

オープンで信頼できるリポジトリの条件 ～FAIR原則とCoreTrustSeal～

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
八塚 茂

ORCID: 0000-0002-6891-5229

自己紹介

- 八塚 茂（やつづか しげる）
 - 学部は人文系
 - たまたまSEになる
 - たまたまバイオ系システムの部門に配属される
 - バイオサイエンスデータベースセンター(NBDC)の担当になる
 - 生命科学系データベースアーカイブの立ち上げから参画
 - とうとうNBDCの「中の人」になる
 - 生命科学系データベースアーカイブを担当

「研究データ流通業」

「オープンで信頼できる」リポジトリ

- 自称「オープン」、自称「信頼できる」リポジトリでは信用してもらえない
- より客観的な条件が必要
 - FAIR原則
 - CTS

FAIR原則

スポーツの話じゃないの？

データの話です！

- データは、
 - 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)

でなければならない

渡る世界はFAIRばかり

■ 生命科学分野



Big Data to Knowledge

Making Data FAIR: Findable, Accessible, Interoperable, and Reusable

Members of the data science community recently published a comment article in *Scientific Data* describing guiding principles for scientific data management and stewardship. [Read the article](#) to learn more about this important effort and how the NIH Big Data to Knowledge (BD2K) initiative is contributing to it.

<https://commonfund.nih.gov/bd2k/highlights>

■ その他の分野



Towards FAIRer RDA Outputs - RDA 11th Plenary BoF meeting

How expensive is FAIR compliance and how expensive is it to not be FAIR compliant - RDA 11th Plenary BoF meeting

Improving the FAIRness of scientific data with DCAT-AP - RDA 11th Plenary BoF meeting



<https://www.dtls.nl/>



<https://fairsharing.org/>

Menu

- [About FAIR Data](#)
- [FAIR tools and more](#)
- [FAIR Events](#)
- [FAIR projects and initiatives](#)
- [Research Data management](#)
- [News about FAIR Data](#)

FAIR Principles (1) Findable

To be Findable:

見つけられるために

F1. (meta)data are assigned a globally unique and eternally persistent identifier.

(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. data are described with rich metadata.

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. (meta)data are registered or indexed in a searchable resource.

(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. metadata specify the data identifier.

メタデータが、データの**識別子 (ID)**を明記していること。

以下、日本語訳は
NBDC研究チーム
『データ共有の基準としてのFAIR原則』
DOI:10.18908/a.2018041901
から引用

FAIR Principles (2) Accessible

To be Accessible:

アクセスできるように

A1. (meta)data are retrievable by their identifier using a standardized communications protocol.
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 the protocol is open, free, and universally implementable.

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 the protocol allows for an authentication and authorization procedure, where necessary.

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. metadata are accessible, even when the data are no longer available.

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。

FAIR Principles (3) Interoperable

To be Interoperable:
相互運用できるように

11. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
(メタ)データの知識表現のため、形式が定まっています、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。
12. (meta)data use vocabularies that follow FAIR principles.
(メタ)データがFAIR原則に従う語彙を使っていること。
13. (meta)data include qualified references to other (meta)data.
(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。

FAIR Principles (4) Re-usable

To be Re-usable:

再利用できるように

R1. meta(data) have a plurality of accurate and relevant attributes.

メタ(データ)が、正確な**関連属性情報**を豊富に持つこと。

R1.1. (meta)data are released with a clear and accessible data usage license.

(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. (meta)data are associated with their provenance.

(メタ)データが、その**来歴**と繋がっていること。

R1.3. (meta)data meet domain-relevant community standards.

(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。

CTS

千歳空港？

Core Trust Sealです！

- 「信頼できるデータリポジトリを認定するための中核的な統一要件」を満たすリポジトリに対して付与される認証
 - Data Seal of Approval (DSA) と国際科学会議 (ICSU) 世界科学データシステム (WDS) がRDAのワーキンググループで策定
- 各リポジトリに対して16項目の自己評価を実施
- 各項目は5段階 (0～4) で評価
- 概ね3以上のレベルであれば認証される (らしい)



南山『信頼できるデータリポジトリの中核的な統一要件』
<http://current.ndl.go.jp/e1888>
から引用

CTSの評価（5段階）

- 0 - 適用不可（Not applicable）
- 1 - まだ考慮されていない
（The repository has not considered this yet）
- 2 - 概念上は存在する（The repository has a theoretical concept）
- 3 - 実装フェーズにある
（The repository is in the implementation phase）
- 4 - 実装済（The guideline has been fully implemented in the repository）

南山、前掲記事から引用

CTSの要件（16項目）

■組織的な基盤（Organizational Infrastructure）

■R1. ミッション／視野（Mission/Scope）

研究領域におけるリポジトリのミッションを明らかにするもの

■R2. ライセンス（Licenses）

データ提供者との合意事項、データ利用者に対するライセンス等

■R3. アクセスの継続性（Continuity of access）

保有するデータへのアクセスとデータの保存を保証する計画等

■R4. 機密保持／倫理（Confidentiality/Ethics）

システムのセキュリティ（特に機密保持）、スタッフへの倫理教育等

■R5. 組織的な基盤（Organizational infrastructure）

ファンディングとガバナンス

■R6. 専門家によるガイダンス（Expert guidance）

専門家の指導とフィードバック

参考：

南山前掲記事および

絹谷・南山 RDUF小委員会「国内の分野リポジトリ関係者のネットワーク構築」検討資料

CTSの要件（16項目）

- **デジタルオブジェクト管理（Digital Object Management）**
 - **R7.データの完全性・信頼性（Data integrity and authenticity）**
データの変更管理
 - **R8. 評価（Appraisal）**
データの受け入れ管理
 - **R9.保管手続きの文書化（Documented storage procedures）**
データの保存ポリシーやバックアップ
 - **R10.保存計画（Preservation plan）**
 - **R11.データの品質（Data quality）**
 - **R12. ワークフロー（Workflows）**
データの受け入れから流通までのワークフロー
 - **R13.データの発見と識別（Data discovery and identification）**
データの検索方法と永続的な識別子
 - **R14.データの再利用（Data reuse）**
メタデータ

CTSの要件（16項目）

■技術（Technology）

- R15.技術的な基盤（Technical infrastructure）
リポジトリのインフラ（ハードウェア・ソフトウェア）
- R16.セキュリティ（Security）
特にシステム障害への耐性

謝辞

■ バイオサイエンスデータベースセンター(NBDC)

<https://biosciencedbc.jp/>

■ 研究データ利活用協議会(RDUF)

小委員会 「国内の分野リポジトリ関係者のネットワーク構築」