2018 年 4 月 19 日

データ共有の基準としての FAIR 原則

DOI: 10.18908/a.2018041901

NBDC 研究開発チーム

FAIR 原則とは

近年、研究データの適切な公開について様々な話題が世界を駆け巡っています。特に「オープンデータ」の潮流は明確に推進されつつあり、例えば 2013 年には G8 サミットにてオープンデータ憲章に対する合意が締結され、国際社会が共同してオープンデータの推進に取り組んでいくこととなりました。さらに 2016 年に日本で開催された G7 の科学技術大臣会合「つくばコミュニケ」では、オープンデータとそれを含むオープンサイエンスを推進する実際の方針について共同声明が出されました。内閣府の定める第 5 期科学技術基本計画（2016 年～2021 年）においてもオープンサイエンス推進が明記され、国内の科学技術研究データの共有・公開について基準が定められました。研究データの公開と共有は、大量データを容易に産出可能な新時代の研究推進方法であるとともに、研究に携わる者の責務として広まりつつあります。

こういった中で良く言及される原則が「FAIR Data Principles 」――「FAIR 原則」です。FAIR は、「Findable（見つけられる）、Accessible（アクセスできる）、Interoperable（相互運用できる）、Reusable（再利用できる）」［1］の略で、データ公開の適切な実施方法を表現しており、データ共有の原則として広まりました。

NBDC では FAIR が提唱される以前から同様の理念で、生命科学系研究データを公開する活動を実施して参りました［2］。以下、FAIR 原則の歴史やその詳細についてご紹介します。

1. FAIR 原則の成り立ち

FAIR 原則の起源は 2011 年に遡ります。2011 年 8 月にドイツの Dagstuhl で開催された、Future of Research Communication (FoRC)のワークショップにて、FORCE11 (The Future of Research Communications and e-Scholarship)と呼ばれるコミュニティが発足しました。11 は 2011 年に由来します。研究者、図書館員、出版関係者、その他の活動家から構成されるこの任意団体では、知の創造と共有推進のため、複数のワーキンググループに分かれて活動を開始しました。

続く 2014 年 1 月にオランダの Leiden で、「Jointly designing a data FAIRPORT」というワークショップが開催され、eScience に関するエコシステムを推進させる方法について検討がなされました。ここで、「コミュニティに受け入れられる最低限の原則や実例があれば、人や機械が大量データに対して、発見やアクセス、相互運用、再利用を適切な引用と共に簡便に実施できる」との着想を得て、9 箇条の「DATA FAIRPORT フレームワーク」がまとめられました。

ワークショップには生命科学分野のメンバーが多く、生命科学データの複雑性や需要を鑑みて、生命科学分野への適用を最優先に実施すると定めていました。最終的にこの 9 箇条の原則は、2014 年 9 月の第 4 回 Research Data Alliance 定例会議 (RDA 4th plenary)にて Force11 のワーキンググループによって修正されて 4 つに統合され、「GUIDING PRINCIPLES FOR FINDABLE, ACCESSIBLE, INTEROPERABLE AND RE-USABLE DATA PUBLISHING VERSION B1.0」として Force11 のサイトから公開され周知されることとなりました。

2. FAIR 原則の原文

FAIR 原則の原文は絶えず議論や指摘を受けて更新が繰り返されています。初版が公開された後も、例えば 2015 年 9 月に NBDC と DBCLS が日本で共催した Biohackathon2015 において、FAIR 原則の内容に関する議論が行われたことが知られています［3］。FAIR 原則は現在、Original Draft 版、Living Document 版、議論反映版の

Licensed under a Creative Commons 表示 4.0 国際
© 2018 国立研究開発法人科学技術振興機構バイオサイエンスデータベースセンター

3 種類が存在します。それぞれの版について解説します。

1. **Original Draft 版**
   https://www.force11.org/fairprinciples
   2014 年に作成された「meeting's original draft」とされる文書です 。正式名称は「GUIDING PRINCIPLES FOR FINDABLE, ACCESSIBLE, INTEROPERABLE AND RE-USABLE DATA PUBLISHING VERSION B1.0」で、以降の 2.及び 3.の文書とは順番や内容が一部異なります。他との最も大きな違いは、外部から検索できるようにするための条項、「(meta)data are registered or indexed in a searchable resource.」が含まれていないことです。
2. **Living Document 版**
   https://web.archive.org/web/20170917212359/http://www.datafairport.org:80/fair-principles-living-document-menu
   DATA FAIRport のサイトから公開されている、コミュニティで議論中の内容を含む文書です。3.とほぼ同じ内容ですが、Findable 条項の 3 番目と 4 番目の順番が逆になっていて、表現がやや異なっています。2017 年 12 月時点で最終更新日は 2016 年 1 月 25 日となっていましたが、2018 年 3 月にはサイトがリンク切れとなっているため、上記には Internet archive のリンク情報を記述しています。
3. **議論反映版**
   https://www.force11.org/group/fairgroup/fairprinciples
   Force11 のサイトから公開されている、コミュニティでの議論の結果を反映した版です。2.とほぼ同じ内容が公開されています。項目ごとの解説はありませんが、下部にコメント欄が設けられており、質問や回答が記入できるようになっています。一方、説明欄には「The FAIR principles have now been published.」と記述されて Scientific Data の 2016 年 3 月の論文 [3]がリンクされており、論文にも本サイトの URL が記載されているため、本文書ではこの版の内容について次の項で説明します。

3. FAIR 原則の日本語訳

NBDC で作成した、FAIR 原則の議論反映版の各項目の日本語訳を以下に記述します。訳文は武田（2016）［1］を参考としました。2018 年 4 月現在、原文のサイトには他言語への翻訳機能が存在しますが、機械翻訳のために日本語として十分に意味の通じる訳文となっていないようです。F1.の冒頭のようにメタデータとデータ両方に適用する内容である場合は、文中で（メタ）データと記述されています［3］。図中の A2.の内容については、オランダの Dutch Techcentre for Life sciences (DTL)にて、「コストの問題でデータ公開終了となったとしても、検索のインデックスや恒常的な URL に情報を残し、リンク切れが起きないようにするということを示す」との解釈がなされています。また I1.の「広く適用可能な記述言語」とは、セマンティックウェブのためのオントロジーやデータで用いられるフォーマットの、OWL や JSON-LD、RDF などが、DTL での例として挙げられています。各項目の詳細については、長大になるため本文書では記述しませんが、適用例と共に解説されている 2016 年の Scientific Data 誌の論文［3］と、解説文が豊富に記述されている DTL のサイトが有用であることを付け加えておきます。

■To be Findable: (見つけられるために)

F1. （メタ）データが、グローバルに一意で永続的な識別子（ID）を有すること。
F2. データがメタデータによって十分に記述されていること。
F3. （メタ）データが検索可能なリソースとして、登録もしくはインデックス化されていること。
F4. メタデータが、データの識別子（ID）を明記していること。

■To be Accessible: (アクセスできるために)

Licensed under a Creative Commons 表示 4.0 国際
© 2018 国立研究開発法人科学技術振興機構バイオサイエンスデータベースセンター

A1．標準化された通信プロトコルを使って、（メタ）データを識別子（ID）により入手できること。
A1.1．そのプロトコルは公開されており、無料で、実装に制限が無いこと。
A1.2．そのプロトコルは必要な場合は、認証や権限付与の方法を提供できること。
A2．データが利用不可能となったとしても、メタデータにはアクセスできること。

■To be Interoperable:（相互運用できるために）

I1．（メタ）データの知識表現のため、形式が定まっていて、到達可能であり、共有されていて、広く適用可能な記述言語を使うこと。
I2．（メタ）データが FAIR 原則に従う語彙を使っていること。
I3．（メタ）データは、他の（メタ）データへの特定可能な参照情報を含んでいること。

■To be Re-usable:（再利用できるために）

R1．メタ（データ）が、正確な関連属性を豊富に持つこと。
R1.1．（メタ）データが、明確でアクセス可能なデータ利用ライセンスと共に公開されていること。
R1.2．（メタ）データが、その来歴と繋がっていること。
R1.3．（メタ）データが、分野ごとのコミュニティの標準を満たすこと。

4. 政策における FAIR 原則

日本における研究データの共有や公開は、地球科学の分野で古くから実施され［4］、生命科学の分野でも遺伝情報や生体分子に関するデータが増加するにつれて他の分野にさきがけて実施されてきました［5］。近年では、研究不正の監視や再現性の確保、データ駆動型研究の推進を目的として、あらゆる分野の学術データに対し、適切な公開の必要性が叫ばれています。この議論の場としては、半年に 1 度会議が開催されている Research Data Alliance（RDA）が国際的に重要と考えられ、日本では 2003 年から国立情報学研究所の国際学術情報流通基盤整備事業（SPARC Japan）がオープンアクセスの話題を中心に議論の場を提供しています。また、科学技術振興機構(JST)の Japan Link Center (JaLC)、物質・材料研究機構（NIMS）、情報・システム研究機構国立情報学研究所（NII）、国立国会図書館（NDL）、産業技術総合研究所（AIST）、情報通信研究機構（NICT)が 2016 年に研究データ利活用協議会（Research Data Utilization Forum, RDUF）を設立し、RDA のフィードバックや研究データ利活用に関する分野横断型のコミュニティとして始動しつつあります。

政策的な動きとしては、ヨーロッパでは研究データ公開のための情報基盤として European Open Science Cloud（EOSC）が設置され、その研究データ公開基準として欧州委員会が FAIR 原則をデータ共有と管理における基礎と位置付けています［6］。アメリカの生命医薬分野では National Institute of Health（NIH)が資金提供する Big Data to Knowledge（BD2K)プロジェクトにおいて、研究の過程で得られたデータは FAIR 原則に従って公開することが推奨されています。また英国 FAIRsharing プロジェクトでは、世界中のこのような公開レポジトリの情報を収集し公開しています。日本では大学などの機関レポジトリ環境提供サービスとして、JAIRO Cloud が構築されているほか、組織ごとの公開基盤も整備されつつあります。特に NBDC の生命科学系データベースアーカイブは、日本の生命科学分野のデータ公開レポジトリとして以前から精力的に整備されています。このため NBDC は、日本医療研究開発機構（AMED）や JST、JSPS といった研究助成機関の一部プロジェクトから、研究データ公開基盤の一つとして明記されています。

このように 2018 年現在、日本のプロジェクトや組織ごとのポリシーで FAIR 原則に言及する例はまだ見当たりませんが、オープンデータへの対応や基盤の形成は着実に始まっています。

5. FAIR 原則の未来

FAIR 原則は FAIRPORT principles として最初に議論がなされた際、インターネットにおける IP レイヤーのように、最低限でありながら広範囲に通用する原則として利用されるべきとされていました。まさに現在そのような存在として広ま

Licensed under a Creative Commons 表示 4.0 国際
© 2018 国立研究開発法人科学技術振興機構バイオサイエンスデータベースセンター

りつつあります。日本の組織や各研究者が FAIR 原則に則ったデータ公開および適切な研究データのマネジメントを理解し、データ共有について考えることで、オープンサイエンスにおける責務を果たしていけるのではないでしょうか。

参考文献

［1］武田英明：研究データ利活用に関する国内活動及び国際動向について．第 2 回 SPARC Japan セミナー，https://www.nii.ac.jp/sparc/event/2016/pdf/20161026_doc6.pdf（2016）

［2］高木利久：いかにして「使える」データベースを維持し続けるか？ ライフサイエンス統合データベースプロジェクトから学ぶこと．ConBio2017，http://dbcls.rois.ac.jp/wp-content/uploads/2017/12/20171206ConBio_takagi.pdf（2017）

［3］Wilkinson MD. et al.: The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data，doi:10.1038/sdata.2016.18（2016）

［4］渡邉堯：科学データの長期保全とグローバルな共有 ―ICSU 世界データシステムの構築― ICSU 世界データシステム(WDS)について．学術の動向，doi: 10.5363/tits.17.6_11（2012）

［5］Kodama Y et al.: DNA Data Bank of Japan: 30th anniversary. Nucleic Acids Res，doi: 10.1093/nar/gkx926（2017）

［6］European Commission. Directorate-General for Research. H2020 programme guidelines on FAIR data management in horizon 2020，http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf（2016）

* 大波純一，八塚茂，信定知江，箕輪真理，三橋信孝，畠中秀樹 ［↑］

Licensed under a Creative Commons 表示 4.0 国際
© 2018 国立研究開発法人科学技術振興機構バイオサイエンスデータベースセンター