
学術データ利活用のための バックアップとアーカイブ

2021年12月15日

AXIES年次大会

RDM部会企画セッション(その2)

バックアップとアーカイブ (このセッションでの文脈)

- 「バックアップ」とは、システムやデータの可用性を維持するために今の状態を保存する行為
- 「アーカイブ」とは、データの分類、メタデータの追加等、第三者が利用できるようデータを整理し、長期にわたり保存する行為
- 大学等組織が「データを預かる」行為は、本質的には「アーカイブ」
→ それが「誰の」「何のために」の明示が必要

データ利活用シーンの拡大

- いたるところに「データ」、アカデミアでは
 - 研究データ
 - 教育データ
 - IRデータ
 - :
- 目的にあった、データ保存のソリューションが必要
 - RDM部会としては「研究データ」が出发点だが...

組織的RDMの技術的側面

- 永続的(デジタル)識別子の普及(Persistent Identifier)
論文・データ: DOI 人物: ORCID 組織: RoR 研究助成金: FundRef
研究資料: RRID, IGSN 研究・観測機器: PIDINST ...
- データや人物、組織がこれらPIDと結びつき、流通
→ 研究者(大学)の役割は、PIDにつながる実体を確保し、プレゼンスを高める
- 特に研究データについては参照可能性→FAIR原則
F(indable), A(ccessible), I(nteroperable), R(eusable)
- 更に実体を預かるメディアへの信頼性→TRUST原則
T(ransparency), R(esponsibility), U(ser Focus), S(ustainability),
T(echnology)
- ようやく「大学ICT」要求の出番
ただ、「持続するサービスモデル」への道のりは険しい

誰に対しFAIR, TRUSTであるか

- 誰に対し?
 - 「いきなり公衆」ではなく、「まずは身近な関係者」が現場の肌感覚
 - 少なくとも「昨日の自分」に対しFAIRであること、TRUSTなメディアを利用することは必要
- これを大学等、「組織のインフラ」へと拡大するには?
 - いきなり「運用体制」「維持費用」「利用率」といった問題となってしまうがち
- 組織としてどの様な目的や利用シーンを定め、ストレージソリューションを選定するか

プログラム

- Google Cloud におけるマルチクラウド・データファブリック
木内満歳 (グーグル・クラウド・ジャパン合同会社)
- NetAppが実現するData Fabric
～スピードが尺度になる時代の次世代デジタルインフラとは～
脇昌弘 (ネットアップ合同会社 パートナーアライアンス営業本部
ビジネス開発推進NetApp DXセンター)
- 富士フイルム: 大容量／長期間のデータ保存を実現するアーカイブシステム
大月英明 (富士フイルム株式会社 記録メディア事業部)
- 日立製作所: データドリブンの時代を支える次世代ストレージ基盤
長堀隆史 (株式会社日立製作所 プロダクツビジネス本部
ストレージビジネス推進部)

発表: 15分
質疑応答: 5分

フォローアップ、議論の拡大

「RDM日本コミュニティ」の掲示板を立ち上げています。ご活用ください。

wa.rdm-japan.org

書き込みにはユーザ登録が必要
グループ毎の「閉じた」掲示板も作れます(応相談)

FAIR原則、TRUST原則

- いずれも、「オープンサイエンス」実現のための要件としてまとめられたもの
- FAIR 原則 (オープンな研究データに求められる要件)
<https://dx.doi.org/10.18908/a.2018041901>
- TRUST原則 (オープンな研究データを預かるメディアに求められる要件)
<https://current.ndl.go.jp/e2296>

FAIR 原則

(オープンな研究データに求められる要件)

- DOI:10.18908/a.2018041901
- 見つけられる(Findable)
- アクセスできる(Accessible)
- 相互利用できる(Interoperable)
- 再利用できる(Re-usable)

TRUST 原則

(オープンな研究データを預かるメディアに求められる要件)

- <https://current.ndl.go.jp/e2296>
- 透明性 (Transparency)
- 責任 (Responsibility)
- ユーザ志向 (User Focus)
- 持続性 (Sustainability)
- 技術 (Technology)