

令和 6 年 6 月 19 日現在

機関番号：62615

研究種目：基盤研究(C)（一般）

研究期間：2019～2023

課題番号：19K11895

研究課題名（和文）オーバーレイクラウドによる計算実験環境再現性

研究課題名（英文）Reproducibility of computational experiment environment using overlay cloud

研究代表者

横山 重俊（Yokoyama, Shigetoshi）

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・特任研究員

研究者番号：10600968

交付決定額（研究期間全体）：（直接経費） 2,000,000円

研究成果の概要（和文）：本研究では実験環境の再現性をコンテナ技術によるオーバーレイクラウド方式の仮想クラウド技術と実行可能な構築手順書構成技術を組み合わせることで、データ駆動型研究実験環境の流通性を飛躍的に高めた。Control DataとしてはJupyter Notebookによる実行可能構築手順書を用いた。Controlled Environmentとしては仮想クラウドを動的に構築できるコンテナ技術を利用したOverlay Cloudアーキテクチャおよび仮想クラウド生成を担うVirtual Cloud Providerを用いた。

研究成果の学術的意義や社会的意義

オーバーレイクラウド技術を用いて実現したデータビリティプラットフォームを使って現実に研究者の間で計算実験環境の流通が進むことに関する実証を行なった。具体的には研究者自らが再現可能な計算実験環境を構成できること。また、その環境を流通させ受け取った研究者がそれを再現すると同時にさらに派生研究をし、変更後の計算実験環境をさらに流通させられることの実証を行なった。これらの実証はオープンサイエンス推進に不可欠な事項であり学術的・社会的意義がある。

研究成果の概要（英文）：In this study, we dramatically improved the distribution of data-driven research experimental environments by combining virtual cloud technology with an overlay cloud method using container technology and executable construction procedure configuration technology to reproducibly create an experimental environment. As control data, we used executable construction procedures using Jupyter Notebook. As a controlled environment, we used an overlay cloud architecture that uses container technology to dynamically create a virtual cloud, and a Virtual Cloud Provider that is responsible for generating the virtual cloud.

研究分野：情報学

キーワード：研究再現性 オープンサイエンス クラウド Notebook

1. 研究開始当初の背景

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。

1. データビリティプラットフォームの必要性

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 オープンデータのみ研究スタイルに示すように進んでいる。この動きをより促進ために不可欠な要素がある。それは共有・流通しているデータを活用する計算実験環境の再現性向上である。データビリティの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 データビリティのある研究スタイルに示すように各実験環境をMicro-abilityとして定義し、それを共有・流通させる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ビリティプラットフォームが必要である。

2. データビリティプラットフォームの実現性

クラウドコンピューティング資源をクラウドプロバイダに依存しない方法で活用することのできるオーバーレイクラウド技術の開発に伴い研究者自らが計算実験環境を流通させることのできるデータビリティプラットフォームの原型が実現出来ている。

2. 研究の目的

実験環境を共有・流通させるための仕組みとして、実験環境は「各分野で標準化された環境テンプレート」と「各実験環境を生成するための再現情報 (環境再現Dataと呼ぶ)」より構成する。前者はControlled Environment であり、後者はそれを制御するControl Dataである。例えばゲノム解析分野の場合、ゲノム解析パイプライン実行環境雛形が標準化された環境テンプレートであり、各研究者が使う各実験環境を構成するための必要なツール群の指定情報等が環境再現Dataである。各分野においては標準化された環境テンプレートが共有化されているので論文と同時に環境再現Dataを共有・流通することで論文の再現性が確保される。DataのSustainabilityやResponsibilityはこの環境再現Dataについても適用されるべきである。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に示すように研究のみならず実践的人材育成やデータビリティエンジニアがデータ活用を実践する領域への展開も視野に入れる。さらには、分野を跨るデータ駆動型研究の推進のためにMicro-abilityを連携し、組み合わせること

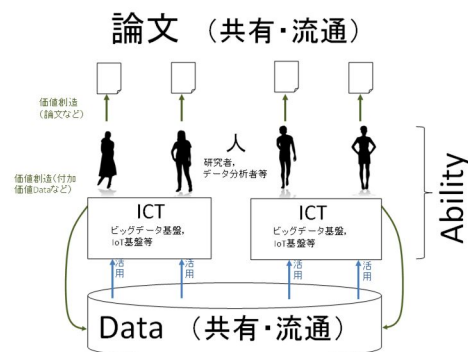


図1 オープンデータのみ研究スタイル

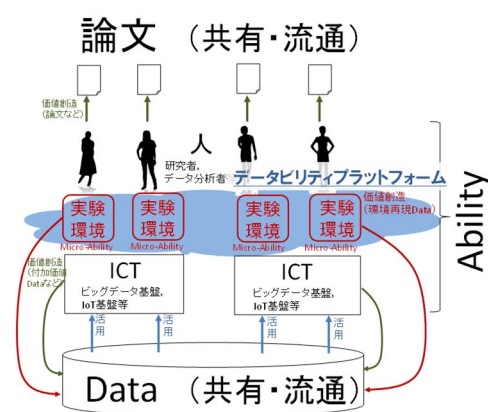


図2 データビリティプラットフォームのある研究スタイル

でOpen Innovation Network Eco-system形成を加速することができる。

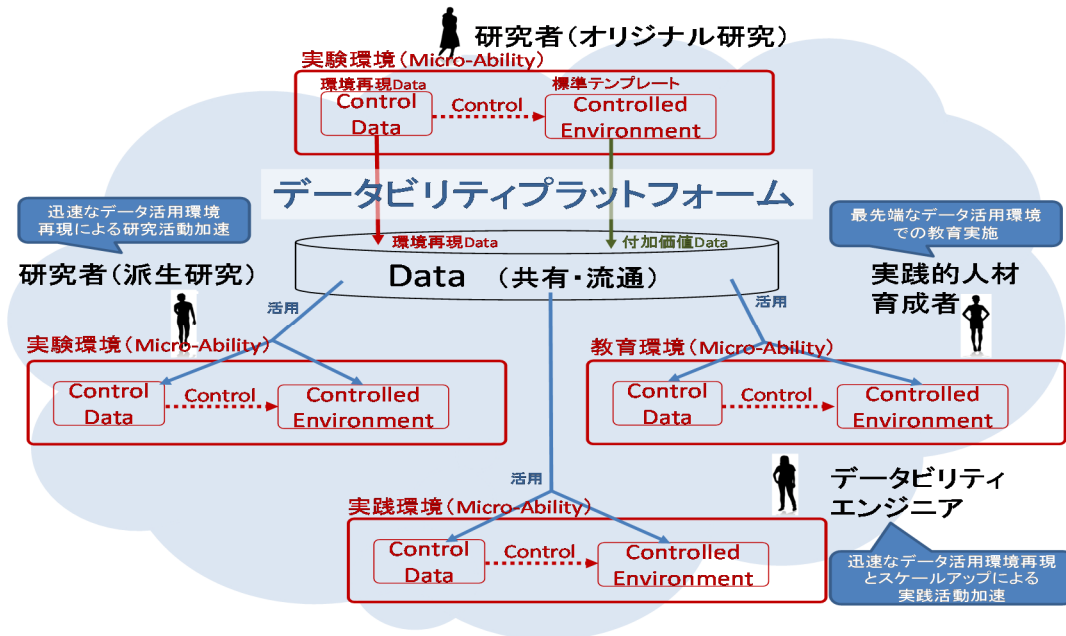


図3 データビリティプラットフォームによる実験環境流通

3. 研究の方法

本研究の基本構成要素であるControl DataとControlled EnvironmentからなるMicro-abilityを実現する。Control DataとしてはLiterate Computing for Infrastructureの研究成果であるJupyter Notebookによる実行可能構築手順書を用いる。これはインフラ構築のための実行スクリプトと構築手順同時に記述するだけでなく、実行結果の保存も可能としている。Controlled Environmentとしては各種クラウド上に跨った仮想クラウドを動的に構築できるコンテナ技術を利用したOverlay Cloudアーキテクチャおよび仮想クラウド生成を担うすでに開発済みのミドルウェア Virtual Cloud Providerを用いる。それらを図3 に示すように統合することでMicro-abilityを実現する。

ゲノム解析分野での実験環境の流通実証をゲノム解析ツール

galaxy(<https://galaxyproject.org/>)の利用者コミュニティメンバの協力を得て実施する。具体的には現在クラウド内のマシンイメージとして流通しているGalaxy実験環境のデータビリティプラットフォームへ移行可能であることを実証する。

数学分野での実験環境の流通実証を数式処理ツールSageMath (<http://www.sagemath.org/>)の利用者コミュニティメンバの協力を得て実施する。上記二つの環境の連携について両コミュニティのメンバの協力を得て取り組み、分野を跨った実験環境の流通にデータビリティプラットフォームが活用できることを実証する。

4. 研究成果

本研究では、ICT基盤上に構築された実験環境の再現性をコンテナ技術によるオーバーレイクラウド方式の仮想クラウド技術と実行可能な構築手順書構成技術(Literate Computing for Infrastructure)を組み合わせることで、データ駆動型研究実験環境の流通性を飛躍的に高めた。2019年度には本研究の基本構成要素であるControl DataとControlled EnvironmentからなるMicro-abilityのプロトタイプを実現した。Control DataとしてはLiterate Computing for

Infrastructureの研究成果であるJupyter Notebookによる実行可能構築手順書を用いた。これはインフラ構築のための実行スクリプトと構築手順同時に記述するだけでなく、実行結果の保存も可能としている。Controlled Environmentとしては各種クラウド上に跨った仮想クラウドを動的に構築できるコンテナ技術を利用したOverlay Cloudアーキテクチャおよび仮想クラウド生成を担うすでに開発済みのミドルウェア Virtual Cloud Providerを用いた。

以下の二つの適用分野でのこのMicro-abilityを用いた実証実験を実施した。

【ゲノム解析分野】実験環境の流通実証をゲノム解析ツール Galaxyの利用者コミュニティメンバーの協力を得て実施する。具体的には現在クラウド内のマシンイメージとして流通しているGalaxy実験環境のデータリティプラットフォームへ移行可能であることを実証した。

【数学分野】実験環境の流通実証を数式処理ツールの利用者コミュニティメンバーの協力を得て実施した。

5. 主な発表論文等

〔雑誌論文〕 計15件（うち査読付論文 2件 / うち国際共著 1件 / うちオープンアクセス 0件）

1. 著者名 Takaki Osamu, Hamamoto Nobukuni, Takefusa Atsuko, Yokoyama Shigetoshi, Aida Kento	4. 巻 207
2. 論文標題 Reasonable Setting Values for Anonymization Algorithms for Online Educational Data Analysis Support System	5. 発行年 2022年
3. 雑誌名 Procedia Computer Science	6. 最初と最後の頁 2556 ~ 2566
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.procs.2022.09.314	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 佐賀一繁, 三浦信一, 丹生智也, 竹房あつ子, 竹房あつ子, 横山重俊, 合田憲人	4. 巻 HPC-179
2. 論文標題 VCPIによるCPUアーキテクチャが異なるHPCシステムとクラウドの連携	5. 発行年 2021年
3. 雑誌名 情報処理学会研究報告	6. 最初と最後の頁 1,8
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 竹房あつ子, 佐賀一繁, 横山重俊, 丹生智也, 合田憲人	4. 巻 HPC-179
2. 論文標題 インタークラウド環境構築システムVCPIによるHPCクラスタ環境構築支援	5. 発行年 2021年
3. 雑誌名 情報処理学会研究報告	6. 最初と最後の頁 1,6
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 浜元信州, 横山重俊, 竹房あつ子, 合田憲人	4. 巻 CLE-33
2. 論文標題 クラウドを利用したMoodleのログ解析環境の実装	5. 発行年 2021年
3. 雑誌名 情報処理学会研究報告	6. 最初と最後の頁 1,7
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 横山重俊, 浜元信州, 長久勝, 藤原 一毅, 政谷好伸, 竹房 あつ子, 合田憲人	4. 巻 50
2. 論文標題 派生研究を支援する再構成容易な実験数学環境	5. 発行年 2020年
3. 雑誌名 情報処理学会 IoT研究会	6. 最初と最後の頁 1,8
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 横山重俊, 浜元信州, 長久勝, 藤原 一毅, 政谷好伸, 竹房 あつ子, 合田憲人	4. 巻 2
2. 論文標題 オープンサイエンス推進のためのデータ分析プロセス共有	5. 発行年 2020年
3. 雑誌名 情報処理学会 RDM研究会	6. 最初と最後の頁 1,6
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 横山重俊, 浜元信州, 長久勝, 藤原 一毅, 政谷好伸, 竹房 あつ子, 合田憲人	4. 巻 3
2. 論文標題 データ分析プロセス共有による研究再現例	5. 発行年 2020年
3. 雑誌名 情報処理学会 RDM研究会	6. 最初と最後の頁 1,7
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 横山重俊, 浜元信州, 長久勝, 谷沢智史, 藤原 一毅, 政谷好伸, 竹房 あつ子, 合田憲人	4. 巻 2020
2. 論文標題 実験数学をJupyter Notebookでもっとやってみる	5. 発行年 2020年
3. 雑誌名 RIMS共同研究「数学ソフトウェアとその効果的教育利用に関する研究」	6. 最初と最後の頁 1,10
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 横山 重俊, 浜元信州, 政谷好伸	4. 巻 2021
2. 論文標題 Jupyter Notebookを活用した実験数学環境の高度化	5. 発行年 2021年
3. 雑誌名 2021年度 数学教育学会 春季例会予稿集	6. 最初と最後の頁 1,3
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 横山重俊, 浜元信州, 長久勝, 政谷好伸, 竹房 あつ子, 合田憲人	4. 巻 2019
2. 論文標題 Jupyter Notebookによる計算実験の再現	5. 発行年 2019年
3. 雑誌名 インターネットと運用技術シンポジウム論文集	6. 最初と最後の頁 98,99
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 横山 重俊, 浜元 信州, 長久 勝, 政谷 好伸, 合田 憲人	4. 巻 2019
2. 論文標題 Jupyter Notebookを活用した情報教育実践	5. 発行年 2019年
3. 雑誌名 情報教育シンポジウム論文集	6. 最初と最後の頁 2,9
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 横山 重俊, 浜元 信州, 長久 勝, 政谷 好伸, 合田 憲人	4. 巻 2019
2. 論文標題 実験数学をJupyter Notebookでやってみる	5. 発行年 2019年
3. 雑誌名 実験数学をJupyter Notebookでやってみる	6. 最初と最後の頁 79,91
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 横山 重俊, 浜元信州, 政谷好伸	4. 巻 2019
2. 論文標題 Jupyter Notebookを活用した実験数学環境に関する提案	5. 発行年 2019年
3. 雑誌名 2019 年度 数学教育学会 秋季例会論文集	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 横山 重俊, 浜元信州, 政谷好伸	4. 巻 2019
2. 論文標題 Jupyter Notebookを活用した実験数学におけるリアルタイム進捗収集ツール	5. 発行年 2019年
3. 雑誌名 2019 年度 数学教育学会 夏季研究会 (関東エリア) 論文集	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 横山 重俊, 浜元信州, 政谷好伸	4. 巻 2019
2. 論文標題 Jupyter Notebookを活用したアクティブラーニングへのトライアル	5. 発行年 2019年
3. 雑誌名 2019 年度 数学教育学会 夏季研究会 (関西エリア) 論文集	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計5件 (うち招待講演 0件 / うち国際学会 0件)

1. 発表者名 横山重俊
2. 発表標題 実験数学を教育から研究までやってみる
3. 学会等名 京都大学数理解析研究所 数学ソフトウェアとその効果的教育利用に関する研究
4. 発表年 2022年

1. 発表者名 南山泰之;河合将志;横山重俊;林正治;大波純一;藤原一毅;朝岡誠;古川雅子;込山悠介
2. 発表標題 京都大学数理解析研究所 数学ソフトウェアとその効果的教育利用に関する研究
3. 学会等名 情報処理学会研究報告
4. 発表年 2022年

1. 発表者名 高木理;浜元信州;竹房あつ子;横山重俊;合田憲人
2. 発表標題 オンライン教育データ解析支援システムの開発に向けた匿名化アルゴリズムに関する妥当な設定値の検討
3. 学会等名 情報処理学会研究報告
4. 発表年 2022年

1. 発表者名 横山重俊, 浜元信州, 長久勝, 中川 晋吾, 政谷好伸, 竹房 あつ子, 合田憲人
2. 発表標題 オンライン情報教育における講義・演習環境の構成例
3. 学会等名 大学ICT推進協議会
4. 発表年 2020年

1. 発表者名 横山重俊, 浜元信州, 長久勝, 政谷好伸, 竹房 あつ子, 合田憲人
2. 発表標題 データサイエンス教育における講義・演習環境の構成例
3. 学会等名 大学ICT推進協議会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------