

# データリテラシー教育におけるデータライフサイクルマネジメントに関する調査

## Survey on Data Lifecycle Management in Data Literacy Education

中鉢直宏<sup>†</sup>

Naohiro Chubachi

<sup>†</sup> 高崎商科大学 商学部

<sup>†</sup> Takasaki University of Commerce

### 要旨

この本研究では、日本の大学におけるデータリテラシー教育の現状と、データリテラシーの教育内容とデータライフサイクルおよびその管理についての必要性について紹介し、大学1年生へのアンケート調査を通じて学生がデータライフサイクルについてどのような意識を持っているかを知ることを目指した。調査を通じて、データの収集、保存、共有、セキュリティなどのデータ管理の各側面に関する学生の意識と理解について報告した。特に自分の作成したデータの保存について、長く保存する意識を持っている学生が多かったが、必要なくなったらデータを削除してしまう学生も一定程度いたことから、データライフサイクルの統合を意識した教育の必要性について述べた。

### 1. はじめに

現在、多くの大学においてデータリテラシー教育は文理を問わず、すべての学部において行われるようになった。2016年度に「大学の数理・データサイエンス教育強化方策について」が打ち出されて以来、「一般教育では、数理・データサイエンスが社会における課題発見・解決や新たな価値の創出や様々な学問分野の発展に役立つものであることを認識し、それがどのような原理で行われているのかを理解できることを目指す。」[1]とあるように、原理の理解など日本のデータ教育は理系色が強いデータサイエンスを目指した教育であった。その後平成29年度の情報通信白書[2]では、データ主導経済（data-driven economy）について取り上げられ、データは「21世紀の石油」とたとえ、多種多様なデータの生成・収集・流通・分析・活用を徹底的に図る必要性が報告されている。また、社会経済活動を再設計し、社会の抱える課題の解決のために、データの利活用の重要性について取り上げている。データ利活用では、データを集めることよりも、データから得られる様々な意味や知見の活用に徐々に重点が置かれるようになった。

その後「AI戦略2019」[3]「AI戦略2022」[4]においては、文理を問わず、全ての大学・高専生（約50万人卒/年）が、課程にて初級レベルの数理・データサイエンス・AIの習得を目指し、「数理・データサイエンス・AI教育プログラム認定制度（リテラシーレベル）認定制度」[5]を設けた。これを契機に、様々な日本の大学でデータリテラシー教育が本格的に取り入れられた。この制度でデータリテラシーレベルのモデルカリキュラム[6]は、図1の内容となっている。モデルカリキュラムは、1. 社会におけるデータ・AIの利活用、2. データリテラシー、3. データ・AI利活用における留意事項、4. オプションで構成されている。

導入	<b>1. 社会におけるデータ・AI利活用</b> 1-1. 社会で起きている変化 1-3. データ・AIの活用領域 1-5. データ・AI利活用の現場	1-2. 社会で活用されているデータ 1-4. データ・AI利活用のための技術 1-6. データ・AI利活用の最新動向
基礎	<b>2. データリテラシー</b> 2-1. データを読む 2-3. データを扱う	2-2. データを説明する
心得	<b>3. データ・AI利活用における留意事項</b> 3-1. データ・AIを扱う上での留意事項 3-2. データを守る上での留意事項	
選択	<b>4. オプション</b> 4-1. 統計および数理基礎 4-3. データ構造とプログラミング基礎 4-5. テキスト解析 4-7. データハンドリング 4-9. データ活用実践（教師なし学習）	
	4-2. アルゴリズム基礎 4-4. 時系列データ解析 4-6. 画像解析 4-8. データ活用実践（教師あり学習）	

図1 数理・データサイエンス・AI（リテラシーレベル）モデルカリキュラム

「AI戦略2022」では、AIやその他のデータ利活用において、膨大なデータを持つことが重要であることは認めつつも、デジタル化された状況においてはサービスの提供を通じてデータを取得されたデータをサービスの向上につなげ、AIを強化するデータ収集等を行うような持続的なサイクル（ループ）を形成するように配慮することが重要であるとあり[3]、データが全てではなく、データのループの形成が重要とされた。筆者は、このループを形成するためには、データサイエンスよりも情報システムの考え方に近いアプローチが必要であり、そのためデータリテラシー教育には情報システムの要素を取り入れるべきだと考えた。そこで、データライフサイクルに着目し、既存の統計・データサイエンスの基礎としてのデータリテラシーを狭義のデータリテラシー、情報システムを含むデータリテラシーを広義のデータリテラシーとして定義した。そして、この広義のリテラシーの基盤となる考え方は、データライフサイクルとその管理にあると考えた。図2に狭義と広義のデータリテラシーと分け、データライフサイクルとの関係を示す[7]。

このデータライフサイクルに基づいた内容をモデルカリキュラムで検討する際、共有や管理の部分でセキュリティなどを含む「3. データ・AI利活用における留意事項」、データベースなどを扱う「4-7. データハンドリング」、データライフサイクル全体として「1-5. データ・AI利活用の現場・データサイエンスのサイクル」が該当すると考えられる。

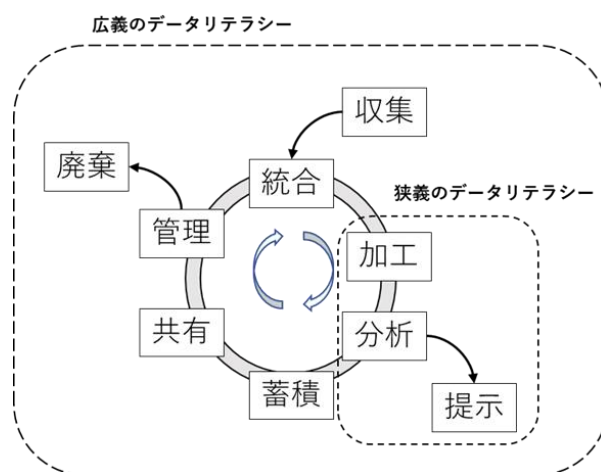


図2 データライフサイクルと狭義と広義のデータリテラシーとの関係図

## 2. データライフサイクルに関する調査について

そこで、本研究では、データリテラシーを学ぶ学生に、データライフサイクルに関する潜在的な意識調査を行った。高崎商科大学の大学1年生に対して、第一回の授業のオリエンテーション時に、補足資料にあるようなアンケート調査を、Google フォームを使用して行った。2022年と2023年の授業期間中にそれぞれ3回の調査を行い、調査に協力した学生170名分のデータを分析した。調査の内容は、データライフサイクルにおいて「収集」「蓄積」「共有」「管理」について、学生の身近な状況でどのようなルールや考えを持っているかを把握するための設問とした。「そう思わない」を得点1とし、「そう思う」を得点5として五段階評価を得点化して、平均点を求めたものを表1に示す。設問3から7はデータ収集に関する質問で、紙と電子デバイスの使用に関して、適切なデータ収集の仕組みを理解できるかどうかについて尋ねていた。中央値が3点で、平均点はこれを少し上回る結果だった。自由記述には、メリットとして年配者への配慮や電子機器を必要としないなどの答えが見られ、デジタルデバイスへの意識が示されていた。紙のデメリットに関して、集計に関する問題が多く指摘されていた。電子機器の利用については、集計の迅速化や入力の手軽さなどがメリットとして挙げられた。特に、スマートフォンなどで回答する際の漢字入力の利便性を、紙に比べて簡単だと感じる学生が多かったことがわかった。デメリットとしては、デバイスがなければ回答できないという意見が多く見られた。

質問番号8、9は保存に関するルールの有無を尋ねる内容で、平均点は3.08であり、中央値とほぼ同じであった。

自由記述では、フォルダ分けのルールについてはいくつかの言及があったが、ファイル名に関する言及は少なかった。

設問 10、11 は削除に関する内容であり、こちらの平均点も 3.04 で中央値に近かった。自由回答では、必要性が無くなったデータについては、十分に確認してから削除するという回答が多く見られた。また、データが不要になったら期限を設けて削除する、授業終了後や単位取得後に削除するといった、不要になったらすぐに削除するという意見も見られた。期待していたデータの再利用の可能性についての言及や、なるべく削除しないという回答は見られず、これは削除のルールと限定して質問したところの影響だと思われる。

設問 12、13 は、セキュリティ対策に関する意識調査である。この設問は、平均点 2.56 で中央値を下回った。セキュリティに関するルールを持っていない学生が多いことが明らかになった。自由記述では、パスワードをかけるやセキュリティソフトを導入するという回答が多かった。特に、クラウド関連やバックアップについての言及は少なかった。

設問 14、15 は、データの公開や提供に関するルールについての設問で、平均点 3.48 で中央値を上回っており、データに関する公開や提供についての意識は高いことがわかった。この設問では、データが漏れないようにする方法についても尋ねた。自由記述では、SNS などには鍵をかける、公開は身内に限るなどの回答があった。また、顔を公開するときに許可を取る、顔などが特定されないようにするなどの回答が多かった。データの公開については、写真ではなく、データやファイルなどへの意識について質問したかったので、質問文に肖像権という文言が入っており、それに引っ張られた可能性があり、あまり目的に合った質問にはなっていなかったと考える。

設問 16、17 は、不慮のデータ削除に関する質問で、平均点は 2.37 となり、バックアップや復元方法などに関する意識が低いことがわかった。

表1 質問項目と平均点

設問番号	質問	平均
3	アンケートなどのデータを取得する際に、紙とペンを使ったアンケートを行う場合、PC やスマートフォンなどの使用してアンケートを行う場合と比較してそのメリットとデメリットを説明することが出来る。	3.24
8	写真のデータ、作成したレポート、アンケートのデータなどを取得して保存する際に自分の中で決めているルールや考え（どのように保存するか、どこに保存するか、バックアップはどのようにしているか等）がある。	3.08
10	写真のデータ、作成したレポート、アンケートのデータなどを取得して保存したものを削除する際に自分の中で決めているルールや考え（削除する判断、その方法、注意点等）がある。	3.04
12	写真のデータ、作成したレポート、アンケートなどの取得したデータを保存したものに対してセキュリティ対策（意図しないデータを他人に取得されるなど）のためのルールや考えがある。	2.56
14	自分のデータを公開や提供（SNS へのアップロードやレポートの提出、ほかの人への提供）するについて自分の中でルール（肖像権、著作権、データの信頼性、セキュリティ、データの使用しやすさ）や決まりがある。	3.48
16	自分の意図しない形でデータが消去もしくは削除されたとき復旧することが出来る	2.37

設問 18 は、自分の重要なデータを保存する主な手段に関する質問であった。一番多かった回答は自分の PC に直接保存するという項目で、本学の学生は PC の必携を求められているため、多くの学生が自分の PC を持っており、このような結果になった。次に多かったのはクラウドへの保存で、37 件であった。重要なデータ保存において、クラウドを利用する学生がある程度いることがわかった。多いと予測されていた USB の利用については、セキュリティ付きが 9、セキュリティなしで 5 であり、合わせて 14 件とクラウドよりも少なかった。

設問 20 は、データを利用してレポートなどを作成する際、元になるデータの扱いについての質問であった。最も

多かったのは、一定期間を設けて保存しておくという回答で 56 件であった。次に多かったのは、速やかに元データを削除することであった。元データはレポート等の根拠資料となるため、なるべく保持してほしいという考えを期待していたが、個人情報や流出のリスクを懸念して削除すると考える学生もある程度いた。できる限り長く保存すると一定期間を設けて保存するという回答を合わせた数は 91 件、元データや本データを削除すると回答した数は合わせて 79 件であった。保存するという回答の総数は削除すると回答した数を上回っていたが、削除すると回答した数もかなり高く、一定数の学生はデータを削除すべきと考え、それらの学生はデータライフサイクルに関する意識が低い可能性があることが示唆された。

表 2. 設問 18.の回答項目と回答数

回答項目	回答数
スマートフォンのアプリ	1
主にクラウド（インターネット上に）保存	37
主に共用の PC に直接保存	2
主に自分の PC に直接保存	116
主に自分の USB（セキュリティ機能あり）に保存	9
主に自分の USB（セキュリティ機能なし）に保存	5

表 2. 設問 20.の回答項目と回答数

回答項目	回答数
データ流出について配慮しつつ元データは一定の期間（5 年等）を決めて保存しておく	56
データ流出について配慮しつつ元データは消さずにできる限り長く保存しておく	35
レポートを作成した後は必要がないため、すべてのデータは目的達成後、個人情報や流出防ぐため速やかに削除しておく	37
レポート作成に使用したデータのみ保存して、元データは個人情報や流出防ぐため速やかに削除しておく	42

### 3. まとめ

この研究では、日本の大学におけるデータリテラシー教育の現状と、リテラシーレベルのモデルカリキュラムについて紹介した。さらに、狭義と広義のデータリテラシーとデータライフサイクルの関係を探るために、大学 1 年生を対象としたアンケート調査を実施した。この調査を通じて学生がデータライフサイクルについてのどのような意識を持っているかを知ることを目指した。調査項目はデータ収集、保存、共有、セキュリティなど、データ管理の各側面への学生の意識と理解に関するものだった。データ収集の項目では、デバイスの違い（紙と PC など）についての理解を尋ねた。自由記述からは、多くの学生がこれらの違いを説明できることがわかった。また、保存に関しては、バックアップへの意識が低い一方で、多くの学生が自分の PC に重要なデータを保存していることが明らかになった。共有の項目では、データ共有に関する質問にもかかわらず、個人情報の公開についての言及が多く、質問の表現に欠点があることが示された。削除に関しては、特に自分が作成したデータの保存についての質問し、長期保存する意向を持つ学生が多い一方で、不要になったデータは削除するという学生もいることがわかった。新たなデータと過去のデータを統合し、新しい知見を得るといったデータライフサイクルの完成に向けた意識を学生に持たせるための学習内容の必要性が明らかになった。

### 参考文献

- [1] 文部科学省, 大学の数理・データサイエンス教育強化方策について,  
[https://www.mext.go.jp/b\\_menu/shingi/chousa/koutou/080/gaiyou/\\_icsFiles/afieldfile/2016/12/21/1380788\\_01.pdf](https://www.mext.go.jp/b_menu/shingi/chousa/koutou/080/gaiyou/_icsFiles/afieldfile/2016/12/21/1380788_01.pdf), 2016
- [2] 総務省, 情報通信白書平成 29 年版,

- <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/pdf/29honpen.pdf>, 2017
- [3] 統合イノベーション戦略推進会議決議, 「AI戦略 2019～人・産業・地域・政府全てにAI～」2019, <https://www8.cao.go.jp/cstp/ai/aistrategy2019.pdf>, 2019 (2023年11月30日閲覧)
- [4] 統合イノベーション戦略推進会議決議, 「AI戦略 2022」, [https://www8.cao.go.jp/cstp/ai/aistrategy2022\\_honbun.pdf](https://www8.cao.go.jp/cstp/ai/aistrategy2022_honbun.pdf), 2022(2023年11月30日閲覧)
- [5] 文部科学省, 数理・データサイエンス・AI教育プログラム認定制度(リテラシーレベル), [https://www.mext.go.jp/a\\_menu/koutou/suuri\\_datascience\\_ai/00002.htm](https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00002.htm), (2023年11月30日閲覧)
- [6] 数理・データサイエンス教育強化拠点コンソーシアム, 数理・データサイエンス・AI(リテラシーレベル)モデルカリキュラム, [http://www.mi.u-tokyo.ac.jp/consortium/pdf/model\\_literacy.pdf](http://www.mi.u-tokyo.ac.jp/consortium/pdf/model_literacy.pdf), 2020
- [7] 中鉢直宏, データライフサイクルマネジメントに基づいたデータリテラシー教育の提案, ISSJ2022第18回 情報システム学会 全国大会・研究発表大会, 2022

#### 補足資料 アンケート調査紙 (Google フォームにて実施)

### データリテラシーに関する調査

次のアンケートに答えてください。記入されたデータに関しては、個人情報などをすべて削除したうえで、研究などの重要なデータとして使用します。調査へのご協力をお願いします。

- 学籍番号
- この質問の回答を研究データとして提供してもよい(個人情報などは必ず配慮いたしますので、データ提供については是非前向きに協力していただけますようよろしくお願いいたします。)  
はい(ご協力をお願いします)  
いいえ
- アンケートなどのデータを取得する際に、紙とペンを使ったアンケートを行う場合、PCやスマートフォンなどを使用してアンケートを行う場合と比較してそのメリットとデメリットを説明することが出来る。  
そう思わない(1-2-3-4-5) そう思う
- 自分が思う紙の用紙を使用する場合のメリットについてか書いてください
- 自分が思う紙の用紙を使用する場合のデメリットについてか書いてください
- 自分が思う PC やスマートフォンなどを使用してアンケートを行う場合のメリットについてか書いてください
- 自分が思う PC やスマートフォンなどを使用してアンケートを行う場合のデメリットについてか書いてください
- 写真のデータ、作成したレポート、アンケートのデータなどを取得して保存する際に自分の中で決めているルールや考え(どのように保存するか、どこに保存するか、バックアップはどのようにしているか等)がある。  
そう思わない(1-2-3-4-5) そう思う
- 写真のデータ、作成したレポート、アンケートのデータなどを取得して保存する際に自分の中で決めているルールや考え(どのように保存するか、どこに保存するか、バックアップはどのようにしているか等)について書いてください。
- 写真のデータ、作成したレポート、アンケートのデータなどを取得して保存したものを削除する際に自分の中で決めているルールや考え(削除する判断、その方法、注意点等)がある。  
そう思わない(1-2-3-4-5) そう思う
- 写真のデータ、作成したレポート、アンケートのデータなどを取得して保存したものを削除する際に自分の中で決めているルールや考え(削除する判断、その方法、注意点等)について書いてください。
- 写真のデータ、作成したレポート、アンケートなどの取得したデータを保存したものに対してセキュリティ対策(意図しないデータを他人に取得されるなど)のためのルールや考えがある。  
そう思わない(1-2-3-4-5) そう思う
- 写真のデータ、作成したレポート、アンケートなどの取得したデータを保存したものに対してセキュリティ対策(意図しないデータを他人に取得されるなど)のためのルールや考えについて書いてください。
- 自分のデータを公開や提供(SNSへのアップロードやレポートの提出、ほかの人への提供)するについて自分の中でル

ール（肖像権、著作権、データの信頼性、セキュリティ、データの使いやすさ）や決まりがある。

そう思わない（1-2-3-4-5） そう思う

15. 自分のデータを公開や提供（SNSへのアップロードやレポートの提出、ほかの人への提供）するについて自分の中でルール（肖像権、著作権、データの信頼性、セキュリティ、データの使いやすさ）や決まり について書いてください。

16. 自分の意図しない形でデータが消去もしくは削除されたとき復旧することが出来る

そう思わない（1-2-3-4-5） そう思う

17. 自分の意図しない形でデータが消去もしくは削除されたときどうしますか。

18. 自分の重要なデータへの保存する主な手段（スマートフォンのアプリ、PC（ローカルもしくはインターネット）はどれですか

- ・主に共用のPCに直接保存
- ・主に自分のUSB（セキュリティ機能なし）に保存
- ・主に自分のUSB（セキュリティ機能あり）に保存
- ・主にクラウド（インターネット上に）保存
- ・その他:

19. 自分の重要なデータへのアクセスしたことがある方法はどれですか（複数回答可）

- ・自分のPCから
- ・共用のPCから
- ・不特定多数が使用するPCから（学校の貸出PC以外、ホテル、ネットカフェ等）
- ・学校の貸し出しPCから
- ・スマートフォンから（ロックのかかっていない）
- ・スマートフォンから（ロックのかかっている）
- ・その他:

20. データを利用してレポートなどを作成するとき、元になるデータをどのようにすることが適切だと思いますか。

- ・レポートを作成した後は必要がないため、すべてのデータは目的達成後、個人情報や流出防ぐため速やかに削除しておく
- ・レポート作成に使用したデータのみ保存して、元データは個人情報や流出防ぐため速やかに削除しておく
- ・データ流出について配慮しつつ元データは一定の期間（5年等）を決めて保存しておく
- ・データ流出について配慮しつつ元データは消さずにできる限り長く保存しておく
- ・その他:

21. この調査への感想について書いて下さい