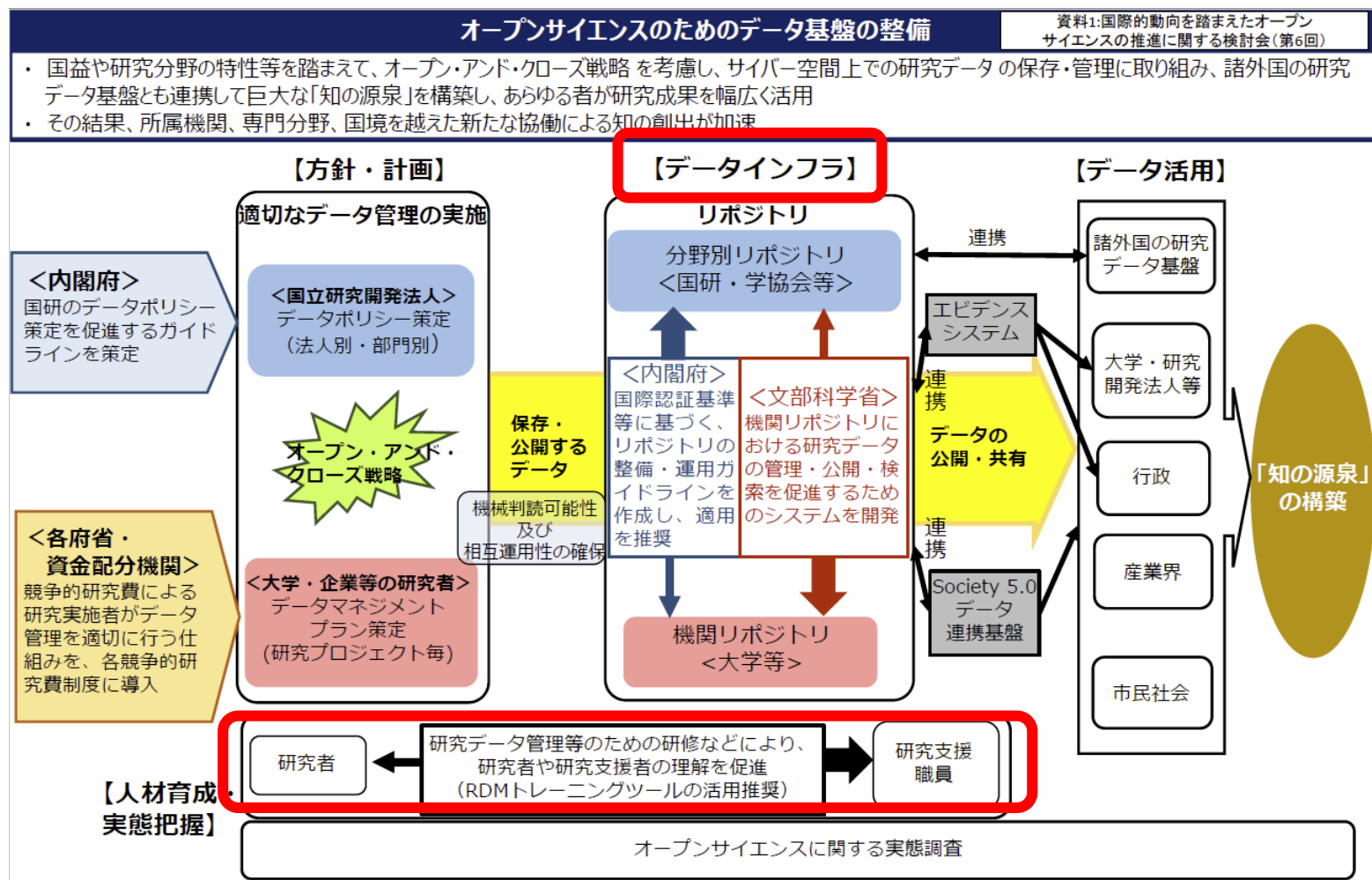


データリポジトリガイドライン

人的基盤とは？

JDARN（ジャパンデータリポジトリネットワーク）

オープンサイエンスのためのデータ基盤整備



【出典】国際的動向を踏まえたオープンサイエンスの推進に関する検討会(第6回)資料

<https://www8.cao.go.jp/cstp/tyousakai/kokusaiopen/6kai/siryoy1.pdf>

4.3.1 データリポジトリにおける職務

管理的業務

- 組織のミッションに即したリポジトリの運営・企画
- ポリシー作成
- 広報・コミュニケーション
- 利活用支援

コンテンツ運用 業務

- データ受け入れ
- クオリティ・コントロール
- メタデータ管理
- データ提供
- データキュレーション

システム運用 業務

- システム・ストレージ等の情報基盤の構築と運用
- サーバ保守
- インシデント対応
- セキュリティ対策
- 情報資産管理

4.3.2 スタッフの配置・教育 ～専門性～

- ・職務ごとに必要な技術・知識が異なる

管理的業務

- ・経営・法律・コンプライアンス
- ・研究における国内・国際連携
- ・コミュニティ参加・運営
- ・広報活動
- ・教育スキル

コンテンツ運用 業務

- ・データ特性の知識
- ・分野毎のメタデータ・品質基準
- ・コンテンツの更新
- ・アクセス管理
- ・データ提供者、利用者との綿密なコミュニケーション
- ・リポジトリ間連携

システム運用 業務

- ・システム開発・メンテナンス
- ・保守・バックアップ
- ・ハードウェア、ソフトウェア、ネットワーク管理
- ・インシデント分析
- ・セキュリティマネジメント

4.3.2 スタッフの配置・教育 ～配置～

- 複数人による業務分担例

管理的業務

- 組織管理部門
- 知財管理者
- URA
- 図書館員

コンテンツ運用 業務

- 研究者
- 研究支援者
- データ専門家

システム運用 業務

- 技術職員
- 情報部門スタッフ

4.3.2 スタッフの配置・教育 ～教育～

- ・配置された人材の専門性の高度化

- ◆各研究データの特性の理解
- ◆機関ごとの情報基盤の相違
- ◆RDMポリシー



・OJT

マニュアル作成
文書管理含む

- ◆基本知識の習得
- ◆スタッフの認識ギャップ解消



・オンライン教材

- ◆国際動向の把握
- ◆リポジトリの異分野・国際連携
- ◆データの利活用促進



・国際会議・セミナー等
への参加
・人事交流

【質 問】

1. データリポジトリに必要な業務の分類は
 - 管理的業務
 - コンテンツ運用業務
 - システム運用業務で良いでしょうか。他に必要な業務があればご指摘ください。
2. データ専門家はどのような仕事をする人だと思いますか？
3. データリポジトリ運営業務においては、コンテンツ運用業務がその中核となると考えられます。
 - コンテンツ運用業務に携わっている人は皆さんの周りにいますか？
 - どんな立場（職種）で、どんなお仕事をされていますか？
4. 皆さんがデータリポジトリの運用にかかわることになった場合、どのような教育やトレーニングを受けたいですか。

データキュレーション例 1 : DIAS

CEOP Flag Definitions

G: Good

I : Interpolated

D: Dubious/Questionable

B: Bad

C: minus precipitation or Abnormal value

M: Missing

U: Unchecked

品質検査を行いフラグを付けたいが、品質検査のプログラムに元データのままで読み込めない

データキュレーション例 1 : DIAS 元データファイル例(エクセル)

Microsoft Excel - Tiksi_EOP4_2.csv

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D) ウィンドウ(W) ヘルプ(H) 質問を入力してください

	A	B	C	D	E	F	G	H	I	J	K	L
1		Element Information										
2												
3												
4	Date & Time	-25.6	-26	-26.8	-27.5	73.8	73.1	74.7	72.7	1.4	0.6	
5		-26.2	-26.3	-26.7	-27.4	73.7	74.8	74.3	72.2	1.1	0.8	
6		-26.7	-26.6	-27	-27.3	74.3	74.6	74.3	72.5	2.8	2.6	
7		-26.5	-26.6	-27	-27.2	73.9	73.7	74.2	72.5	1.9	2	
8		-26.5	-26.2	-26.6	-26.7	73.6	73.8	74	72.4	2.9	2.7	
9		-25.9	-25.6	-26.1	-26.4	72.7	73.1	73.3	72.1	1.5	1.5	
10		-25.8	-24.8	-25.9	-26.3	73.2	70.5	73.6	72	1.4	1.4	
11		-25	-22.3	-25.2	-25.4	71.5	59.2	71.6	70.6	0.5	0.6	
12		-24.4	-20.1	-24.1	-24.2	72.7	54.3	72.4	70.6	0.6	0.6	
13		-22.6	-19.2	-22.4	-22.8	72.3	56.3	72.4	72.3	2.3	2.6	
14		-22	-19.5	-21.5	-22.3	71.6	60.9	70.7	72.2	3.4	3.4	
15		-21.5	-18.8	-20.9	-21.7	71.5	59.8	70.5	72.2	3.4	3.6	
16		-21.1	-18.9	-20.6	-21.4	72	62.1	70.9	72.6	3.6	3.7	
17		-21	-19.3	-20.5	-21.2	73.4	66.2	72.6	73.5	3.6	3.7	
18		-21.1	-19.3	-20.6	-21.3	73.1	65.7	72.3	73.4	3.6	3.5	
19		-21.1	-19.4	-20.7	-21.4	72.7	65.7	71.8	72.9	3.3	3.4	
20		-21.8	-20.1	-21.6	-22.2	74.6	67.1	74.3	74.3	3.3	3.3	
21		-22.5	-21.4	-22.4	-22.8	75.3	70.9	75.3	74.2	3.5	3.6	
22		-23.3	-22.8	-23.5	-23.9	75.2	73.6	75	74	3.4	3.5	
23		-24.3	-24.1	-24.8	-25.4	75.9	75.5	75.3	74.1	3.2	3	
24		-24.7	-24.6	-25.2	-25.5	74.8	75.3	75.9	74.4	3.7	3.8	
25		-26.3	-26	-26.8	-27.2	75.3	74.3	75.1	73.8	3.5	3.5	
26		-27.1	-27	-27.6	-28	74.8	75.2	74.8	72.9	4.1	3.8	
27		-27.7	-27.4	-28.1	-28.6	74.6	74.5	74.4	72.1	3.2	3.1	

Tiksi EOP4 2

データキュレーション例 1 : DIAS 元データファイル例 (CSV)

PRP-Global_20040227-064139.RAW - ワードパッド

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ヘルプ(H)

Date & Time

\$PRP0,1,35.21,-0.4,0.1,-45.0,46.2,121.0,0.3,185.6,416.4,-70.8,31.63,31.33,14.3,
\$PRP0,1,35.19,-0.4,0.1,-45.0,46.2,121.0,0.3,186.7,416.6,-70.3,31.60,31.31,14.3,
\$PRP0,1,35.18,-0.4,0.1,-45.0,46.2,121.0,0.4,187.7,416.8,-69.9,31.58,31.29,14.3,
\$PRP0,1,35.16,-0.5,0.2,-45.0,46.2,121.0,0.4,188.5,416.9,-69.4,31.55,31.27,14.3,
\$PRP0,1,-99.00,-99.0,-99.0,-99.0,-99.0,-99.0,-99.0,-99.0,0.0,-99.0,-99.00,-99.00,-99.0,
\$PRP0,1,-99.00,-0.4,0.1,-45.0,52.0,121.0,0.4,191.3,418.4,-65.9,31.39,31.15,14.3,
\$PRP0,1,35.01,-0.4,0.2,-45.0,49.3,121.0,0.4,191.3,418.2,-66.0,31.38,31.14,14.3,
\$PRP0,1,35.01,-0.5,0.1,-45.0,48.1,121.0,0.4,191.1,418.3,-65.9,31.37,31.12,14.3,
\$PRP0,1,34.99,-0.4,0.1,-45.0,47.4,121.0,0.4,191.0,418.5,-65.8,31.36,31.11,14.3,
\$PRP0,1,34.97,-0.5,0.1,-45.0,47.0,121.0,0.4,190.8,418.7,-65.5,31.35,31.10,14.3,
\$PRP0,1,34.97,-0.5,0.1,-45.0,46.7,121.0,0.4,190.6,418.9,-65.3,31.34,31.09,14.3,
\$PRP0,1,34.96,-0.5,0.1,-45.0,46.5,121.0,0.4,190.4,419.0,-65.1,31.33,31.08,14.3,
\$PRP0,1,34.95,-0.5,0.1,-45.0,46.3,121.0,0.3,190.2,419.2,-64.8,31.32,31.07,14.3,
\$PRP0,1,34.94,-0.5,0.1,-45.0,46.2,121.0,0.3,190.1,419.4,-64.4,31.31,31.06,14.3,
\$PRP0,1,34.92,-0.5,0.1,-45.0,46.2,121.0,0.3,189.9,419.7,-63.9,31.29,31.04,14.3,
\$PRP0,1,34.89,-0.5,0.1,-45.0,46.2,121.0,0.3,189.9,420.0,-63.4,31.26,31.03,14.3,
\$PRP0,1,34.87,-0.5,0.1,-45.0,46.2,121.0,0.3,190.1,420.4,-62.8,31.25,31.01,14.3,
\$PRP0,1,34.84,-0.5,0.1,-45.0,46.2,121.0,0.3,190.6,420.8,-62.2,31.23,30.99,14.3,
\$PRP0,1,34.84,-0.5,0.1,-45.0,46.2,121.0,0.4,191.3,421.1,-61.7,31.21,30.98,14.3,
\$PRP0,1,34.84,-0.5,0.0,-45.0,46.2,121.0,0.3,192.3,421.3,-61.3,31.19,30.97,14.3,
\$PRP0,1,34.83,-0.5,0.1,-45.0,46.2,121.0,0.4,193.5,421.6,-60.8,31.17,30.95,14.3,
\$PRP0,1,34.83,-0.5,0.1,-45.0,46.2,121.0,0.4,195.0,421.8,-60.4,31.15,30.94,14.3,
\$PRP0,1,34.83,-0.5,0.1,-45.0,46.2,121.0,0.4,196.8,422.0,-60.0,31.14,30.92,14.3,
\$PRP0,1,34.82,-0.4,0.1,-45.0,46.2,121.0,0.4,198.8,422.1,-59.9,31.12,30.91,14.3,
\$PRP0,1,34.82,-0.4,0.1,-45.0,46.2,121.0,0.4,201.0,422.2,-59.7,31.11,30.89,14.3,
\$PRP0,1,34.82,-0.4,0.1,-45.0,46.2,121.0,0.4,203.2,422.3,-59.5,31.09,30.87,14.3,
\$PRP0,1,34.82,-0.4,0.1,-45.0,46.2,121.0,0.4,205.4,422.4,-59.4,31.07,30.85,14.3,
\$PRP0,1,34.82,-0.4,0.1,-45.0,46.2,121.0,0.4,207.6,422.4,-59.2,31.06,30.84,14.3,

データキュレーション例 1 : DIAS 元データファイル例(エクセル)

Microsoft Excel - 13_001_2009090115505813.xls

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D) ウィンドウ(W) ヘルプ(H) Adobe PDF(B)

	C	D	E	F	G	H	I	J	K	L	M	N	O	P
5	Other Information													
6														
7	Element Information													
8														
9														
10	Year													
11	Day	Month												
12		0	0	0	10.4	0	0	1	4.9	Trace	0	0	0	
13		0	0	0	0	0	0	0	0	0	0	0	0	
14		0	0	0	0	0	0.1	1.8	0.2	0	0	0	0	
15		0	0	0	0	0	6	0	0	0	0	0	0	
16		0	0	0.4	1.1	0.1	0	0	0	0	0	0	0	
17		0	0	0	0	0	0	0	0	Trace	1.3	0	0	
18		0	0	0	0	0	0	0	0	0	0.7	0	0	
19		0.2	0	0	0	0	0	0	0	0	0	0	0	
20		0.1	0	0	Trace	0	0	0	0	0	0	0	0	
21		0	0	0	0	0	5.3	0	0.5	0.7	0	0	0	
22		0	0	0	0	0	0	0	0.8	0	0	0	0	
23		0.4	0	0	0	0	0	0	0	0	0	0	0	
24		1.8	0	0	0	0	0	0	6	0	0	0	0	
25		0.1	0	0	1.2	0	0	3.6	0.2	0	0	0	0	
26		0.3	0	0	0.6	0	0	0	0	0	0	0	0	
27		0	0	0	0	0	0	0	0	0	0	0	0	
28		0	0	0	0	0	0	0	0	0	0	0	3.5	
29		0.3	0	0	Trace	0	0	0	0	0	0	0	0.5	
30		0.7	0	0	Trace	0	0	0	4	0	0	0	0	

Precipitation / Temperature (max.) / Temperature (min.) / Wind Speed(00Z) / Wind Direction(00Z) / Wind Sp

データキュレーション例 1 : DIAS 元データファイル例(ワード)

05_001_2009102013280905.DOC - Microsoft Word

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) 罫線(A) ウィンドウ(W) ヘルプ(H) Adobe PDF(B) Acrobat コメント(C)

Element Information													
Other Information													
Year	Month												
	4.3	5.2	4.3	4.4	4.3	4.2	4.2	5.0	4.2	4.5	4.5	4.3	
	3.5	4.5	5.1	4.6	4.5	4.7	4.6	4.4	5.1	5.3	5.0	4.5	
	4.1	4.7	4.2	5.0	4.6	4.3	3.8	3.3	4.4	4.5	4.4	4.3	
	4.1	4.5	4.2	4.3	4.3	4.1	3.6	4.2	4.7	4.1	4.5	4.3	
	4.3	4.3	4.0	4.3	4.6	4.2	3.6	3.7	4.2	4.0	4.8	4.4	
	4.5	4.7	4.4	4.3	4.2	4.0	3.8	3.9	4.0	4.5	5.2	4.4	
	4.3	4.6	4.5	4.7	4.5	4.4	4.2	4.3	4.9	4.5	4.9	4.9	
	4.5	4.9	4.4	4.4	5.0	4.8	4.6	4.4	5.0	4.9	4.8	4.7	
	5.1	4.9	4.7	5.3	4.8	5.1	4.2	4.2	5.1	4.8	5.1	4.6	
	4.6	4.7	4.5	5.3	4.8	4.8	4.2	4.4	4.7	4.5		4.5	
	4.3	4.7	4.4	4.7	4.6	4.5	4.1	4.2	4.6	4.6	4.8	4.5	
	4.8	5.0	4.7	4.8	4.8	4.6	4.4	4.6	4.8	7.1	5.1	4.5	
	4.8	4.8	4.6	5.2	5.0	4.7	4.4	4.3	4.5	4.9	5.4	4.8	
4.6	4.9	4.7	5.2	4.5	4.5	4.1	4.6	4.5		4.9	5.0		

3 ページ 1 セクション 3/5 位置 行 桁 記録 変更 拡張 上書 英語 (US)

データキュレーション例 1 : DIAS

様々な日付時刻表現

Time Format	Example
(Serial Value of EXCEL)	36527.12784
AAA MMM DD hh:mm:ss YYYY	Sun Jan 02 03:04:05 2000
MM/DD/YYYY,hh:mm:ss.sss	01/02/2000,03:04:05.006
YY/MM/DD,hh:mm:ss.sss	00/01/02,03:04:05.006
YYYY M D h	2000 1 2 3
YYYY M/D hmm	2000 1/2 304
YYYY,DOY,hmm	2000,2,304
YYYY,DOY,hmm,s.sss	2000,2,304,5.006
YYYY,M,D,hmm	2000,1,2,304
YYYY,MDD,hmm	2000,102,304
YYYY/M/D h:mm	2000/1/2 3:04
YYYY/M/D,hh:mm	2000/1/2 03:04
YYYY/MM/DD hh:mm	2000/01/02 03:04
YYYY-MM-DD h:mm	2000-01-02 3:04
YYYYMMDD hh:mm:ss	20000102 03:04:05
YYYYMMDD,hmm	20000102,304

データキュレーション例 1 : DIAS

様々な欠損値表現

-99999.	99999.	(Blank)
-9999.	9999.99	-9.9M
-8190.	9999.9	DNA
-6999.	9999.	Miss
-999.99	6999.	NA
-999.9	999.9	NaN
-99.9	999.	NR
-99.	99.9	
-9.		

データキュレーション例 1 : DIAS カラムずれ

	A	B	C	D	E	F
1	2003/4/1 0:00	2.993	99	4.909	3.607	
2	2003/4/1 1:00	2.014	99	4.942	3.506	
3	2003/4/1 2:00	1.684	99	4.876	3.473	
4	2003/4/1 3:00	1.051	99	4.809	3.406	
5	2003/4/1 4:00	0.249	99	4.742	3.339	
6	2003/4/1 5:00	→	-0.446	99	4.742	3.306
7	2003/4/1 6:00	→	-1.224	99	4.675	3.273
8	2003/4/1 7:00	→ ?	-2.31	99	4.608	3.139
9	2003/4/1 8:00	→	-3.19	99	4.542	3.106
10	2003/4/1 9:00	→	-2.925	99	4.542	3.072

データキュレーション例 1 : DIAS

データカラムが途中から増えている

	A	B	C	D	E	F
1	2003/4/1 0:00	2.993	4.422	4.909	3.607	
2	2003/4/1 1:00	2.014	3.585	4.942	3.506	
3	2003/4/1 2:00	1.684	3.05	4.876	3.473	
4	2003/4/1 3:00	1.051	2.54	4.809	3.406	
5	2003/4/1 4:00	0.249	1.882	4.742	3.339	
6	2003/4/1 5:00	-0.446	1.328	4.742	3.306	2.822
7	2003/4/1 6:00	-1.224	0.715	4.675	3.273	2.656
8	2003/4/1 7:00	-2.31	0.077	4.608	3.139	2.45
9	2003/4/1 8:00	-3.19	-0.681	4.542	3.106	2.222
10	2003/4/1 9:00	-2.925	-1.092	4.542	3.072	1.975

データキュレーション例 1 : DIAS

データの不自然な結合

	A	B	C	D	E	F
1	2003/4/20 7:00	-1.87	-0.036	9.28	10.45	1.768
2	2003/4/20 8:00	-2.007-10.316	8.82	9.28	1.603	
3	2003/4/20 9:00	-2447368438	8.88	8.35	1.418	
4	2003/4/20 10:00	-0.535N9284	8.28	1.232		
5	2003/4/20 11:00	0.016	-0.074	9.95	8.55	1.088

?

データキュレーション例 1 : DIAS タイムスタンプ不連続や重複

	A	B	C	D	E	F
1	2003/4/1 0:00	2.993	4.422	4.909	3.607	3.359
2	2003/4/1 1:00	2.014	3.585	4.942	3.506	3.318
3	2003/4/1 2:00	1.684	3.05	4.876	3.473	3.235
4	2003/5/1 8:00	1.051	2.54	4.809	3.406	3.111
5	2003/5/1 9:00	0.249	1.882	4.742	3.339	2.987

?

	A	B	C	D	E	F
1	2003/4/1 0:00	2.993	4.422	4.909	3.607	3.359
2	2003/4/1 1:00	2.014	3.585	4.942	3.506	3.318
3	2003/4/1 2:00	1.684	3.05	4.876	3.473	3.235
4	2003/4/1 2:00	1.051	2.54	4.809	3.406	3.111
5	2003/4/1 3:00	0.249	1.882	4.742	3.339	2.987

?



2017/12/6 ConBio2017

データキュレーション例 2 : NBDC

<https://doi.org/10.6084/m9.figshare.5682847.v3>

生命科学における オープンデータの理想と現実

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

データキュレーション例 2: NBDC データは右から左へ受け流せない



データを作成する・利用する

■データを作成する

- 実験機器などから産生されたデータを、作成者が自分の観点でまとめる

■データを利用する

- データの作成者が利用する
- データの作成者以外の人を利用する
 - 観点は作成者と必ずしも同じではない

再利用

データキュレーション例2: NBDC アーカイブ作成: 3つのプロセス



アーカイブ作成: 3つのプロセス



データキュレーション例 2: NBDC 元データの例

NBDC

解体前のデータのイメージ

main

ID	名称
P1	タンパク質A
P2	タンパク質B

mainのIDと
同じなのだ
ろうか？

ex_cond

ID	実験条件	フラグ	文献ID
P001	条件1	00	123456
P002	条件2	01	234567

このフラグ
の意味は？

sequence

ID	配列
P1	111-111.fsa
P2	human.fsa

配列そのもので
はなく、ファイル名
のようだ

このフラグ
の意味は？

reference

文献ID	文献名	フラグ
123456	文献1	1
234567	文献2	0

ex_res

ID	実験結果	文献ID
P001	結果1	123456
P002	結果2	234567

mainのIDと
同じなのだ
ろうか？

データキュレーション例 2: NBDC データ内容の解析



「考古学」的考察の結果

main

ID	名称
P1	タンパク質A
P2	タンパク質B

mainのIDが0詰めされていた

00:赤で表示
01:緑で表示
10:青で表示
11:黒で表示

ex_cond

ID	実験条件	フラグ	文献ID
P001	条件1	00	123456
P002	条件2	01	234567

sequence

ID	配列
P1	111-111.fsa
P2	human.fsa

配列ファイル名
だった

0:独自のID
1:Pubmed ID

reference

文献ID	文献名	フラグ
123456	文献1	1
234567	文献2	0

ex_res

ID	実験結果	文献ID
P001	結果1	123456
P002	結果2	234567

mainのIDが0詰めされていた

データキュレーション例 2: NBDC データ再生



そして再生

main

ID	名称	配列
P1	タンパク質A	MPLGLIGEKVGMTRVLLK
P2	タンパク質B	MYALLVISLYLQRFYNLSIIPQL

Sequenceテーブル
と統合ファイル中の配列を
テーブルに書き出し

フラグを廃止

reference

文献ID	文献名	URL
123456	文献1	https://www.ncbi.nlm.nih.gov/pubmed/123456
234567	文献2	-

Pubmedに含まれ
るものにはURLを
記載

experiment

ID	実験条件	実験結果	文献ID
P1	条件1	結果1	123456
P2	条件2	結果2	234567

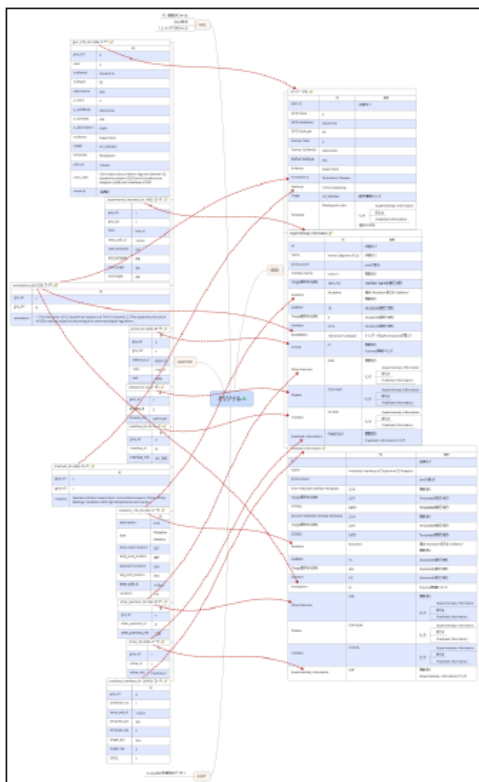
フラグを廃止

2つの"ex_..."
テーブルを統合

IDをmainと揃える

データキュレーション例 2: NBDC 元データの解体と再生

実際の解体～再生の例



メタデータのない
14テーブル



メタデータ付き
5テーブル