

[ポスター発表] 研究報告

# Jupyter Notebookによる計算実験の再現

横山 重俊<sup>1,3,a)</sup> 浜元 信州<sup>1</sup> 長久 勝<sup>2</sup> 政谷 好伸<sup>3</sup> 竹房 あつ子<sup>3</sup> 合田 憲人<sup>3</sup>

## How to Reproduce Research Environments with Jupyter Notebook

バイオインフォマティクス分野などデータ駆動型研究分野では、論文発表された計算実験の再現性確保に対する要請が強く、DNA 塩基配列の公共データベース構築によるデータ共有、およびデータ解析ソフトウェアのオープンソース化による共有が進んでいる。しかし、計算環境再構築技術の未成熟さから、実験の再現性確保には課題が残存する [1]。

本研究ではこの課題解決のために、データ駆動型研究分野の計算実験の再現性を推進するデータビリティプラットフォームの普及を目指す。具体的には、ICT 基盤上に構築された計算実験(以後 実験)の再現性をコンテナ技術を用いたオーバーレイクラウド方式の仮想クラウド技術 [2], [3] と実行可能な構築手順書構成技術 (LC4RI: Literate Computing for Infrastructure)[4] の組合せで実現する。

データ駆動型研究分野の研究プロセスを単純化すると、(1) データの蓄積, (2) 実験 (データ分析と発見), (3) 論文化となり、(1) と (3) についてはオープン化が進んでいる。残る (2) についてのオープン化を進める手段を提供することが本研究の対象となる。

データビリティ (Datability) とは、Data と Ability を合わせた造語で「大規模なデータを持続可能 (Sustainability) かつ責任ある形 (Responsibility) で活用する能力」であると定義されており、データ駆動型の研究推進において鍵となる概念である。オープンデータ化によるビッグデータの共有・流通に基づくデータ駆動型研究の推進やデータ駆動型研究人材の育成が、図 1「オープンデータのための研究スタイル」に示すように進んでいる。論文とともにその際生成される中間データも付加価値 Data としてオープン化される。データビリティプラットフォームはこのオープン化の動きを促進するために不可欠な要素である。

その要素は、共有・流通しているデータを活用する計算

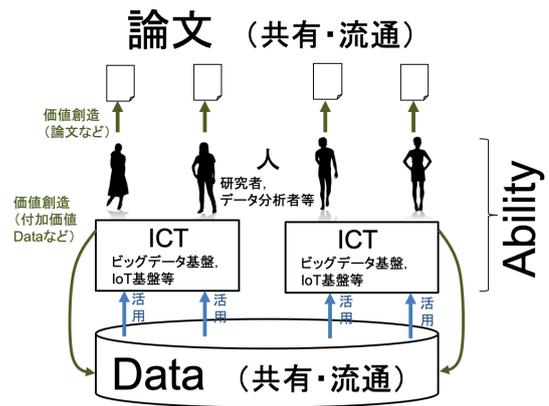


図 1 オープンデータのための研究スタイル

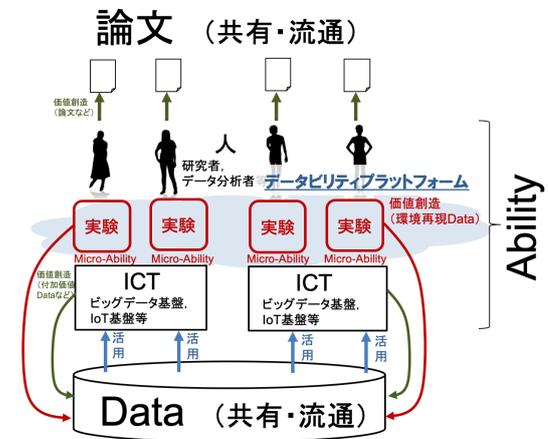


図 2 データビリティプラットフォームのある研究スタイル

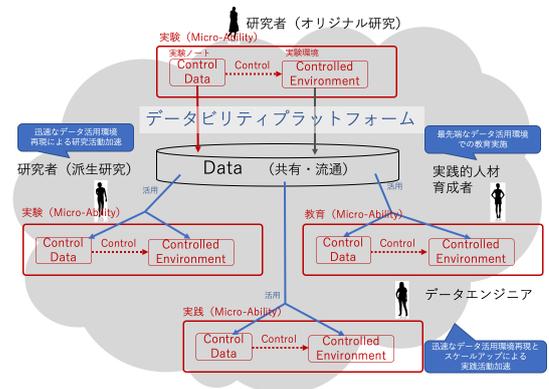


図 3 データビリティプラットフォームによる実験環境流通

1 群馬大学  
Gunma University  
2 ライフマティクス株式会社  
Lifematics  
3 国立情報学研究所  
National Institute of Informatics  
a) yoko@gunma-u.ac.jp

実験の再現性向上である。データバリエーションの Ability 軸で言うと、アカデミッククラウドを代表とする ICT 基盤とそれを活用する人という Ability に加えて、図 2「データバリエーションプラットフォームのある研究スタイル」に示すように、各実験を Micro-ability として定義し、実験再現のための情報も付加価値 Data に加えて、それを共有・流通させるこのような仕組みをアカデミックコミュニティとして持つことが、データ駆動型の研究推進を成功させる鍵となる。

実験環境に使われるデータ分析基盤は多様であり、それを許容するデータバリエーションプラットフォームが必要となるため、この点にも留意したデータバリエーションプラットフォームを用意する。

そのような柔軟性を持った実験を共有・流通させるための仕組みとして、まず実験を「各分野で標準化された実験環境テンプレート（以後 実験環境テンプレート）」と「各実験を実行するための再現情報（以後 実験ノート）」より構成する。前者の実験環境は Controlled Environment であり、後者の実験ノートはそれを制御する Control Data である。論文と同時に各分野においては標準化された実験環境テンプレートが共有化され、実験ノートを共有・流通することで、実験の再現性が確保される可能性がある。Data の Sustainability や Responsibility はこの実験環境テンプレートや実験ノートについても適用されるべきである。

さらに図 3「データバリエーションプラットフォームによる実験環境流通」に示すように、付加価値 Data として加えられた実験再現のための情報を、研究フェーズの実験のみならず、実践的人材育成やデータエンジニアがデータ活用を実践する領域へ流通させ、データバリエーションプラットフォームのさらなる展開も視野に入れたい。

今回は、教育実践研究分野をデータ駆動型研究の一分野とし、その分野の一つの研究を対象とした実験再現について報告する。この分野の実験環境テンプレートとして、国立情報学研究所で教育向けに特化して JupyterHub[5] を機能拡張して開発している CoursewareHub[6], [7] を採用した。再現対象の実験は群馬大学で実施された情報教育とした [8]。実験ノートとしては、この教育実践で用いられた教材としての Jupyter Notebook と講師が講義時に使用した実験実施用の Jupyter Notebook とする。

再現対象であるオリジナル教育実践実験は、国立情報学研究所のクラウド内に構築された CoursewareHub を利用して実施された。この CoursewareHub を教育実践実験環境として商用クラウドやプライベートクラウド上に構築するために、学認クラウドオンデマンド構築サービスのアプリケーションテンプレートとして CoursewareHub 用のものを開発した。このアプリケーションテンプレート（以後 CoursewareHub テンプレート）を使って、各実験環境の要求条件（構築先クラウド/ノード数/マシンサイズ/ホスト名/サーバ証明書類 など）に合わせた Jupyter Notebook

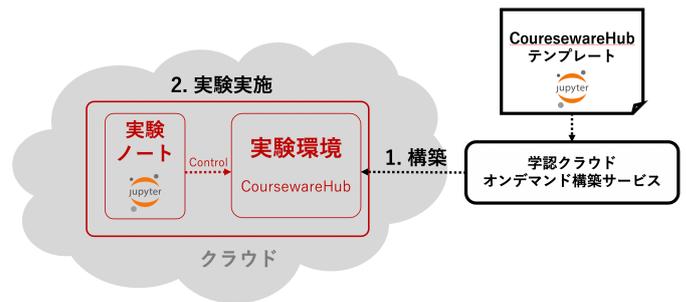


図 4 教育実践実験の再現構成

実行環境をクラウド上に動的に構築することで、実験ノートを含んだ実験環境である Micro-ability を図 4 の様に実現する。本方式においては、実験環境構築段階で、Jupyter Notebook を LC4RI に沿って学認クラウドオンデマンド構築サービスを利用して実験環境構築を行うために活用すると同時に、実験実施時に実験手順を記述する実験ノートのためにも利用する。これら二つの異なるレベルで Jupyter Notebook を利用することに注意していただきたい。

以上説明した実現方式で、筆者らの論文 [8] について、上記方針に沿って商用クラウド上に再現することが可能であることを確認できた。実験実施手順が記述された Jupyter Notebook（実験ノート）を github から構築した実験環境内に clone することも Jupyter Notebook で記述し実行することで、実験ノートを実験環境に持ち込んだ。

今後、CoursewareHub テンプレートおよび今回利用した実験を実施する手順を記述した Jupyter Notebook を github などで開示する予定であり、これらの情報と国立情報学研究所の提供する学認クラウドオンデマンド構築サービスを利用し、今回報告した実験の再現性を体験できる。

## 参考文献

- [1] Practical Computational Reproducibility in the Life Sciences, Bjorn Gruning et al., Cell Systems volume 6, issue 6, pp. 631-635, 2018
- [2] 竹房あつ子, 佐賀一繁, 丹生智也, 横山重俊, 合田憲人, 学認クラウドオンデマンド構築サービスの推進, AXIES 2018 年度年次大会
- [3] 学認クラウドオンデマンド構築サービス, <https://cloud.gakunin.jp/ocs> (2019-10-10)
- [4] 長久勝, 政谷好伸, 谷沢智史, 中川晋吾, 合田憲人, Notebook を介した作業ノウハウの継承・移転を分析するための基盤, インターネットと運用技術研究会, 情報処理学会, 2019 年 2 月 28 日
- [5] Project Jupyter and JupyterHub, <https://jupyter.org>, <https://jupyter.org/hub> (2019-10-10)
- [6] 長久勝, 政谷好伸, 合田憲人. Notebook による講義・演習環境の開発. 第 27 回教育学習支援情報システム研究会, 情報処理学会, 2019 年 3 月 22 日
- [7] CoursewareHub, <https://github.com/NII-cloud-operation> (2019-10-10)
- [8] 横山重俊, 浜元信州, 政谷好伸, 合田憲人, Jupyter Notebook を活用した情報教育実践, 情報処理学会 情報教育シンポジウム SSS2019