

デジタルアーカイブ学会
第5回研究大会ワークショップ

デジタルデータの保存・管理

ー現場視点からの共通課題を考える

2020年10月24日(土) 14:00～15:30

早稲田大学 坪内博士記念演劇博物館 中西 智範

ワークショップ企画趣旨

(第5回研究大会 HPより)

文化・知識資源を長きにわたり保存する使命をもつ（長期保存、長期利用保証の観点を含む）、図書館をはじめ、博物館・美術館、文書館等は、デジタルデータの保存・管理・活用にあたり、様々な課題に直面している。

国内では、MLAなどをキーワードに、利用者サービスの面など、活用に関する連携協力は盛んに議論されるものの、ビットブリザベーションやマイグレーションなど、デジタルアーカイブの基礎となるデジタルデータそのものの保存や管理については、各組織で多くの問題を抱え、個別に対応がなされている状況であろう。

2019年3月、このデジタルデータの保存や管理の課題に着目し、図書館・博物館・美術館・文書館などの業務従事者や、有識者で構成されるワーキンググループ『デジタルデータ管理勉強会(仮)』が早稲田大学演劇博物館の呼びかけをきっかけに、活動をスタートした。

本企画では、ワーキンググループの立ち上げの経緯やこれまでの活動の中で議論された保存・管理のための議論などを紹介(第一部)するとともに、ヒト・モノ・コストの不足といった組織共通の課題に着目し、今後の展望や連携協力の意義などについて、フロアの参加者を交えながら、ディスカッション(第二部)を行う。

進行内容

第1部

イントロダクション(10分) 中西智範

- ワークショップの趣旨
- 「勉強会」の活動について
- “デジタルデータの保存・管理”におけるフォーカス
- 前提となる問題意識

事例報告(30分) ライトニングセッション

第2部

テーマディスカッション(50分)

モデレータ 中村覚（東京大学情報基盤センター）

イントロダクション

ワークショップの趣旨

ひとりで悩まない
ために・・・

“デジタルデータの保存・管理”をめぐることは・・・

＜文化資源を保有する諸機関の現状＞

デジタルアーカイブをはじめとし、機関が保有するデジタル資源が増加するなか、機関の担当者はその保存・管理に苦心している

＜ワークショップで行うこと&期待＞

まず必要なこと⇒

現場の課題意識を共有

～デジタル資源における「保存」の最優先課題・任務を考えるきっかけ
の場となること～

期待されること⇒

国内での問題解決に向けた議論や活動のきっかけとなるとともに、現状の課題について、解決の糸口が見いだせること

「勉強会」の活動について

- 早稲田大学演劇博物館での、顕在するデジタルデータ保存・管理の課題があった
- 企画書をもとに、課題意識の共通する組織に声かけを行い、勉強会としてスタート

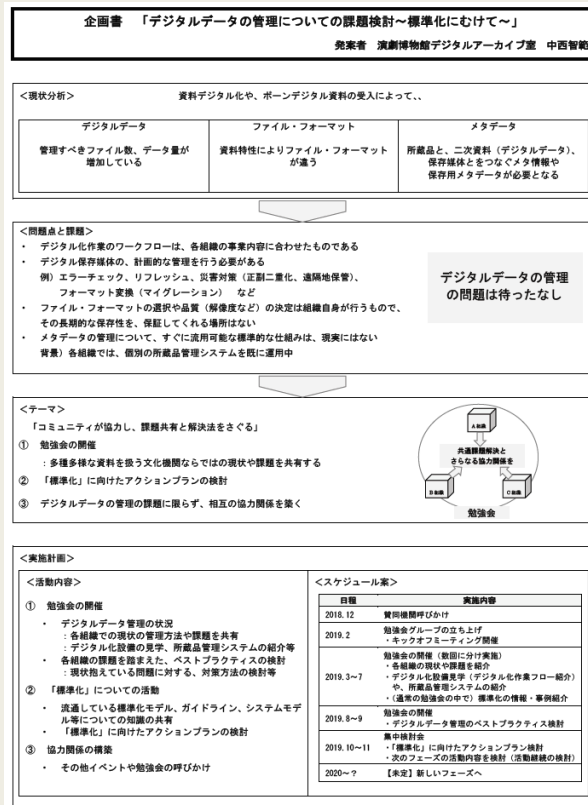
2018年12月 企画提案

2019年3月 初打ち合わせ

現在までに 計 8 回開催

- 各組織の現状や課題の共有
- アクションプラン、検討課題を議論

本ワークショップのモチベーションは、勉強会の中から生まれたもの



11機関、19名で活動（2020年10月24日時点）

“デジタルデータの保存・管理”におけるフォーカス

建物の構造のように、基礎部分からの対応が求められる

技術

- Bitstream Preservation
 - インテグリティの保証
- 保存システム(IT Infrastructure)
 - デジタル保存システム、ストレージ
- フォーマット
- メタデータ
- マイグレーション

コンテンツ

- Content Preservation
 - コンテンツの選定、オーセンティシティ、品質保障
 - アキュジション、インジェスト、アーカイブ、トランスファー
- 検索とアクセス
 - 権利 ・ アクセスライツ、著作権

保存

マネジメント

- デジタル保存・管理サービス ・ データアドミニストレーション、OAIS参照モデル
- IT capability (情報化能力) ・ 保守運用、ファシリティ ・ ITリソース管理、情報セキュリティ
- 技術管理(テクノロジー・モニタリング)

ケイパビリティ(組織力)

- ポリシーと戦略 ・ 保存戦略と予算計画、ビジネスケース、リスク管理、アセスメント
- 持続可能性 ・ 資金調達、エコ
- 標準化・ドキュメント化 ・ 用語集、ガイドライン
- 人的資源と教育
- コミュニティ

管理

“デジタルデータの保存・管理”におけるフォーカス

管理

保存

デジタル保存・管理		現在地		1年 ^{評価}		3-5年 ^{/更新・入替}		10-30年 ^{/現実目標}		永年保存 ^{/理想}	
管 理	ケイパビリティ(組織力)			ポリシー策定	ポリシー等評価	ポリシー等更新				このよ 話し	
	ポリシーと戦略			リスク管理		リスク管理					
	持続可能性			予算計画	資金調達						
	標準化・ドキュメント化			ガイドライン							
	人的資源と教育				人材確保 職員教育	雇止め 引継ぎ	世代交代				
	コミュニティ			コミュニティ連携	コミュニティ成果						
	マネジメント										
	デジタル保存・管理サービス			OAIS参照モデル				設備老朽化			
	IT capability (情報化能力)			バックアップ							
	技術管理(テクノロジー・モニタリング)			セキュリティ ITリソース	保守料						
			モニタリング方針	モニタリング	耐用年数	技術陳腐化					
保 存	コンテンツ										
	Content Preservation			リスト化	コンテンツ棚卸						
	検索とアクセス			選定	廃棄選択						
	権利			アクセス権							
				著作権・許諾	法改正			許諾再確認			
	技術										
	Bitstream Preservation			技術全般 現状把握							
	保存システム(IT Infrastructure)			インテグリティ保証		データチェック					
	フォーマット			フォーマット種類	処理能力評価	システム更新	フォーマット陳腐化				
	メタデータ				メタ標準化						
マイグレーション			マイグレーション計画		マイグレーション						

このような話題を勉強会で
話し合ったりしています

問題意識と課題

2つの問題意識と、4つの課題

<問題意識>

- | |
|-------------------|
| ①問題意識の共有の必要性 |
| ②現場作業者の協力体制構築の必要性 |

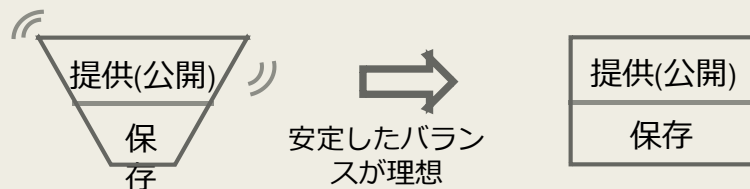
<課題>

- | |
|----------------------------|
| ③“デジタルアーカイブ”の再定義 が必要 |
| ④保存すべき対象を明確化することが必要 |
| ⑤デジタル保存の基本は“メンテナンス”をとらえること |
| ⑥保存にかかるコストの可視化が必要 |

前提となる問題意識

①問題意識の共有の必要性

- 各機関は「保存・管理は万全にできている」という誤った認識は無い
か？
 - デジタルアーカイブを公開 ≠ 保存できている
 - 公開コンテンツ以外にも保存・管理すべき対象は多い
- 国内での「デジタル保存」の議論の場、コミュニティが無い!?
 - 国際会議「iPRES」で取り上げられるテーマ類
- “デジタルアーカイブ”という用語の再定義
 - ～様々なデジタル情報資源を収集・保存・提供するための仕組み
の総体～内閣府知財戦略推進「国におけるデジタルアーカイブの
取り組み(H29)」



前提となる問題意識

②現場作業者の協力体制構築の必要性

- 人材とコストの問題
 - “保存”はお金を生まない、という共有意識
→“収集”と“利活用”のエリアに事業予算や補助金が割り当てられる
- 今対策しなければ、失われていくデジタル資源がある
 - 各組織の担当者は、個別に対策をとっている＝マネジメントのための知識と経験が蓄積されている（はず！）



- 協力体制を築き、できることをやる必要があるのではないか？
（最低限レベルからの対策方針の検討などでもいい）

まず対応したい課題

A.“デジタルアーカイブ”の再定義

- Wikipedia「デジタルアーカイブ(ja.)」→「Digital Preservation(en.)」

B.保存すべき対象を明確化することが必要

- 相互の組織のミッションの違い・残すべき対象を認識すること
- 共通的には「中身・コンテンツが重要」という点
- 保存期間について(次の世代に託す：人材のライフサイクル)

C.デジタル保存の基本は“メンテナンス”

- 保存に適した環境＋主体性のある管理の継続(正倉院)
- 保存とバックアップを混同しないこと

D.保存にかかるコストの可視化が必要

- 持続可能で計画的な対応を行うために重要

事例紹介 ライトニングセッション

※事例紹介については、
ワークショップ当日の発表のみに限っています

事例紹介 ライトニングセッション

- 各機関から以下のテーマについて発表

- ①組織のミッション
- ②具体的な対策
- ③コストと人員
- ④困っていること

<セッションのポイント>

- デジタル保存というテーマの下では、MLAという枠組みは意識する必要はない！？
- 共通課題を協力し対策していくことが重要ではないか！？

発表者

慶應義塾大学アート・センター（本間友）

国立映画アーカイブ（岡本直佐）

国立近現代建築資料館（飛田ちづる）

国立公文書館（吉田敏也）

国立国会図書館（木下貴文）

東京国立博物館（村田良二）

東京大学附属図書館（中村美里）

早稲田大学演劇博物館（中西智範）

第2部 テーマディスカッション

モデレータ

中村 寛 東京大学情報基盤センター

流れ

- 事例報告に対する質疑応答
- テーマディスカッション

事例報告に対する質疑応答

【事例紹介】

事例紹介(イントロダクション)

〇〇のミッション

- 主な収集対象/所蔵資料
 - ××
 - ××
- 長期保存の対象・保存期間
 - 〇〇を××の期間保存
- 所有しているデジタルコンテンツの量
 - ××

長期保存の取組

事例紹介(イントロダクション)

- 方針・計画など
 - 計画はないが、××などの内容を担当者間で共有
- 資料デジタル化
 - ××コレクションのデジタル化
 - ××のデジタルアーカイブで公開（又は構築中）
- デジタルデータ長期保存対策
 - 提供用とは別に保存用データを××の媒体で保存、〇年ごとに媒体交換
 - 長期アクセス保証のため、永続的識別子（DOI）を付与

モノ

予算と体制

事例紹介(イントロダクション)

- 予算（システムの場合はリプレースの確保も）
 - デジタル化予算
 - ×円
 - データベース運営予算
 - ×円 リプレース経費の確保はなし
 - 長期保存対策経費
- 体制
 - 〇名のスタッフで兼務で実施

課題—困っていること

事例紹介(イントロダクション)

- ××がない
 - ××！
- 〇〇がおかしい
 - ××
- △△に対応できない
 - ××

ヒト

カネ

テーマディスカッション

事前アンケート結果の共有

テーマ案（事前アンケート結果：10/23時点）

1. 長期保存のための技術：66.7%
2. メタデータ：43.3%
3. 長期保存のための媒体：43.3%
4. 人的資源と教育：40%
5. 失敗事例：40%

技術

Bitstream Preservation

- ・ インテグリティの保証

保存システム(IT Infrastructure)

- ・ デジタル保存システム、ストレージ

フォーマット

メタデータ

マイグレーション

コンテンツ

Content Preservation

- ・ コンテンツの選定、オーセンティシティ、品質保障
- ・ アクイジション、インジェスト、アーカイブ、トランスファー

検索とアクセス

- 権利 ・ アクセスライツ、著作権

保存

マネジメント

デジタル保存・管理サービス ・ データアドミニストレーション、OAIS参照モデル

IT capability (情報化能力) ・ 保守運用、ファシリティ・ITリソース管理、情報セキュリティ

技術管理(テクノロジー・モニタリング)

ケイパビリティ(組織力)

ポリシーと戦略 ・ 保存戦略と予算計画、ビジネスケース、リスク管理、アセスメント

持続可能性 ・ 資金調達、エコ

標準化・ドキュメント化 ・ 用語集、ガイドライン

人的資源と教育

コミュニティ

管理

管理	ケイパビリティ(組織力)		ポリシー策定	ポリシー等評価	ポリシー等更新		
	ポリシーと戦略		リスク管理				
	持続可能性		予算計画	資金調達			
	標準化・ドキュメント化		ガイドライン				
	人的資源と教育			人材確保 職員教育	雇止め 引継ぎ	世代交代	
	コミュニティ		コミュニティ連携	コミュニティ成果			
	マネジメント		OAIS参照モデル				
	デジタル保存・管理サービス		バックアップ			設備老朽化	
	IT capability (情報化能力)		セキュリティ ITリソース	保守料			
	技術管理(テクノロジー・モニタリング)		モニタリング方針	モニタリング	耐用年数	技術陳腐化	
保存	コンテンツ		リスト化	コンテンツ棚卸			
	Content Preservation		選定	廃棄選択			
	検索とアクセス		アクセス権				
	権利		著作権・許諾	法改正		許諾再確認	
	技術		技術全般 現状把握				
	Bitstream Preservation		インテグリティ保証		データチェック		
	保存システム(IT Infrastructure)			処理能力評価	システム更新		
	フォーマット		フォーマット種類			フォーマット陳腐化	
	メタデータ			メタ標準化			
	マイグレーション		マイグレーション計画		マイグレーション		

補足スライド

テーマ案

1. 長期保存全般について
2. 長期保存のための技術
3. 保存にかかる費用
4. コミュニティ形成

1. 長期保存全般について

デジタル保存連合 ラピッド・アセスメントモデル (DPC RAM)

Digital Preservation Coalition Rapid
Assessment Model (2019年9月公開)

DOI: <http://doi.org/10.7207/dpcram19-01>

Q. どのようなもの？

- 組織がデジタル保存を行うに当たっての体制や技術、機能における成熟度の評価を行うための基準を示し、自己評価できるよう開発されたツール

「組織レベルの能力」と「サービスレベル（＝運用・技術レベル）の能力」のパートで構成され、11の要素について段階的に評価を行う。

0- 最低限認識している

1- 認識している

2- 取り組まれている

3- 管理されている

4- 最適化されている

組織レベルの能力		
A	組織的活力	デジタル保存活動のためのガバナンスや、組織構造、人材配置、リソース配分
B	方針と戦略	デジタルアーカイブの運用や管理を行うための方針、戦略、手続き
C	法的基盤	デジタルコンテンツの取得や保存、アクセスの提供に関わる、契約上の、あるいは許諾その他の法的権利と責任の管理（例えば、許諾、著作権、使用条件、データ保護規定等）
D	IT能力	デジタル保存活動を支える情報技術についての能力
E	継続的改善	現在のデジタル保存能力のアセスメント、目標の定義と進捗のモニタリングのプロセス
F	コミュニティ	広範なデジタル保存コミュニティとの関わりと貢献
サービスレベルの能力		
G	取得、移管、取込み	コンテンツを取得、または移管し、デジタルアーカイブに取込むプロセス
H	ビット列保存	保存対象であるデジタルコンテンツのストレージとその完全性が保たれていることを保証するプロセス
I	コンテンツ保存	デジタルコンテンツの意味内容または機能を保存し、長期にわたって継続的なアクセス性と使用性を保証するプロセス
J	メタデータ管理	保存されたデジタルコンテンツの保存、管理、利用のために十分なメタデータを作成、維持するプロセス
K	発見とアクセス	ユーザーのためにデジタルコンテンツを発見し、アクセスの提供を可能にするプロセス

Q. 本モデルの特徴（本文より）

- どのような規模や分野の組織にも応用できる
- 長期的価値のある全てのコンテンツに応用できる
- 保存についての戦略や解決方法に依存しない
- 既存のグッドプラクティスに基づく
- わかりやすく速やかに実施できる

Q. どのように使えそう？

成熟度レベルに関する情報共有を促進し、個々の組織が、その現状を評価結果に基づいて、DPC全体またはDPCメンバーの類似機関（美術館・図書館・文書館・博物館など）と比較しあうことができる、とされる

標準化 : Levels of Digital Preservation

- 米・国家デジタル管理連盟 (NDSA)
- 組織がデジタル保存に取り組む際の推奨事項を4つのレベルに分けて表形式に整理したもの
- <https://ndsa.org/publications/levels-of-digital-preservation/>

Functional Area	Level			
	Level 1 (Know your content)	Level 2 (Protect your content)	Level 3 (Monitor your content)	Level 4 (Sustain your content)
Storage	Have two complete copies in separate locations Document all storage media where content is stored Put content into stable storage	Have three complete copies with at least one copy in a separate geographic location Document storage and storage media indicating the resources and dependencies they require to function	Have at least one copy in a geographic location with a different disaster threat than the other copies Have at least one copy on a different storage media type Track the obsolescence of storage and media	Have at least three copies in geographic locations, each with a different disaster threat Maximize storage diversification to avoid single points of failure Have a plan and execute actions to address obsolescence of storage hardware, software, and media
Integrity	Verify integrity information if it has been provided with the content Generate integrity information if not provided with the content Virus check all content; isolate content for quarantine as needed	Verify integrity information when moving or copying content Use write-blockers when working with original media Back up integrity information and store copy in a separate location from the content	Verify integrity information of content at fixed intervals Document integrity information verification processes and outcomes Perform audit of integrity information on demand	Verify integrity information in response to specific events or activities Replace or repair corrupted content as necessary
Control	Determine the human and software agents that should be authorized to read, write, move, and delete content	Document the human and software agents authorized to read, write, move, and delete content and apply these	Maintain logs and identify the human and software agents that performed actions on content	Perform periodic review of actions/access logs
Metadata	Create inventory of content, also documenting current storage locations Backup inventory and store at least one copy separately from content	Store enough metadata to know what the content is (this might include some combination of administrative, technical, descriptive, preservation, and structural)	Determine what metadata standards to apply Find and fill gaps in your metadata to meet those standards	Record preservation actions associated with content and when those actions occur Implement metadata standards chosen
Content	Document file formats and other essential content characteristics including how and when these were identified	Verify file formats and other essential content characteristics Build relationships with content creators to encourage sustainable file choices	Monitor for obsolescence, and changes in technologies on which content is dependent	Perform migrations, normalizations, emulation, and similar activities that ensure content can be accessed

調査報告（海外事例なども）

<https://www.ndl.go.jp/jp/preservation/dlib/research.html>



[本文へ](#) [サイトマップ](#) [よくあるご質問](#) [Language](#)

ホーム	サービス・国会関連情報	来館される方へ	事業紹介	関係者の方へ	国立国会図書館について
---------------------	-----------------------------	-------------------------	----------------------	------------------------	-----------------------------

[ホーム](#) > [資料の保存](#) > [電子情報の長期的な保存と利用](#) > 電子情報の長期利用保証に関する調査研究

資料の保存
ニュース
所蔵資料の保存
資料のデジタル化について
電子情報の長期的な保存と利用
保存協力
IFLA/PACアジア地域センター
マニュアル・パンフレット・翻訳資料

電子情報の長期利用保証に関する調査研究

国立国会図書館では、平成14年度から電子情報の長期的な保存と利用保証のための調査研究を進めています。具体的には、文献やアンケートに基づく国内外の先行事例・動向の調査のほか、当館が所蔵するCD-ROMやフロッピーディスクの再生可能性の調査、マイグレーションやエミュレーションに関する調査、長期保存のためのシステムに関する調査などを行っています。

調査報告書（令和元年度）

国立国会図書館が収集・保存している電子情報の長期保存対策の具体化の検討に資するために、最近の関連論文、海外機関（図書館、文書館、博物館等）の長期保存に関する取組、電子情報の長期保存に関連する国内製品について、調査を実施しました。本報告書では、電子情報の長期保存に係る最新動向を多面的に概観できるよう、調査結果を取りまとめて報告します。

[電子情報の長期保存対策に係る調査報告書（令和2年8月）](#)

[エグゼクティブ・サマリー（PDF: 106KB）](#)

国がまとめた長期保存ガイドライン

- デジタルアーカイブのための長期保存ガイドライン（2020年版）
 - https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/pdf/guideline2020.pdf
- デジタルアーカイブジャパン推進委員会及び実務者検討委員会（内閣府知的財産戦略推進事務局）で作成
 - https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/index.html
- 「本ガイドラインは、アーカイブ機関がデジタルデータの長期保存のためにどういった観点での検討が必要かを示すためのもの」

目次

1. デジタルデータ管理の基本的考え方
2. デジタルデータの保存方針・計画策定
3. デジタルデータ管理
4. デジタルデータ管理
5. 安定稼働できるシステム
6. メタデータ管理
7. アクセス保証の仕組み
8. デジタルデータ移行性の担保
9. 人材育成

2. 長期保存のための技術

長期保存のための技術

- Bitstream Preservation
- 保存システム（IT Infrastructure）
 - デジタル保存システム、ストレージ
- フォーマット
- メタデータ
- マイグレーション

長期保存のための媒体

- メディア・媒体

- [デジタルデータ保存のためのメディア | BDCブログ](#)
- HDD, BD, LTO, 光ディスク, クラウドストレージ

- ファイルフォーマット

- [国立国会図書館が採用するファイルフォーマット | 国立国会図書館-National Diet Library](#)
- 静止画 : TIFF, JPEG 2000, PDF, ...
- 動画 : MP4, ...

長期保存のための技術

- OAIS参照モデル
 - [情報の長期保存に有力なOAIS参照モデル 第1回（全2回） | BDCブログ](#)
 - [情報の長期保存に有力なOAIS参照モデル 第2回（全2回） | BDCブログ](#)
- [入澤道場](#)：アーカイブズ学に基づくデジタルアーカイブを考える会
 - archivematica: open-source digital preservation system
 - [BitCurator](#)

メタデータ

- PREMIS: Preservation Metadata: Implementation Strategies
 - “Data Dictionary for Preservation Metadata”（保存メタデータのためのデータ辞書）
- METS: Metadata Encoding and Transmission Standard
- Dublin Core
- MODS : Metadata Object Description Schema

[MODS, METS, and other metadata standards](#)

METS resource and metadata bundle

Resource Space	METS Information
Audio File (WAV, etc.)	Description MODS MARCXML
Video File (MPEG, etc.)	Technical MIX (image)
Text File (TEI, etc.)	Rights (METSRights, etc.)
Image File (TIFF, etc.)	Preservation PREMIS
Web File WARC	Structure Map File Section

OPEN SOURCE SOFTWARE FOR DIGITAL PRESERVATION REPOSITORIES: A SURVEY

Table 3. Standards and protocols supported by each of the surveyed software solutions.

- International Journal of Computer Science & Engineering Survey (IJCSES) Vol.8, No.3, June 2017
- <https://arxiv.org/pdf/1707.06336.pdf>

[illegible]

3. 保存にかかる費用

保存にかかる費用

組織での保存活動を円滑（円熟）にするには、保存活動・事業を組織内で重要な位置まで引き上げなければならない。

→保存ミッションの明文化、コストの可視化が重要？

→だが、その前に「保存活動が必要&重要なこと」と組織を説得する必要がある。

- 事例の共有：データ消失、管理できなくなった（資料との紐付が失われた）
 - 初動管理が不十分だと、後々に困る。
- コストの可視化：松竹梅のアイデア

- Digital Preservation Coalition

- [Business cases, benefits, costs, and impact - Digital Preservation Handbook](#)
- [The Costs of Inaction: advocating for digital preservation](#)

- 参考 : Amazon S3 Glacier Deep Archive

- 月額0.002USD/GB
- 10TBを1年間保存 => $0.002\text{USD} * 12\text{ヶ月} * 10,000\text{GB} = 240\text{USD} \Rightarrow 25,000\text{円}$
- ただし利用条件等について要確認

4. コミュニティ形成

コミュニティ形成

組織毎に予算や体制は違うが（潤沢な予算があるのは、ごく限られた組織）、共通課題についての、知識や知恵は共有できるはず。

→？協働できる部分はどこか　？どんなメリットがあるか

→どんな組織にも共通して、効果が期待できるのはどんなことか？

- 必要な人材、スキルとは？
- 三種の神器的なもの：これは必ず必要！といえるもの
- 「完全性」、可用性、機密性：ファイル同一性（ハッシュ値）の管理
- 識別子の管理方法のアイデア共有
- データ管理の単位

コミュニティ形成（学協会）

- 関連学会
 - [iPRES - Intl.Conf. on Digital Preservation](#)

その他

標準化：DPC Rapid Assessment Model

- 英・電子情報保存連合（DPC：Digital Preservation Coalition）
- デジタル保存の各要素に関する11のセクションからなり、各セクションについて自機関の状況に応じ5段階で評価を行う仕組み
- デジタル保存に携わる組織であれば、種類や規模、保存するコンテンツの種類を問わずに使用できるように、また、短時間で容易に実施できるように設計
- DPCでは、このツールを使用して毎年進捗を確認し、目標の達成状況を評価することを推奨