「株式会社技術情報協会 月刊 研究開発リーダー 2026年1月号」掲載記事
URL:https://www.gijutu.co.jp/index.htm

◆連　載◆

《隔月連載 全 5 回》　　第 3 回

R & D 部門における機械学習・AI・生成 AI 活用へのデータ共有の重要性

上島 豊　　（株）キャトルアイ・サイエンス　代表取締役

《PROFILE》

略歴：

1992 年　3 月	大阪大学工学部　原子力工学科　卒業
1997 年　3 月	大阪大学大学院工学研究科　電磁エネルギー工学専攻　博士課程修了
1997 年　4 月	日本原子力研究所　博士研究員
2000 年　4 月	日本原子力研究所　研究職員
2006 年　3 月	日本原子力研究開発機構（旧日本原子力研究所）　退職
2006 年　4 月	キャトルアイ・サイエンス設立　代表取締役　就任

主な参加国家プロジェクト：

文部科学省 e-Japan プロジェクト「ITBL プロジェクト」,「バイオグリッドプロジェクト」
総務省 JGN プロジェクト「JGN を使った遠隔分散環境構築」
文部科学省リーディングプロジェクト「生体細胞機能シミュレーション」

主な受賞歴：

1999 年　6 月	日本原子力研究所　有功賞 「高並列計算機を用いたギガ粒子シミュレーションコードの開発」
2003 年　4 月	第 7 回サイエンス展示・実験ショーアイデアコンテスト文部科学大臣賞「光速の世界へご招待」
2004 年 12 月	第 1 回理研ベンチマークコンテスト　無差別部門　優勝

主な著作：

培風館「PSE book ーシミュレーション科学における問題解決のための環境（基礎編）」ISBN：456301558X
培風館「PSE book ーシミュレーション科学における問題解決のための環境（応用編）」ISBN：4563015598
培風館『ペタフロップス コンピューティング』ISBN978-4-563-01571-8
臨川書店『視覚とマンガ表現』ISBN978-4-653-04012-5

7　データ蓄積，共有で研究の何が改善できるのか？

前章では，「R & D 部門におけるデータ蓄積，共有，利活用の実情」と題し，データ蓄積，共有，利活用の実情に関して紹介し，蓄積データを使って AI，機械学習等を行う場合の注意点の説明を行った。本章では，「データ蓄積，共有で研究の何が改善できるのか？」と題し，そもそもデータ蓄積，共有で，「どのようなことが改善できるのか？」，逆に言うと「どのような改善は期待してはいけないのか？」を論じる。

「データ蓄積，共有で研究の何が改善できるのか？」を論じる前に，まず，「研究」というもの自身がどういう要素から成っているかを考えてみる。研究は，閃きや発想という独創的な力と，データを分析し，傾向を把握する力という 2 つの力から成っている。閃きや発想は，属人的，主観的なもので，それが研究者の個性的能力ともいえる。一方，データを分析し，傾向を把握する力は，一見，閃きや発想と同じように属人的，主観的なもののように感じるが，よく考えてみると数学や統計などの客観的方法で確立されるべきものである。実際，同じデータを使って，分析し，傾向を把握したとすると分析結果の解釈以外は，同じ結果になるべきものである。また，分析結果の解釈も単なる閃きや発想と比較すると属人的，主観的というよりも結果解釈時に何に重きを置くのか？何を無視するのか？という取捨選択のパターンという客観的なものとして考えることも可能である。つまり，データを分析し，傾向を把握する力は，主観的でなく，客観的なものだということである。

前述のことを念頭に置きながら，「研究とは何か」を少し考えてみる。大きく想像を膨らませてほしいのだが，「大航海時代に冒険家が世界の海を渡り，次々に新発見をしていった過程」と，「研究者が研究で新しい発見をしていく過程」は近いものがあるのではないだろうか？そう考え，研究を航海に例えてみると，「航海士の経験，勘は，研究者の閃きや発想」に対応し，「航海での海図と羅針盤とその利用技法が研究でのデータとデータ抽出，分析手法」に対応するのではなかろうか？どれだけ優秀な航海士もでたらめの海図と羅針盤とその利用技法では，まともな航海はできないように，研究もデータ

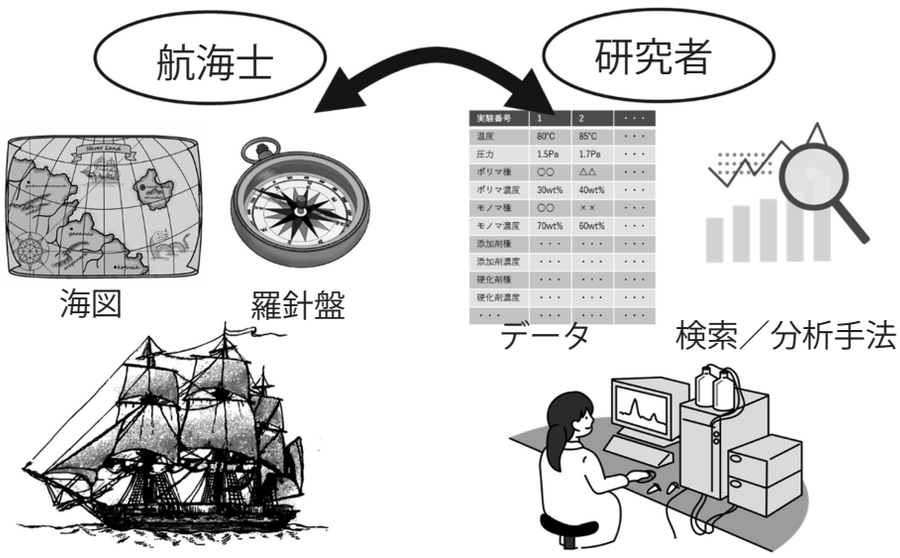

図 15　研究者のデータと検索／分析手法は，航海士の海図と羅針盤と同じ

とデータ抽出，分析手法がでたらめでは，まともな研究はできない。実際，海図と羅針盤とその利用技法が航海士で共通のものであるべきであるように，データとデータ抽出，分析手法も研究者で共通であるべきで，それによって研究の客観性を担保するのである。優秀な航海士は，その客観的情報（海図と羅針盤とその利用技法）に，自身の主観的な経験，勘を組み合わせることで，他の航海士より，危険な海をより安全に航海できる。当然研究者も客観的情報（データとデータ抽出，分析手法）に，自身の主観的な閃きや発想を組み合わせることで，他の研究者ではできない発見が可能なわけである。「研究と航海」，非常によく似た対応関係になっているように見えないだろうか。

本題に戻って，「データ蓄積，共有で研究の何が改善できるのか？」について，考察を進めよう。研究は，閃きや発想という独創的な力とデータを分析し，傾向を把握する力という客観的な力の２つの力から構成されているので，「データ蓄積，共有で研究の何が改善できるのか？」は，「データ蓄積，共有で閃きや発想が改善できるのか？」と「データ蓄積，共有でデータ分析や傾向把握が改善できるのか？」に言い換えることができる。「閃きや発想」は，属人的，主観的なものなので，改善できるとしても研究者ごとにその改善内容や改善度合いは異なってしまう。また，「閃きや発想」というものは，定量的に評価しにくいものであり，改善されたかどうかを判定すること自体が非常に難しい。一方，「データ分析，傾向把握」は，根本的には数学的，統計的データ処理やグラフ化などの客観的方法で実施されるべきもので，主観ではなく客観的な対象データ自体によって決定づけられるものである。つまり，数学的，統計的データ処理やグラフ化の結果解釈以外，すなわち，数学的，統計的データ処理やグラフ化自体は主観的でなく，客観的なものだということである。

原理的には，データ蓄積，共有で「閃きや発想」も「データ分析や傾向把握」も改善は可能ではあるが，主観が絡み合った部分は，改善指針や改善評価自体を研究者個人個人に合わせこむ必要があるので，負担が大きく，取り組みにかかる負担が改善による効用を上回ってしまう可能性が高く，最初に取り組むべき課題ではない。したがって，改善指針や改善評価が人によらない数学的，統計的データ処理やグラフ化部分をデータ蓄積，共有での最初の改善対象と考え，その改善に道筋ができたあとで，数学的，統計的データ処理やグラフ化の結果解釈部分をデータ蓄積，共有での改善対象として，取り組むべきである。

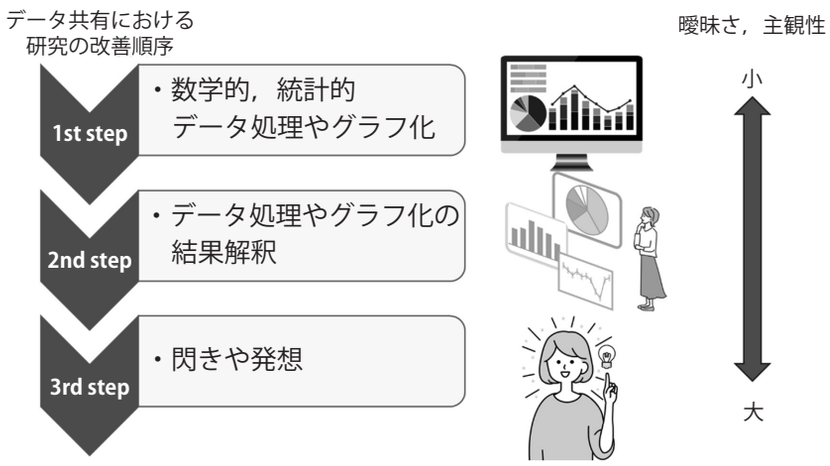

図 16　データ蓄積，共有における研究の改善順序

「閃きや発想」は，直接的な改善を期待するのではなく，上記の改善で「閃きや発想」にかけられる時間的余裕ができ，新しい「閃きや発想」が生まれる機会が増えると考えておくのが現実的である。もちろん，直接的な改善は絶対ないということではない。もしそういうことがあれば，蓄積，共有したデータをどのように扱えば「閃きや発想」が改善されるのかを客観的，統計的に検証し，他の人でも「閃きや発想」の改善が確実にできるような方法を確立していけばよい。実際，こういう客観的，統計的検証を経るまでは，「閃きや発想が改善している」ということ自体が主観的な思い込みの可能性が高いので，信用してはいけない。

ここで，少し逆説的な話をしておく。ここまでで，データ蓄積，共有での最初の改善対象は，データ処理やグラフ化部分と考えるべきと述べた。それは本来客観的に扱える部分を主観的に扱っていたものを客観的に扱えるようにすることで，研究の質的改善を促そうというものであった。航海との比較でもこの質的改善の意義は理解して頂けたと思う。しかし，実はデータ解析やグラフ化，AI，機械学習を使って研究を高速化（目標の性能改善が図れるまでの時間を短縮）することを目的として，データ蓄積，共有は行ってはいけないのである。航海の話に例えると，高速に航海をするということを目的として，海図データを増やそうとしてはいけないということである。海図データを蓄積，共有するのは，航海を客観的に行えるようにし，航海の安全性，安定性を確保するという観点で航海を質的に改善するものである。実験データの蓄積，共有も実験自身に客観性を付与し，実験の再現性や客観的知見を共有することで，研究を質的に改善するものなのである。

この言明は，「蓄積されたデータを使って，データ解析やグラフ化，AI，機械学習を行ってはいけない」ということを言っているのではない。この言明で何を伝えたいかというと，「蓄積されたデータを使って，データ解析やグラフ化，AI，機械学習を行ったところで，研究が高速化される保証はない」ので，「データ解析やグラフ化，AI，機械学習による研究の高速化を目的として，データ蓄積，共有を行おうと考えてはいけない」ということである。研究の質的な改善が高速化というような量的な改善にはつながらないということを言っているのでもない。データ蓄積でデータ解析やグラフ化，AI，機械学習も改善することは間違いない。しかし，データ解析やグラフ化，AI，機械学習による研究の改善が研究自身の高速化に影響を及ぼす程度の量的改善になる保証は何もないということなのである。研究だけに限らず，こういう因果関係の話を単純に考えすぎて，間違った判断をしてしまうことは多いので，これを機に心に留めておいてほしい。「研究を高速化しろ」と言われて取り組みを行っているのに，そんなことを言われても･･･。そもそも，「データ解析やグラフ化，AI，機械学習を使わずに研究を高速化することなんて，できないのでは？」と今思ってしまっている人は，思考が固定パターンに嵌り過ぎている状況で，研究を進める上では危険な状態である。

図 17　データ蓄積，共有が不完全だと研究に様々な無駄が発生

私が様々な組織でデータ蓄積，共有を推進してほしいと頼まれるときに，最初に行うことがある。それは，現在の研究，実験において，困っていること，時間がかかっていることをブレインストーミング的に洗い出すことである。もちろん，「目標の特性改善を達成するまでに時間がかかり過ぎる」，「安定した計測ができる装置の使い方が分からない」というようなものが出てくると思う。ただ，その他にも「過去にこんな実験をしたと思うのだけど，Windows のフォルダを探し回ってもなかなか見つからない」とか，「レシピを記録したと思われるファイルを見つけたのだけれど，材料名が曖昧に書かれていて，どんな材料かわからなかった」とかもあるはずである。このような単純な困りごとは，そういうことが年に何回ぐらいあって，1 回あたりどれくらいかかっているかを積算してみると，結構な時間がかかっていることに気づくはずである。

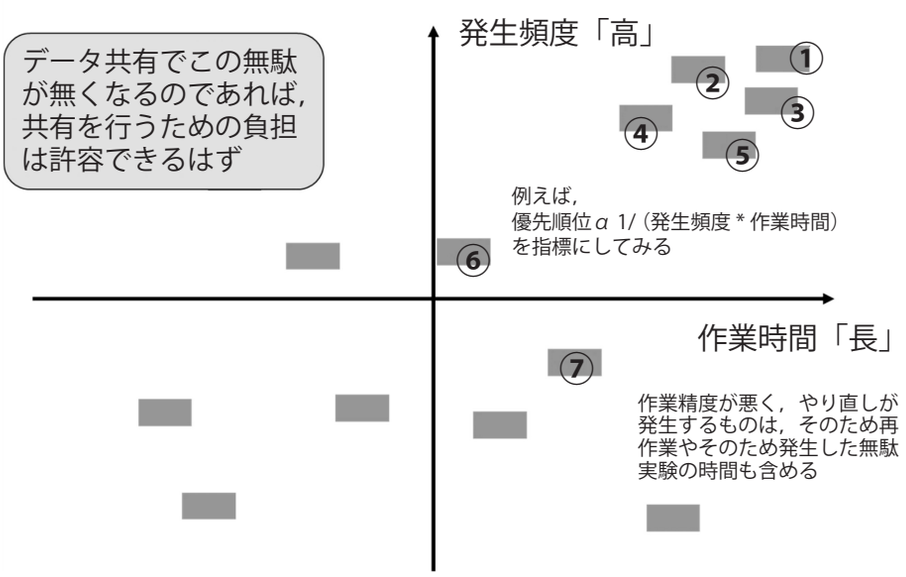

図 18　困っていること，時間がかかっていることを定量化

データ蓄積，共有をしっかりして，データベースなどにデータを格納しておき，いつでも素早く探し出せるようにしておくと，これらの時間が大幅に短縮される。これは，確実に「目標の特性改善を達成するまでの時間が短縮」される時間になるのである。コロンブスの卵のようなことではあるが，特性改善をするのに解析や AI，機械学習などの高度な技法を用いなくとも，確実に改善できる方法は山のようにあるわけである。「特性改善をする高度な技法」は，そういうコロンブスの卵で節約できた時間で挑戦すればいいわけである。宝くじは買わなければ当たらないのと同じで，「特性改善をする高度な技法」もその開発に挑戦しなければ手に入らない。しかし，「特性改善をする高度な技法」は，必ずそういう技法が見つかる保証はない。だから，生活に支障のでない程度で購入する宝くじのようにコロンブスの卵で節約できた時間を超えない範囲で，挑戦し続ければいいわけである。

最後にもう一言。「目標の特性改善を達成するまでに時間がかかり過ぎる」という課題はどうしようもないのだろうか？「特性改善をする高度な技法」を見つけようとする前にやるべきことがある。「目標の特性改善を達成するまでに時間がかかり過ぎる」のは，何故なのか？をしっかりと問うべきなのある。いわゆる，「なぜなぜ分析」の要領で，要因を分解していくのである。そうすると，その要因のいくつかは，「過去にこんな実験をしたと思うのだけど，Windows のフォルダを探し回ってもなかなか見つからない」とか，「レシピを記録したと思われるファイルを見つけたのだけれど，材料名が曖昧に書かれていて，どんな材料かわからなかった」とかのように，確実に改善策が見つかる課題に分解される可能性がある。そういう確実なものを確実に改善していき，確実とは言えないけれど，改善できれば大きな改善になりうる挑戦すべきものを明確化，先鋭化させていくのである。明確化，先鋭化されないと挑戦しようにも何をどう挑戦すればよいかが分からないはずで，右往左往するだけになってしまう。したがって，挑戦すべきものは他の要因を削ぎ落とすことで挑戦方針を明確化，先鋭化してから，他の改善で節約できた時間で挑戦すべきなのである。

8 R & D 部門におけるデータ蓄積を Excel で行えるか？

前章では，「データ蓄積，共有で研究の何が改善できるのか？」と題し，そもそもデータ蓄積，共有で，研究の「どのようなことが改善できるのか？」，逆に言うと「どのような改善は期待してはいけないのか？」を論じた。本章では，「R & D 部門におけるデータ蓄積を Excel で行えるか？」と題し，R & D 部門においてデータを Excel で蓄積していくことが可能か？課題があるのであればどんな課題があるのか？を解説する。

R & D 部門におけるデータ蓄積が他の部門のデータ蓄積と大きく異なるのは，項目（名）の多さと項目（名）の追加，変更頻度の高さである。十数名レベルの R & D でさえ，全実験パラメータ，計測パラメータ及び計測値すべての項目を数え上げると数千項目に上ってしまう。また，数週間毎に，頻度が高い場合は毎日のように項目（名）の追加，変更が必要となる。R & D 部門では，常に新しい材料を評価し，新しいプロセスを考案し，日々試作物の性能向上を目指すという業務の特性上，致しかたないところである。通常の事務系や営業系や生産ライン系の業務では，こういうことは起こらない。そういう意味で，データ蓄積，共有が一番難しい部門と考えて，間違いではない。

R & D 部門におけるデータ蓄積，共有，利活用がうまくいっていない第一の理由は，本連載の最初に話をした「第三者が実験及び解析を再現するために必要な情報を記録」がなされていないことである。ただ，データ蓄積，共有に挑戦しようとした組織では，この問題はクリアされおり，問題はこれだけではないことは明らかである。データ蓄積，共有，利活用がうまくいっていない第二の理由は，項目名が統一されていないことである。項目名の統一には，利用者皆での合意形成と項目名の定義辞書のようなものが必要ではあるが，それができるのであれば Excel 程度でも項目名が統一の仕組みを作ることは可能である。

研究者皆が参照可能な共有ファイルサーバに項目名を列挙したテキストファイルを置いておき，実験データを記録する Excel には，この項目名を列挙したテキストファイルを読み込み，項目名のドロップダウンリストを自動生成する仕組みをマクロで組み込めばいいのである。そして，ファイル保存時に項目名として認められ

ない（項目名を列挙したテキストファイルに存在しない＝ドロップダウンリストに存在しない）項目名を書き込んでいたら Excel を保存できない仕組みにしてしまえばよい。項目名が追加，変更された場合は，項目名を列挙したテキストファイルを変更すればそれを参照する全 Excel も項目名が追加，変更される。そうであるならば,「Excel だけでデータ蓄積は問題ないのではないか？」と思われるかもしれない。それは，半分正解で，半分は間違いである。

研究者が各自の Excel に実験データを書き込む運用の場合，研究者ごと別ファイルになるためデータを探すときに結局すべてのファイルを開いて確認するしかない。また，ファイルごとの項目並びも一定でないので，1 ファイルへまとめることも困難で，データの蓄積はできてもデータを探すのは難しい。結局，自分のデータだけの利活用は進むが，他者のデータを含むデータ共有，利活用は進まない。

共有ファイルサーバに 1 個のマスター Excel を置き，研究者がマスター Excel に実験データを直接書き込む運用のケースもある。その場合，マスター Excel には同時に書き込めないので，書き込み待ちが頻繁に発生する。研究者は，実験をしながら入力をしたいはずであるが，それをすると長時間マスター Excel を占有することになるので，禁止されることが多い。一時的に個人の Excel に実験データを入力し，後からマスター Excel へ書き込むようにすべきだが，結局，面倒でマスター Excel への書き込みをしないままになってしまうことが多い。また，その使いにくさを避けるために，マスター Excel を自 PC にコピーし，自 PC で入力後，共有ファイルサーバのファイルを上書きする人も出てくるのだが，多くの人がそのようなことをしてしまうと，一部データが欠損しまったり，複数のマスターファイルができて，収拾がつかなくなってしまう。

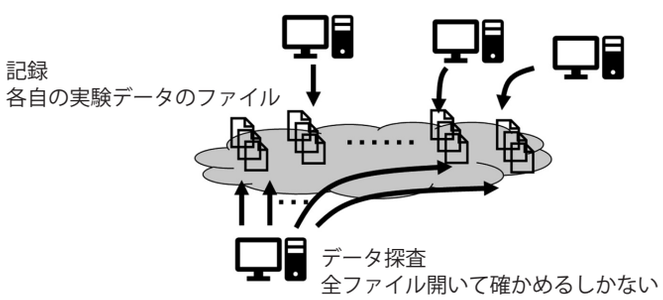

図 19　研究者が各自の Excel に実験データを書き込む運用のイメージ

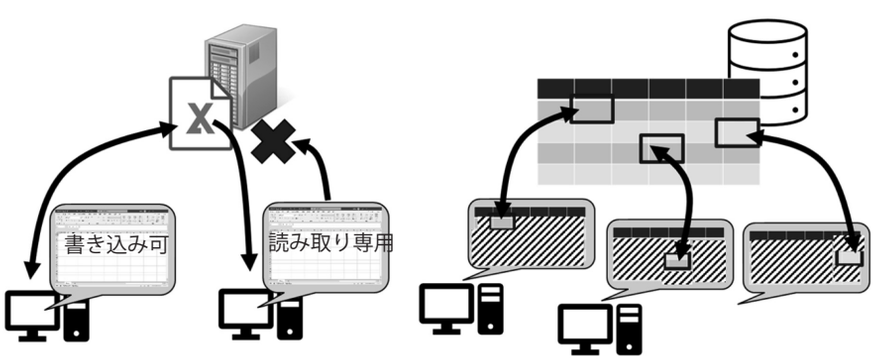

図 20　研究者がマスター Excel に実験データを直接書き込む場合のイメージ

Sharepoint を使えば，マスター Excel には同時に書き込みは可能になり，上記状況は幾分改善するが，それでもマスター Excel に自分の実験データを入力したり，転記するのは，思った以上に大変である。R & D だと 1 実験当たりの項目数が百近くになり，そういう実験が数百集まると実は数千もの項目になる。マスター Excel を作り，そこに入力するということは，列数が数千もの Excel から入力する百程度の項目を探しだし，入力しなければならないのである。また，変更する必要のない項目が数千もあり，常に表示されているので，誤入力，誤編集リスクも大きく，他者の実験データを間違って上書きしてしまうことも頻繁に発生する。さらに，データ蓄積が進んでくると Excel を開いたり，編集する処理がどんどん重くなり，使用に耐えられない状況になってしまうのである。

「変更や確認する必要のない実験（行）や項目（列）を表示せず，変更，確認する実験や項目のみ表示し，複数の人が同時に値を変更できるようにすることで誤入力，誤編集リスクを抑制する」は，事務系業務でも必要なことである。実は，事務系業務で，これらを実現するためにデータベースというものが作られたのである。

参考文献

1）川田重夫，田子精男，梅谷征雄，南多善，上島豊，他
PSE book －シミュレーション科学における問題解決のための環境（応用編），川田重夫，田子精男，梅谷征雄，南多善　共編，培風館，（2005），p69-82

2）谷啓二，奥田洋司，福井義成，上島豊
ペタフロップスコンピューティング，
矢川元基　監修，培風館，（2007），p183-202

3）牧野圭一，上島豊，視覚とマンガ表現，臨川書店，（2007），p1-5，221-229

4）上島豊，月刊「研究開発リーダー」6 月号，技術情報協会，（2023），p63-68

5）上島豊，月刊「研究開発リーダー」7 月号，技術情報協会，（2023），p86-91

6）上島豊，月刊「研究開発リーダー」8 月号，技術情報協会，（2023），p78-82

7）上島豊，月刊「研究開発リーダー」7 月号，技術情報協会，（2024），p77-84

8）上島豊，月刊「研究開発リーダー」9 月号，技術情報協会，（2024），p73-81

9）上島豊，月刊「研究開発リーダー」11 月号，技術情報協会，（2024），p85-96

10）上島豊，月刊「研究開発リーダー」1 月号，技術情報協会，（2025），p74-82

11）上島豊，月刊「研究開発リーダー」3 月号，技術情報協会，（2025），p68-73

12）上島豊，他
研究開発部門への DX 導入による R & D の効率化，実験の短縮化，技術情報協会，（2022），p195-221

13）上島豊，他
ケムインフォマティックスにおけるデータ収集の最適化と解析手法，技術情報協会，（2023），p39-74

14）上島豊，他
実験の自動化・自律化による R & D の効率化と運用方法，
技術情報協会，（2023），p159-199

15）上島豊，他
少ないデータによる AI・機械学習の進め方と精度向上，説明可能な AI 開発，技術情報協会，（2024），p112-127