

研究データ利活用促進のための暗黙知継承に関する 着眼点応用への検討

甲斐尚人¹

概要：オープンサイエンスの動きは世界的な潮流であり、国内においてもその動きの一環である研究データのオープン化に向けた各種取り組みについて活発に議論がなされている。研究公正のための研究データ公開の他に、研究者間での研究データの共有が進むことで新たなイノベーションの創出も期待されている。研究データの利活用が促進されるためには、研究者にその価値を理解してもらうことが重要である。同分野の研究者間だけではなく、異分野間の研究融合も期待され、異分野の研究者を想定したデータ利活用に資する適切なメタデータ付与も重要であると考えられる。このような視点から、利用者に有用なメタデータとは何かを考察する過程は、熟練技術者の暗黙知をどのようにして若手技術者に伝達すべきかを考察する過程に共通していると考えられる。本稿では、経験によってのみ獲得される熟練技術者の暗黙知伝達要素である「気づき」が研究データ利活用促進に必要な着眼点であるかを考察を行う。具体的には、先行研究による研究者へのインタビュー結果を再利用し、そこに含まれるデータ利活用の問題点に対して「気づき」が効果的であるかを考察を行う。

キーワード：研究データ、利活用、暗黙知、気づき

Examination of Application of The Viewpoint on Tacit Knowledge Succession to Promote The Utilization of Research Data

NAOTO KAI^{†1}

Abstract: The movement of open science is a global trend, and various efforts toward the openness of research data, which is part of the movement, are being actively discussed in Japan. In addition to the disclosure of research data for research integrity, it is expected that new innovations will be created by promoting the sharing of research data among researchers. In order to promote the utilization of research data, it is important for researchers to understand its value. It is expected that research will be integrated not only between researchers in the same field but also between different fields, and it is important to assign appropriate metadata that contributes to data utilization assuming researchers in different fields. From this point of view, the process of considering what metadata is useful to users is common to the process of considering how to express the tacit knowledge of skilled engineers. In this paper, we focus on "awareness," which is an element that conveys the tacit knowledge of skilled engineers that has been acquired only through experience, and what kind of effects can be expected based on the results of interviews with researchers from previous research.

Keywords: Research data, utilization, tacit knowledge, awareness

1. はじめに

AI, IoT, ビッグデータ解析などの先端技術が、金融分野や農業分野と融合し「フィンテック」や「アグリテック」などを生み出し、産業や社会に変革を起こしている。このような先端技術やオープンデータなどによる変革は「暗黙知」を「形式知」に変換することにも繋がっており、技術者不足を解決するものとして期待されている。具体的には、例えば熟練技術者の視線をトラッキングした動画データなどを蓄積・分析し、若手技術者が身に付けていない着眼点を表出させる技術開発などが挙げられ、データをもとに新しい知の継承がなされようとしている。[1][2]

研究活動においても、研究データを公開し利活用するオープンサイエンスの考え方が徐々に浸透しつつある。しかし、研究データの利活用を阻む課題として、データを提供

する研究者が重要と考えるデータ情報と、それらを利活用する研究者が必要とするデータ情報とが異なっていることが挙げられる。これによって、研究データの共有、公開が行われたとしてもその利活用が進まないことが想定される。これらの課題を克服するための共通メタデータの枠組みなどが整えられつつあるが、多種多様な研究データを説明する上で重要なデータ記述の記述内容に関する議論はまだ十分に行われていない。

本稿の目的は、これらの問題点の克服に向けて、熟練技術者が若手技術者に暗黙知を伝達する過程で重要な「気づき」に着目し、「気づき」が研究データ利活用を促進させる重要な要素と成り得るか明らかにすることである。

上記の考察には、Yoon によるデータ再利用が進まない原因を明らかにしたアンケート結果の一部を用いる。[3]この調査アンケートは、データ再利用経験のある研究者に対してデータ再利用の失敗経験をインタビューするもので、研究者の生の声を報告している。これに対して、多くの研究

¹ 大阪大学
Osaka University

者が様々な分野で指摘している暗黙知継承に重要な「気づき」について、筆者の鉄道車両メンテナンスの事例研究を紹介しながら、インタビューで指摘される再利用が進まない原因に対して、「気づき」がこの解消に効果的か考察を行う。[4]

2. 関連研究

2.1 オープンサイエンスに関する研究

中村裕一らはフィールドワークにおける全方位カメラを活用し仮想的なフィールドワークに繋げる研究を行っている。[5]このようなデータの共有は研究者の再体験や他研究者などによる新たな視点を生み出すなど、研究推進には重要な動きである。

また、伊藤智明は起業家との対話を逐語記録として残すこれまでにない研究データの保存、共有の研究を行っている。[6] 起業家が試行錯誤し、悩みや課題をどのように克服してきたかなど、組織が出来上がる過程を具体的に記録したものであり、これらのデータの価値を他研究者が理解することは、伊藤とは異なった視点の研究につながる。

2.2 データ再利用に関する研究

Heidi J.Imker らはデータ再利用を阻む障壁について、研究者に直接アンケートを行うことでその一端を明らかにした。[7] Imker らはイリノイ大学の8つの学問分野の各分野における高被引用論文を引用した著者に対して、引用した研究論文で使用されているデータを再利用したかどうか、また再利用しなかった場合その理由は何かについてアンケートを行っている。回答があったもののうち、全体の15%はデータ再利用を行ったとの回答があった。一方データを再利用しなかったと答えた研究者は、主に「引用論文に特化したデータであり、自身の論文に有効なデータではない」(58%)といった理由を挙げた。さらに「新しい研究になぜ過去のデータを利用する必要があるのか」といったコメントもあり、研究者がデータを再利用する価値を見出せていない、つまり他研究者が作成した研究データに無関心である、という現状を報告した。

Jinfang Niu はデータ記述の不足がデータ利活用の促進を妨げるものであるとした。[8] データ記述が不十分な理由として、第一に、データ作成者とデータ利活用者との利害関係がデータ作成者の情報提供の意欲を減衰させること、第二に、直接的なコミュニケーションの減少によって必要な情報の優先順位を決める必要があり、データ作成者のみを知る暗黙な情報は優先度が下がるためとしている。Niu はこれらの克服方法として、データ利活用者の周辺知識を増加させることやデータ作成者とデータ利活用者のコミュニケーションチャンネルを増加させることなどが必要であると説明している。

これらの研究はデータ再利用の現状を明らかにした上で重要な指摘を行っているが、データを共有する側の研究者に求められる要素の考察までは行っていない。池内有為はエディンバラ大学の研究データ管理におけるメタデータ事例に触れ、研究データの共有において、メタデータはエディンバラ大学のような基本的な要素にとどめ、メタデータに寄らない自由記述が可能な「データ記述 (data description)」を充実させるべきであるという提案を行っている。[9] 本稿で行う考察がデータ記述の共通したノウハウを明らかにすることに繋がることを期待する筆者と考え方が一致する。

3. 技術継承と研究データ利活用の考察

3.1 暗黙知の伝達要素

まず、筆者がこれまで行った暗黙知抽出のための研究について紹介する。[10][11]

まず、この研究を行った背景について説明する。多くの企業は団塊世代技術者の大量退職による技術継承の断絶を防ぐために、マニュアルの整備に励んだ。しかし、急速に整備されたマニュアルは技術継承に必要な不可欠である暗黙知という要素を欠いたものが多く、実用的でなく使用されていないものも多い。筆者が経験した鉄道車両メンテナンスの分野では、この問題を認識するものは多くなく、若手社員の経験不足による作業ミスが度々生じてきた。また、団塊ジュニア世代の退職が迫っており実用的なマニュアルの作成が急務であることから、過去の鉄道車両の故障事例や既存のマニュアルを比較・考察することによって、認識すべき暗黙知伝達の着眼点や暗黙知の所在を明らかにする研究を行ってきた (図1)。

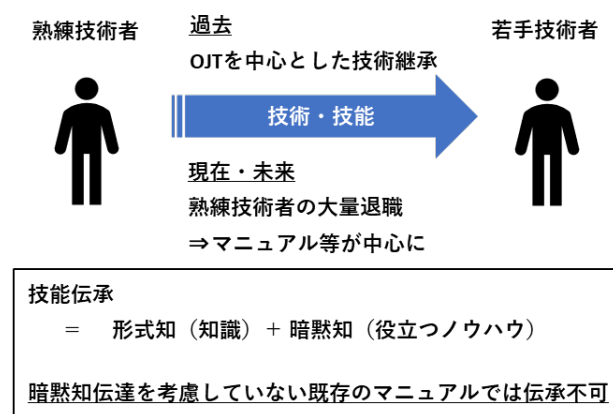


図1 暗黙知の技術継承

具体的には、車両故障事例について取り上げ、車両部品の故障原因を分析する過程に着目した。真の原因が究明されるまで誤った対策が行われ、再度同様の故障を発生させた事例である。これは、部品の折損原因を通常考えられる金属疲労として単純に解釈するのではなく、車両の設計変

更によって部品に過負荷が生じた、という例外として解釈することが重要であった事例である。これは破断面分析結果に対する「先入観」が、根本原因である「例外」的故障の存在を隠し、「気づき」を遅らせることを示していた。ここから、車両故障分析および対策には、「気づき」を与える要素として「先入観」、「例外」に関する情報が必要であると言えることができる。（図2）

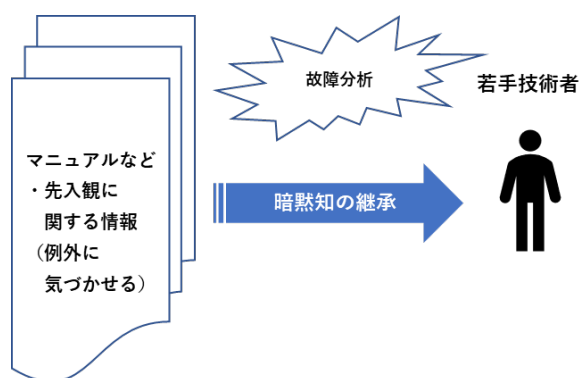


図2 「気づき」を与える技術継承

3.2 研究データが利活用されない要因

データ再利用に関する以下の先行研究に着目した。Yoonは、他の研究者が生成したデータについて、その再利用を断念した経験をもつ研究者にインタビューを実施した。[3]これらの経験は、既存データを再利用して意図した研究成果（ジャーナル出版、プレゼンテーション、研究報告書など）に至らなかった経験と、研究者が何らかの理由でデータ再利用を断念せざるを得なかった経験と定義されている。研究の方向性の変更やより良いデータセットの探索だけでは再利用を断念した経験としてみなしていない。調査対象である研究者は、少なくとも1度はデータを再利用した経験をもつ23人の定量的社会科学データ利用者であり、彼ら彼女らからデータ再利用に失敗した経験をインタビューすることで、その原因を考察している。以下に、論文中で報告されているインタビュー結果の中で、本研究と関連するものについて示す。

【I05】“I usually start considering data that I already know what's in there..., but the only times that [not using data] happened that when I found the data didn't have the variables or sample size that I thought the dataset had.”

【I08】“[The original investigators] did not capture all like [ethnic group A] people.”
“did not think all other scholars are aware of [the problem with samples though] because [you'd know] if you are studying a specific party. So I was really focusing on the same statistics [as

well as populations]; that's why I got to know.”

【I11】“I didn't have a lot of faith in this particular dataset [because] for the [variables] that I was particularly interested, there was just too much missing data.”, “it's a lot of time wasted”

【I13】“looking at the data [themselves] was kind of challenging as well, cause you don't know unless you really dig into [them].”

【I16】“as I looked at that more and how they coded things, I thought it was a bit off.”

“They're saying this group is both [A identity] and [B identity]. Well, that's kind of a contradiction. ... So at that point I ended up kind of rejecting [original] codes.”

【I22】“sometimes, people have very short timetables they're working on for whatever reason, it might not be worth it to invest all that time and energy to gaining access and getting set up with the data.”

3.3 技術継承と研究データ利活用の関係

データ利活用を技術継承に当てはめると、図3のようになる。データやメタデータは形式知に分類され、それらでは得られないデータ提供者が有する情報（データ記述）が暗黙知に該当する。3.1項、3.2項をもとに、インタビュー結果が示唆する研究者が感じるデータ再利用の障壁に対して、「気づき」の視点が有効であるか4章で考察を行う。

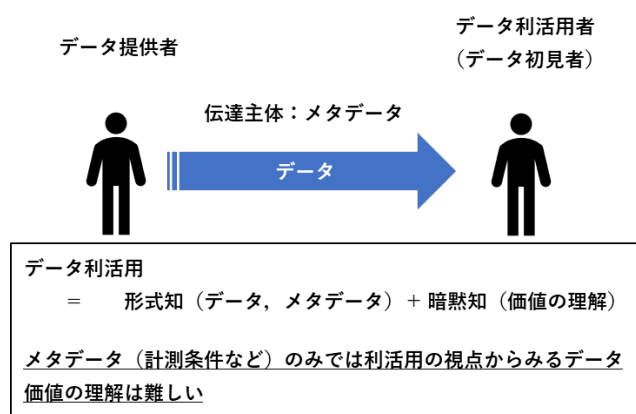


図3 データ価値伝達のイメージ

4. 「気づき」の研究データ利活用促進への応用の考察

研究データ再利用を断念した経験についてインタビューを行った Yoon の報告にはプロセスの具体的な内容までは含まれていない。[3] そのため、表1に各インタビュー内

容 (I05, I08, I11, I13, I16, I22) のデータ再利用を断念した理由から想定される問題点について、表 1 にまとめる。

表 1 インタビュー結果から想定される問題点

	データ再利用断念の理由	想定される問題点
I05	データセットに必要な要素が不足している	提供者のデータ公開に対する意識
I08	研究範囲が異なることでデータに偏りがある	データ説明の不足
I11	データが欠損している	入力ミス
I13	データ理解に時間を要する	データ説明の不足
I16	データが矛盾している	収集ミス
I22	データ利用までの準備に時間を要する	データ説明の不足

I11, I16 のデータの欠損、データの矛盾については研究者のミスである可能性もあるため、メタデータの正確な入力の啓発、もしくはシステムのチェックの導入などが解決策として考えられる。またメタデータ作成を支援するデータライブラリアンやキュレーターの役割も今後大きくなると考えられる。一方で、Yoon によると上記のようなデータ再利用の躓きに対して、これまでのデータ再利用の経験を活かして解決したという研究者のコメントがあったことを紹介しており、データ再利用の経験の差が影響している可能性も示唆した。[3]

4.1 研究データ理解に要する時間

I08, I13, I22 のインタビュー結果に加えて、研究者間でデータの再利用を諦めた時点について、早い段階で諦めた研究者もいれば、データ再利用の検討にかなりの時間を費やした後に諦めなければならなかった研究者もいるなど、研究者によってデータを理解する時間にばらつきがあることが報告されている。ばらつきの可能性の一つとして、データ再利用経験をどのくらい積んでいるかに影響されていると考えられる。また実際に経験豊富な研究者はデータに不明な点を感じた際、早い段階でデータ作成者と直接コミュニケーションを取り、その障害となる問題を取り除いており、データ再利用の経験が浅い研究者にとってデータ記述は大きな障害と成り得ると指摘している。

さらに、I13 のインタビュー結果には、データを掘り下げて調べない限り詳細がわからないとある。自力で他研究者が作成したデータを深く理解しようと試みることは非常に時間を要する。その労力が必要な場合も当然考えられるが、掘り下げるに値するデータであるのか、つまり自身の研究に必要な基準を満たしたデータであるかを早期に判断できる要素がデータ記述に含まれていることは、研究者の次のデータ再探索の行動を促し、データ再利用の好循環を生むと考える。(図 4)

このような点において、利活用を促進する「気づき」を補助するデータ記述は暗黙知を有しない研究者の判断の時間を短縮する効果があると考えられる。

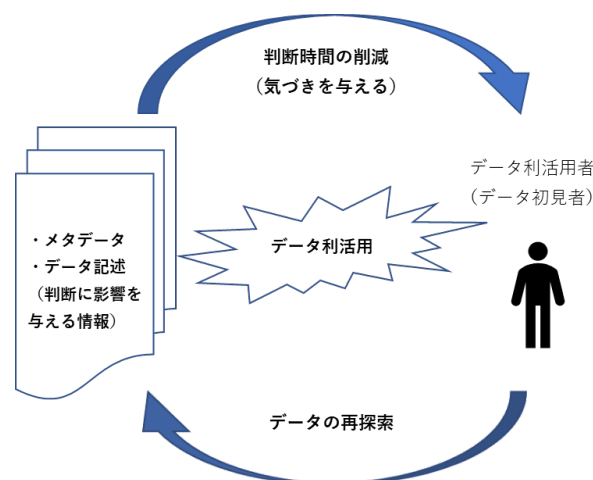


図 4 データ探索の好循環

4.2 データ記述の現状

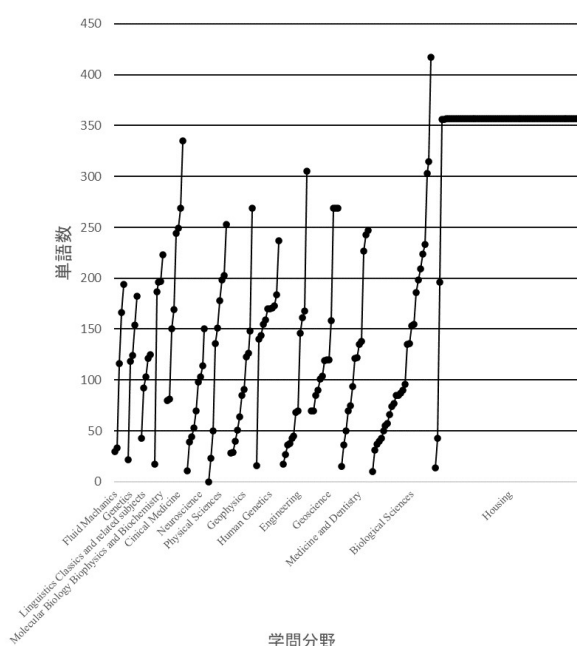
研究者がデータ再利用を途中で断念する要因として、データ再利用の可能性の判断までに時間がかかるという点に注目した。これは、判断に必要な情報が十分でないことを示唆している。データ利活用を考える際、計測機器や実験条件などのメタデータが正確であることが当然求められるが、研究の背景や目的に応じて収集されたデータが、データ利活用を検討する研究者が求めるデータと異なる状態であることは十分に考えられる。

では、どのようなことが想定し得るか考える。有機化学分野における「気づき」の情報について考察する。化学合成物を生成した際、そのメタデータには実験環境や条件が示されるはずである。実際に、国立研究開発法人物質・材料研究機構が運営するデータリポジトリ Materials Data Repository (MDR) で付与可能なメタデータを見てみると、入力が必要である共通メタデータ（識別子、記入者、物質、装置、データ起源など）、分野別のメタデータである計測データ（計測法、測定環境など）、試料データ（物質タイプ、構造的特徴など）、特性メタデータ（特徴的性質）、合成・プロセスメタデータ（処理日、処理温度）、計算メタデータ（計算機、ソフトウェアなど）があり、計測主要パラメータ、試料主要パラメータ、特性主要パラメータ、合成・プロセス主要パラメータ、計算主要パラメータと続く。[12][13] 例えば実験の化合物抽出過程で使用した溶剤などはメタデータ項目としてないが、水溶性を鍵とする研究をしている研究者には実験途中の溶剤などの情報も必要かもしれない。このような情報に関して、MDR では任意デー

タが入力できるようになっている。現在 MDR には 1,869 個のデータセットが格納されている（2022 年 1 月 21 日現在）が、任意のデータ記述は確認した範囲では見られなかった。またデータには偏りがあり、具体的な分析はまだ難しい。

データ記述について、池内の事例紹介を参考に、エディンバラ大学が運営するデータリポジトリ Data Share に登録されたデータのデータ記述の登録状況などについて確認を行った。[14] 2021 年中に登録されたデータセットは 298 件であり、それらのデータ記述について調査を行った。データ記述が登録されている 298 件のうち、分類された各学問分野で 5 件以上データが登録されている学問分野について分析を行った。各学問分野毎にデータ記述単語数が少ない順にプロットしたものがグラフ 1 である。ほとんどのデータにおいてデータ記述がなされていたが、データ記述された単語数は同じ分野であってもかなりのばらつきがある。またデータに関する記述ではなく、データを掲載した論文の抄録をそのまま載せたものもあれば、データに関する背景について詳細に記述されたものもあった。さらに、「Housing」に関してはモダン・ムーブメントにかかわる建物と環境形成の記録調査および保存のためのデータであり、データ登録者が同一で、データ記述が同じ内容であることが課題であると感じられた。

上記のように、単語数に着目するとほとんどの分野でばらつきがある一方で、全く同じデータ記述も確認された。これらの情報や背景から「気づき」を与える要素について具体的に分析を行うことが今後の課題である。



グラフ 1 登録されたデータのデータ記述の単語数

5. おわりに

本稿では、データ利活用の促進のために技術継承における暗黙知伝達に必要な「気づき」の要素が適用可能かについて考察を行った。先行研究のインタビューを通して、研究データの利活用には既存のメタデータのみでは、その利活用を判断できるまでに時間を要し、「気づき」の要素を含むデータ記述は研究者の利活用を検討する時間削減の効果がある可能性があることがわかった。

今後、特にオープンサイエンス過渡期においては、データ利活用の判断に多くの時間を費やす研究者が生じることが想定され、多くのデータにアクセスする時間が確保できない可能性がある。研究者の中には少ないメタデータから自身の要求に合致したデータか否かについて自身の経験から予測する術を持っているかもしれないが、特に異分野研究者やデータ初見者にとっては「気づき」を導く情報は重要である。このような記述は、データ探索に費やす時間を確保することを可能にし、関連研究で紹介した Imker の調査が示した研究者のデータ利活用への無関心さに対しても対処できると考える。[7]

このように、「気づき」を意識したデータ記述の充実、データ利活用の検討、データ探索の好循環を生む可能性があり、今後「気づき」を与える要素のデータ記述に向けた分析が課題である。

参考文献

- [1] 宇田川洋一, 木村崇. 建築設備の技術伝承・技能伝承に関する研究 (第 1 報) 施工品質検査におけるアイトラッキングの活用検証. 令和 2 年度大会 (オンライン) 学術講演論文集, 2020, vol. 1, p. 129-132.
- [2] 熟練技術者の勘・コツ・暗黙知を可視化する技術伝承ソリューション.
<https://www.acuity-inc.co.jp/pickups/knowhow/docs/20201012/>.
 最終アクセス日 2021.1.20.
- [3] Ayoung Yoon. Red flags in data: Learning from failed data reuse experiences. Proceedings of the Association for Information Science and Technology, 2016, vol.53, issue1, p.1-6.
- [4] 江藤香, 松居辰則, 椋田實, 樺澤康夫. 初心者への気づきによるケアプラン策定過程におけるノウハウ情報表出の方法とその評価. 人工知能学会身体知研究会 SIG-SKL, 2010, p.1-8.
- [5] 中村裕一, 近藤一晃, 赤石大輔, 徳地直子. フィールド科学教育・研究のためのフィールドワーク体験蓄積とサイバーフィジカル教育研究支援. 情報知識学会誌, 2021, 31(4), p. 428-433.
- [6] 伊藤智明, 起業家との対話の逐語記録の作成と蓄積についての現状と課題. 情報知識学会誌, 2021, 31(4), p.434-439
- [7] Heidi J.Imker, Hoa Luong, William H.Mischo, Mary C.Schlembach, Chris Wiley. An examination of data reuse practices within highly cited articles of faculty at a research university, The Journal of Academic Librarianship, 2021, vol.47.
- [8] Jinfang Niu. Overcoming inadequate documentation. Proceedings of the American Society for Information Science and Technology, 2010, vol.46, issue1, p.1-14.
- [9] 池内有為. 大学図書館による研究データ管理の最前線～研究力を強化するエディンバラ大学の事例～. 現代の図書館, 2014, vol.52, No.4. p.227-236.

- [10] 甲斐尚人. 鉄道車両の故障分析による暗黙知の抽出に関する研究. レコードマネジメント, 2020, No.79, p. 19-33.
- [11] Naoto Kai, Kota Sakasegawa, Tsunenori Mine, Sachio Hirokawa. Machine Learning of Ambiguous Sentences and Analysis of relation between Ambiguous Sentences and Diagrams in Technical Manual for Tacit Knowledge Acquisition, International Journal of Smart Computing and Artificial Intelligence, 2021, vol.5, no.1, p. 19-32.
- [12] MDR(Materials Data Repository)
<https://dice.nims.go.jp/services/MDR/>
- [13] 松田朝彦. 材料データリポジトリにおける共通メタデータ・分野別メタデータ. Japan Open Science Summit 2019.
- [14] Data Share. <https://datashare.ed.ac.uk/>