

大学における日々の研究活動の保存・活用に関する一検討

松浦 健二^{1,a)} 佐野 雅彦¹ 大平 健司¹ 谷岡 広樹¹ 上田 哲史¹ 西村 友基¹

概要：学術論文のオープンアクセス化に加えて、オープンサイエンスの流れにより、論文以外にもその共有可能なデータの流通が促進されている。一方で、研究の説明責任や追跡性・再生可能性などを目的とする、研究データの一定期間保存が義務付けられるようになっている。ただし、その実装に関する具体的事例の収集やそれらの整理に関してはまだ十分議論されておらず、特に高等教育機関における研究では、学生教育としての側面からも、保管に関する運用設計、技術実装などの方法論が必要であり、これは事例を通じた知見を共有していく必要がある。本稿では、このような背景から、大学等の機関における研究資源の保存と活用方法について考察し、著者らの所属機関での現状からこの実装に関する考察を行う。

Discussion on Management and Application of Daily Academic Activity at a University

MATSUURA KENJI^{1,a)} SANO MASAHIKO¹ OHIRA KENJI¹ TANIOKA HIROKI¹ UETA TETSUSHI¹
NISHIMURA YUKI¹

1. はじめに

実世界、仮想世界の様々な事象に対する電子化が行われ、そのインターネットを介したオープン化が進んでいる。オープンソース、オープンデータ、オープンエジュケーション、オープンサイエンス等の各方面で広がりを見せている。これは、デジタル化、データ化、情報化、知識化といった段階的な電子化事象の取り扱いを考慮しても、いずれかの段階に特化した動きではなく、いずれの段階にも適用される活動である。すなわち、加工する前の一次(生)データも、高度に加工された二次データであっても、それらがオープン化に適した表現がなされれば、その表現された対象を解釈する人間やソフトウェアが活用できる可能性を検討し、積極的にオープンにできる。オープン化の概念自体は新しいものではなく、様々な文脈で実施されてきたが、ここ数年特に顕著に前述のような文脈で取り上げられるようになった。

例えば“オープンデータ”の概念においては、自治体の

データや地域コミュニティ等の多様な情報提供者からのデータが、自由に利用できるようなインターネット上で提供される活動として国内外で注目され、ビジネスモデルの一つとなりつつある。それまで、暗黙に非公開と思われていた様々な公開可能データを、自治体を中心に積極的に公開されることで、様々なアプリケーション開発も行われる。特徴的なアプリケーションには、地図と関連づけられた交通情報の提供やゴミ収集に関するリアルタイム情報提供などもあり、日々の生活で使われる利便性の高い分野が先行している。このように、データをオープンにすることによって、市民参加型のアプリケーション開発や市民のためのソフトウェア開発といったビジネス以外の効能も見られる。

一方で、学問・学術の分野においては、知的生産活動、とりわけ研究に関する活動は、インターネットを通じて公開・共有されることにより、学術発展や実践適用などの貢献が期待される。このような活動は、研究主体や研究組織の主宰者が、従前想定していなかった応用や適用範囲の拡大などに繋がることも期待される(図1参照)。ただし、大学における研究活動として企業等におけるそれと異なるのは、中間生産物を含む研究活動成果が、教員と学生の間で

¹ 徳島大学
Minamijosanjima2-1, Tokushima, 770-8605, Japan
^{a)} ma2@tokushima-u.ac.jp

の協働の場が生み出す生産物となる場合、そこには教育という側面が少なからず含まれるということである。

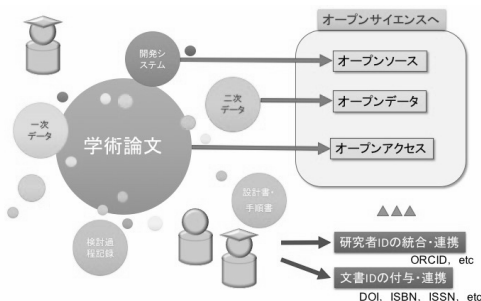


図 1 研究活動とオープンサイエンス化の概念
Fig. 1 Research activities and open science

さらには、研究活動に対する信頼性や再現性の確保など、後に真正性を保証するための仕組みも重要である。日本学術会議では、文部科学省からの審議依頼に対して、答申「科学研究における健全性の向上について」[1]の中で、慎重に述べられており、その基本精神については電子的な方法での研究記録は従来の製本されたノートと同じと位置づけているが、「電子データ化されているか、紙媒体等の資料かによって、扱いが異なる」とも述べられている。

そこで、本稿では大学における研究データの保存・活用に関する特徴を論じながら、研究活動記録の保存と活用に関する検討を行う。特に、流通することを考慮したメタデータ設計、承認やアクセス制御といった技術的側面、およびポートフォリオ/CMSといった実装的側面に注意しながらの設計指針について述べる。

2. 研究データの保存・活用

2.1 研究データの保存要件

オープンサイエンスの活動を通じた社会還元的前提として、その公開対象が事実、真正である必要がある。研究活動の記録と記録結果の保存は、後に展開研究につながる手がかりになる貴重な財産であると同時に、研究者にとって正当な研究がなされていることを示す役割や、研究不正を防止する役割もある。

以前から実験ノートの記録の重要性は、文部科学省の研究不正防止ガイドラインとして、平成 18 年に整理・公開されている [2]。ただし、旧来の実験ノートではなく、デジタル化に根ざした研究活動の記録手段や標準化については、まだ十分整理されているとは言い難い。平成 26 年 8 月「研究活動における不正行為等への対応に関するガイドライン」でも研究機関が研究者に対してデータ保存の義務付けの規定を求める旨の記載はあるが、保存に関する期間、方法等については、「データの性質や研究分野の特性等を踏まえることが適切である」としている。これに呼応した形では、前述の日本学術会議の答申 [1] は、研究者個人、研

究室主宰者、研究機関の長それぞれの役割に対して、下記 4 項目に対する指針を示している。

- (1) 安全管理
- (2) 研究倫理・行動規範遵守
- (3) 資料等保存
- (4) 試料等保存

中でも、「(3) 資料等保存」に関しては、研究者個人には「研究記録やメタデータの整理により、検索・抽出可能な形で整理・保管、適正なバックアップの作成」が謳われ、主宰者には、「教育・指導、メタデータ管理、研究室の統一フォーマットの作成など」が謳われ、さらに研究機関の長には「データ・バックアップ用サーバーの提供など、インフラ整備」とされている。

大学等がこれに従って整備しようとする際には、組織的にはインフラ整備が求められるところであるが、予算面や人的資源投入といった複雑な問題もあり、少なくとも個々の研究室主宰者や研究者個人が、それぞれの立場で、これらの指針に準拠した対応をとっておくことが望ましい。

一方で、文書記録に関する規格としては、ISO15489 *Information and documentation – Records management* – [3] が継続的にメンテナンスされており、それには、記録やメタデータ、制御、プロセスなどが細かく規定され、その *authoritative records* のデータ記録要件として、以下が挙げられている。

- (1) Authenticity
- (2) Reliability
- (3) Integrity
- (4) Useability

これは、研究データの保存に限られてはいないが、文書の保存管理に対する要件として参考になるものである。実装例では、例えば島津製作所の研究開発記録システム「源蔵」もこの要件を参考に構築されている [4]。ただし、大学における研究主宰者や研究機関と学生等の組織構成に対して、このシステムをそのまま適用する際には、組織構成や役割の相違に注意を要する。いずれにせよ、これらを満たす実装を具体化するには、記録が正確で一定の論拠となるものであり、かつ再現可能な程度の粒度で記載されていることや、それが主宰者など承認者によって認められている形式であること、あるいは主宰者自身が作成したものであることなどが挙げられる。また、記録は不足なく残されており、改変・改竄不能状態であることや、可用性および利便性の高いシステムによって提供されることも必要になる。

なお、ISO15489 ではさらに、その *records systems* の特徴として、システム実装上の要件も整理されており、そこには、(a) Reliable, (b) Secure, (c) Compliant, (d) Comprehensive, (e) Systematic が挙げられている。多くの CMS はこれらに何らかの形で対応付けられる機能提供がなされ

ているが、組織の様態に具体は委ねられる。

2.2 権限管理とアクセス制御要件

研究資料保管のガイドラインは、組織内、あるいは研究室内での共有に止まるレベルでの対象の保管については、共通化され難いため、個別検討が必要と考えられる。大別すると、電子化された研究記録は、すでに成果として公知とみなされる記録と、公知となる前の記録とに分かれるが、それぞれに検討を要する。一概に規定できるケースは少ないが、概ね、図2のように整理可能である。

公知となる例は、論文等の学術団体が発行する雑誌等を通じた公開が典型であり、所謂オープンアクセスジャーナルに限らず、研究組織のレポジトリ等を通じても実現される。無論、オープンアクセス化に付随して、オンライン文書に対する識別子としての DOI (Document Object Identifier) [5] や、研究者に対する ResearcherID、例えば ORCID *1 や日本の研究者に対する研究者番号の付与がなされていれば、その利活用上の利便性は飛躍的に高まる。本稿では、主に研究室主宰者以降の立場を前提に検討するものであるが、いずれの場合にもメタデータの設計および付与が重要である。

論文に限らず、学術文書には DOI が付与可能であるため、学会発表や雑誌掲載によって、公知となりうる文書は多い。これらは、研究成果をまとめたものであって、開発システムのプログラムソース等はこのような学術界でのオープン化の手続きとは別に、開発者の意思によって Git 等で公開されることもあり、それは研究データの一部とすることも可能である。また、研究者識別子を取得するか否かに関わらず、学会での発表や論文を通じて、著者に学生が含まれる場合には、その氏名は著者として掲載される。

2.3 研究データの活用例

前述の研究に対する正当性を高めるための記録・管理に対して、それらを社会で共有することにより、新たな創造的研究が可能となる。このためには、対象に対するポイントとなる情報の標準化や検索時の容易性が要件となり、それにはメタデータを用いる。武田らは、DOI 登録機関として日本国内の RA (Registration Agency) である JaLC による研究データの DOI 登録を試みている [5]。DOI もまた標準化されており、ISO 26324:2012 がそれにあたる。JaLC の試みからは、情報収集と標準化の動きは、データ層に限らず、メタな情報の層、さらにはメタな情報のスキーマ層、および識別子といったそれぞれの層での課題が整理されつつある。

また、国立情報学研究所での様々なプロジェクトの中には、オープンアクセスジャーナルに寄与する学術機関レポ

ジトリのプラットフォームサービスに繋がる JAIRO や、機関レポジトリの高次サービス活用を目的としたメタデータ・フォーマット junii2、オープンサイエンスに向けた研究データレポジトリモジュール、WEKO (NetCommons2 上で動作) など開発されている。同様の思想は国内にとどまらず、例えば EU では、EUDAT *2 が研究データの保管・利活用を進めており、世界的潮流となっている。なお、流通に際しては、OAI-PMH *3 といったプロトコルに準拠することなども研究の価値を高める手段として、一定の効果を見込める。

これらは、ネットワークの組織境界を意図せず研究データを公開・活用しようとする場合に主眼がおかれ、公開可能データが前提である。したがって、論文になる前段階や、特許出願、加工前の一次データに個人情報が含まれる場合など、一定の非公開データをどのように合理的に扱うかは研究機関、研究室主宰者、研究者個人といったそれぞれの立場で検討していくことになる。特に、非公開ながら組織内共有されるデータは、意外と多いことが想像され、そのメタデータ設計に関しては、将来的に上記のような公開が可能となることを前提とすれば、それに容易にコンバート可能なメタデータの枠組みを用意し、コンバータを実装することになる。しかし、その検討はさほど多くの事例が共有されている訳ではなく、大学組織固有の検討が必要になる。

2.4 研究活動のメディア・CMS

大学での研究活動、特に著者らの属する情報工学関係の分野では、研究者としての教員の研究活動は、学生教育と密接に関係することが多い。例えば、卒業研究や特別研究といった名目で単位化されうる。研究室の運営自体は、実態としては研究室主宰者の運営方針と設計・実装に依拠する所ではあるが、その記録は電子化され、WEB サイト等を通じて整理・共有されることも一般化しつつある。CMS、あるいはポートフォリオシステムの選択次第で、その運営上の設計が変化する場合もあり、そのような場合、システム選択と運用・運営指針は相互依存の関係にある。

一例として、著者らは、eポートフォリオシステムとして多用される Mahara を研究室運用し、ゼミ記録、報告資料保存、資料配布等に利用していると同時に、研究成果の類似リンクや引用関係ネットワークなどの研究も展開している [6]。研究データ保管には、NetCommons 等の CMS を利用することも可能であるが、アーティファクトとグループの紐付けや公開のための仕組みが利点として議論され、結果として Mahara を採用している。なお、徳島大学ではその後、Mahara を全学のポートフォリオとして採用・運用するに至っている。

*1 <http://orcid.org/>

*2 <https://www.eudat.eu/>

*3 <http://www.openarchives.org/pmh/>

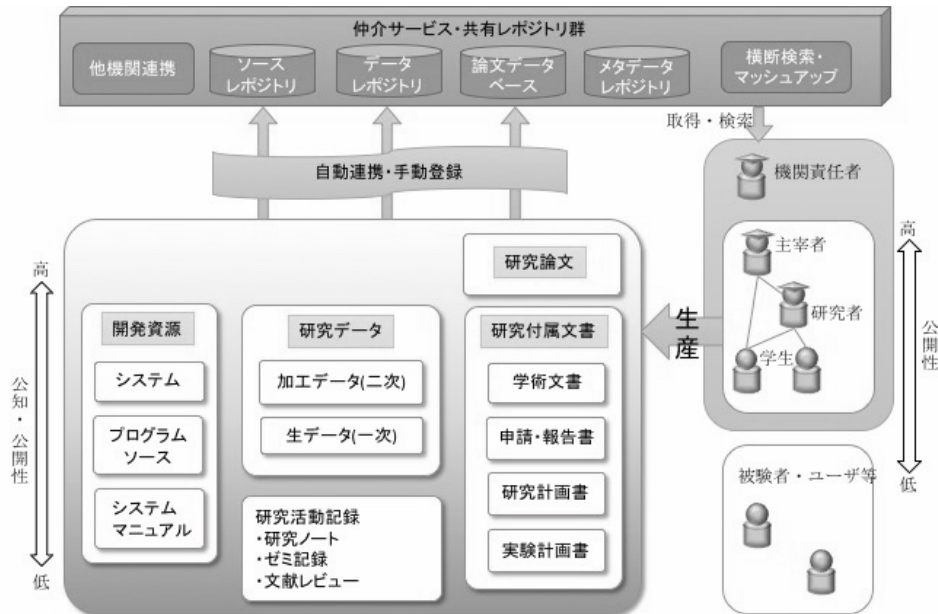


図 2 研究活動と研究組織におけるオープン・クローズ
Fig. 2 Research activities and organization



図 3 研究活動の Mahara での保管
Fig. 3 Mahara for Stroing Research activities

例えば、図 3 にあるように、ゼミ記録としての“議事録”が本稿執筆時点で 86 件蓄積されており、参加者リストや、報告会等での決定事項、その意思決定過程などが閲覧できる。また、ゼミで報告のあった全てのファイルや、サーベイ記録なども研究グループ内で共有されている。プログラムソースなどは、サブグループ単位での共有であったり、教員との間でのみ共有されることもあり、コンテンツによ

る開示範囲の柔軟性についても運用対処されている。このように運用することで、年度を跨いで学生が入れ替わる際に、資料の引き継ぎを相対的に容易にしたり、開発実装上の困難性の敷居を下げるなど、様々なメリットがある。宮寺らが研究室内の蓄積コンテンツを基に、研究活動の構造化を試みているが [7]、この目的に類似した効果も期待される。例えば、構造を予め用意するなど可能と考えるが、それには定型フォームを要した編集プラグインを作るなどが必要になり、現時点では実装には至っていない。

3. メタデータ設計

3.1 フレームワークと適用範囲

メタデータを用いた登録・流通・検索等については、学術コンテンツ以前にも教育用のコンテンツとして、かつては例えば NICER がその中継の役割を担っていた [8]。教育コンテンツの流通に大きな意味のあるサイトであったが、現在はその後継サイトに機能を引き継がれている。NICER では、LOM(学習対象メタデータ)による検索を狙い、全ての教育情報に LOM 付与することができるものであり、その標準化は IEEE 標準との対比によって整理されていた。

NII のメタデータ登録サービスや、現在の JaLC の試みは、基本的には上記のようなオープンエジュケーションにおける LOM 導入の構造と似ており、各組織の保有する公開コンテンツに対するメタデータ付与が図 2 における仲介サービス・共有レポジトリの位置には重要である。一方で、非公開コンテンツの扱いは、これには陽には規定されない。著者らの属する研究室のように、ポートフォリオ／

表 1 語彙空間

Table 1 Namespaces for common terminologies

略称	名前空間
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
dc	http://purl.org/dc/elements/1.1/
dcterms	http://purl.org/dc/terms/
bibo	http://purl.org/ontology/bibo/
foaf	http://xmlns.com/foaf/0.1/
org	http://www.w3.org/ns/org

CMS による研究リソースの共有化においては、アーティファクトの扱う蓄積リソースと共有手順について、事前に整理しておく必要がある。

例えば、著作権の扱いに注意を要する学術論文は保存しない、研究データにおいては、個人情報を含むような加工前の生データについても同様に保管しない(別管理)など、WEB 共有の不向きな素材に関する説明、教育を実施している。また、ゼミ記録としての議事録などは、定型の雛形を決めて、それを埋める形で記録している。ファイル命名規則としては、“YYYYMMDD-著作者名. 拡張子”のような形式を一応のルールとしている。

本サイトは、研究グループ共有のサイトとして5名の教員と、教員に指導を受ける学部生・大学院生等がメンバーになるが、サブグループとして、さらに小規模の三つのグループに分かれて運用されている。このため、これら三つのサブグループ毎のアクセス権設定を行う運用も実施している。研究室保存データに対するメタデータ設計に際しては、これらの要件を考慮して行う。

3.2 ポートフォリオを基盤とするメタデータ付与

メタデータの語彙集には、DublinCore や FOAF などの従来の基本的な語彙を用いる。表 1 に、本研究のメタデータにおける語彙空間のリストを挙げる。ここで、これら語彙を用いて Mahara に蓄積保管する研究活動記録をいくつかのカテゴリに分類して、それぞれにメタデータの属性を定義することを考える。例えば、図 2 には、研究データ、開発資源、付属文書といったカテゴリのほか、一般の研究活動記録も示している。この図の全てが Mahara にて保管される訳ではなく、Mahara 上のアーティファクトとして保管される際には、アーティファクトの実装種別とのマッピングも必要である。

表 2 は、メタデータ付与の対象が卒業論文等の場合の属性セットの例である。また、表 3 は、Mahara に一般のファイルが投稿された際に、半自動取得される属性である。例えば、著者らの所属する研究グループでは、Mahara のログインには Shibboleth に対応させており、作成者の ID は環境変数から自動取得する。記述属性については、Mahara のインタフェース上の注釈等から取り込むことで、自動化

表 2 対象が論文の場合の属性

Table 2 A set of attribute for academic paper

名前空間	語彙	内容	必須
dcterms	URI	URI	Y
dc	type	リソースタイプ	Y
dc	title	タイトル	Y
dc	creator	著者	Y
dc	date	公開日	Y
dc	language	言語コード	Y
dcterms	description	記述	Y
dcterms	format	形式	Y
dcterms	references	参考文献	Option
dcterms	accessRight	アクセス権	Y
dc	subject	キーワード	Option

表 3 Mahara での半自動付与の語彙集

Table 3 Terms of semi-automatic attribute in Mahara

語彙集	語彙	内容
dcterms	URI	URI
dc	type	種類
dc	title	タイトル
dc	creator	作成者
dcterms	created	作整日
dcterms	modified	更新日
dc	language	言語
dcterms	description	内容記述
dcterms	format	記録形式
dcterms	accessRight	アクセス権

を行うが、これらについては編集可能な属性とする。

前節の通り、2.1 節記載の要件を概ね満たすメタデータ設計を行っている。しかし、一般の CMS と異なり、決裁を伴うワークフローまでの実装とは現時点では至っていない。これは、Authnticity を実現する機能要件となりうる機能である。すなわち、ラボノートと同様に、研究者(学生)、研究者(教員)、研究室主催者(研究室レベルの責任者)、研究機関の責任者(多くの場合は、研究担当理事)、研究機関の長(学長、総長)といった階層に応じた承認者が承認を電子的に行う機能である。対象データがファイルである場合には、ファイルに対する暗号化(オプション)、電子署名と、メタデータに対する電子署名のためのワークフローを実装することになる。なお、Mahara のアーティファクトとして、システムの保有する単純なテキストデータについては、このような暗号技術の応用ではなく、一般的な Web フローでの実装となるが、これについては今後検討していく必要がある。

3.3 関連システムへの展開・連携の可能性

徳島大学全体で、前述までの仕組みを展開しようとする場合、いくつかの関連システムがあり、これらとの関係を考慮していくことになる。ここでは、それらの機能的分掌

を記載するとともに、全学展開の可能性を示すこととする。ただし、下記それぞれ運用組織が異なるなど、システム面だけでなく一定の運用指針も同時に検討していく必要がある。

- (1) 全学ポータルフォリオシステム先にも述べたように、徳島大学では全学構成員が一様に Mahara を利用することができる。学生に対しては、学習履歴を主眼とした利用を狙っており、教員に関しても授業コンテンツ等の蓄積を目的とするものではあり、単純に著者らのシステムを展開できる訳ではない。しかし、無論、同じプラットフォームとしての展開の容易さはあるため、今後の適用可能性はあると考えている。
- (2) 機関レポジトリ図書館を中心とした学術コンテンツのリポジトリシステムである。本学では、博士論文等学術論文の機関レポジトリ登録が進められているところである。本システムには、公知となったものを中心に学術論文の本体が一定の許諾の下で公開されるため、連携時にはコンテンツ本体へのリンク先としての活用が見込まれる。
- (3) 教員業績等管理システム機関レポジトリが学術コンテンツの本体を保管するシステムとすれば、その書誌情報や、機関レポジトリには含まれない文献リスト等、本学構成員が関与した研究業績のデータベースシステム (EDB: 徳島大学教育・研究者情報データベース) である。本システムは、登録されるデータの単位をオブジェクトとして扱い、継承等の可能なオブジェクト指向のデータベースとなっている。例えば、著者名や組織名が全てオブジェクトであり、業績登録者による入力時の揺らぎや誤入力を防ぐ利点がある。

研究活動に関する様々な関連システムは、これら機関の長に推進されるシステムと、研究室主宰者によるシステム、および研究者個人が保管する研究資源といった具合に分類できる。本稿で述べたような研究室主宰者レベルでの非公式なデータが、これら機関全体のシステムと連携して機能するような枠組みが実現できれば、学際研究に繋がったり、展開研究につながる可能性も高まることが期待されるが、今後の検討課題である。さらには、国立情報学研究所や JaLC 等との連携につながるような仕組みの実現に向けての課題も洗い出す必要があると考えている。

4. おわりに

本研究では、研究室主宰者レベルでの、主に非公式・非公開記録を含む研究データの保管に関する検討を行った。メタデータによる研究データに対する活用方法の設計には、メタデータ自体の設計・スキーマ設計が必要であるが、本稿では、現時点でのアドホックな可能性を示したに過ぎない。今後は、前節に述べたような、研究室主宰者から上位層へのアプローチも検討していきたい。

謝辞 本研究は JSPS 科研費 JP15K12168 の助成を受けたものです。

参考文献

- [1] 日本学術会議 : 「科学研究における健全性の向上について」(online), 入手先 (<http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-k150306.pdf>) (2016.10.17).
- [2] 文部科学省: 研究活動の不正行為への対応のガイドラインについて (online), 入手先 (http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu12/houkoku/06082316.htm) (2016.10.17).
- [3] ISO15489 : Information and documentation – Records management – (online), available from (<https://www.iso.org/obp/ui/#iso:std:iso:15489:-1:ed-2:v1:en>) (2016.10.17).
- [4] 水戸康敬, 石橋功 : 研究開発向け記録管理システム “源蔵” の開発 (online), 入手先 (http://www.genzou.jp/images/menu_siru.gif), 島津評論, Vol.64, No.3.4, pp.205-214, 2007.
- [5] 武田英明, 村山泰啓, 中島律子 : 研究データへの DOI 登録実験, 情報管理, Vol.58, No.10, pp.763-770, (2016).
- [6] 関陽介, 野村卓哉, 曾我部紗也香, 佐野雅彦, 松浦健二, 大平健司, 上田哲史 : 研究活動の持続的成長を支援する e ポータルフォリオシステムの提案, 情報処理学会研究報告, Vol.2016-CLE-18, No.7, pp.1-6, (2016).
- [7] 宮寺庸造, 中村勝一, 横山節雄, 夜久竹夫 : 研究情報推移グラフによる情報の個人管理・共有手法, 電子情報通信学会論文誌, Vol. J91-D, No.3, pp.639-653, (2008).
- [8] 清水康敬, 岩田裕美, 榎本聡 : 授業で使えるコンテンツを提供する総合 Web サイト NICER, 学習情報研究, pp.13-20, (2004).