

# J-STAGE DataからのFAIR 原則をみたすデータ公開

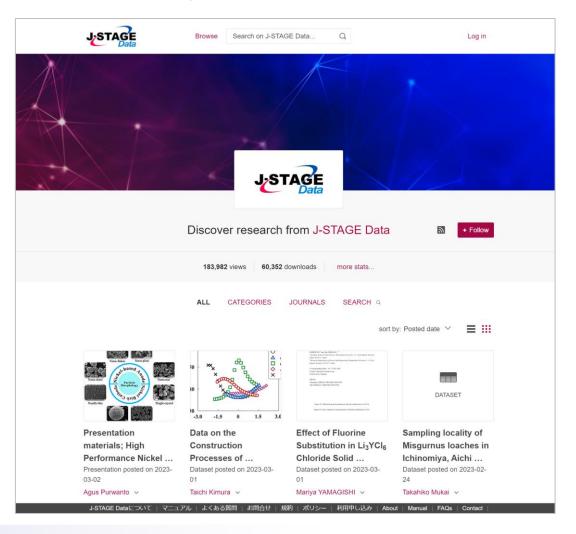
2023年3月17日 J-STAGEセミナー

J-STAGE Dataサポート 井津井豪

#### J-STAGE Dataとは

#### J-STAGEの論文に付随する研究データを公開するデータリポジトリ

- CCライセンスを明示し オープンアクセスで公開
- 2021年3月本運用開始
- 2023年2月28日時点
  - 申込ジャーナル: 72
  - 公開ジャーナル: 33
  - 公開データ: 479



#### J-STAGE Dataのデータ画面



## J-STAGE Dataでのデータ公開状況

- 公開データ件数:479
  - データ種類内訳
  - 一つのデータで公開されているファイル数分布

| データ種類           | 件数  |
|-----------------|-----|
| Presentation    | 208 |
| Dataset         | 204 |
| Figure          | 42  |
| Media           | 16  |
| Software        | 6   |
| Online resource | 3   |

| ファイル数   | 件数  |
|---------|-----|
| 1       | 324 |
| 2       | 78  |
| 3-10    | 67  |
| 11-50   | 4   |
| 50-100  | 2   |
| 101-200 | 1   |
| 201-    | 1   |

#### J-STAGE Dataの目指すところ

- J-STAGE登載誌の研究根拠データ公開の場 論文にあわせて、論文の根拠となるデータを公開することにより研究 の透明性を高める
- 研究データの二次利用推進 労力・費用をかけて得られた貴重な研究の生データが独立してオープ ンアクセスで公開されることで、だれもが自由に二次利用でき、別の 観点から新たな知見が生み出される

#### FAIR原則 一データ共有の基準一

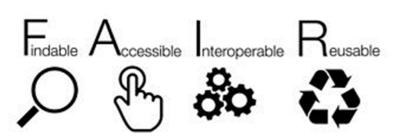
#### FAIR原則

• **F**indable (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

データ公開の適切な実施方法を表現している。

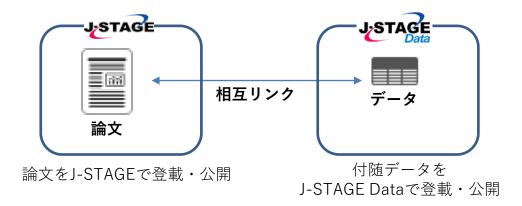
## FAIR 15の原則

| 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<br>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)<br>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、<br>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<br>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<br>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<br>こと。                    |
|  | R1.3 (メタ) データが、分野ごとのコミュニティ<br>の標準を満たすこと。             |

表内文章は以下を引用: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

#### FAIR原則 J-STAGE Dataでの対応

- J-STAGE Dataでのデータ公開
  - データ自体に独自のDOIがある
  - DOIでデータにアクセスできる
  - データのメタデータがある
  - データ自体が検索できる
  - 利用ライセンスが明示されている
  - 元論文と関連づいている
- 電子付録でのデータ公開 FAIR原則には従っていない





#### FAIR原則 J-STAGE Dataでの対応

| 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<br>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)<br>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。 デー                                 | A1.1 そのプロトコルは公開され 無料で、 mxicDOIを付与する CDOIでデータにアクセスできる |
| 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 ク ケの学型フ (ID) を明コレイいること データがGoogle Dataset Searchなどから検索できる | 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。 データのCCライセンスを明示する       |
| 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<br>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。                       | R12 (メタ) データが その来歴と繋がっている<br>データが関連する論文とDOIでリンクしている  |
|   | 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。                 |

表内文章は以下を引用: 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

#### FAIR原則に従うデータ公開のポイント

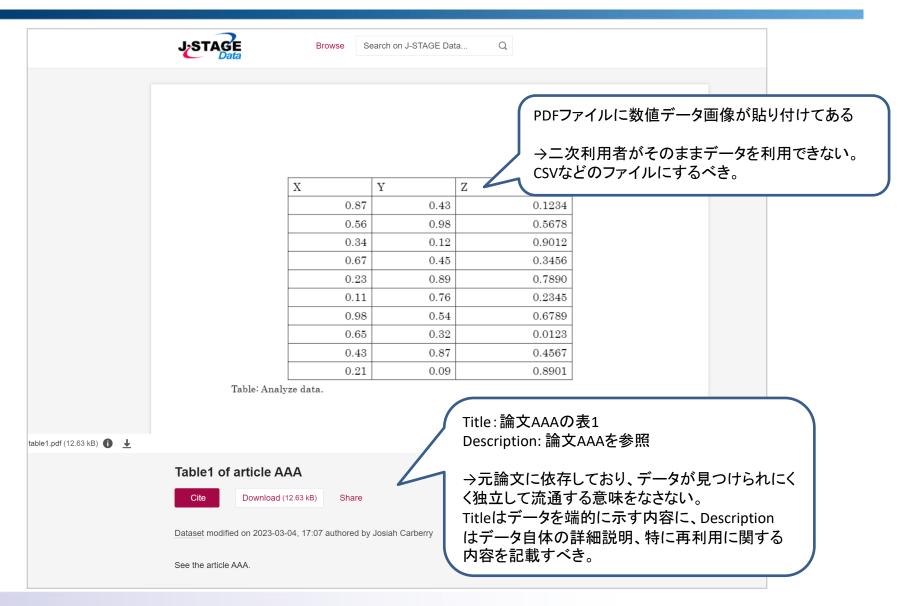
自動的に対応する項目とは別に、個々のデータ公開において 著者や発行機関が以下項目を意識して工夫する必要がある:

- 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。
- R1.メタ(データ)が、正確な関連属性を豊富に持つこと。

具体的には、**人間にも機械にも見つけられやすく再利用され やすいデータ**にすること。

- ファイル:人・機械が可読な標準的なファイル形式にする
- メタデータ:人・機械が可読な詳細内容を記述する

#### FAIR原則に従わないデータ



#### FAIR原則に従ったファイル

- 表データはCSV形式(Excelより汎用的)
- 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
- 必要に応じて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

# FAIR原則に従ったメタデータ

| Title*                             | このデータを端的に示し、検索されやすいタイトル       |
|------------------------------------|-------------------------------|
| Authors*                           | このデータの著者(元論文著者と異なる場合もある)      |
| Categories*                        | 分野                            |
| Journal*                           | ジャーナル名                        |
| Item type*                         | データ種類                         |
| Keywords*                          | このデータのキーワード                   |
| Description*                       | 元論文を見なくてもわかる詳細な説明や再利用に関する内容   |
| Funding                            | このデータが関連する助成金情報               |
| Resource title*                    | 元論文のタイトル                      |
| Resource DOI*                      | 元論文のDOI                       |
| License*                           | このデータの二次利用に関するライセンス           |
| References                         | このデータが関連するURL                 |
| Corresponding author email address | このデータに関連する連絡先著者メールアドレス        |
| Translated title                   | Titleの日本語表記                   |
| Translated description             | Descriptionの日本語表記             |
| Translated manuscript title        | Resource titleの日本語表記          |
| Translated authors                 | Authorsの日本語表記 (一字 ナル オー オントン) |
| Copyright                          | 発行年、著作権保持者を明示 (元論文に依存しない)     |

(元論又に依存しない) **そのデータに関する内容** であることが重要

#### J-STAGE Dataの課題

メタデータが十分に記述されていない例が多い

- 個々のデータの整備支援で対応
  - データを公開する前に発行機関から連絡をもらい内容を確認する
  - データが公開された後に内容を確認して発行機関に修正を依頼する
- 今後の対応
  - FAIR原則に従うポリシーの明文化
  - 見つけられて再利用されるデータの作り方マニュアル整備
  - システムでの対応検討(任意項目の必須化など)

#### 参考資料

- ➤ FAIR Principles GO FAIR

  <a href="https://www.go-fair.org/fair-principles/">https://www.go-fair.org/fair-principles/</a>
- ➤ J-STAGEニュースNo. 51

  J-STAGE Dataユーザ会を初めて開催してみた

  https://doi.org/10.34344/jstagenews.2023.51.2



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