J-STAGE Dataの公開状況と課題

2023年11月30日
J-STAGE Dataユーザ会

J-STAGE Dataサポート 井津井豪

データ公開状況①

- 2023年11月23日時点
  - 公開ジャーナル数：39
  - 公開アイテム（データ）数：634

データ種類	件数
Dataset	300
Presentation	210
Figure	91
Media	21
Software	8
Online resource	3
Journal Contribution	1

言語	件数
日本語タイトルなし	338
日本語タイトルあり	296

データ公開状況②

ライセンス	件数
CC BY	403
CC BY-SA	9
CC BY-ND	41
CC BY-NC	51
CC BY-NC-SA	35
CC BY-NC-ND	90
CC0	3
GPL	1
MIT	1

著者人数	件数
1	206
2	90
3	103
4	75
5	59
6	31
7	28
8	20
9	6
10人以上	16

Fund数	件数
0	417
1	99
2	44
3	39
4	21
5	11
6件以上	3

データ公開状況③公開ファイル

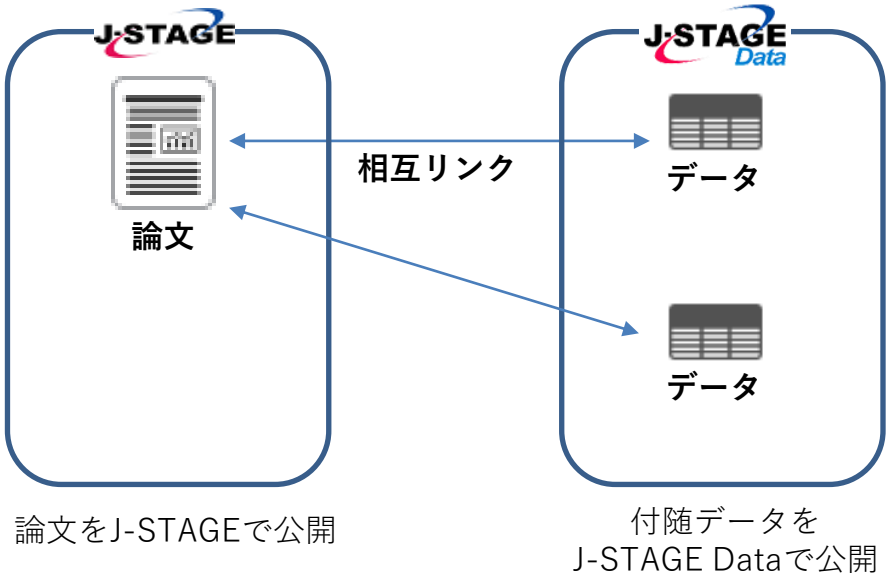
1データのファイル数	件数
1	412
2	120
3-10	89
11-50	7
50-100	2
101-200	1
201-	1
0(Retraction)	2

ファイル拡張子	dataset	presentation	figure	media
csv, tsv	**200**			
xlsx	**176**			
tif, jpg, png, ai	143		242	
pdf	125	208	35	
nc	136			
ctl	117			
grd	80			
gz, zip, tgz	79			
docx	54		13	
txt	27			
mp4, avi	13			23

（10件以上のみ集計）

データ公開状況④論文とデータの対応

一つのJ-STAGEの論文に紐づくJ-STAGE Dataのデータ件数



紐づくデータ数	件数
1	525
2	11
3	4
4	3
5	1
6	2
13	1
33	1

データごとに異なるDOI・メタデータで公開

データ登載・公開の課題①

- J-STAGE記事でDASが公開されていない

J-STAGE Dataでのデータと J-STAGE記事でのDASは同時に公開する。

※ただし、J-STAGE DataとJ-STAGEでのシステム仕様上の公開タイミングの差異があるため、発行機関での運用としては、J-STAGE記事（DAS設定済）の公開日の前日にJ-STAGE Dataでデータを公開する流れ（「よくあるご質問」を参照）。

記事でDASが公開されないと…
・その記事にデータが付随されていることがわからない
・記事からデータへリンクされない

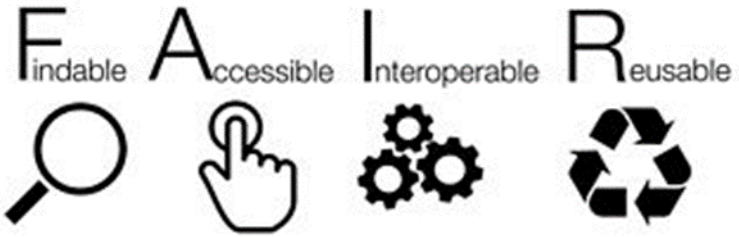
データ登載・公開の課題②

- FAIR原則に従っていない
  - ファイル
    - 二次利用できない、しにくい
    - 数値データが機械判読不能
  - メタデータ
    - 元論文に依存している
    - 記述が不十分

FAIR原則　―データ共有の基準―

FAIR原則

- **F**indable　（見つけられる）
- **A**ccessible　（アクセスできる）
- **I**nteroperable（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

- データ自体が検索できる
- 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
- 元論文と関連づいている
- データのメタデータがある
- 利用ライセンスが明示されている

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。

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

J-STAGE Data　Browse　Search on J-STAGE Data...

X	Y	Z
0.87	0.43	0.1234
0.56	0.98	0.5678
0.34	0.12	0.9012
0.67	0.45	0.3456
0.23	0.89	0.7890
0.11	0.76	0.2345
0.98	0.54	0.6789
0.65	0.32	0.0123
0.43	0.87	0.4567
0.21	0.09	0.8901

Table: Analyze data.

table1.pdf (12.63 kB)

Table1 of article AAA

Cite　Download (12.63 kB)　Share

Dataset modified on 2023-03-04, 17:07 authored by Josiah Carberry

See the article AAA.

PDFファイルに数値データ画像が貼り付けてある

→二次利用者がそのままデータを利用できない。CSVなどのファイルにする。

Title：論文AAAの表1
Description: 論文AAAを参照

→元論文に依存しており、データが見つけられにくく独立して流通する意味をなさない。
Titleはデータを端的に示す内容に、Descriptionはデータ自体の詳細説明、特に再利用に関する内容を記載する。

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
  - 表タイトル、注釈は含めない（Descriptionに記載）
  - 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
  - スペースや改行等で体裁を整えない
- 複数ファイルの場合ファイル命名規則やDescription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
Copyright	発行年、著作権保持者を明示

元論文に依存しない
そのデータに関する内容
であることが重要

その他

- Fund情報
  - 論文に記載があるのに入力していない
  - 該当Fund情報が検索されるのに手入力している

    「JP+8桁の課題番号」ではなく「8桁の課題番号」で検索を