

研究開発支援総合ディレクトリ(ReaD)  
データ交換 データ仕様書(XML/CSV・TSV)

第 2.0 版

平成 22 年 2 月 19 日

## 目次

1. 目的	p.1
2. データ交換概要	p.1
3. 対象情報	p.1
4. データ交換実施の前提条件	p.1
5. 受入時のデータ更新	p.2
6. 研究活動情報ファイル	p.2
6.1 ファイルについて	p.2
6.2 ファイル仕様	p.3
6.3 データチェック仕様	p.5
6.4 日英連動項目	p.6
6.5 各項目における公開・非公開フラグについて	p.7

別紙 1: 研究活動情報ファイル一覧

別紙 2: チェックパターンファイル設計書

別紙 3: 学校機関コード表

別紙 4: 学校機関以外のコード表

## 1. 目的

本仕様書は、研究開発支援総合ディレクトリ(以下、ReaD と呼ぶ)において実施している JST と研究機関との間の ReaD データのやり取り(以下、データ交換と呼ぶ)において使用する研究活動情報のデータ定義を定めるものである。

## 2. データ交換概要

データ交換とは、ReaD と各研究機関で構築している研究活動情報に関するシステムとの間の「ReaD データのやり取り」である。以下に模式図を示す。

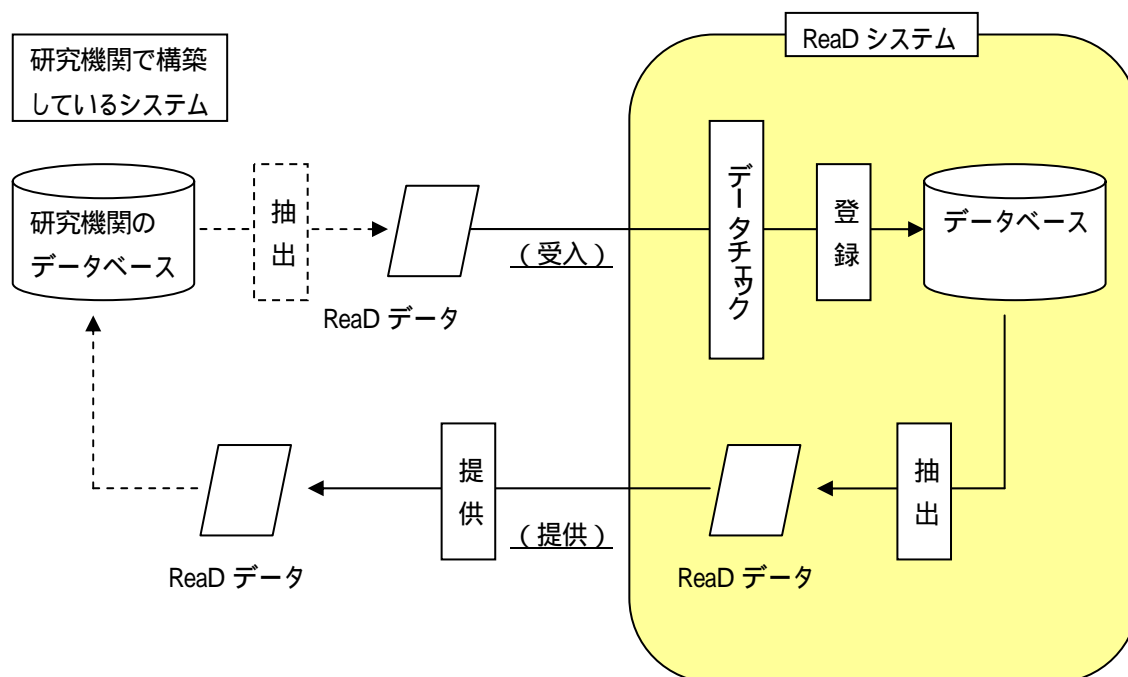


図 2-1 データ交換模式図

## 3. 対象情報

データ交換の対象となる情報は研究機関情報、研究者情報、研究課題情報、研究資源情報の 4 種類である。ただし、研究者情報については、データ交換実施機関に所属する研究者の情報、その他の情報については研究機関が保有あるいは実施している情報に限るものとする。

## 4. データ交換実施の前提条件

データ交換の実施にあたり、以下の条件を成立させておく必要がある。

1) データ交換実施機関に対して機関コードが採番されていること。

- ・機関コードは機関からの要請にしたがって、JST にて追加、変更等を管理する。
- ・コード体系は合計 10 桁で構成し、以下のように階層化して割り当てる。

機関コード(10 桁) = 大学等機関(4 桁) + 学部・部等(3 桁) + 学科・課等(3 桁)

2) 個々のデータ交換対象情報について、各コードが採番されていること。

- ・研究者コード 研究者に付与されるユニークな番号である。
- ・研究課題コード 研究課題に付与されるユニークな番号である。
- ・研究資源コード 研究資源に付与されるユニークな番号である。

上記のコードは各情報を ReaD 側で新規登録する際に自動付番される。

## 5. 受入時のデータ更新

公開するデータのみを受入する。ReaD に登録されているが、受入データに存在しない研究活動情報については何も行われない。

登録データの削除を希望する場合は、別途 ReaD 事務局に相談すること。

## 6. 研究活動情報ファイル

以下にデータ交換に使用する研究活動情報ファイルおよびそれらに含まれる項目の仕様を記載する。

### 6.1 ファイルについて

データ交換において使用する研究活動情報ファイルのフォーマットは **XML**、**カンマ区切りテキスト(CSV)**、**もしくはタブ区切りテキスト(TSV)**のいずれかである。

#### 1)ファイル名

データ交換において使用する研究活動情報ファイルのファイル名の命名規則は以下の通りである。

R e a d \_ A A A 9 . 拡 張 子

Read_	大文字の"R"ではじまり、"e"から"d"までは小文字、"_"を必ず付けることとする。
AAA	研究活動情報の種類ごとに以下の通りとする。 ・研究機関 ...kkn ・研究者 ...ksh ・研究課題 ...kdi ・研究資源 ...sgn
9999999999	機関コード(10桁)
_99	親子ファイル識別 ・_00 ...CSV、TSV ファイルの親ファイル、XML ファイル ・_01 ~ _15 ...CSV、TSV ファイルの子ファイル XML ファイルに子ファイルは存在しない。 付番については 2)ファイル一覧を参照のこと。
.拡張子	拡張子は以下の通りとする。(全て小文字) XML ファイル ....xml CSV ファイル ....csv TSV ファイル ....txt

#### 2)ファイル一覧

別紙 1:研究活動情報ファイル一覧を参照のこと。

### 3)ファイル構成について

- ・XML ファイル、CSV、TSV ファイルのいずれの場合について、差分データのみで構成されるファイルを使用して差分更新を行うことはできない。
- ・XML ファイルにおいては特定の研究活動情報に関する項目がすべて一つのファイルに記述されるため、子ファイルは存在しない。
- ・CSV ファイル、TSV ファイルにおいては、研究活動情報の項目ごとにファイルが分かれている。データ登録に際しては、**データの有無にかかわらず全てのファイルを作成しなければならない。**

## 6.2 ファイル仕様

以下に研究活動情報ファイルの仕様を説明する。

### 1)使用文字コード

- データ交換において使用する文字コードは **UTF-8** とする。

### 2)要素名とフィールド名

- XML ファイルの最下層の要素名と CSV、TSV ファイルのフィールド名は一致する。

### 3)XML ファイルについて

- ・データを設定しない項目については、空要素タグの設定をしてもしなくても良い。
- ・データ中に次の文字「&」、「'」、「<」、「>」を使用する場合はそれぞれ次の文字参照を使用すること。  
「&amp;」、「&quot;」、「&lt;」、「&gt;」  
文字参照を使用せずにそのまま記述されている場合はエラーとなる。
- ・上記 4 パターン以外の文字参照は使用しないこと。使用した場合はエラーとなる。
- ・制御コードを含めないこと。
- ・データ部の改行コードは変換不要でそのまま設定する。改行が設定できない項目は別紙 2:チェックパターンファイル設計書を参照のこと。
- ・ルート要素は「row」とする。
- ・行末の改行コードは、CR、LF、CRLF のいずれかとする。

### 4)CSV、TSV ファイルについて

- ・項目間のデリミタは、CSV ファイルの場合は「,(半角カンマ)」、TSV ファイルの場合は「タブ」とする。
- ・データ部は「”(ダブルコーテーション)」で囲むこと。
- ・空データの場合もデータ部は「”(ダブルコーテーション)」で囲むこと。
- ・データ部に「”(ダブルコーテーション)」を設定する場合は、「””」のように二重化すること。
- ・ファイルの一行目はフィールド名を記述すること。
- ・データ部の「改行コード」は「&#0d0a;」(半角小文字)に置き換えること。改行が設定できない項目は別紙 2:チェックパターンファイル設計書を参照のこと。
- ・文字参照は、「&amp;」、「&quot;」、「&lt;」、「&gt;」のみ使用可能。4パターン以外の文字参照は使用しないこと(使用した場合は、そのまま「&auml;」のように表示される。)
- ・制御コードを含めないこと。
- ・行末の改行コードは、CR、LF、CRLF のいずれかとする。

### 5)byte 数

ファイル項目で定義しているデータサイズは byte 数であり、文字数ではない。

### 6)属性とフラグ

XML ファイルの属性(attribute)には以下のものが存在する。

- ・language 属性            研究活動情報が日本語データか英語データかを指定する (JP:日本語、EN:英語)。
- ・public 属性              指定した項目の公開可否を指定する (true:公開、false:非公開)。  
この属性は研究者情報と課題情報に存在する。

CSV、TSV ファイルでは、上記の属性はそれぞれ「言語区分」、「非公開フラグ」という項目で管理する。

### 7)必須条件

必須条件として、以下の二種類の要素が存在する。

- ・必須項目                研究活動情報ファイル中に必ずデータが存在しなければならない項目。この項目が存在しないファイルはデータチェック時にエラーとなる。  
*例) 研究者情報における「研究者コード」、「所属機関コード」等。*
- ・条件付必須項目        必須項目以外の項目を登録する場合に、その項目を構成する下位項目のうち必ず存在しなければならない項目。  
*例) 研究者情報の『研究者業績(論文・解説)』における「論文名」等。  
『研究者業績(論文・解説)』にデータが1件も存在しない場合でもエラーにはならないが、データが存在する場合は必ず「論文名」が必要となる。*

### 8)データ型

データ型は以下のものが存在する。

表 6-2-1 データ型一覧

名称	説明
半角数字	0~9
半角英字	a~z、および A~Z
半角英数字	0~9、a~z、および A~Z
半角英数字記号	基本ラテンとラテン補助(U+0020-U+00ff)
全半角	表 6-2-2 を参照

表 6-2-2 全半角で使用可能な文字範囲 (UTF-8)

バイト数	範囲			
	1 バイト目	2 バイト目	3 バイト目	4 バイト目
1 バイト文字	09,0a,0d,20-7f			
2 バイト文字	c0-df	80-bf		
3 バイト文字	e0-ef	80-bf	80-bf	
4 バイト文字	f0-f7	80-bf	80-bf	80-bf

但し efbfbe, efbfbf を除く。

#### 9)ファイル項目一覧

別紙 2: チェックパターンファイル設計書を参照のこと。

### 6.3 データチェック仕様

データ登録を行う前に以下のデータチェックを行う。

#### 1)ファイルチェック

ファイル名が規則どおり付与されているか。必要なファイルが全てそろっているか否か確認する。

#### 2)単項目チェック

ファイルの各項目(要素)について以下のチェックを行う。

必須チェック 必須項目にデータが存在しているか否かのチェックを行う。

byte 数チェック byte 数のチェックを行う。固定長、可変長の二種類存在する。

データ型チェック 1 データ型が正しいか否かチェックを行う。

データ型チェック 2 上記のデータ型チェックに加えて、日付、メールアドレス、URL については各々、日付、メールアドレス、URL の形式チェックを行う。

コード存在チェック コード類において、指定されたコード値の DB との整合性チェックを行う。

#### 3)項目間チェック

データ項目間の依存関係のチェックを行う。

#### 4)チェックパターン一覧

各項目に対して行われるチェックの内容は別紙 2: チェックパターンファイル設計書を参照のこと。

#### 6.4 日英連動項目

下記の項目は日本語サイトと英語サイトの連動項目になる。

日・英両方のデータを受入した場合は、後に受入した方の値が登録される。

研究活動情報の種類	項目名
機関研究活動情報	・研究分野コード
研究者研究活動情報	・文部科学省科学研究費研究者番号 ・研究分野コード ・ホームページURL ・海外での国際会議、学会への出席回数及び発表回数 ・公開非公開
課題研究活動情報	・研究機関 ・研究分野コード ・研究予算 ・公開非公開
資源研究活動情報	・研究分野コード

## 6.5 各項目における公開・非公開フラグについて

研究活動情報の種類	項目名	公開非公開を指定できる項目	非公開項目 2	原則公開項目 3
研究者研究活動情報	一般公開の可否【必須】			
	氏名特記事項			
	所属部署名			
	職名【必須】			
	連絡先住所			
	TEL			
	FAX			
	メールアドレス	1		
	生年月日【必須】			
	性別			
	研究分野キーワード【必須】			
	ホームページURL			
	研究テーマ【必須】			
	出身大学院			
	出身学校			
	取得学位			
	研究職歴			
	委員歴・役員歴			
	受賞学術賞			
	所属学会			
	研究業績(論文・解説)			
	研究業績(著書)			
	研究業績(口頭発表・ポスター)			
	研究業績(特許)			
研究業績(その他の活動)				
海外活動のための渡航回数				
海外の会議、学会出席回数 / 発表回数				
使用外国語				
課題研究活動情報	一般公開の可否【必須】			
	プロジェクト名			
	プロジェクト代表機関名			
	研究代表者名			
	研究予算			
	概要			
	キーワード			
	研究期間(研究開始年度、研究終了年度)			
	研究成果			

1 「公開」を指定した場合、検索画面にはメールアドレス(文字列)ではなく「研究者に問合せ」ボタンが表示される。

2 「公開」を指定しても検索画面には表示されない。

3 原則公開項目は公開が原則であり、項目毎の公開・非公開フラグは「公開」を指定することとする。

なお、「非公開」を指定してもデータ受入は可能であり、検索画面でも公開されないが、ID・PWでオンライン更新を行うとデータは公開される。