローや判断プロセスの詳細が他人によっても実行できることが確認できている。
データの共有と公開	研究成果を共有しているが、その根拠となるデータについては共有できていない。	リクエストがあった場合にのみ研究データの共有を行っている。	自身の研究成果や根拠となるデータは、第三者が利用できる形式で共有できる。	優れたデータマネジメントを実践し、自身のデータをいつでもだれとでも効率的に共有できる。

行為者に関する情報を有する。

UC1-3 ストロークの状態として、描画済み、および、消去済みがある。

UC1-4 ストロークに対する操作として移動、複写、拡大・縮小、消去ができる。

ユースケース 2:

UC2-1 LRS に蓄積された手書きプロセスデータを LRS 間で流通・共有する。

ユースケース 3:

UC3-1 対面授業での討論等、教員・ティーチングアシスタント・学生で構成される複数のユーザが同時に手書きプロセスデータを生成し活用する同期型アプリケーションで利用する。アプリケーションとしては、共有ホワイトボードに教員やティーチングアシスタント、学生が書き込むものを想定している。

UC3-2 手書きストロークデータは手書き端末からアプリケーションに送出され利用される。

UC3-3 アプリケーションでの利用状況とともに手書きプロセスデータは、LRS に送信・蓄積される。

ユースケース 4:

UC4-1 教員・ティーチングアシスタント・学生で構成される複数のユーザが異なる時間に手書きプロセスデータを生成し活用する非同期型アプリケーションで利用する。アプリケーションとしては、一般的な LMS が有する課題ツールに手書きで書いた課題レポートを学生が提出したり、学生が提出した課題レポートに手書きで採点・コメントを記入するものを想定している。

UC4-2 オフライン状態で利用するアプリケーションでの手書きプロセスデータの生成も想定する。

UC4-3 アプリケーションでの利用状況とともに手書きプロセスデータは、LRS に送信・蓄積される。

UC4-4 提出時はファイルをアップロードすることにより提出することが一般的であり、非同期型アプリケーションやオフラインアプリケーション間での手書きプロセスデータのやりとりは何らかの方法でパッケージ化されている方が望ましい。

4. CDL 研究データマネジメント成熟度モデルに沿った評価

前節でまとめた京都大学における研究データマネジメント環境および手書きプロセスデータのコアユースケースを、CDL の研究データマネジメント成熟度モデルの各項目に沿って評価を試みる。

4.1 データマネジメント計画

我々は、提案している Handwriting Profile の妥当性を検証するため、のべ 47 名の京都大学学生による手書きプロセスデータ収集実験を行い、電気回路、行列証明、一般教養などの課題についての擬似課題演習による手書きプロセスデータを収集した (図 6 参照)。これを用いて、公開・共有を前提に最もよい方法を検討している「常に見直す」状態にあるが、「将来スムーズにデータが利用できるように計画している」とも言えるので、「再利用のために最適化」に向けて具体化を行っている段階と言える。

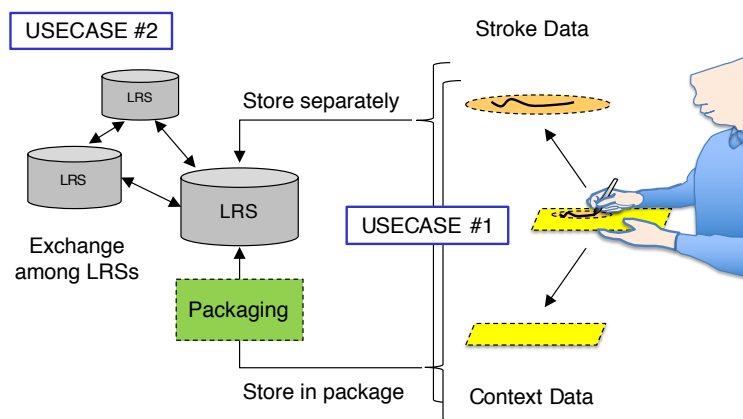


図 4 手書きプロセスデータの取得・蓄積・流通の観点から想定するコアユースケース [3].

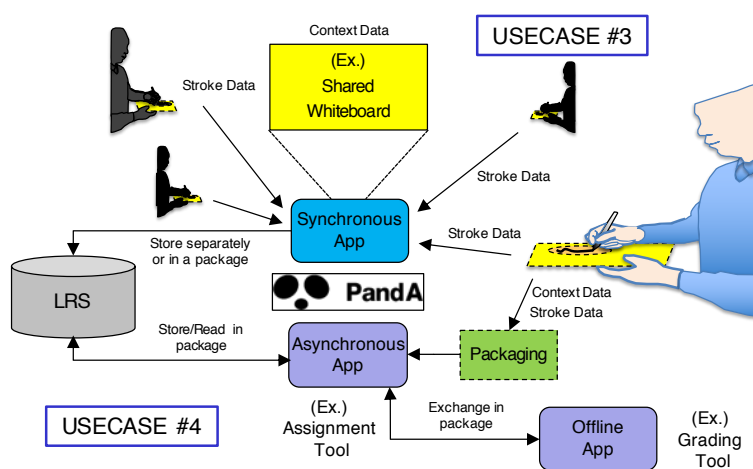


図 5 手書きプロセスデータの活用の観点から想定するコアユースケース [3].

4.2 データの体系的整理

すでに文献 [3] で提案したとおり、IMS Global Consortium Caliper Workgroup での標準化に向けて準備を行っていることから、「再利用のために最適化」に向けて具体化を行っている段階と言えよう。

4.3 データの保存とバックアップ

現在、国立情報学研究所の学認 RDM [9] の利用を試行しており、機関利用している Google Drive または学内の Nextcloud をアタッチして利用する予定であることから、「再利用のために最適化」に向けて具体化を行っている段階と言えよう。

4.4 データを解析可能な状態にする

データ処理方法については、提案しようとしている Handwriting Profile に従って実際のデータを記述するところまではできていないが、Handwriting Profile に従った記述の改善を進めていくことで、誰もが利用しやすい「再利用のために最適化」された段階に持っていく予定である。

4.5 データ解析と成果の取り扱い

取得した手書きプロセスデータの解析やその成果の取り扱いについては、現段階では全く検討できていない事項であるので、「その都度」の段階にあると言えよう。今後、手書きプロセスデータの利用者コミュニティで情報共有がしやすくなるよう、標準的な書式に従ってデータ解析やその成果が共有できる仕組みを検討する必要がある。

4.6 データの共有と公開

研究成果の共有についても全く検討できていない事項であるので、「その都度」の段階にあると言えよう。京都大学では附属図書館が機関リポジトリを通じた文献情報だけでなく、研究データの登録業務も開始している [10]。また、情報環境機構がデータ保存サービス [11] を試行している。京都大学関係者であればこれらの利用も可能であるが、手書きプロセスデータを取り扱うオープンサイエンスコミュニティづくりを指向するのであれば、学認 RDM 等、機関非依存のサービスを利用した方がよいと考えている。

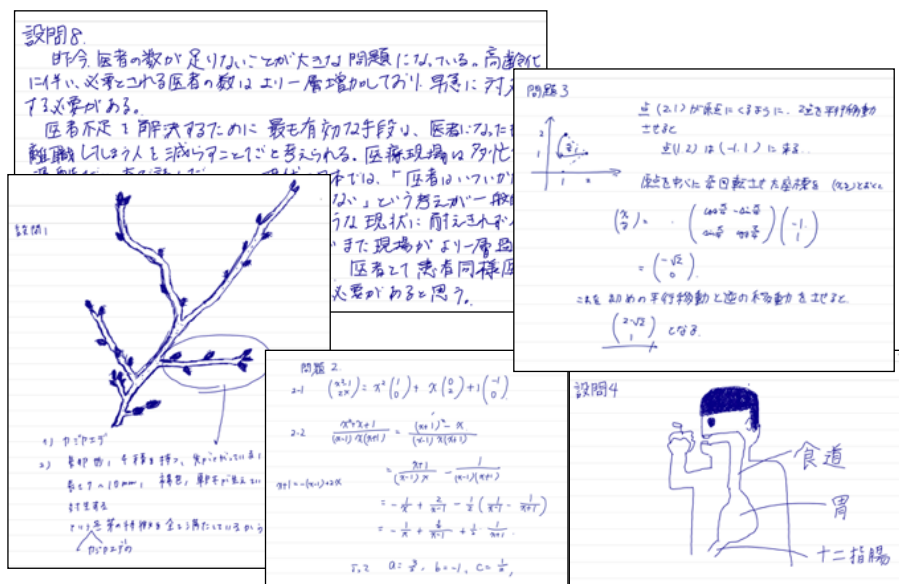


図 6 収集した手書きプロセスデータの例 [4].

5. まとめと今後の課題

本報告では、California Digital Library が開発した研究データマネジメントに関する成熟度モデルを参照しながら、京都大学における研究データマネジメントに関係する支援環境の利用を前提とした手書きプロセスデータ流通基盤の構築について議論した。その結果、IMS Caliper による手書きプロセスデータの標準化の取り組みは RDM 成熟度を上げるためには極めて重要であるとともに、データ解析やその成果の方法についても標準的な方策を検討する必要があるようになった。

すでに述べたとおり、手書きプロセスデータを入力端末から取得する方法の技術基盤がなく、教育学習活動に係る手書きプロセスデータは集積されていない。今後は、IMS Caliper Workgroup での標準化を進めながら、我々の取り組みを核とした手書きプロセスデータに関する国際的なオープンサイエンスコミュニティを形成したいと考えている。

謝辞

本研究は、JSPS 科研費 17H06288 の助成を受けたものである。また、本研究の一部は、株式会社 Linfiny Japan の協力の下、富士通クライアントコンピューティング株式会社との共同研究により実施されている。この場をお借りして感謝致します。

参考文献

[1] 岩恵裕夢, 永井孝幸, “ペンタブレットを用いた集団学習環境での筆記遅れ学習者リアルタイム検知システムの試作”, 情報処理学会研究報告教育学習支援情報システム (CLE), Vol. 2019-CLE-27, No. 11, pp. 1-8 (2019 年 3

月 13 日)
[2] IMS Global Learning Consortium Caliper Analytics, <https://www.imsglobal.org/activity/caliper> (2019 年 8 月 19 日アクセス)
[3] 梶田将司, 喜多一 “IMS Caliper プロファイルによる手書きプロセスデータの記述”, 情報処理学会研究報告教育学習支援情報システム (CLE), Vol. 2019-CLE-28, No. 5, pp. 1-6 (2019 年 5 月 25 日)
[4] 梶田将司, “手書きプロセスデータの国際標準規格の策定に向けて”, 情報処理学会研究報告教育学習支援情報システム (CLE), Vol. 2018-CLE-26, No. 2, pp. 1-5 (2018 年 11 月 30 日)
[5] “Purdue University Research Repository”, <https://purrr.purdue.edu> (2019 年 8 月 19 日アクセス)
[6] John A Borghi, Stephen Abrams, Daniella Lowenberg, Stephanie Simms and John Chodacki, “Support Your Data: A Research Data Management Guide for Researchers”, Research Ideas and Outcomes, Vol.4, pp. e26439, doi:10.3897/rio.4.e26439 (May 2018)
[7] 梶田将司, 青木学聡, “研究ライフサイクルに沿ったアカデミックデータマネジメント支援環境による研究基盤強化”, 情報処理学会研究報告インターネットと運用技術 (IOT), Vol.2017-IOT-38, No.4, pp.1-6, 徳島大学, 徳島, 2017 年 6 月 24 日
[8] 青木学聡, 梶田将司, 元木環, 家森俊彦, 川口朋子, “研究データマネジメントループによる自己評価と共通認識の醸成”, DEIM Forum 2019 C2-1, 第 11 回データ工学と情報マネジメントに関するフォーラム (第 17 回日本データベース学会年次大会), ホテルオークラ JR ハウステンボス, 佐世保, 長崎, 2019 年 3 月 4 日~6 日
[9] 国立情報学研究所 Gakunin RDM, <https://rcos.nii.ac.jp/service/rdm/> (2019 年 8 月 19 日アクセス)
[10] 京都大学附属図書館, “公開支援サービス”, <https://www.kulib.kyoto-u.ac.jp/researchdata/1380893> (2019 年 8 月 19 日アクセス)
[11] 京都大学情報環境機構, “データ保存サービス”, <http://www.iimc.kyoto-u.ac.jp/ja/services/archive/system/> (2019 年 8 月 19 日アクセス)