

リポジトリの 信頼性を高める為に必要なアイテム (試案)

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

本発表の目的

- リポジトリの運営者（またはファウンダー）の方へ
 - リポジトリの運営方法を見直すヒントを提供したい
 - リポジトリがCTSなどの認証を獲得するためのヒントを提供したい
- リポジトリの（潜在的な）利用者の方へ
 - 利用するリポジトリ選びのヒントを提供したい
 - 利用しているリポジトリを叱咤激励する材料を提供したい

CTSについておさらい

- 「信頼できるデータリポジトリを認定するための中核的な統一要件」を満たすリポジトリに対して付与される認証
 - Data Seal of Approval (DSA) と国際科学会議 (ICSU) 世界科学データシステム (WDS) がRDAのワーキンググループで策定
- 各リポジトリに対して16項目の自己評価を実施
- 各項目は5段階 (0~4) で評価
- 概ね3以上のレベルであれば認証される (らしい)



南山『信頼できるデータリポジトリの中核的な統一要件』
<http://current.ndl.go.jp/e1888>
から引用

CTSの評価（5段階）

- 0 - 適用不可（Not applicable）
- 1 - まだ考慮されていない
（The repository has not considered this yet）
- 2 - 概念上は存在する（The repository has a theoretical concept）
- 3 - 実装フェーズにある
（The repository is in the implementation phase）
- 4 - 実装済（The guideline has been fully implemented in the repository）

南山、前掲記事から引用

CTSの要件（16項目）

■組織的な基盤（Organizational Infrastructure）

■R1. ミッション／視野（Mission/Scope）

研究領域におけるリポジトリのミッションを明らかにするもの

■R2. ライセンス（Licenses）

データ提供者との合意事項、データ利用者に対するライセンス等

■R3. アクセスの継続性（Continuity of access）

保有するデータへのアクセスとデータの保存を保証する計画等

■R4. 機密保持／倫理（Confidentiality/Ethics）

システムのセキュリティ（特に機密保持）、スタッフへの倫理教育等

■R5. 組織的な基盤（Organizational infrastructure）

ファンディングとガバナンス

■R6. 専門家によるガイダンス（Expert guidance）

専門家の指導とフィードバック

参考：

南山、前掲記事および

絹谷・南山 RDUF小委員会「国内の分野リポジトリ関係者のネットワーク構築」検討資料

CTSの要件（16項目）

- **デジタルオブジェクト管理（Digital Object Management）**
 - **R7.データの完全性・信頼性（Data integrity and authenticity）**
データの変更管理
 - **R8. 評価（Appraisal）**
データの受け入れ管理
 - **R9.保管手続きの文書化（Documented storage procedures）**
データの保存ポリシーやバックアップ
 - **R10.保存計画（Preservation plan）**
 - **R11.データの品質（Data quality）**
 - **R12. ワークフロー（Workflows）**
データの受け入れから流通までのワークフロー
 - **R13.データの発見と識別（Data discovery and identification）**
データの検索方法と永続的な識別子
 - **R14.データの再利用（Data reuse）**
メタデータ

CTSの要件（16項目）

■技術（Technology）

- R15.技術的な基盤（Technical infrastructure）
リポジトリのインフラ（ハードウェア・ソフトウェア）
- R16.セキュリティ（Security）
特にシステム障害への耐性

ここまでの経緯（小委員会の活動）

- CTS要件の日本語訳
- CTS認証済のリポジトリから、3つの自己評価例を抽出
 - DRUM（ミネソタ州立大学の機関リポジトリ）
 - Finnish repository（フィンランド社会科学分野の研究データリポジトリ）
 - NGDC（英国 地球科学データリポジトリ）
- 各リポジトリについて、16項目の自己評価を分析
- 分析結果からリポジトリに必要なアイテムを抽出（八塚による試案、小委員会内で議論・検討中）

アイテムとは

- 自己評価事例を分析すれば、**どのように**書けば良いか（認証を受けられるか）がわかる
- しかし、自己評価を書く以前にリポジトリとして**何を**保有（あるいは公開）している必要があるのかわからない
- 「何を」にあたるのがアイテム

抽出したアイテム（試案）

■組織

- 安定した運営組織
- 運営組織内でのフォロー委員会または外部諮問委員会

■文書・ポリシー・マニュアル類

- リポジトリのミッションや組織での承認レベルを記載した文書またはWebサイト
- リポジトリの長期運営計画
- データ提供者との契約文書
- データ利用者との契約文書
- データ保存ポリシー
- データ開示ポリシー
- 個人情報取り扱いポリシー
- データコレクションポリシー（品質、メタデータ、ファイルフォーマット等）
- データキュレーションマニュアル
- データ管理マニュアル
- 障害復旧マニュアル

抽出したアイテム（試案）

■情報インフラ（ハードウェアを含む）

- リポジトリのインフラ
- データ管理システム
- データバックアップシステム
- データ検索システム

■教育

- スタッフへの教育（データ保護）
- スタッフへの教育（ITスキルやデータのキュレーション）

■その他

- PID

最後に

- リポジトリの運営者（またはファウンダー）の方へ
 - 「アイテムは全部そろっている」というところは少ないのでは？
 - アイテムそのものはなくても、同等のものは既にあるかもしれません
 - それらを整理・公開するだけでリポジトリ運営は改善されるはず
 - それでもまだアイテムが足りないかもしれません
 - 私たちも同じです
 - 一緒に活動しませんか⇒RDUF小委員会「国内の分野リポジトリ関係者のネットワーク構築」
- リポジトリの（潜在的な）利用者の方へ
 - これらのアイテムを少しでも多く保有しているリポジトリを利用してください
 - 利用しているリポジトリにこれらのアイテムがなければ（愛をこめて）叱咤激励してください
 - リポジトリ側にはきっと（良い意味で）プレッシャーになるでしょう

謝辞

- 研究データ利活用協議会(RDUF)
小委員会「国内の分野リポジトリ関係者のネットワーク構築」