◇◇◇ シリーズ 「会員のつぶやき」◇◇◇

データは誰のもの？

山口大学

山本　修一

9年前の「つぶやき」にメカニステイックモデルとデータ駆動モデルについて書きました（日本食品工学会誌，**17**（4），A-2, 2016）．自分自身の研究は，最初から，メカニステイックモデルを利用しており，昔は，モデルと言えばメカニステイックモデルでした．

最近は機械学習（ML），人工知能（AI），データ駆動モデル，ハイブリッドモデル等を使用している研究が増えており，生成AIもさまざまな目的で研究に活用されています．生成AIは，多様なデータを使用していますが，論文については著作権や出版社の制約によりフリーあるいはオープンアクセス（OA）のみしか利用できないことが多いようです．

論文については公的資金を利用した研究成果はOA公開が原則になりつつあります．同時に論文に使用されているデータもOA公開が進められています．また，データのみを掲載するdata journalというカテゴリーの雑誌もできています．

このようなOAデータを利用することにより研究が促進することはまちがいがありませんが，生成AIも容易に学術論文データにアクセスできるので，自分のデータが研究者以外にも知らないうちに利用されることになります．学術論文では引用が明記されますが，生成AIでは，自分のデータがどのように解釈されて利用されているのかはわかりません．

さて，データは誰のものなのでしょうか？

同じような疑問を持つ人は多いようで，“who owns the data”あるいは“データは誰のもの”と検索すると多数のサイトが見つかります．多くは知的財産権と著作権に関する内容ですが，中には“データ”そのものについて議論しているものもあります．「論文には著作権があるが，その論文のデータは著作権・特許の範囲外であるという概念が一般的であったものの，遺伝子配列データが，その概念をくつがえした」とも指摘されています．また，さまざまな遺伝子データベースやタンパク質構造データベースも，データの公開・共有（public domain）を前提としています．

このような動きは2000年代からはじまり，現在のオープンアクセス論文およびdata journalにもつながっていきます．多くのデータが発表されずに埋もれていくのはもったいないという意見も強いです．2005年の論文（*Genome Biology*, **6**, 107, 2005）では，“Who owns the data”の回答として，“誰のものでもない”あるいは“みんなのもの”と締めくくっています．

もちろん，当時は，現在のような能力を持つAIの出現は想定されておらず，いまはAIが使用するデータは誰のものという議論が活発です．データは実体がない（無体物）ため，民法上は，所有権や占有権などの対象とならないとされていますが，多くの議論は個人情報あるいは個人の保有するデバイスから得られるデータについてです．学術論文データに関する議論は少なく，調べてもよくわかりませんでした．生成AIが既存のデータから新しい知見を見出したときの権利については未だ見解がないようです．

一方，AIが有用な学術情報を引き出すためには，データはFAIR原則（findability, accessibility, interoperability, and reusability）に基づいて管理することが望ましいと指摘されています（*Nature*, **604**, 635, 2022）．それ以前にも，博士課程の学生が膨大な時間をデータの変換や整理に無駄に費やしておりFAIRが重要であること，さらには，データの管理・活用“データスチュワード”をする担当者“データスチュワード”も必要であると述べられています（*Nature*, **578**, 491, 2020）．日本食品工学会誌もJ-STAGE Dataでデータ公開することになりましたが，J-STAGE DataはFAIR原則に基づいています．

今後も“データは誰のもの”および“データをどう共有するか”に関する議論が続いていくのでしょう．

山本修一（Shuichi YAMAMOTO）
京都大学農学部食品工学科卒業，同大学院博士課程修了後，
山口大学工学部着任
化学工学科助手，助教授を経て1998年より教授．
2001年から2016年まで医学系研究科担当
2019年より教授（特命）・名誉教授

〒755-8611　宇部市常盤台2-16-1
E-mail: shuichi@yamaguchi-u.ac.jp

© 2025 Japan Society for Food Engineering