

公開データの二次利用を 促進するメタデータの付与

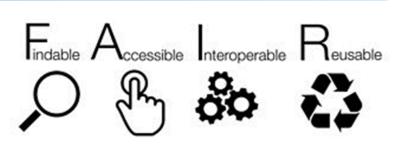
2022年12月5日 J-STAGE Dataユーザ会

J-STAGE Dataサポート 井津井豪

FAIR原則 一データ共有の基準一

FAIR原則

- Findable (見つけられる)
- Accessible (アクセスできる)
- Interoperable (相互運用できる)
- **R**eusable (再利用できる)



FORCE11: THE FAIR DATA PRINCIPLES (2016). https://www.force11.org/group/fairgroup/fairprinciples, NBDC研究チーム(訳), "FAIR原則(「THE FAIR DATA PRINCIPLES」和訳)"(2019). https://doi.org/10.18908/a.2019112601

の略。データ公開の適切な実施方法を表現している。

J-STAGE DataはFAIR原則を達成可能なデータリポジトリ

FAIR原則を達成するために、データを標準的な形式で公開すること、メタデータの標準化や充実などが重要になってくる

FAIR 15の原則

To be Findable:(見つけられるために)	To be Accessible: (アクセスできるために)
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。	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。
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。	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。
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。	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。
F4. メタデータが、データの識別子(ID) を明記していること。	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)	To be Re-usable:(再利用できるために)
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。	R1. メタ(データ)が、正確な関連属性を豊富に持つこと。
I2. (メタ) データがFAIR原則に従う語彙を使っていること。	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。
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。	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。

FigshareでのFAIR原則対応

Findable

The Findable principle is about making sure that data is stored with a persistent identifier, well described with rich metadata and indexed in as many relevant searchable resources as appropriate. This allows data to always be able to be found by a single URL despite and changes or migrations and promotes its discovery across multiple platforms.

- All data are assigned a Datacite DOI which can be custom branded with the publisher's name aiding discoverability in citation explorers
- All data are described with rich metadata with schemas that can be customised to the specific needs of each journal
- All data in Figshare is indexed by major search engines, Google Dataset Search, as well as a number of community and national data registers

Accessible

The Accessible principle ensures data can be accessed via standard non-proprietary protocols and metadata is stored in perpetuity. This allows the simple access of data now and future proofs the provenance of data in the event that it can't be stored long term due to storage limitations.

- All data on Figshare can be retrieved by their identifier using a standardised communication protocol
- Publisher portals are accessible via https and all data can be viewed in a web browser or can be accessed in bulk via the API
- · No sign up or account is required to view data
- No data will ever be behind a paywall
- Figshare support a variety of cases where data can't be made publicly available for privacy or commercial sensitivity concerns, but metadata is still published
- Figshare can integrate with a number of long term preservation systems

Interoperable

The interoperable principle promotes the use of formal, accessible, shared and broadly accepted language for knowledge representation. This allows meta(data) to be accessed and retrieved in a number of different formats and promotes the free flow of data between systems.

- (Meta)data on Figshare is represented in a number of languages including but not limited to html, JSON-LD, Schema.org, OAI-PMH interface, Dublin core, Datacite XML, RDF and VIVO
- Metadata schemas on Figshare can be customised to use vocabularies that follow the FAIR principles
- All Figshare items can be linked to the peer reviewed publication as well as citing any datasets used in the creation of the supporting data

Reusable

The Reusable principle encourages researchers to upload (meta)data richly described to the domain-relevant community standards. This should provide enough context to allow a peer to understand, reproduce and build on top of the original work.

- All meta(data) on Figshare have a clear and accessible usage license
- The Figshare metadata schema can be extended to meet community standards
- The Figshare review functionality can be used to check files for detailed provenance before publication

https://knowledge.figshare.com/ publisher/fair-figshare

FigshareでのFAIR原則対応① (拙訳)

見つけられる

Findable の原則は、データが永続的な識別子と共に保存され、豊富なメタデータで適切に記述され、適切な数の関連する検索可能なリソースでインデックス化されることを保証することです。これにより、変更や移行に関係なく、常に単一の URL でデータを見つけることができ、複数のプラットフォーム間での発見が促進されます。

- すべてのデータには DOI が割り当てられます。この DOI は、引用エクスプローラーで見つけやすくするために発行者の名前でカスタム ブランド化できます。
- すべてのデータは、各ジャーナルの特定のニーズに 合わせてカスタマイズできるスキーマを含む豊富な メタデータで記述されます
- Figshare のすべてのデータは、主要な検索エンジンである Google Dataset Search だけでなく、多数のコミュニティおよび国のデータ レジスタによってインデックス化されています。

アクセスできる

Acecesibleの原則により、データは標準の非独占的なプロトコルを介してアクセスでき、メタデータは永続的に保存されます。これにより、現在および将来のデータへの簡単なアクセスが可能になり、ストレージの制限によりデータを長期間保存できない場合でも、データの出所を証明できます。

- Figshare 上のすべてのデータは、標準化された通信 プロトコルを使用して識別子で取得できます
- パブリッシャーポータルは https 経由でアクセスでき、すべてのデータは Web ブラウザーで表示したり、API 経由で一括アクセスしたりできます。
- データを表示するためにサインアップやアカウント は必要ありません
- ペイウォールの背後にデータが存在することはありません
- Figshare は、プライバシーや商業上の機密性の懸念 からデータを公開できないが、メタデータは引き続き公開されるさまざまなケースをサポートしています。
- Figshare は、多数の長期保存システムと統合できます

FigshareでのFAIR原則対応② (拙訳)

相互運用できる

Interoperableの原則は、知識表現のために、正式で、アクセスしやすく、共有され、広く受け入れられている言語の使用を促進します。これにより、メタ(データ)にアクセスしてさまざまな形式で取得できるようになり、システム間のデータの自由な流れが促進されます。

- Figshare上の(メタ) データは、html、JSON-LD、 Schema.org、OAI-PMHインターフェース、Dublin core、Datacite XML、RDF、VIVOなど、多くの言 語で表現されていますが、これらに限定されるもの ではありません。
- Figshare のメタデータ スキーマは、FAIR の原則に 従う語彙を使用するようにカスタマイズできます
- すべてのFigshareアイテムは、サポートデータの作成に使用されたすべてのデータセットを引用するだけでなく、査読付き出版物にリンクすることができます。

再利用できる

Reusableの原則により、研究者は、ドメイン関連のコミュニティ標準に合わせて詳細に記述された (メタ) データをアップロードすることが奨励されます。これは、他者がオリジナルの作品を理解し、再現し、構築できるようにするのに十分なコンテキストを提供する必要があります。

- Figshare のすべてのメタ (データ) には、明確でア クセス可能な使用ライセンスがあります
- Figshare メタデータ スキーマは、コミュニティの 基準を満たすように拡張できます
- Figshare レビュー機能を使用して、公開前にファイルの詳細な来歴をチェックできます

システムにより自動的にFAIR原則が守られるが、 個々の共有データはFAIR原則を意識して登録・公開する必要がある。

FAIR 15の原則

To be Findable:(見つけられるために)	To be Accessible:(アクセスできるために)
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。	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。
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。	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。
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。	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。
F4. メタデータが、データの識別子(ID)を明記していること。	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)	To be Re-usable:(再利用できるために)
	10 50 110 404510. (13/13/13 C C & /2/2/2/
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。	R1. メタ(データ)が、正確な関連属性を豊富に持ってと。
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	R1. メタ(データ)が、正確な関連属性を豊富に持
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	R1. メタ (データ) が、正確な関連属性を豊富に持っこと。 R1.1 (メタ) データが、明確でアクセス可能な

見つけられ再利用されやすいデータ

研究データの(メタ)データを詳細に記述することで FAIR原則を守れる

…のですが、どのようなデータだとふさわしくないか、 J-STAGE Dataで実際に見られたケースを紹介するので、 今後のデータ公開の参考にしていただけると幸いです。

ファイル

- ふさわしくない例
 - 表データがPDFファイル内に記載されている
 - Excel形式
- 望ましい
 - 表データはCSV形式(Excelより汎用的)
 - 機械判読可能な表記(後述)
 - READMEファイルの配置
 - 複数ファイルの場合ファイル命名規則で違いを明確に

ファイル -機械判読可能な表記-

- 機械判読可能な表記
 - 1セル1データ
 - 数値データは数値属性にし、単位・区切り・注釈記号などの文字列を入れない
 - セル結合しない
 - スペースや改行等で体裁を整えない

参考:統計表における機械判読可能なデータ作成に関する表記方法

https://www.soumu.go.jp/main_content/000723626.pdf

J-STAGE Dataのメタデータ

Title	タイトル
Authors	著者
Categories	関連分野
Journal	関連ジャーナル名
Item type	種類
Keywords	関連キーワード
Description	説明文
Funding	助成金情報
Resource title	リンクすべきJ-STAGE記事のタイトル
Resource DOI	リンクすべきJ-STAGE記事のDOI
License	二次利用に関するライセンス
References	関連するURL
Corresponding author email address	連絡先著者メールアドレス
Translated title	タイトル(日本語表記)
Translated description	説明文(日本語表記)
Translated manuscript title	リンクすべきJ-STAGE記事のタイトル(日本語表記)
Translated authors	著者(日本語表記)
Copyright	著作権表示

メタデータのふさわしくない例①

- Title
 - 「(ジャーナル名、巻号ページ)のデータ」→見つけられにくく独立して流通する意味をなさない
 - ・ 論文と同じタイトル→同じ場合もあり得るが、データを端的に示す内容を記述
- Description
 - 論文の抄録と同じ内容
 →論文の説明ではなくデータ自体の詳細説明、特に再利用に関する内容を記述

メタデータのふさわしくない例②

- Funding
 - 論文には記載されているのに未登録
 - →Dimensionsと連携すればアクセスされやすくなる
- Corresponding author email address
 - 未登録
 - →二次利用者にとってデータ作成者への連絡先は事実上必須
- Copyright
 - 未登録
 - 「© 公開年 著作権者」の形式になっていない
 - →標準的な形式での著作権の明示は事実上必須

メタデータのふさわしくない例③

- Translated 項目
 - ジャーナルが和文誌だが未登録
 - →英語だけより見つけられやすくなる 特にTranslated Manuscript Titleは、Resource title(英語) より正式記事タイトルに相当
- Translated Title
 - Title (英語) と内容が一致していない
- Translated Description
 - Description (英語) と内容が一致していない

メタデータのふさわしくない例サマリ

メタデータ項目	ふさわしくない例
Title	「(ジャーナル名、巻号ページ)のデータ」
	論文と同じタイトル
Description	論文の抄録と同じ内容
Funding	論文には記載されているのに未登録
Corresponding author email address	未登録
Translated title	和文誌なのに未登録
	Title(英語)と内容が一致していない
Translated description	和文誌なのに未登録
	Description(英語)と内容が一致していない
Translated manuscript title	和文誌なのに未登録(Resource titleよりこちらが正式記事タイトルに相当)
Translated authors	和文誌なのに未登録
Copyright	未登録
	「© 公開年 著作権者」の形式になっていない

入力に関する補足

Funding

検索は「JP+8桁の課題番号」ではなく「8桁の課題番号」 ※謝辞での記載ルールとして「JP8桁の課題番号」が2016年度に定め られているが、J-STAGE Data連携サービスDimensionsには8桁の課 題番号でしか登録されていないため

• Copyright 論文の公開年に関わらず、データの公開年を設定する

参考資料

- Figshare and the FAIR data principles https://doi.org/10.6084/m9.figshare.7476428.v1
- fair figshare
 https://knowledge.figshare.com/publisher/fair-figshare



お問い合わせはこちらまで data-support@jstage.jst.go.jp