

【データ連携を支える次世代データ蓄積基盤】

2020年2月12日

産業競争力懇談会 **COCN**

【エグゼクティブサマリ】

1. 本プロジェクトの基本的な考え方

IoT/AI の浸透や、5G、6G 等の最新技術で実現が加速されるデータ駆動社会では、膨大なデータが生成・蓄積・活用されることが見込まれながらその対応が遅れている。一方でデータの蓄積・活用が進むことでデータセンターの消費電力が大幅に増加すると温暖化の一因にもなる。本プロジェクトでは、「階層型ストレージ」システムを有効利用した次世代のデータ蓄積基盤構築を推進し、消費電力の抑制を通じて SDGs（特に CO₂ 削減）に貢献し、データ連携の加速と拡大を図りながらデータ駆動社会を支えることで関連産業の競争力強化につなげる。

2. 検討の視点と範囲

(1) 現状・課題認識

データ保存は、①生成時の一時保存、②活用頻度の多い短期保存、③長期保存に分類できるが、昨今は利活用を前提としたデータの長期保存（アーカイブ）の重要性が高まっている。また AI 等データの活用にはデータ数が多いほど精度の高い結果が得られるため大量のデータが望まれ、さらにエッジ・デバイスの高性能化によりデータサイズもより大きくなる。このように大量で大容量なデータを扱うため、容量と性能のスケールアウト可能なオブジェクトストレージを採用したパブリッククラウドサービス（GAFA 等）の利用が進んでいるが、データアクセス方式がクラウドベンダーごとに異なっている。より広範囲なデータ活用を進めるにはマルチクラウド環境が前提となるが、その際、アプリケーション開発が非効率となりデータ利用拡大の妨げとなる。更に国際的には CO₂ の排出量削減が求められており、データ量の増大が見込まれる中での省エネルギー／CO₂ 削減という、相反した対応も求められる。このような認識の下、課題を以下の通り整理する。

（課題 1）データ蓄積基盤において、増え続ける膨大なデータを物理的にどのように保存し、如何に効率的に管理するか、という検討が遅れている。例えば、アクセス性能に優れるホットストレージ（HDD、SSD）は、データの増大に伴い、電力の消費やデータの保存・利用コストが大きな課題となる。一方、電力・コストでは優位なコールドストレージ（テープや光ディスク）は、アクセス性能などでは劣る。

（課題 2）分野によってデータに求められる特性（データ蓄積要件）が異なるため、データ特性の洗い出しとそれに対応した蓄積の仕組みが必要であるが、その検討が遅れている。

（課題 3）大手クラウドサービスの普及が進む一方で、異なるサービス間のデータ連携の仕組みの検討は進んでおらず、サイロ化（ベンダーロックイン）が進む。

加えて、異なる分野間でのデータ流通の仕組みも発展途上にあり、多様なデータの価値を活用するには、ユーザーが欲しいデータに適時・適切に検索・アクセスできる仕組みも重要である。

(2) 検討範囲

本プロジェクトでは、ホットストレージとコールドストレージを組み合わせた「階層型ストレージ」およびそのゲートウェイを含むデータ蓄積基盤を検討の対象とした（図 1）。

これまでの調査・検討の結果では、蓄積基盤に求められる要件は分野ごとに多種多様であること

があらためて認識された。当初の課題認識に加え、活動を通して得られた知見も含め、改めて課題と必要な対応や活動を下記のようにまとめる。

観点	課題	必要な対応や活動
データ蓄積	データ保全性の確保	確実なデータ保全方法の提示（ガイドライン化・標準化の整備）とその普及啓発
	データの長期保存性の実現	長期保存に適した階層型ストレージの要件や指針の提示、実現事例・ノウハウの普及活動
	コストやCO ₂ 排出量の削減	階層型ストレージによるコールドストレージの有効活用推進
データ連携	異なるシステム間連携やベンダーロックインの回避	データ蓄積基盤間をシームレスかつ効率的に連携させて活用する「統一的なデータ管理手法の仕組み」の提示と普及活動
実現化	分野ごとに異なる要件	各分野との継続的な協調・協業活動による要件の整理と検証

上記の活動を進めるにあたっては、具体的な課題を有する分野との協業が必要であり、統合イノベーション戦略の中で強化すべき主要分野、データ蓄積基盤に先行して取り組んでいる分野を優先する。第一弾としてはバイオ戦略 2019 の 4 重点分野の 1 つである『バイオとデジタルの融合を実現するデータ基盤』との協業を進める。主な課題としては、①データベースが散在し連携が取れていないこと、②ゲノム等のデータ量増大に対応した長期保全可能なデータ蓄積基盤が必要であることが指摘されており、バイオ戦略 2020 ではデータ蓄積基盤を含めたシステム全体の議論を進めることとなっている。

3. 産業競争力強化のための提言および施策

以下の 2 点を提言とする。

- ・ データ蓄積基盤側でのアクセスおよびデータ管理手法の統一を図ることで、異なる分野間のデータ連携のさらなる加速と拡大を目指すべきである。
- ・ データ駆動社会では、膨大なデータがあふれる社会になるため、その維持・運営には多くの電力が必要になる。このため、これらを支えるデータ連携・蓄積基盤は、SDGs（特に CO2 削減）に適合するよう低消費電力化を目指していくべきである。

ソフトウェア階層におけるデータ連携のみならず、それを支える「データ蓄積基盤」を含めたシステム全体での検討が必要で、今後も関係各組織が協調・協業して継続的に活動する必要がある。今後の活動の場としては、本プロジェクトの母体となった IDEMA JAPAN（日本 HDD 協会）を活用し、その発展形として「次世代データ蓄積・連携協議会」（仮称）を設立し（図 2）、大学や研究機関の協力も得ながら以下の検討を行う。

■グリーンイノベーションとして期待される階層型ストレージの標準化とその横展開

- － データの長期保存のためのシステム構築事例のカタログ化
- － データ管理を統一的に行うためのメタデータの標準化方式の提示
- － コールドストレージ利用促進による CO2 削減への貢献と関連産業の活性化
- － 将来的な海外展開・連携の模索

■「データ保全」「長期保存」を実現する指針の提示（ガイドラインや標準化）と普及活動

関係省庁には、「ガイドライン・標準の普及推進のサポート」、「データ連携を推進している分野との関係作り」、「国際連携」など引き続き様々なサポートをお願いする。

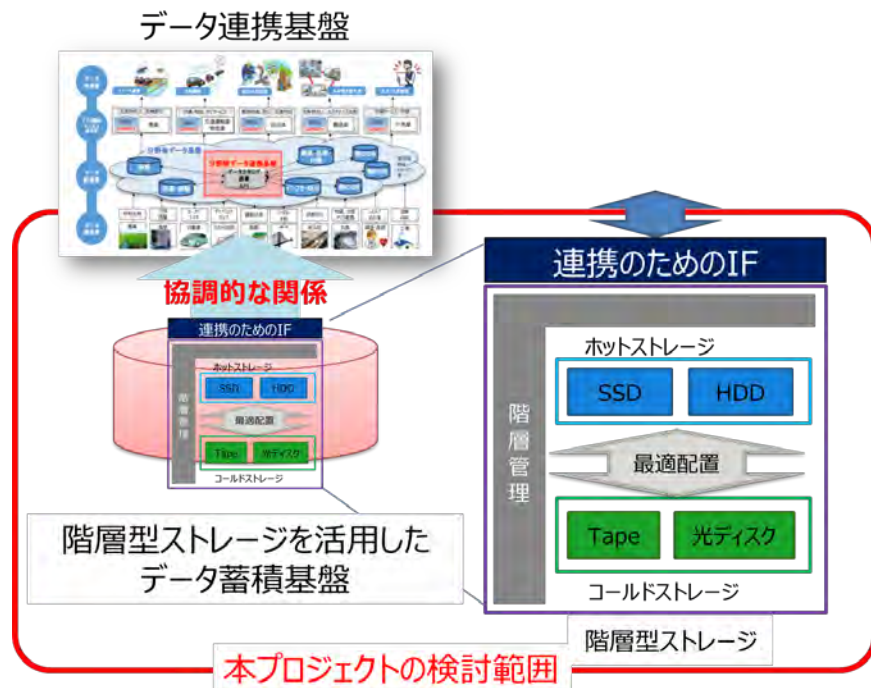


図1 本プロジェクトの検討範囲

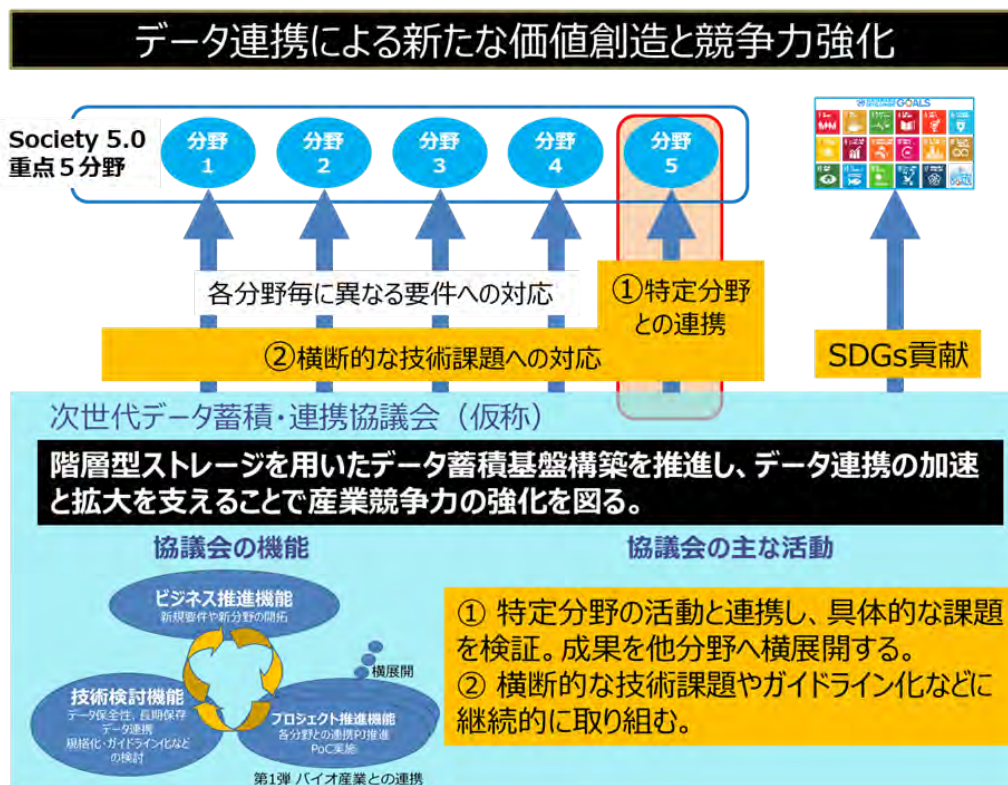


図2 協議体の機能と活動概要

【目 次】

はじめに	2
プロジェクトメンバー	3
1. プロジェクトの背景	5
2. 次世代データ蓄積基盤に求められる要件	6
2-1 次世代データ蓄積基盤に向けた課題の洗い出し	
2-1-1 現行の問題点（全般）	
2-1-2 各分野における課題抽出のためのインタビュー調査	
2-1-3 各分野における課題のまとめ	
2-2 次世代データ蓄積基盤に求められる要件	
2-2-1 ユーザー調査結果のまとめ	
2-2-2 CO ₂ 削減（省電力）について	
3. データ連携を実現するためのデータ管理手法	16
3-1 データ連携の課題	
3-2 クラウドサービスの普及と階層型ストレージの変化	
3-3 各種アーカイブストレージ群の連携 ～ メタデータの活用	
3-4 データ連携において必要となる要件	
3-5 コールドストレージ活用に向けての考え方	
3-6 ユースケース	
3-7 メタデータの具体例	
4. 今後の活動について	21
4-1 今後の活動の場について	
4-2 バイオ産業との連携活動	
4-3 今後の検討項目について	
4-3-1 「データ保全性」に関する要件	
4-3-2 「長期保存」に適した階層型ストレージについて	
5. まとめ	28

【参考資料 1】 データ蓄積基盤に係る技術動向

【参考資料 2】 次世代データ蓄積基盤に求められるシステム要件詳細一覧

【はじめに】

内閣府が第5期科学技術基本計画で掲げる Society5.0 では、サイバー空間とフィジカル空間が高度に融合するため、扱われるデータ量、データを生成する機器の数やそれぞれの通信速度、分野システム間の連携度合などはこれまでと次元を異にするものである。

このため、様々な分野において、国、学术界や産業界を巻き込んで Society5.0 実現に向けたシステムアーキテクチャの議論やデータ連携の活動が急ピッチで進められてきており、データ連携については、進捗がみられている。

しかし、複雑で、老朽化、ブラックボックス化した IT システムでは、デジタルトランスフォーメーションを阻害してしまうのではないかと危惧（2025 年の崖）が経済産業省の DX レポートの中でも指摘されている。システムとしての 2025 年の崖に加えて、データ蓄積基盤側でも課題を抱えている。すなわち、従来型階層型ストレージでもファイルベースで 1 つの組織内でデータを蓄積・利用することしか考えられておらず、また、長期間に渡るデータの保存に関しても統一的なルールやガイドラインが整備されてこなかった。

本プロジェクトでは、データの保全、長期保管に関する課題をストレージ基盤側で解決することを目指している。新しい階層型ストレージについての検討、実現可能性の検討、PoC の実施を行い最終的には機能の社会実装、機能の標準化を図っていく。これにより、データ駆動社会の実現に貢献していくものである。

【プロジェクトメンバー】

プロジェクトメンバー

■リーダー：

岩本 哲也（ソニーイメージングプロダクツ&ソリューションズ株式会社）

■ワーキンググループ主査・副査

標準データ蓄積基盤 WG 主査：脇 昌弘（ソニービジネスソリューション株式会社）

同 WG 副査：荒木 純隆（富士通株式会社）

データ蓄積エコシステム WG 主査：鶴見 利章（クラウドファン株式会社）

同 WG 副査：熊沢 忠志（富士通株式会社）

将来データ蓄積基盤技術 WG 主査：今井 拓司（技術ジャーナリスト）

■プロジェクト事務局

竹島 秀治 三菱ケミカル株式会社

金井 浩之 株式会社三菱ケミカルホールディングス

柴垣 奨 ソニーイメージングプロダクツ&ソリューションズ株式会社

河村 和彦 ソニーイメージングプロダクツ&ソリューションズ株式会社

■メンバー：

松本 信幸 NTT データ先端技術株式会社

中村 俊一 NTT データ先端技術株式会社

川村 直毅 NTT データ先端技術株式会社

角野 弘明 有限会社オプティカルエキスパート

中澤 道治 さくらインターネット株式会社

吉田 浩 大学共同利用機関法人 国立情報学研究所

山地 一禎 大学共同利用機関法人 国立情報学研究所

合田 憲人 大学共同利用機関法人 国立情報学研究所

安田 哲二 国立研究開発法人 産業技術総合研究所

来見田 淳也 国立研究開発法人 産業技術総合研究所

小笠原 泰弘 国立研究開発法人 産業技術総合研究所

田中 良夫 国立研究開発法人 産業技術総合研究所

妹尾 義樹 国立研究開発法人 産業技術総合研究所

時実 象一 東京大学大学院

柳 与志夫 東京大学大学院

沖野 剛史 株式会社東芝

安達 三郎 有限会社 日本 HDD 協会

青山 宏	パイオニア株式会社
谷口 昭史	パイオニア株式会社
立川 篤	富士フイルム株式会社
村田 悠人	富士フイルム株式会社
長谷川 文彦	富士フイルム株式会社
富井 大介	富士通株式会社
山下 智也	富士通株式会社
田村 雅寿	株式会社富士通研究所
野村 貴彦	株式会社ボウラインマネジメント
溝上 卓也	株式会社ボウラインマネジメント
千葉 善二郎	株式会社三菱ケミカルホールディングス
橋爪 宇生	三菱ケミカル株式会社
佐藤 嘉記	三菱ケミカル株式会社
大漣 宏明	三菱ケミカルメディア株式会社
柴田 高広	株式会社三菱総研
西郷 貴洋	株式会社三菱総研
鶴 薫	三菱電機株式会社
濱口 雄彦	株式会社 HGST ジャパン
竹内 晃一	沖電気工業株式会社
伊加田 恵志	沖電気工業株式会社

■COCN 実行委員：

江村 克己	日本電気株式会社
田中 幸二	株式会社日立製作所

■COCN 企画小委員：

大久保 進之介	富士通株式会社
武田 安司	株式会社日立製作所
田中 克二	株式会社三菱ケミカルホールディングス
菊地 達朗	株式会社日立製作所

■COCN 事務局長：

中塚 隆雄	
-------	--

■COCN 副事務局長：

五日市 敦	株式会社東芝
佐藤 桂樹	トヨタ自動車株式会社

【本 文】

1. プロジェクトの背景

昨今の IoT/AI の発展が牽引するデータ利活用は加速度的に増大しているが、日々増え続けるデータの保存・管理に関しては、その議論が国家単位で検討され始められた段階であり、諸外国に大きく遅れをとっている状況にある。本プロジェクトでは、データ駆動社会を支えるデータ蓄積基盤のあるべき姿を明らかにし、将来必要となる「階層型ストレージ」システムを利用した次世代のデータ蓄積基盤構築を推進し、データ流通の加速と拡大を通じて産業競争力強化を図る。

データ保存は、①生成時の一時保存、②活用頻度の多い短期保存、③長期保存に分類できるが、昨今は利活用を前提としたデータの長期保存（アーカイブ）の重要性が高まっている。また、AI 等データの活用にはデータ数が多いほど精度の高い結果が得られるため大量のデータが望まれ、さらにエッジ・デバイスの高性能化によりデータサイズもより大きくなる。このように大量で大容量なデータを扱うため、容量と性能のスケールアウト可能なオブジェクトストレージを採用したパブリッククラウドサービス（GAFA 等）の利用が進んでいる。更に、国際的には、CO₂ の排出量削減が求められており、データ量の増大が見込まれる中での省エネルギー／CO₂ 削減という、相反した対応も求められている。このような認識の下、課題を以下の通り整理する。

- （課題 1）データ蓄積基盤において、増え続ける膨大なデータを物理的にどのように保存し、如何に効率的に管理するか、という検討が遅れている。例えば、アクセス性能に優れるホットストレージ（HDD、SSD）は、データの増大に伴い、電力の消費やデータの保存・利用コストが大きな課題となる。一方、電力・コストでは優位なコールドストレージ（テープや光ディスク）は、アクセス性能などでは劣る。
- （課題 2）分野によってデータに求められる特性（データ蓄積要件）が異なるため、データ特性の洗い出しとそれに対応した蓄積の仕組みが必要であるが、その検討が遅れている。
- （課題 3）大手クラウドサービスの普及が進む一方で、異なるサービス間のデータ連携の仕組みの検討は進んでおらず、サイロ化（ベンダーロックイン）が進む。

加えて、異なる分野間でのデータ流通の仕組みも発展途上にあり、多様なデータの価値を活用するには、ユーザーが欲しいデータに適時・適切に検索・アクセスできる仕組みも重要である。

2. 次世代データ蓄積基盤に求められる要件

2-1 次世代データ蓄積基盤に向けた課題の洗い出し

本プロジェクトでは、まずは、代表的な機関、団体、企業などのユーザーへのインタビューに基づきデータ活用を前提としたデータ蓄積に関する課題抽出と、データ蓄積基盤に求められる要件の整理を実施した。

2-1-1 現行の問題点（全般）

近年、生成されるデータ量が益々膨大になっている中、確実なデータ保全だけでなく、個人情報などのセキュリティ対応や、より広範囲なデータ活用に当たってアクセス性の担保など、システムに求められる要件は高度化しているが、経済性の追及も重要な課題である。

一方で従来の大量データ保存システムの多くは、組織内での利用が前提となっているうえ、特定のデータ管理ソフトウェアの古い仕様に縛られ、十分な拡張性や俊敏性に乏しい。また、外部とのデータ連携に対する機能が十分ではなく、データ活用による新たなビジネス創出などには対応が困難である。

そのため、外部とのデータ連携を求めるケースではパブリッククラウドを利用することが増えているが、システムの稼働率、性能およびコストがユーザーのニーズと一致しないことも多い。また、大量データのパブリッククラウド間での移行に関しては、性能やコスト以外にもアクセス・インターフェースの違いなどによる障壁がありユーザーの不利益となっている。

これらの背景に加え、データの種類、ユーザーの属性により優先順位が違うことから、複数の記録メディアを活用する階層型ストレージシステムが必要であると考えられる。階層型ストレージシステムは従来から一部のユーザーでは利用されてきたが、運用性や導入の複雑さ、環境整備の遅れなどから普及が進んでいないため、これらの課題を解決し、大量のデータを各々のユーザー要件に合わせて効率的に管理するソリューションを用意する必要があると考えられる。

階層型ストレージシステムとは、異なる特性のメディア(*)を組み合わせることでアクセス性とコストの両立を図るシステムであり、一般には SSD/HDD といったホットストレージとテープ/光ディスクといったコールドストレージを組み合わせる。また、メディア間のデータ移動の機能を持つソフトウェアで構成される。特にデータ量の増加が大きく、アクセス頻度が低いデータについては、読出し時間が許容されれば、コールドストレージを活用することで、コスト削減と消費電力の削減が可能となる。

* 異なる特性のメディア

ホットストレージ

SSD : 高速データ処理に優れるため、ホットストレージとして適しているが、比較的高コストである。

HDD : 高速データ処理が可能でホットストレージとして普及しているが、耐衝撃性、消費電力に課題がある。

コールドストレージ

Tape : 大容量、低コストかつ低消費電力であるため、一般的なコールドストレージとして普及している。また、データ圧縮技術や暗号化技術も標準搭載している。

Optical : 長期保存特性、災害耐性、改ざん防止に優れ、かつ低消費電力であることから、特にデータ保全が必要なコールドストレージとして適している。

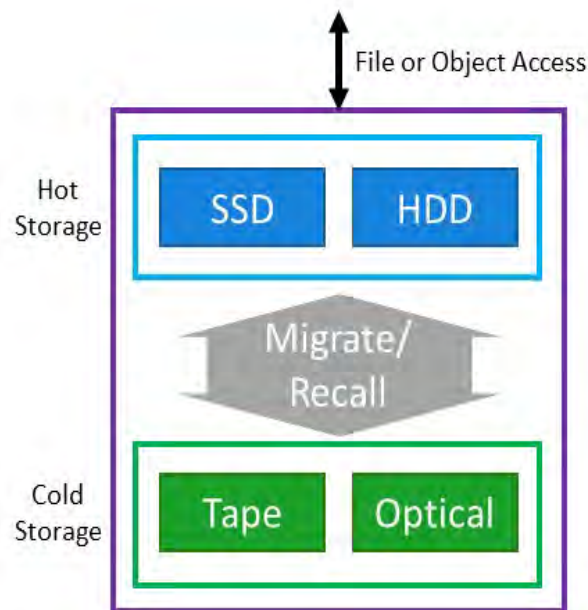


図 2-1 階層型ストレージシステムの概念図

2-1-2 各分野における課題抽出のためのインタビュー調査

- ✓ 調査内容：データ種別、データ量／増加量、保存要件など
- ✓ 調査対象の団体・企業：データ種ごとの主なユーザー 51 団体
- ✓ 結果概要

ユーザーカテゴリごとの要件の中で優先度の高いものを以下にまとめた。

ユーザー カテゴリ・データ種	保存 年数	読み出し 許容時間	優先順位		
			1 位	2 位	3 位
図書館/博物館/ 公文書館など 文化財のデジタル コンテンツ	永年	時間確保可の コンテンツ あり	データ保全性	保存環境の 継続性	コスト
建設/土木・図面 など	永年	時間確保可	データ保全性	コスト	性能
企業の会計情報 など	法に基 づく	時間確保可	データ保全性	コスト	
医療・手術映像	20 年	時間確保可	データ保全性	保存環境の 継続性	秘匿性
研究データ	10 年	時間確保可	データ保全性	保存環境の 継続性	コスト
製薬	法に基 づく	時間確保可	データ保全性		
地球・宇宙・自然現 象観測データ	永年	時間確保可	データ保全性	保存環境の 継続性	性能
放送・報道映像素材	永年	時間確保可	保存環境の 継続性	データ保全性	性能
国防、国家機密	永年	時間確保可	秘匿性	データ保全性	
ゲノム・生体データ	永年	時間確保可	秘匿性	データ保全性	改ざん防止
コネクテッドカー (自動運転)	20 年	時間確保可	コスト	省電力	データ保全性

上記のユーザーニーズに対し具体的な解決方法が標準化されていないことが、複数のユーザーへのインタビュー調査で判明した。

ユーザーニーズ	課題と解決手段に関する考察
データ保全性	確実なデータ保全に当たっては、“3 Copies, 2 Different Media Types, 1 Copy Offsite”（3つのコピーを作成、2つの異なるメディアに保存、1つはオフサイトに保存）と言われるように災害、機器障害、メディアエラーなどに関しリスク回避が必要となる。また、災害への備えとして国内においては大規模地震を想定し、データセンター間の距離(100km 以上が推奨)だけでなく、異なる電力会社のエリアにデータセンターを分散配置するなども考慮に入れる必要がある。また、近年増加している外部からの侵入・改ざん・消去などへの対応のためにもデータ復旧が必要となる。 しかし、これらのデータ保全に関する基本的概念も JIS として標準化されているものもあるが、全ての記録メディアが記載されているわけではない。その他のガイドラインなども多数存在するが十分に浸透しているとは言えず、実際にはリスクを含んでいるシステムも見受けられる。この基本的な概念については、もっと周知する活動が必要と考える。 階層制御ストレージでの具体的なデータ保全としては、①クラスタ内でのミラーリング(三重以上が推奨)やイレジャーコーディング(複数のパリティ)により冗長データを持つ。②同一サイト内でも異なるクラスタなど分離された物理的に異なる種類のメディアに置く。③大規模災害に対応するためオフサイトへのコピーをとる。
保存環境 の 継続性	長期保存の観点では、データ保存環境およびデータ再現環境について計画的な対応が必要となるが、これらがユーザーの負担増加につながっている。 ① 機器および記録メディアのマイグレーション(物理的的老朽化対応) ✓ 特にコールドストレージを長期に棚保管している場合など、再生可能なドライブが入手できないというユーザー事例もあり注意が必要。場

	合によってはメディアコンバートサービスなどの利用も検討可能。 ② データフォーマットやビューアーの継続性(論理的老朽化対応) ✓ 建設図面などでは、TIFF 形式での電子的な保存もあるが、数十年に渡る見読性の観点から、拡大することで目視できるマイクロフィルムも多用されている。 ✓ 映像データも作成された時点に普及しているデータフォーマットを使用するが、データ圧縮をしないベーシックなフォーマットでの記録をしているユーザーもある。いずれにしても数十年後の再生に関してはフォーマットの変換や対応するビューアーの継続性が課題となっている。
性能	性能に関しては以下の3つの観点があるが、ユーザーにとっては具体的な実装方法が分かりにくいという課題がある。特にコールドストレージに関する知識が少ないため、実際にはコールドストレージを活用できるケースでも、ホットストレージの単一構成をとっているケースも多く見受けられる。 ① 総書き込み／読出しスループット性能 ✓ データ蓄積基盤として記録要求されるすべてのデータを取りこぼしなく格納することが必要 ✓ 階層型ストレージではホットストレージ層はスケールアウト構成をとりやすく性能担保は比較的容易であるが、コールドストレージ層はメディアプールとドライブの割当てについて十分に考慮する必要がある ② 単一ストリーミングでの書き込み／読出しスループット性能 ✓ 大容量データの高速入出力に関しては、ホットストレージ層でデータ転送時間としてはカバーできるが、コールドストレージのアクセスに関して時間がかかることによって、ホットストレージ層の容量を圧迫することがある ✓ コールドストレージ層の高速化はドライブおよび媒体の並列転送などの機能を持つ必要がある ③ 読出しリクエストからデータ転送開始までの時間 ✓ コールドストレージに関しては、ライブラリ装置のロボットによるメディアの搬送、ドライブでのローディングなどの時間が必要となる。そのため、インターネット経由のアクセスではアプリケーション側のタイムアウトが発生しない仕組みが必要。
コスト	コスト最適化は多くのケースで言われており下記のような観点でコスト最適化が可能であるが、階層制御の複雑さからコールドストレージを活用できていないケースがある。 ① コールドストレージの活用

	✓ アクセス頻度が低いデータを容量単価の低いメディアに配置することでコストを削減 ● 生成されるデータ量に対し、メディアの記憶容量の増加が追いついていないという課題がある ✓ コールドストレージを使うことでの消費電力を削減し運用コストを最適化する ② ソフトウェアライセンス ✓ 既に大規模な階層型ストレージを使っているユーザーからは、制御ソフトウェアのライセンス料・サポート料が高額で規模の拡大がしにくいという課題が挙げられている。また、階層制御ソフトウェアの提供ベンダーの寡占や事業継続性に対する不安もある ③ データリダクション ✓ データ圧縮や重複排除など、物理的に記録されるデータ量を削減する。ただし、メディア単体で再現性が担保されることが前提。
秘匿性	データ活用が大きなテーマとなっているが、一方ではオープンに出来ないデータも多く存在している。 ✓ 国防などの機密情報 ✓ ゲノム、医療情報ほかの個人情報となるもの ✓ 企業の競争領域となる情報 メディアの外部保管や廃棄(消去)の観点では、暗号化機能も検討する必要がある。暗号化については、OSなどに依存しないハードウェアレベルでの暗号化機能とアクセスキーの統合管理機能により容易な管理を提供する必要がある。 一方で、企業の内部情報に関しては仕分けができないため、社外提供されていないが、適切に仕分けされれば社外利用可能なものもあり「知」の共有の阻害要因となっている。これらについて適切なガイドラインなどを設けることが望ましい。
省電力	一般には電力などの費用は、経費項目や管理担当者が異なり十分に意識されないケースが多い。しかし、一部の大規模ユーザーでは電力の総供給能力の問題などもあり、大きな課題として認識されている。 一方、国レベルではCO₂の排出量の削減に取り組むことは国際的に求められる課題であり、データ量の増大が見込まれる中で省エネルギー／CO₂削減に対応するソリューションが求められる。 データ蓄積基盤としては、コールドストレージの活用により電力消費量を大幅に削減することが可能であり、今後のデータ量の増大やエネルギー消費効率の向上の観点で貢献可能と考える。

2-1-3 各分野における課題のまとめ

今回のユーザー調査及び、各参加企業がこれまで行ってきた調査から各分野におけるユーザーの課題を下記の通りまとめた。

分類 ユーザ	分類 データ	データ管理の特徴と課題	ニーズ
大学・研究機関	観測データ	再現不可能な観測データなどについては確実な永年保存が求められる。機関によっては場所を含めデータを多重で持つことでデータ保全を図っているが、コスト負担も大きい。既存で数十 PB 程度のデータを保有している機関も多くあり、マイグレーションの負担を感じている。 また、階層型ストレージ管理ソフトウェアのライセンス料や保守料もコスト課題として挙げられている。	永年保存 コスト削減 容易なマイグレーション (機器、データ)
		研究機関によっては、コストとアクセス頻度から可搬媒体へ記録し棚保管をしている部門も少なくない。不適切な媒体管理に伴うデータ消失のリスクがある。	消去防止、媒体管理 * 将来的にはオンライン アクセスが可能なデータ センターへの集約が必要
	エビデンス情報 論文の	各機関では論文および、その裏付けとなるデータなどの 10～15 年保存を進めている。	改ざん防止・改ざん検知
	研究資料	特に人文系の研究室では、収集された貴重な資料がデジタル化もされず、棚保管されていることが少なくない。	デジタル化、データ保全 データ共有・検索機能 * IT リテラシーの高くない研究者にデータ化と共有のための仕組みの提供が必要
		研究員の退職の際に資料やデータの管理が引き継がれず消失するものも多くある。	
研究機関・創薬	ゲノムデータ	個人を特定可能な情報であり、秘匿化や厳格なデータ管理が必要。	個人情報の秘匿化 アクセスコントロール
		データの保存期間は永年であること、次世代シーケンサによるデータ量の増大への対応が必要。	コスト削減 拡張性

	バイオ	従来、データは各社がクローズド戦略で進めてきたが、今後はバイタルデータなど異業種のデータを含めた連携が成長戦略の鍵となる。	データ連携 個人情報の取り扱い
図書館・博物館・美術館・公文書館など	書籍のスキャン	国立国会図書館、公文書館や自治体の図書館での書籍のスキャンデータ、電子文書の保存においては保存期間が永年または長期にわたること、継続的にデータが増え続けることへの対応が求められている。また、国立国会図書館には、その他の図書館や海外の図書館への送信が求められる。送信要求に対しては、利用者が端末から行うため、ある程度の即時性を求められる。	永年保存 拡張性(データ量増大への対応) ある程度の即時読出し
	画像	地方にある古文書や古地図などのスキャンデータの保存と閲覧による、原本の劣化を防止している。課題としては画像データのフォーマットに関する悩みが多い。また、自治体の予算が限られるなどの問題もある。	永年保存の観点でのデータフォーマットの選択や変換 コスト削減
	映像	地方の無形文化財(祭り、踊り)などを映像化し、観光コンテンツ化することや、後世に残すなどの事例が増えている。高精細映像化によるデータ量の増加や、自治体の予算が限られるなどの問題がある。	永年保存 コスト削減 大容量化への対応
	3D スキャン	有形文化財(仏像、建物など)を3Dスキャンする事業が進められているが、高精細化によるデータ量の増大の課題がある。	永年保存 コスト削減 大容量化への対応

建設・土木	図面	構造物がある限り保存対象となる。数十年に渡るためデジタルデータの場合、可視化するソフトウェアおよびデータフォーマットの継続性が大きな課題となる。そのため、アーカイブ方式として拡大すれば目視できるマイクロフィルムも利用されている。また、紙の図面として倉庫に保管されているものの多く存在する。実際のデジタル化率は1%程度であり、デジタル化のコストおよびデータ量の増大が課題となる。 災害対策のため2か所以上の場所に保管が必要。 読出しはコンプライアンス要件などのため、頻度は非常に低い、ある程度の即時性が必要。データは公開・共有されることはない。	長期保存の観点でデータフォーマット(アプリケーション)の継続性 確実な原本保証 コスト 性能
上場企業	会計情報	会社法、金融商品取引法に基づく監査などに対応するため必要な情報を法令の定める期間保存する義務がある。 データ量は比較的少ない。	監査の観点での真正性 コスト
警備会社など	監視画像	現在は警察、消防などとの連携のみであり、長期保存要件はない。 ただし、別の業界からは人の行動分析などのためにデータを欲しいという話が出ており今後の活用が期待される。	個人情報への配慮

2-2 次世代データ蓄積基盤に求められる要件

2-2-1 ユーザー調査結果のまとめ

データ蓄積基盤としての要件を、ユーザー調査結果をもとに優先順位付けした。

① データ保全

- ✓ ユーザーが簡単な保護レベルを指定するだけで適切なデータ保全環境が提供される
- ✓ これまで以上に長期保存特性、災害耐性、改ざん防止に優れた記録メディアも期待されている

② コスト（省電力）

- ✓ 消費電力および容量単価の低いメディアの活用(コールドストレージ)
- ✓ コールドストレージをホットストレージのように使わせる階層制御機能

- ✓ データマイグレーションコストの削減手法／マイグレーション周期の長期化
- ③ 保存環境の継続性
 - ✓ 永年保存を前提としたハードウェア・記録メディアのマイグレーション機能
 - ✓ 記録フォーマットや再現するビューアーの提供またはフォーマット変換の対応
- ④ 秘匿性
 - ✓ 情報漏えい対策も機器やメディアのライフサイクルに渡った考慮が必要
 - ✓ 外部とのデータ交換時も暗号化の対応
 - ✓ 新たな視点として、データの中に意図せず混入している個人情報保護する機能
- ⑤ 性能
 - ✓ オブジェクトストレージの機能としてデータ管理メタに性能に関する定義
 - ✓ 利用頻度に応じた適切な階層管理機能
- ⑥ 改ざん防止
 - ✓ 更新されないまたは更新してはならないデータの真正性の担保
 - ✓ 改版履歴保存やデータリネージュへの対応

また、ユーザー調査では出てきていない下記機能も必要と考えられる。

- ⑦ 容易なスケールアウト構成
 - ✓ データ量の増大に対応し容易に増設できる仕組み
- ⑧ アクセスログなどの監査機能
- ⑨ CO₂削減（省電力）

上記を踏まえ、大容量データを効率的かつ省エネルギーな形で蓄積・連携させるためには、ホットストレージとコールドストレージを組み合わせた階層型のオブジェクトストレージが必要と考える。従来の階層型ストレージはファイルアクセスでありネットワークを経由した遠隔地からのアクセスには大きな課題があるため、新規にオブジェクト形式の階層型ストレージシステムの提供が必要である。⑨については、2-2-2 にて補足説明をする。

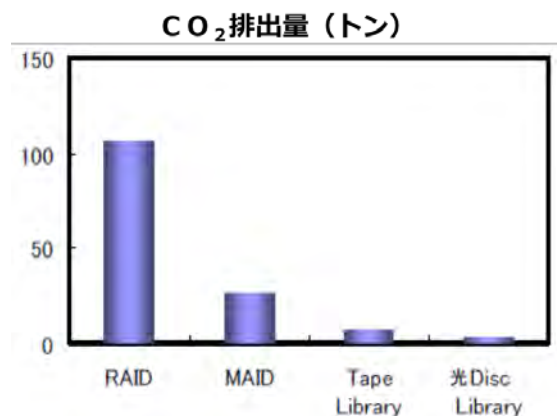
詳細な要件については、『【参考資料2】次世代データ蓄積基盤に求められるシステム要件詳細一覧』にて提示する。

2-2-2 CO₂削減（省電力）について

ユーザー調査から得られたデータ蓄積基盤に求められるシステム要件の他に、昨今の世界的なCO₂削減要求への対応も重要である。JST 低炭素社会戦略センターによる「情報化社会の進展がエネルギー消費に与える影響（Vol.1）－ IT 機器の消費電力の現状と将来予測－」

(<https://www.jst.go.jp/lcs/pdf/fy2018-pp-15.pdf>)によると、「現在の技術でまったく省エネルギー対策がなされないと仮定すると、2030年には年間使用電力量の倍近い電力をIT関連機器だけで消費する予測となる。（中略）また、2050年の電力消費量は、現在と比較して日本、世界

ともに約 200 倍という極端な数字となり、情報関連だけで全てのエネルギーを消費してもまだ不足するという状況になりうる。この情報量の爆発に対しての対策が必要なことは明らかである。」という報告がなされている。消費電力や CO₂ 削減要求へ応えるためには、図 2-2、図 2-3 に示すように消費電力、CO₂ 排出量が少ないテープや光ディスクといったコールドストレージの活用が有効である。保存されるデータには、あまりアクセスはしないが長期保管が必要なデータも多く存在する。現在は、このようなデータもホットストレージ（SSD・HDD）に保存されているが、コールドストレージへ保存することにより消費電力、CO₂ 排出量の削減が可能となる。すなわち、コールドストレージの活用はエネルギー・環境分野におけるイノベーションであり、グリーンイノベーションといえる。



RAID (Redundant Array of Independent Disks) :

複数のディスクに冗長性を付加してデータを分散格納することにより、障害時にデータ再生を可能とする技術

MAID (Massive Array of Idle Disks) :

アクセス待機時にはストレージ内の HDD を停止させ、アクセスが必要になったときにのみ回転させる技術

図 2-2 各記録媒体アーカイブ時の二酸化炭素排出量

(https://www.jiima.or.jp/wp-content/uploads/policy/Opticaldisk_active_guideline201310.pdf)

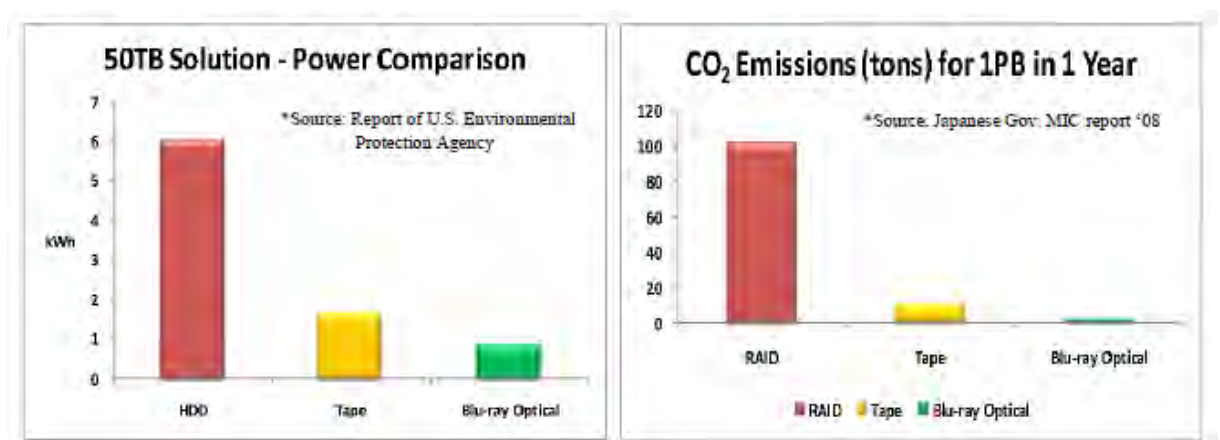


図 2-3 各記録媒体の環境特性比較

(Hitachi-LG Data Storage, Inc. "Introduction of Optical Archive System for Long-term Data Retention")

3. データ連携を実現するためのデータ管理手法

3-1 データ連携の課題

本章では、蓄積されるデータが大量、大容量化していくにつれ、システム運用上の大きな負荷となる「データ管理」の課題について検討した。

本文において「データ管理」とは、以下のようなストレージデータ管理を意味している。

- ✓ データの価値やアクセス統制に応じて格納するストレージデバイスを選択して、データを移動するデータの最適配置管理
- ✓ ストレージデバイスやシステムの故障時にもデータが必要に応じて復元できるバックアップやリカバリのようなデータの保護管理
- ✓ 各データへのアクセスを定められた権利・規約に従って許可するなどのデータのセキュリティ管理など

またデータ管理の運用規約を「データ管理ポリシー」と呼ぶ。例えば、階層型ストレージの最適配置の管理のための運用規約は「階層管理ポリシー」として、システムやデータごとにデータの取り扱い方法が決められる。例えば、アクセス時間が常に短いことが要求されるデータの階層管理ポリシーは「ホットストレージに保存すること」となり、長いアクセス時間許されるデータの階層管理ポリシーは「アクセス頻度が少ない場合は、コールドストレージに移動配置する」などとなる。

システム運用においては、バックアップの数を指定する「バックアップ管理ポリシー」や、組織や個人の権限に応じてデータアクセスの許可／禁止を規定する「アクセス許可ポリシー」など、多くのポリシーが存在する。データが少量である場合には、文書化されたデータ管理ポリシーに従い人手作業で運用可能であるが、データ数が多い場合は、プログラム・コード化したデータ管理ポリシーをデータ管理サーバに入力設定し、自動運用するのが一般的である。

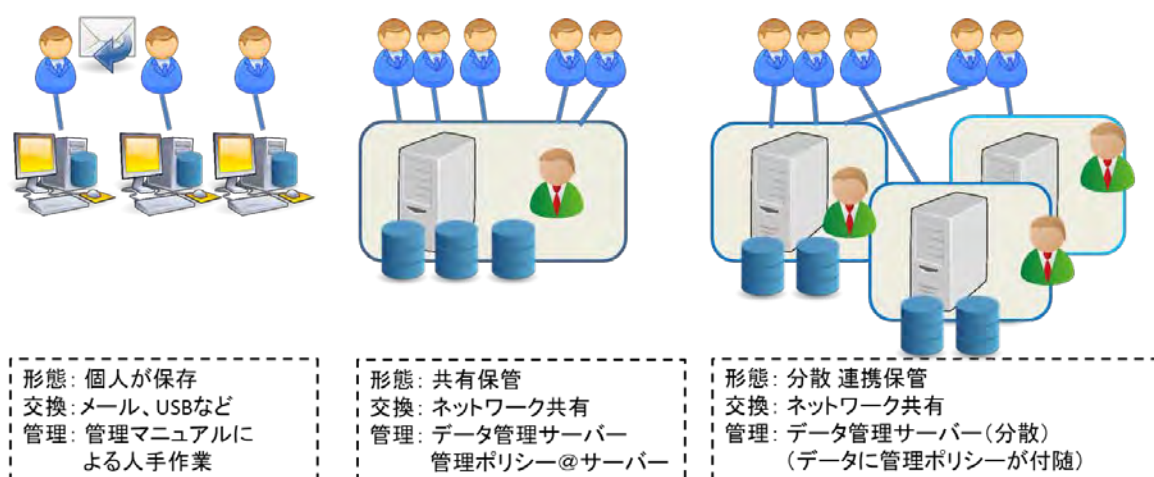


図 3-1 データ保管形態とデータ管理手法

しかし、これらのデータ管理ポリシーの記述や取り扱いの仕様は、データ管理サーバの機種、ベンダーごとに異なっており、システム構成に合わせた個別仕様も多数存在している。大規模シ

システムでは異なるベンダーの異なる装置機種を組み合わせることで構成せざるを得ないことも起きえる。

このような分散環境でデータ管理を統一的行うための仕組みとして、個々のデータの「管理ポリシー」をデータに付随して保存する方法がある。データ管理サーバはデータへのアクセス時あるいは定期的に管理ポリシーを読み取り、そのポリシーに従った管理を実現することが出来る。実際、大規模なクラウドシステムなど、大量のデータを蓄積しているシステムでは、この方法でデータ管理を自動的・効率的に運用している。次節以降に、大規模なクラウドサービスの事例をもとに、データ管理運用で活用されている技術、手法について説明していく。

3-2 クラウドサービスの普及と階層型ストレージの変化

クラウドサービスの一部として、ストレージサービスが各種存在している。その中でもデータ増加に柔軟に対応できる RESTful API（HTTP ベース）で動作するクラウドストレージ（オブジェクトストレージ）が膨大なデータを保持するようになってきた。インターネットを介して利用するクラウドストレージは、低コストだがアクセス時のレイテンシーが大きい傾向にある（コールドストレージと近い側面を持つ）ため、アーカイブとして利用されるケースも多い。さらにインターネットの特性、すなわちどこからでもアクセスできるという点によりコールドストレージの課題の一つを解決している。また、RESTful API を利用したクラウドストレージへデータをアーカイブできるアプリケーション・ソフトウェアが増えている。このため RESTful API に対応したコールドストレージ装置が普及しつつある。

以上の状況から、コールドストレージは、これまで特定の階層型ストレージ管理の元でのみ利用されてきたが、今後はより広範な活用が期待されている。

3-3 各種アーカイブストレージ群の連携 ～ メタデータの活用

RESTful API を利用する観点で考えると、ネットワークの接続性次第で各種コールドストレージやクラウドストレージ（オブジェクトストレージ）は全て、階層アーカイブストレージシステムの対象として連携し易くなる。パブリッククラウドサービスとプライベートクラウドの連携や、各サイト間を跨がるストレージにおいても容易に相互アクセスが可能になる。

しかし、アクセス性が向上するもののコールドストレージでは、読み書きの性能要件がホットストレージとは異なっており、設置場所や構成に依存してコールドストレージ特有の留意すべき要件が想定される。

以上のようにコールドストレージ特有の要件を把握した上でアーカイブストレージシステムが効率良く動作できるよう、共通の管理要件を把握する必要がある。

データ管理ポリシーを保持する場所として、RESTful API で利用可能なメタデータ領域を活用する事が考えられる。特にユーザーメタデータ領域は任意の使い方が許されているため、メタデータ記述に一定の規約を設けることで、異なるアーカイブストレージシステム間でも、特定のデータ管理運用を指示・実行することが可能になる。

3-4 データ連携において必要となる要件

本プロジェクトでは、考慮すべき要件の抽出を優先し、具体的なメタデータの取り決めは実装段階で確定していくこととした。以下に、①オブジェクト（ファイル）の観点と②アーカイブストレージの観点で要件を記載する。前者の要求に後者が応えられなければ、効率を維持できない事を鑑み、特にコールドストレージを構成するアーカイブストレージの要件も記載する。

① オブジェクト（ファイル）の観点

- ✓ オブジェクト（ファイル）名に応じて階層アーカイブストレージを移動
 - 前方一致
 - 中間一致
 - 後方一致（拡張子）
- ✓ サイズに応じて階層アーカイブストレージを移動
- ✓ 時間軸に応じて階層アーカイブストレージを移動
 - 作成日時（作成日を起算とした階層化）
 - 最終アクセス日時（アクセス頻度による階層化）
 - 任意の日時（イベント等終了や期限が明確な階層化）
 - ◇ コールドストレージに持っていきだけでなく、日時が近づいたら事前にコールドストレージから読み出しておくために利用する方法もあり。
- ✓ 参照回数に応じて階層アーカイブストレージを移動
- ✓ 特定階層のアーカイブストレージを移動
 - アーカイブストレージタイプ（オフライン可否や手動/オフライン可否等まで）
 - 読み出し速度（これ以上遅いストレージには移動しない）
- ✓ 特定地理のアーカイブストレージを移動
 - GPS と距離
 - 国、地域
 - サイト名
- ✓ オブジェクト（ファイル）種別による階層アーカイブストレージを移動
 - カテゴリー
 - タグ
- ✓ アーカイブではなくバックアップ扱い（アーカイブストレージのバックアップをオフラインメディアへ取る場合）

コンテナ/バケットと呼ばれるオブジェクトの入れ物の単位でも、上述の内容をまとめて設定できると利便性が増す。

② アーカイブストレージの観点（主にコールドストレージで構成するアーカイブストレージ）

- ✓ ストレージ特性を開示し、前段の階層からの移動の判断基準とする
 - RESTful API 種別（S3, GOOGLE, AZURE, CDMI, SWIFT 等）

- アーカイブストレージ種別（SSD, HDD, OPTICAL, TAPE 等）
- 基本性能（読み書き平均速度）
- 性能統計（高負荷時間帯、低負荷時間帯）
- 状況（通常状態や輻輳状態）
- キャッシュ有無
- 受け入れ可能容量（現在の空き容量）

3-5 コールドストレージ活用に向けての考え方

コールドストレージはアクセスが可能になるまでの時間を短くすることでシステムの経済性と利便性の両方を満足する答えになりうる。そのためにはリクエストに対する①ジョブ管理、②事前読み出し、③キャッシュ領域といった仕組みが必要となる。

ジョブ管理に求められる基本機能は待ち行列上の I/O リクエストの管理であるが、オフラインメディアが不得手とする小さいサイズのファイル群をまとめて書き込む、ファイルの読み出しを効率化できるよう待ち行列上のリクエストを並べ替えるなどの最適化案が考えられる。また、キャッシュ領域に関して容量の拡張が必要な場合、オブジェクトストレージ製品の一部をキャッシュ領域として割り当てるような機能も考えられる。これらの動作を連携させるためにメタデータを活用することも実装段階では検討対象とすべきである。

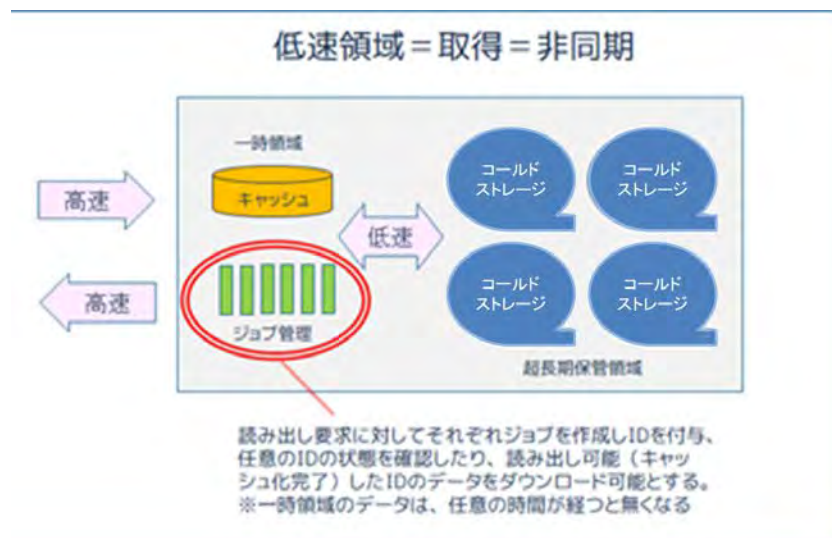


図 3-2 コールドストレージへのアクセス応答短縮のための仕組み

3-6 ユースケース

階層アーカイブストレージの利点を享受できるユースケースには以下のような分野、システムが考えられる。

- ① 研究目的、特に機械学習の分野で、年次や月次といった、ある程度の期間ごとにデータを読み出して解析する事例
- ② WORM（Write Once Read Many）で物理的な削除を防ぐような事例（オブジェクトストレージにも WORM 機能はあるが、ここでは物理的に変更不可な光ディスクを想定）

- ③ 任意の複数拠点へデータをアーカイブしておくといった地域冗長（DR）を踏まえた事例
- ④ 参照頻度の低いデータの保持のために、高価なストレージ装置のみでシステム構成する事例

3-7 メタデータ的具体例

メタデータ的具体例として、クラウドストレージのデータアクセスと管理インターフェースの標準化を目的とした CDMI（Cloud Data Management Interface）仕様の一部を記載する。その他の RESTful API を利用するアーカイブストレージ群と連携する際は、CDMI で提示されたメタデータを基準に対応表を作成していく方法を検討するべきである。

メタデータ名	説明
cdmi_size	オブジェクトの保存されたバイト数
cdmi_ctime	オブジェクトの作成時刻
cdmi_atime	オブジェクトが最後にアクセスされた時刻
cdmi_mtime	オブジェクトが最後に変更された時刻
cdmi_acount	オブジェクトが作成されて以降にアクセスされた回数。アクセスには、読み取り、書き込み、および表示が含まれる。
cdmi_mcount	オブジェクトが作成されて以降に変更された回数。変更には、値とメタデータの変更が含まれる。読み取りの結果メタデータが変更された場合（atime の更新など）は変更としてカウントされない。
cdmi_data_dispersion	データの複数のコピーをサポートするインフラストラクチャ間の必要距離（km）
cdmi_geographic_placement	当該オブジェクトの格納が許される（あるいは許されない）地理的領域（ISO3166 の国コードで表現）のリスト
cdmi_retention_id	当該オブジェクトに適用される特定の保有ポリシー
cdmi_retention_period	当該オブジェクトの保有期間（指定された場合に保有対象となる）
cdmi_retention_autodelete	保有期間満了後の自動削除の指定
cdmi_latency	最初のバイトの到達までの最大時間の希望値（ミリ秒）
cdmi_throughput	データ読み出し時の最大データ転送量の希望値（バイト/秒）
cdmi_santization_method	当該オブジェクトを更新あるいは削除後に完全に不可能とするための方法の指定：セキュアワイプ、DoD 方式など
cdmi_RPO	当該オブジェクトの最大データ復旧時点（復旧時刻からどこまで遡ったデータが保証されるか）の条件（秒）
cdmi_RTO	当該オブジェクトの最大データ復旧時間の条件（秒）

4. 今後の活動について

4-1 今後の活動の場について

データ連携を加速するために、ソフトウェア階層におけるデータ連携だけでなく、それを支えるデータ蓄積基盤を含めた全体の検討加速が必要である。来期以降の活動の場としては、2018年度から本プロジェクトの準備会として機能してきた IDEMA JAPAN(日本 HDD 協会)を活用する。関連する産業界、アカデミアから検討に必要なメンバーを加え、発展形として「次世代データ蓄積・連携協議会」(仮称)を IDEMA 内に設立し活動を行っていく。特に、今後はストレージシステムやデータ管理の検討が必要となるため、この分野の専門家が多く参加している SNIA-Japan との連携や、DAPCON、デジタルアーカイブ学会、JIIMA、情報処理学会、光協会等の関連機関とも連携を取りながら活動を行っていく。

協議会では、主に①プロジェクト推進機能、②技術検討機能、③ビジネス推進機能の3つの機能を持たせる。具体的には下記の活動を行っていく。

① プロジェクト推進機能：バイオプロジェクトの具体化検討

現在、連携を検討しているバイオプロジェクトにおいて、「データ基盤の全体設計」や「健康医療関連データ基盤」の構築を目指すとともに要件を抽出する。

② 技術検討機能：ガイドライン・標準化(JIS/ISO 等)等の検討と普及啓発活動

上述のプロジェクト具体化検討で抽出されるデータの保全、長期保存やデータ連携などを実現するために必要な仕組みについて検討し、ガイドラインの作成や JIS/ISO 等の標準化を検討する。また、それらを普及させるため、啓発活動を実施する。その中で、ユーザーの検討の負担を減らすため、保存期間や更新の有無、アクセススピードなどデータ属性に合わせた具体的な解決方法を示していく。

③ ビジネス推進機能：新規要件や新分野の開拓

バイオ以外の分野との連携を目指し、バイオプロジェクトで得られたベストプラクティスを活用して、新たに「データ連携を支える次世代データ蓄積基盤」検討の実践の場を探索する。府省別懇談会でも増え続けるデータへの対応の必要性が挙げられており、次世代データ蓄積基盤に対するニーズは存在すると考えられる。

新規の実践の場の候補として下記の分野が考えられる。

(a) バイオ戦略 2019 の他の市場領域

特に、大規模コホート・バイオバンク構築などの大容量データの長期保全が必要な領域などでは本プロジェクトで得られる知見が有効活用可能と考えられる。

(b) 統合イノベーション戦略 2019 において最先端(重要)分野として強化すべき応用分野となっている衛星データ/海洋データ分野

衛星データ/海洋データは再現が不可能な観測データであり、永年保管が前提となっていることから、今後、「データ連携を支える次世代データ蓄積基盤」が必要と考えられる。

(c) 統合イノベーション戦略に基づいた研究データ蓄積分野

統合イノベーション戦略では、NIIを中心とした研究データの蓄積が推進され始めている。今後、研究データとその他の産業分野とのデータ連携、及び、研究データの利活用促進が必要となってくると予想される。NII、AXIES等と連携して、新たなデータ連携、データ利活用促進の仕組み作りが必要と考えられる。

上記に加え、議員立法で準備が進められている「デジタルアーカイブ整備推進法案（仮）」が、この数年の内に可決される可能性がある。本法案では、各府省庁にデジタルアーカイブ計画を立案し、定期的に見直すことが求められており、関連団体と協調しながら貢献していくことも視野に入れる。

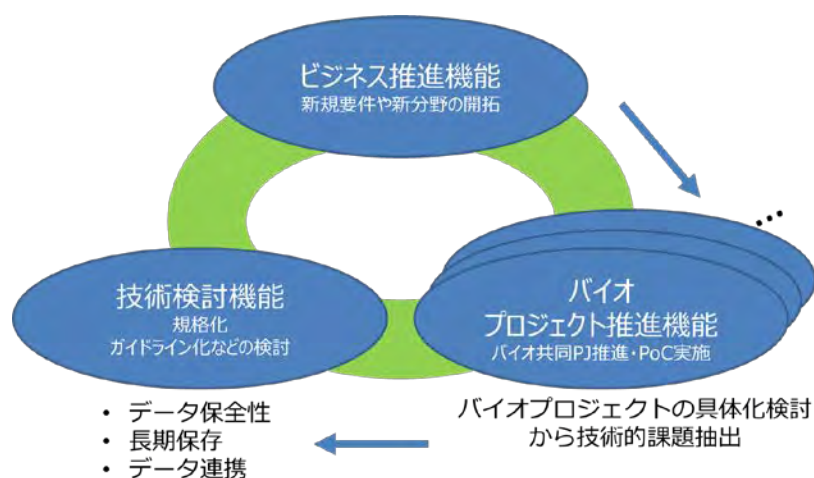


図 4-1 協議会の活動

4-2 バイオ産業との連携活動

「バイオ戦略有識者提言」をもとに総合科学技術・イノベーション会議（CSTI）と省庁が作成した「バイオ戦略 2019」が統合イノベーション戦略推進会議で決定された。この中の「5つの基本方針」の中に「バイオとデジタルの融合」を掲げ、業界が国と一体となり、「バイオデータ駆動社会」を目指すとしている。

バイオ戦略 2019 では、以下の9つの市場領域が設定されている。

- 市場領域① 高機能バイオ素材（軽量性、耐久性、安全性）
- 市場領域② バイオプラスチック（汎用プラスチック代替）
- 市場領域③ 持続的・一次生産システム
- 市場領域④ 有機廃棄物・有機排水処理
- 市場領域⑤ 生活習慣改善ヘルスケア、機能性食品、デジタルヘルス
- 市場領域⑥ バイオ医薬・再生医療・細胞治療・遺伝子治療関連産業
- 市場領域⑦ バイオ生産システム（バイオファウンドリ）
- 市場領域⑧ バイオ関連分析・測定・実験システム
- 市場領域⑨ 木材活用大型建築・スマート林業

本プロジェクトでは、実践の場の1つとして、バイオ戦略で定義されている9つの市場領域の内、上記市場領域⑤の分野でバイオ産業人会議（JABEX）と連携し、「データ基盤の全体設計」や「健康医療関連データ基盤」の構築を実施する。

背景として、これまでのバイオインダストリー協会（以下 JBA）、バイオ産業人会議との意見交換で、「実際にバイオとデジタルの融合を実現するデータ基盤を構築するためにはストレージ基盤の専門家の協力が是非とも必要」、「COCN データ連携を支える次世代データ蓄積基盤プロジェクトとは補完関係になるため、相互協力の方向で活動して頂きたい」といった声を複数の会議参加者から頂いており、今後のデータ急増が想定される市場領域でのデータ基盤構築において、相互協力の必要性が急務であるとの認識で一致している。

検討に当たっては、想定される要件を本プロジェクトで作成し、市場領域⑤のリーダーおよびサブリーダー企業、JABEX 事務局と整合性を確認し、参加企業に対し提案する。JABEX としてのロードマップ作成と合わせ、バイオ戦略 2020 に盛り込む。

	2019年				2020年								
	11月		12月		1月		2月		3月		4月		
	上旬	下旬	上旬	下旬	上旬	下旬	上旬	下旬	上旬	下旬	上旬	下旬	
次世代ヘル スケア 産業協議会	新産業創出WG		▼							▼		健康・医療 新産業協議会	
JABEX会合	ロードマップ検討会												バイオ戦略 2020
ロードマップ 作成	▼		▼		▼	▼	▼		▼	▼			
	経済産業省・ヘルスケア産業課も同席									協調領域、KPI、 ロードマップの提示			

図 4-2 バイオ領域でのロードマップ作成のスケジュール

市場領域⑤で取り扱うデータとしては、これまでの議論から下記のものを想定。

- ① スマートフォン、ウェアラブルデバイスから収集される活動情報（歩数、心拍数、血圧など）
- ② 機能性食品などの検証情報
- ③ アプリから収集される健康情報（服薬状況、血糖値、体重など）
- ④ 外部からのデータ（環境、購買、気象など）

いずれも個人情報とされない加工済みデータとして関連企業間でデータを活用する。これらのデータの活用にあたっての要件としては以下を想定。

- (a) アクセス制御とセキュリティ
 - ✓ 有資格者のみ対象のデータへアクセスを可能とする
 - ✓ 安全なネットワーク転送
- (b) 容易なデータ追加

- ✓ 日々生成されるデータを随時追加
- (c) 長期間(数十年)に渡るデータ保存・提供
 - ✓ 利用者の生涯に渡るデータ蓄積と経年変化分析への対応
- (d) Microsoft Azure とのデータ連携

- ✓ 参加企業の多くが既存で Azure を利用している

当初はこれらの要件からデータ基盤の設計を始めるが、バイオ側の扱うデータ種が増えることが想定されるので随時柔軟に対応できるようにする。

4-3 今後の検討項目について

本プロジェクトで実施してきたインタビュー調査の中で浮かび上がった 2 大キーワードは「データ保全性」と「長期保存」であった。従来は、更新やアクセスが不活性化したファイルをより適切なストレージ装置(コールド・ストレージ)に移動することにより行われてきたが、近年は、オブジェクトストレージの台頭により、HDD を長期保存用途にも用いるケースが増えてきている。これらは比較的短期間(5 年以下)の保存では好都合であっても、10 年以上といった長期保存では必ずしも最適な方法とは言えず、実際はコスト面などからデータの保存を諦めている例も多い。将来価値を生むと考えられるデータが急増していることは広く認知されていても、なかなか長期の「データ保全」に投資されない理由は主に以下が考えられる。

- ✓ 長期の「データ保全」に伴うコストを賄うための利活用の手段が整備されていない
- ✓ システム管理責任者が短期で入れ替わるため、長期視点に立った投資が行われない

長期の「データ保全」に係る課題についても、今後の活動の中で検討を行っていくが、データ保全のための要件と、それを支える階層型ストレージシステムやデータの管理手法に関し今後検討する項目について説明する。

4-3-1 「データ保全性」に関する要件

確実なデータ保全のためには、「3 Copies, 2 Different Media Types, 1 Copy Offsite」(3 つのコピーを作成、2 つの異なるメディアに保存、1 つはオフサイトに保存)という方法が提唱されている。階層型ストレージでは元々、複数のメディア(オンラインストレージとオフラインストレージ)が一体となったストレージシステムであり、この利点を活用することで、消費電力や CO₂ の排出量を削減できる。具体的には、オンラインアクセスが必須なデータであっても、データ保全のためコピーデータは HDD などのオンラインストレージにある必要性はない。つまり、従来の HDD を使ったストレージでは、3 つのコピーを持つとデータ量の 3 倍の消費電力と CO₂ の排出量となるが、本プロジェクトで提唱する階層型ストレージでは 1 倍となる。合わせて、オンラインアクセスが必須でないデータは、全てオフラインストレージで構成するため、データ量が増大しても消費電力や CO₂ の排出量は大幅に少なくすることが出来る。

ユーザーヒアリングで要件の上位を占めた「データ保全性」に関しては、重要性の認識はある

ものの具体的な実施方法についての知見が不足しているユーザーも少なからず見られた。非常に参考になるドキュメントとして、Storage Networking Industry Association (SNIA)の「Storage Security: Data Protection Technical White Paper March 2018」があり、バックアップ、アーカイブおよびディザスタリカバリーだけでなく、より広範囲なデータ保全、セキュリティに関する考え方をISOなどの標準と合わせて解説している。国内でもJIS規格や文書管理に関しては、「日本文書情報マネジメント協会(JIIMA)」が各種ガイドラインを出しユーザーへの啓発活動を実施している。しかし、データ管理技術はエンドユーザー部門に広く浸透しているとは言えない実態がある。また、データの種類や保持期間など複数の要件とコストの観点で適用する手法が変わることもユーザーの理解を難しくしている原因となっている。

- ✓ SNIA : Storage Security Data Protection Technical White Paper
<https://www.snia.org/education/whitepapers/security-wp>
- ✓ JIS Z 6019:2018 磁気テープによるデジタル情報の長期保存方法
日本産業調査会(JIS 検索) : <https://www.jisc.go.jp/app/jis/general/GnrJISSearch.html> (“Z6019”で検索)
- ✓ 公益社団法人 日本文書情報マネジメント協会
ガイドライン <https://www.jiima.or.jp/activity/policy/#guideline>

以上の状況を踏まえ、次年度以降は協議会で、以下の点について活動を実施する。

- ✓ 最新の技術動向を踏まえたデータ保全性に関するガイドライン・標準化(JIS/ISO等)と普及啓発活動
保存データの特性など要件により多様な選択肢があることを踏まえた、適切なデータ保全の分かり易いベストプラクティスを作成し啓発活動を実施する。その中で、ユーザーの検討の負担を減らすため、保存期間や更新の有無、アクセススピードなどデータ属性に合わせた具体的なデータ全方法を示していく。

4-3-2 「長期保存」に適した階層型ストレージについて

「データ保全性」の次に多くあがった「保存環境の継続性」、「省電力・コスト」などの要件は、データを「長期保存」することによって重要度が増す問題である。

現在、データ利活用の有効性を示す成果が得られ始めている分野の多くはビジネス効率化や現実の社会課題をテーマにしているため、現状の可視化、潜在知の発見を目的としており、入力となるデータはリアルタイムデータである。その観点では「長期保存」は、直ぐに顕在化する問題ではない。

しかし、過去の多くの研究テーマが研究の対象を特定の時間帯から時間軸方向に拡大することで新たな発見をしてきた。データ利活用においても年代ごとの記録や経年変化を入力とすることで、異なる視点での発見、知見が得られるものと期待できる。

その意味で「長期保存」はデータ蓄積基盤のロードマップ上に置き、検討しておくべき重要な要件である。

データ蓄積基盤の構築に当たって決定すべきパラメータは多種多様である。それらはベースとするデバイス技術、アクセス利便性とコストを両立させ最適化するソフトウェア技術、データ管理手法などであり、アプリケーションの要件に応じて適宜、選択する必要がある、要件項目などをまとめた文書は幾つか存在している。

光産業技術振興協会オプトニュース Vol. 14, No. 3 (2019)

「大容量データのアーカイブに適したライブラリ装置-データ分析基盤の構築のために」

一方、データ蓄積基盤は、これまでは幾つかの個別システムが構築、運用されているものの大規模かつ長期にわたって運用された事例は、一般的には知られていないため、今後のシステム開発の足枷となりかねない状況である。そこで次年度以降の協議会では、以下の活動を通して、今後必要となるデータ蓄積基盤の構築や運用を担当する部門の人々に共有知識を提供するなどにより、貢献していきたい。

- ① データ蓄積基盤構築のための要件項目、決定指針の充実
- ② データ長期保存のためのシステム構築事例のカタログ化とノウハウの普及活動

4-3-3 「データ連携」を実現する統一的なデータ管理手法の仕組みについて

3章において、分散環境でデータ管理を実現する技術と以下の手法について述べた。

- ① データ管理ポリシーをメタデータとして記述し、データとともに移動する
- ② データ管理サーバは、データへのアクセス時あるいは定期的に管理ポリシーを読み取り、そのポリシーに従った操作を行う。

このような手法は、分散システム構築の設計段階で統一規約を設定することで実現できる。しかし、今後は異なる分野、組織、体制によって、同時並行的に様々なデータ蓄積基盤が構築されていくことになる。実際、幾つかのプロジェクトではデータ蓄積基盤の構築・運用を待たずに、クラウドサービスにデータを蓄積しているプロジェクトも存在している。

このため各システムで設定されている管理ポリシーはシステム間で互換性がない状態になっており、クラウドシステムを跨いでデータを移動した場合には、「不明の管理ポリシー」と認識され、所望のデータ管理を行うことが出来ない。今後、データ蓄積基盤を連携させた場合に、各基盤内では自動化されているデータ管理が機能せず、無駄な管理工数が必要になると予見される。

以上のことから本プロジェクトでは、大規模なデータ蓄積基盤が複数、分散して構築される場合でも、それらの基盤を連携させて活用することを自動的かつ効率的に行える「統一的なデータ管理手法の仕組み」を用意すべきであると考える。

今後、協議会においては、「統一的なデータ管理手法の仕組み」の検討と提言をまとめ、来るべき「分野間でのデータ連携」が円滑に実現できるよう貢献していく。

<検討内容>

① 仕組みの具体的な実現案の比較検討

✓ メタデータ読み替え案

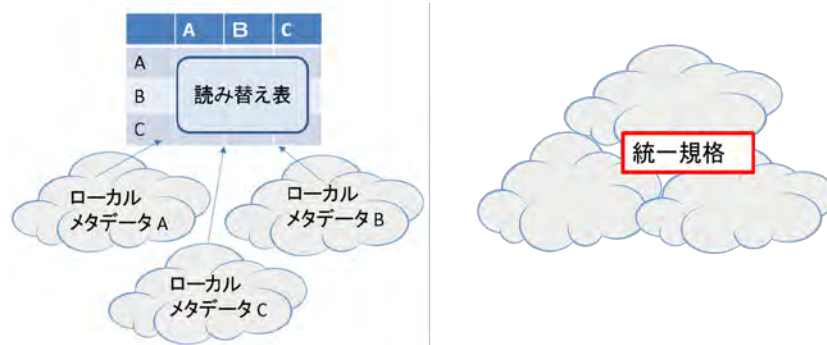
各基盤で使われるローカルメタデータをデータ移動時に読み替え表によって変換する

✓ 統一規格案

統一規格を設定し、予め基盤構築, または連携の際にメタデータを合わせ込む

② 「読み替え」または「統一規格」のベースとなる標準の選定

上記のいずれの仕組みでもメタデータの共通辞書が必要になる。その標準辞書の選定
(CDMI、デファクト標準、etc.)



蓄積基盤のメタデータ読み替え案

蓄積基盤のメタデータ統一規格案

図 4-3 今後の検討内容

5. まとめ

以下の2点を提言とする。

データ蓄積基盤側でのアクセスおよびデータ管理手法の統一を図ることで、異なる分野間のデータ連携のさらなる加速と拡大を目指すべきである。

データ駆動社会では、膨大なデータがあふれる社会になるため、その維持・運営には多くの電力が必要になる。このため、これらを支えるデータ連携・蓄積基盤は、SDGs（特にCO2削減）に適合するよう低消費電力化を目指していくべきである。

ソフトウェア階層におけるデータ連携のみならず、それを支える「データ蓄積基盤」を含めたシステム全体での検討が必要で、今後も関係各組織が協調・協業して継続的に活動する必要がある。今後の活動の場としては、本プロジェクトの母体となった IDEMA JAPAN（日本 HDD 協会）を活用し、その発展形として「次世代データ蓄積・連携協議会」（仮称）を設立し（図2）、大学や研究機関の協力も得ながら以下の検討を行う。

■グリーンイノベーションとして期待される階層型ストレージの標準化とその横展開

- データの長期保存のためのシステム構築事例のカatalog化
- データ管理を統一的行うためのメタデータの標準化方式の提示
- コールドストレージ利用促進によるCO2削減への貢献と関連産業の活性化
- 将来的な海外展開・連携の模索

■「データ保全」「長期保存」を実現する指針の提示（ガイドラインや標準化）と普及活動

関係省庁には、「ガイドライン・標準の普及推進のサポート」、「データ連携を推進している分野との関係作り」、「国際連携」など引き続き様々なサポートをお願いする。

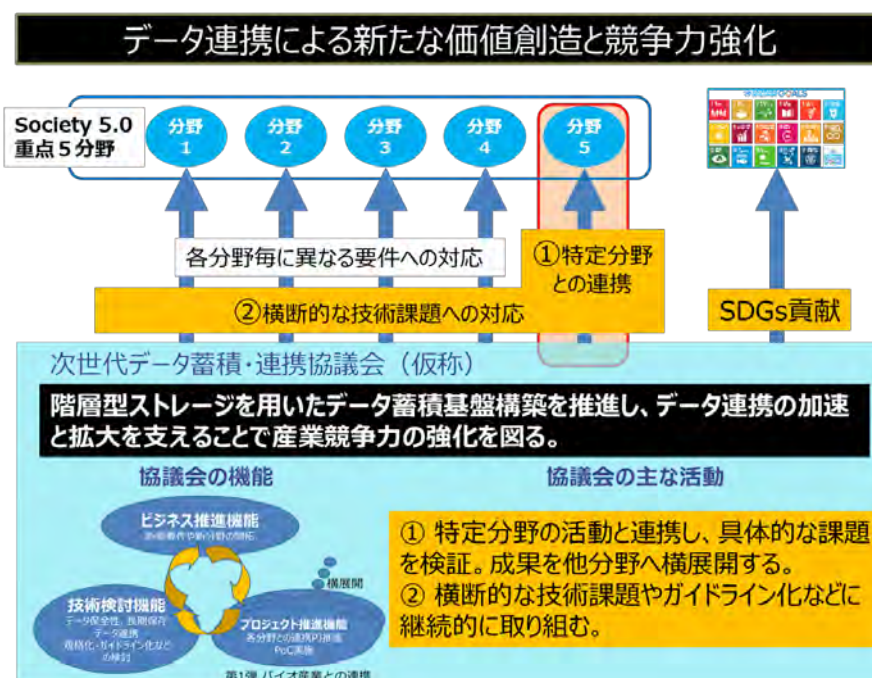


図 5-1 協議体の機能と活動概要

以上

【参考資料 1】 データ蓄積基盤に係る技術動向

0. はじめに

今後の情報システムがオンプレミス、プライベートやパブリッククラウド環境のいずれの形態になっても、それらを支えるハードウェアに対する要求は基本的には同様と考えられる。このため、COCN「データ連携を支える次世代データ蓄積基盤プロジェクト」では、HDD、磁気テープ、光ディスクといったハードウェアレベルの技術動向を調査し、各技術の発展を鑑みて、今後のストレージシステムについて検討を行った。

1. ストレージ技術ロードマップ

大規模なデータを扱う高性能のコンピューティング環境では、データを保管するストレージとして複数の異なる装置が利用されている。以前は HDD が中心的な役割を果たしていたが、現在はいより高速なデータアクセスが可能な SSD や、大量のデータを安価に長期間保存できる磁気テープ装置や光ディスク装置の利用も広がってきた。それぞれの装置には他の装置にはない特徴があり、それらが相互に補完することで、コンピューティング・システム全体のストレージに対する需要を満たす記憶階層を成している。

今後 10 年程度を考えても、これらが適材適所で役割分担する構図は変わらない見込みである。各々の装置のメーカーは大容量化や高速化に向けたロードマップを描いており、製品の進歩が続くことは間違いない。ただし、いずれも技術開発のハードルは高く、他の装置の役割を包含してしまうほどの進化を遂げる装置は見当たらないからである。

例えば SSD と HDD の間の 1T バイト当たりの価格は、当面は 1 桁違う状態が続くとみられる。調査会社のテクノシステムリサーチは、2023 年の時点でも SSD とニアライン向け HDD では、単位容量当たりの価格は 10 倍違うと予測する（図 1-1）。また、米 Western Digital 社が 2017 年に公表した資料は、2028 年の時点でも HDD と SSD の間には 10 倍程度の差があると指摘した（図 1-2）。このため、当分は SSD で HDD を置き換えるよりも、両者を活用した方がコスト面では有利になる。

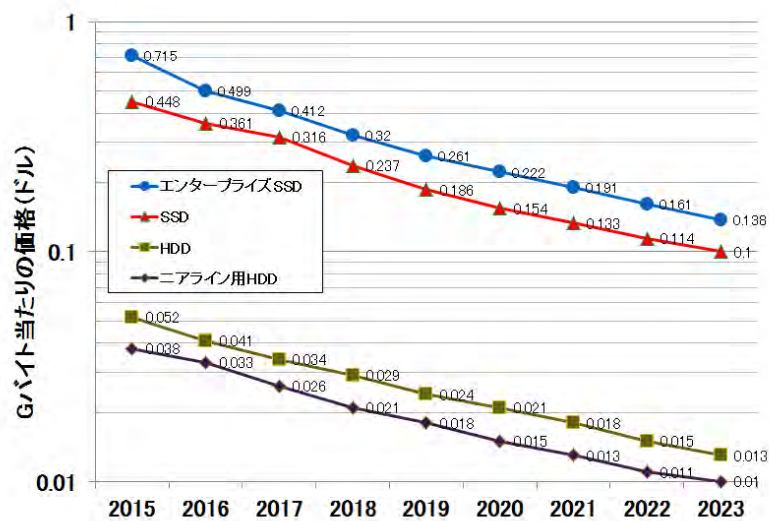
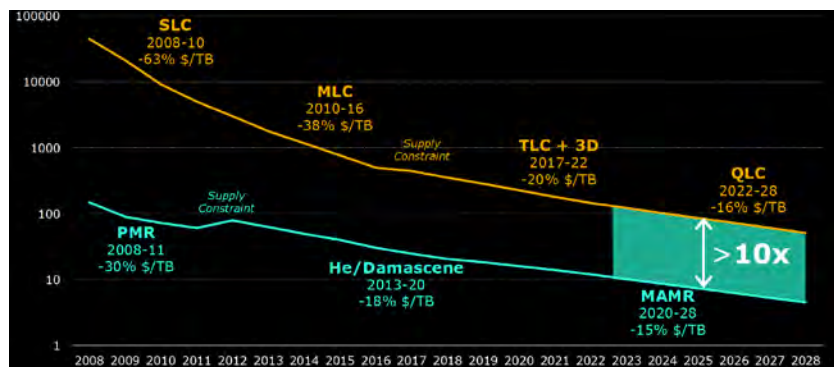


図 1-1 : HDD と SSD の G バイトあたりの価格の推移
(出所 : テクノシステムリサーチのデータをもとに作成)



SLC : Single Level Cell MLC : Multi-Level Cell TLC : Triple-Level Cell QLC : Quad-Level Cell
PMR : Perpendicular Magnetic Recording MAMR : Microwave Assisted Magnetic Recording

図 1-2 HDD と SSD の T バイトあたりの価格の推移
(出所 : Western Digital 社)

このことは、SSD、HDD、磁気テープ装置、光ディスク装置が自動的に連携し、ユーザーからはあたかも 1 つのストレージ装置であるかのように振る舞うことを可能にする技術の重要性も示唆している。ユーザーにとって肝心なのは、必要なデータを必要なときに高速で保存ないしは再生でき、あるいは将来必要になるときまで安全に格納しておくといったストレージ装置の機能であり、個別の装置の使い勝手に煩わされることではないからである。

とりわけデータの長期保存という観点では、今後は HDD、磁気テープ装置、光ディスク装置の連携がますます重要になるだろう。これからの社会の発展においてデータが果たす役割は膨らむ一方であり、それに応じて保存すべきデータの量も指数関数的に増えていく。それらのうち頻繁かつ高速にアクセスする必要があるものは SSD に格納される一方で、アクセスの頻度が少ないも

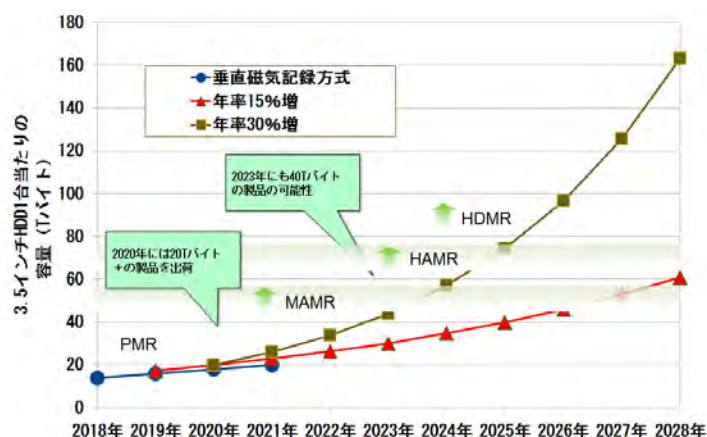
の、今後の利用価値を見込めたり保存の義務があったりするデータは、その下の記憶階層にアーカイブされる。

アーカイブされるデータの中でも、必要に応じて即座に読み出される必要があるもの（アクティブアーカイブ）は HDD、長期保存が何よりも重要でアクセスの即時性はさほど要求されないもの（ロングタームアーカイブ）は磁気テープや光ディスクに保管される。ロングタームアーカイブのうち、容量あたりのコストの低さや高速なデータ転送速度が重要な場合は磁気テープ、数十年にわたる保管や耐環境性の高さなどを要求する場合は光ディスクが適している。こうした役割の違いから、HDD、磁気テープ、光ディスクはそれぞれの特徴をより発揮できる方向で技術開発を続けることになる。

HDD の動向

HDD では、従来と同様に価格あたりの容量の最大化が第一目標となる。2019 年には、これまで容量拡大のけん引役だった垂直磁気記録（PMR：Perpendicular Magnetic Recording）に代わる新方式の実用化が始まり、今後 10 年程度の成長に道筋をつける見通しだ。1 つはマイクロ波アシスト記録（MAMR：Microwave Assisted Magnetic Recording）、もう 1 つは熱アシスト記録（HAMR：Heat Assisted Magnetic Recording）である。前者は米 Western Digital（WD）社、後者は米 Seagate Technology 社が 2019 年中にそれぞれ製品を投入するとしている。日本勢の東芝も媒体メーカーの昭和電工と協力して、まずは MAMR 方式のサンプル品を年内に生産する見込みだ。

新方式の登場は、今後 5～10 年にわたって HDD の継続的な容量拡大を約束する。WD 社や Seagate 社は、2020 年には容量が 20T バイト超の 3.5 インチ HDD を提供するとそれぞれ表明済みである。その後の拡大についても、WD 社は年率 15% 増、Seagate 社は同 30% 増で進むと表明したことがあり、2020 年代半ばから後半にかけて容量の拡大基調は続くとみられる（図 1-3）。



PMR：Perpendicular Magnetic Recording MAMR：Microwave Assisted Magnetic Recording
HAMR：Heat Assisted Magnetic Recording HDMR：Heated-Dot Magnetic Recording

図 1-3 HDD 容量の推移

（出所：各社の発表を基に推定）

ただし、MAMR、HAMR とともに達成できる記録密度には限度がある。前者は WD 社によれば 4T ビッ

ト/インチ²程度、後者は米 ASTC (Advanced Storage Technology Consortium) が 2016 年に公表したロードマップによると 5~6T ビット/インチ² のようだ。記録密度が約 1T ビット/インチ² でディスク 8 枚を内蔵する 3.5 インチ HDD の容量が 14T バイトだったことから、ディスク枚数が一定だとすると MAMR は 60T バイト弱、HAMR は 70~80T バイト程度が装置あたりの容量の上限になる。

もっとも、ガラス基板を用いることで、現在と同じ装置の寸法の中にディスクを 12 枚格納できることを、HOYA が示している。その分のコストを吸収する努力は必要ではあるが、記録密度が同等でも容量を 1.5 倍に引き上げることは可能なのである。ここから、1 台あたり 100T バイト程度の容量までは、実現の見通しが立ったといえそうだ。

ASTC のロードマップでは、5~6T ビット/インチ² 以上の記録密度の実現には、いわゆるビットパターンドメディアと HAMR を組み合わせた HDMR (Heated-Dot Magnetic Recording) 方式が必要になるとされる。技術のアイデアは以前からあるものの、現在までにこの方式で実用的な HDD を構成した例は発表されておらず、今後の研究開発が期待される。

容量の拡大が順調に進む一方で、HDD の性能はさほど向上していない。典型的な例がデータ転送速度で、ここ数年大きな進歩がみられない。2019 年 7 月現在、Seagate 社が業界最大の容量 16T バイトの製品を販売しているが、カタログ上の転送速度は最大 261M バイト/秒で、約 2 年前に登場した 12T バイト品と同じある。HDD の回転速度はいずれも 7200rpm と同一であることから、2 つの製品の線記録密度が同等であることが推測できる。

今後、HAMR や MAMR によって記録密度の向上が加速すれば転送速度も高まりうるが、面記録密度が年率 15~30% 増で高まったとしても、BAR (Bit Aspect Ratio) が同等と仮定すると、線記録密度の伸びは 4~5% 程度に過ぎない。高速化の 1 つの手段は回転速度の向上だが、消費電力を押し上げるため採用しづらい。

今後活用される可能性があるのが、1 台の HDD に独立して動作する複数のアクチュエータを搭載する技術である。実際、Seagate 社は 2018 年 3 月にデュアルアクチュエータを利用する「MACH. 2」と呼ぶ技術のデモを公開し、データ転送速度を 480M バイト/秒まで高めることが可能と表明した。回転速度 1 万 5000rpm の製品と比べても 60% 高速であるという。

複数のアクチュエータの搭載は、同一装置にある異なるデータに並行してアクセスできることにもなり、アクセス時間の短縮にも効果がある。ただし、本質的には容量が 1/2 の製品を 2 台使った場合と効果は同じであるため、複数アクチュエータを搭載する大容量の装置を作る場合と、小容量の装置を複数台作る場合のどちらの製造コストが安くなるかの勝負になる。将来は、単純な構造で容量の小さい HDD を多数並べて、高機能のコントローラで制御するといった構成もありうるかもしれない。

磁気テープの動向

HDD と磁気テープは、今後も競合ではなく棲み分けを続ける見込みである。オンラインでデータにアクセスできる HDD に対して、磁気テープでは媒体の搬送などに時間がかかり、アクセス時間は大きく見劣りする。

一方で、管理やメンテナンス、必要な人的コストまで含めた TCO (Total Cost of Ownership) では磁気テープが有利である。米 Enterprise Strategy Group (ESG) 社の 2018 年の調査によれば、業界標準の LT0 第 8 世代 (LT0-8) を利用した場合で 10 年間にかかる TCO (210 万ドル) は、全てを HDD に保存した場合 (1520 万ドル) の 14% に留まるという (図 4-4)。

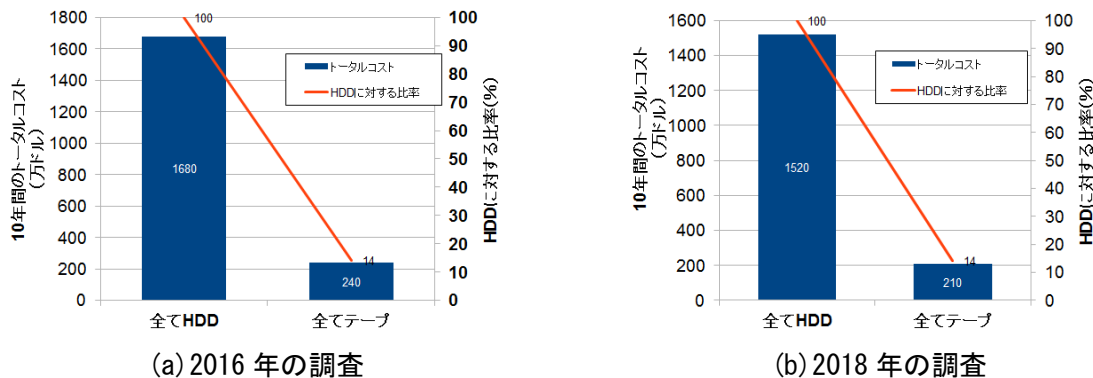


図 1-4 磁気テープ、HDD の TCO の比較
(出所：米 Enterprise Strategy Group 社)

また、米 INSIC (Information Storage Industry Consortium) によれば、2016 年の ESG による同様な調査や、2015 年に米 Clipper Group が実施した類似の調査から、磁気テープと HDD の間の TCO の比率は年々拡大しているという。2015 年には HDD の TCO が磁気ディスクと比べて 6.2 倍大きかったのに対して、2016 年には 7.0 倍、2018 年には 7.2 倍に拡大した。

磁気テープの今後の容量拡大のロードマップからすれば、少なくともこの差が縮まることはなさそうだ。INSIC が 2019 年に公開した技術ロードマップでは、2019 年の 1 巻当たりの非圧縮容量が 25T バイトであるのに対し、年率 40% 増で成長し、2029 年には同 723T バイトに達するという (図 1-5)。HDD の容量が仮に年率 30% 増で拡大したとしても差は開くばかりである。

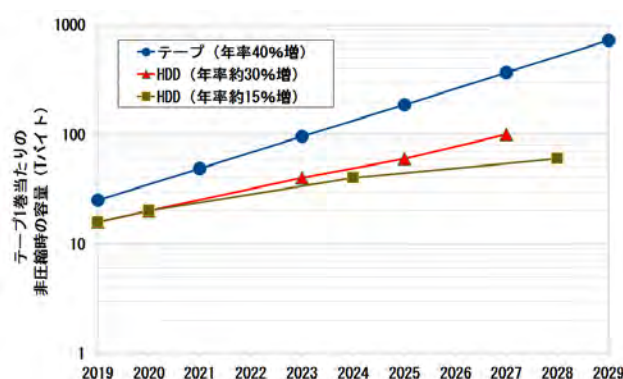


図 1-5 磁気テープの容量の推移
(出所：INSIC のロードマップのデータを基に加筆)

これだけの差が開く理由は、容量拡大の裏付けになる面記録密度の違いにある。記録面積が大きい磁気テープでは、同じ容量を実現するために必要な面記録密度が HDD の 1/100 程度と低い。INSIC のロードマップが示す 2029 年の容量 723T バイトに対応する面密度は 278 G ビット/インチ

²で、HDD 製品が 2009 年に達成しているものと同等であり、磁気記録としては実現可能であることが明らかである。また、最新の磁気テープの技術デモでも現行テープの 10 倍程度の面記録密度（201 G ビット/インチ²）が実証されている。これらの根拠から、今後も磁気テープは面記録密度を高め容量コストを低減することで TCO の優位性を保つ見込みだ。

なお、INSIC のロードマップは「可能な製品の平均的な値を表す技術ロードマップ (a technology roadmap that might represent an average of possible products)」といい、実際の製品が必ずこの通りに登場するわけではない。INSIC のロードマップで 2019 年としている 25T バイトに近い製品としては、米 IBM 社のエンタープライズシステム（3592 シリーズ）の JE テープがあり、1 巻あたりの非圧縮容量は 20T バイトである。オープン規格の LT0 では現行の LT0-8（12T バイト）の 4 世代先の LT0-12（最大 192TB）までのロードマップが示されているが、製品化の記事は未公表である。

性能面でも HDD よりも磁気テープの方が有利な点がある。大量のデータを連続的に転送する場合の速度である。複数チャネルを利用してデータを一気に転送できる磁気テープ装置は、現在の LT0-8 で非圧縮時に最大 300M バイト/秒以上の転送速度を実現しており、上述した最新のニアライン用 HDD の最高速度を凌いでいる。INSIC の技術ロードマップは、2019 年の 475M バイト/秒を起点に、その後は年率 15%増で高速になるという将来像を描いている。

なお磁気テープの高速転送能力は、あくまでも大容量のデータを転送する際のみに威力を発揮することに注意が必要である。富士フイルムのウェブサイトでは、データの容量が 30G バイトを超えると優位性が現れると指摘している。

光ディスクの動向

日本で独自の進化を遂げてきたストレージ装置が光ディスクであり、中でもデータのアーカイブに狙いを定めた規格が「Archival Disc（アーカイバル・ディスク）」である。民生機器の Blu-ray Disc の技術を基に高密度化を進め、ソニーとパナソニックが 2014 年に共同で発表した。既に容量 300G バイトのディスクを 11 枚内蔵するカートリッジを用いた製品をソニーが販売しているうえ、2019 年内には第 2 世代の容量 500G バイトのディスクを用いてカートリッジの容量を 5.5T バイトに高めた製品が登場する見込みである。

Archival Disc の特色は、温度・湿度に対する媒体の耐性が高いことであり、世界中のほとんどの都市で、空調設備なしでも長期保存が可能な性能を有する（図 4-6）。究極の状況を想定して、熱湯や海水に浸けたり、亜硫酸ガス（SO₂）や硫化水素（H₂S）などの腐食性ガスに一定期間暴露したりするなどの過酷なテストを行った場合も、その後にディスク表面をクリーニングするだけで問題なくデータが再生可能であることが確認されている。このため、災害などの不測の事態にも耐える保存手段が必要なケースには、Archival Disc が最も適している。

また、Archival Disc はライトワンス方式で記録膜の中の金属酸化物の不可逆変化を用いて書き込んだデータは物理的に書き換え不能であるため原理上改ざんが不可能という特徴もある。特に中国市場ではこの特徴に注目して、Archival Disc に対する需要が高まっている。光ディスクでは、寿命推定試験方法やデータの長期保管における運用管理方法が ISO、JIS で整備されている

ことも大きな特徴である。これらの規格により、長期間、改ざんなく確実に保存しておきたいデータには Archival Disc をはじめとした光ディスクが適しているといえる。

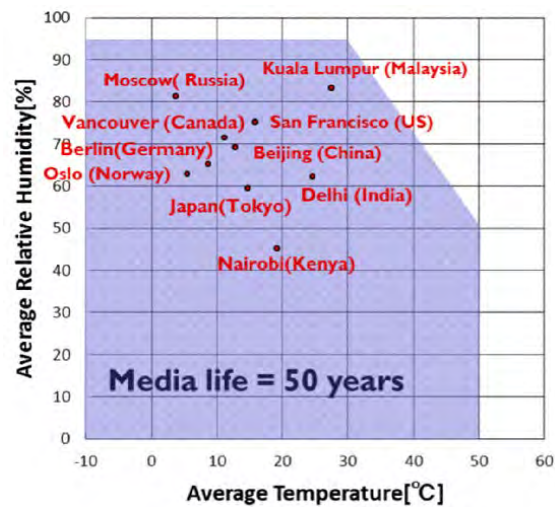


図 1-6 Archival Disc の耐環境性

(出所 : Archival Disc Technology の White Paper)

Archival Disc はディスク 1 枚あたり 1TB を実現する第 3 世代までのロードマップが示されている (図 1-7)。11 枚のディスクが収納されたカートリッジの容量は 11TB に到達する。

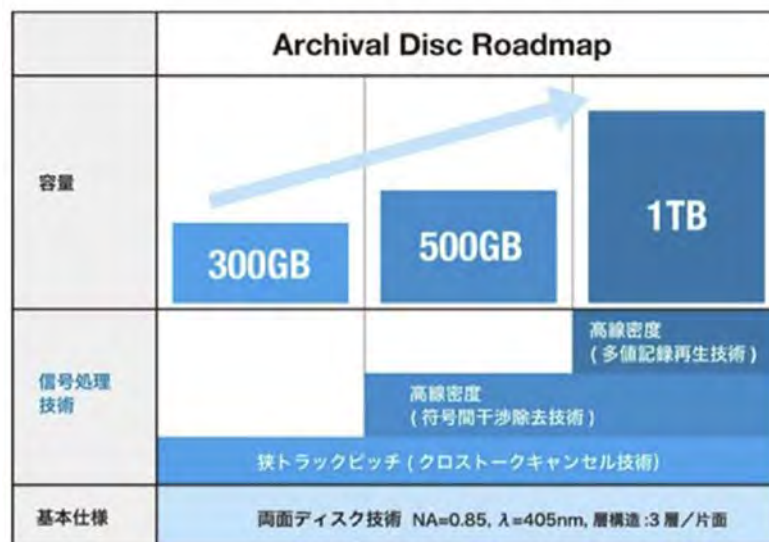


図 1-7 Archival Disc の容量の推移

(出所 : Archival Disc Technology の White Paper)

さらに、Archival ディスクにとどまらず、光ディスクの記録密度を向上させる技術開発は多くの企業や大学で活発に行われている。一例をあげるとクッション材と蛍光体材料を埋め込まれたフィルムをラミネート構造化することで 16 層の超多層化を実現し、容量 10TB を達成する技術 (Folio Photonics 社) や超高速レーザーを使って石英ガラスに直接データを記録することによ

り 360TB の容量を目指す研究（Microsoft 社／University of Southampton）などがある。これらの技術開発により光ディスクは大容量化を実現しつつ、データの長期保全性といった強みが発揮できる領域を中心に活用の広がりが期待される。

2. 先端技術開発動向

新方式のストレージ技術の有力候補としては、新たな記録原理に基づく半導体メモリ技術がある。MRAM、ReRAM、PCRAM といった技術の開発が進んでいる。ただし最初に結論を書くと、これらは記憶階層の中でも DRAM と NAND フラッシュメモリの間を狙うものといえ、HDD より下の階層の装置の置き換えを狙ったものではない。例えば高速かつ書き換え可能回数が多い MRAM は SoC の主記憶やキャッシュメモリ、容量あたりのコストが低い PCRAM はいわゆる SCM（ストレージクラスメモリ）を狙っており、既に実用化も始まっている。

日本国内では、これらの新型メモリの性能を大幅に改良する研究開発が進んでいる。実際、産業技術総合研究所が開発した Ge 原子の配位変化を用いる PCRAM は、既存の PCRAM と比べて消費電力を桁違いに改善できる。もっとも、これらの先端技術をもってしても、NAND フラッシュメモリのコストを下回することは困難であり、データ蓄積基盤のメインの記録媒体として採用することは困難な状況である。

このほか、生物の DNA を記憶媒体に利用する DNA ストレージなどのエキゾチックな技術の開発が進んでいるが、2019 年現在では実用化へのメドが立ったとはいえない。上記の検討から、今後 10 年程度は新型半導体メモリ、新方式のストレージ技術ともに、次世代データ蓄積基盤の要素技術にはなりにくいと結論づけられる。

3. データ蓄積基盤向けストレージシステムの今後

前節までに示した様に、ストレージデバイスのロードマップや、新デバイスの学術研究を調査した結果、現在のストレージの全階層をカバー可能な超汎用的なデバイスの出現を予見させる技術には当たらなかった。また過去のストレージデバイス研究について振り返ると研究段階の技術の実用化には 10 年を超える期間を要している。よって、本文に述べている階層型ストレージシステムは、「アクセスの利便性」、「大量・大容量性」、「経済性」を同時に満たす現実なシステム構築方法であり続けると推測できる。

この推測の傍証として、以下に例を挙げる。

- ① クラウドサービスが始まった当初のサービスプロバイダーのストレージシステムは全て HDD によって構成されていたが、データ容量の増大、特にアーカイブデータの容量増大に対応するためライブラリ装置を活用し始めている。
- ② 中国では放送局、政府機関のデータ利活用基盤、博物館などのデータセンターにおいて光ディスクライブラリ装置の導入が積極的に行われている。

以上のように、階層型ストレージシステムはデータ利活用で先行する各企業、各国が採用している標準的な考え方であり、増大するデータを受け止めることができる唯一の現実解である。

【参考資料 2】 次世代データ蓄積基盤に求められるシステム要件詳細一覧

COCN「次世代データ蓄積基盤プロジェクト」では、次世代データ蓄積基盤としての要件を、ユーザーヒアリング結果や専門家の知見を集約し下記の通りまとめた。

機能カテゴリー	機能項目	備考	ユーザー ニーズ
階層型ストレージ機能	階層制御の基本機能と複数のポリシー設定	- データのアクセス特性に合わせた複数の記録メディアの階層制御定義 - マイグレート条件(日数/オブジェクトサイズ/拡張子/パラメータ：即時指定など) - リコール時のキャッシュ保持期間 - High/Low Watermark 制御 - データのグルーピングと一括マイグレート/リコール	コスト削減 (低コストメディアの活用)
	オフラインメディアのアクセス方式	ファイルまたはオブジェクト形式でのアクセス技法 生のライブラリ制御はベンダー側提供ソフトで隠ぺい	
	外部クラウド連携	パブリッククラウドとのデータ連携機能	
	リソース制御	マイグレーション/リコールで使用するリソースの上限設定など 媒体・ドライブ割り当ての最適化/排他制御	
	インターナルジョブ管理	マイグレーション/リコールおよびデータ復旧などの内部ジョブ管理機能	
	スループット性能の確保	シングル・ストリーミングでの高スループット性能の確保 複数ドライブ(媒体)の並列動作機能など	アクセス性能の担保
		キャッシュディスクのデフラグ	
	アクセス状況に応じた最適配置	アクセスログの分析から格納領域の自動最適化 ポリシーや API 設定からの変更を通知	

データ保全	レプリケーション	一般には三重化以上のレプリケーション技術	確実な データ保全
	イレージャー コーディング	データ保存コスト削減と保全性のバランス ● 適切なパフォーマンスの確保	
	遠隔地 レプリケーション	センター間データ同期手法とデータプロテクション	
	データ改ざん防止／ 削除不可	光ディスク/磁気テープ(WORM 媒体)の利用	真正性の 保証
		完全な改ざん防止はメディアだけでは不完全 一般には電子認証およびタイムスタンプなどを用いる ● 基盤側だけであれば、API レベルでも実現可能 ● ネットワーク上の転送に関しては、ハッシュ値 確認機能なども検討要(「周辺機能：接続ネット ワーク関連」の項) 電子認証やタイムスタンプは認証局の方式変更があり、この際に認証情報の再付与が必要となる。 ● WORM メディアのデータは全て新メディアへの 格納が必要(無駄な(廃棄)メディアが発生) ● データ量が多くなると、認証情報の再付与のため の I/O 負荷がシステムリソースを圧迫する ● 認証情報の再付与したものをオリジナルとして 扱うのか、改版として保持するのかの検討要 (元データは読めない)	
	媒体上の暗号化と 統合的な 暗号鍵管理システム	記録媒体に依存しない暗号化と鍵管理システム Trusted Computing Group(TCG)/Opal Storage Specification 準拠など 媒体を持ち出すと読めない／再組み込みで読める HDD/SSD の SED および LTO は対応可 * 媒体廃棄と合わせて検討が必要	①暗号化による 媒体の 外部保管時の 情報漏えい 防止
	暗号危殆化対策	ハードウェア暗号化の場合、機器のリプレイス(5～7 年)により暗号方式の変更が可能*である。 *:適切な暗号化方式に対応したハードウェアが提供 されている前提。システム提供側は機器の入れ替え およびデータマイグレーションを速やかに実施する 必要あり。また、媒体の外部保管(棚管理)している 場合もデータマイグレーションが必要。	②機器・媒体の 廃棄時の データ消去 代替機能

		ユーザー側で暗号化する場合は対策が必要となる。 データはユーザーによって、「読出し→復号化→新暗号方式での暗号化→書き込み」の必要がある。合わせてユーザーによる、暗号鍵管理が前提。 ● 過去の暗号方式のデータの完全消去が必要 (改版として保存すると解読されてしまう) ● 基盤側としての対応 上記のデータ消去に伴い、記録されていた WORM メディアは破棄。同一メディア上の別データは別のメディアへ移動させる必要あり。	
	未アクセスデータの定期的なチェック・訂正機能	長期保存のための媒体の経年劣化への対応 ● 書き込み時以外のベリファイ	長期のデータ保証
システム管理	認証機能		
	アクセス権限管理	利用者属性に応じたデータアクセス機能 ● Owner/User Group が不在となったオブジェクトの処置の考慮	アクセスコントロール
	管理・監視機能	稼働状況・障害・性能監視、構成管理 寿命管理(媒体、ドライブ、ロボット)／交換指示 ハードウェア版数、BIOS／ファーム版数、ソフト版数管理 負荷状況によるリソースの自動再配置(性能平準化)	システムの安定稼働 リソースの有効活用によるコスト最適化
	アクセスログの取得と監査機能	通常のアクセスログだけでなく、特権 ID の操作ログも取得 ● e-Discovery などへの対応も検討要	内部統制
	マルチテナント	セキュアな分割機能。認証、RBAC との連携 複数のサービスレベルの提供 ● SLA を担保する仕組み	基盤のサービス化
	Role-based access control (RBAC)	権限分割と移譲の仕組み	
	課金システム	認証機能・アクセス情報から課金請求／回収	
	外部からの攻撃に対応する機能	FIPS/NIST SP800-53/171 などへの対応 ● 侵入検知／遮断、汚染データの削除と復元 ● ソフトウェア／ファームウェアの保護と復元	情報漏えい防止

外部インターフェース	オブジェクト アクセス	CDMI (Cloud Data Management Interface) 準拠 ISO/IEC 17826 ● 実装レベルの検討 ● ビジネス性を含めた事実上デファクトとの互換性比較検討 *Amazon S3/Microsoft Azure など各種クラウドサービスとのブリッジ機能要	アクセス性の 担保 標準化
	ファイルアクセス	データ流通の仕組みとの連携 同一サイト内 (in B など) での NFS/ SMB (CIFS) でのアクセス * WAN 経由は不可	アクセス性の 担保 (既存アプリから容易なアクセス)
データ管理	保存要件ごとの処理	保存期限設定：期限後の自動削除 (通知機能の考慮) ● 公文書の場合、期限満了後、国立公文書館への移管あり 改ざん防止設定 (削除不可) / 改版可能 (履歴あり)	自動化・ 省力化
	データのリネージュ	データの改版および来歴管理	真正性の保証
	メタ検索機能	利用者権限に基づくデータの検索機能/データ間の関連付け AI によるメタの自動付与についても検討 メタの保存形式についても検討要	データ活用
	個人情報などの秘匿化	意図せず混入している個人情報などの削除 (マスキング) やアクセス制限 ● 書込み直後は読出し不可としておき、クリーニング後問題なければアクセス可とする	個人情報保護
	データフォーマット変換機能 (見読性の担保)	経年によるデータフォーマットの標準が変化することへの対応 ● 改版保存としてオリジナルデータを残す (リネージュとして) ● 例として、天文分野では画像ファイルを FITS フォーマットで保存する取り組みなど ● OAIS など ISO もあり対応を検討する必要あり	データの長期 活用対応

ライフサイクルマネジメント (LCM)	ハードウェア保守	リソース切り離し・組み込み、データ再同期、強制ノード間コピー	システムの安定稼働・信頼性担保 マイグレーションによる長期稼働保守コスト削減
	ソフトウェアアップデート	改版管理、ローリングアップデート	
	ベンダー保守レス運用	サーバ/HDD/ネットワーク機器の使い捨てによる保守コスト削減 磁気テープ/光ディスクライブラリの考慮	
	機器増設・減設	ソフトウェア・ファームウェアの自動適合化 ソフトウェアのコンテナ化と自動配布(配布管理) ハードウェアの機能ブロック化(増設単位)	
	機器・媒体のマイグレーション	ライブラリ・ドライブ・媒体の更新方法 オンライン移行のための階層制御ソフトの機能	
	リプレース機器／媒体のデータ消去／廃棄	暗号化のみの廃棄可否 廃棄業者との契約 廃棄メディアの再資源化の検討	
	ドライブ・クリーニング機能	自動的なクリーニング動作	
	メディアエラー時のデータ復元	冗長データからの復元	
	インシデント管理／リリース管理／修正確認と情報共有	サービス品質維持のための開発・運用者間の各種調整と情報共有	
周辺機能	大量データの移入／移出	短期でPBクラスの移入作業など セキュリティを担保した安全なデータ移入／移出手段の提供	利用者環境に合わせたデータ活用機能
	可搬媒体での移入／移出	ネットワーク転送が困難なケース(データ量/環境) DC環境ではない現場の環境への対応 暗号化による情報漏えい防止／改ざん防止(WORM) 輸送業者との契約 媒体によるデータアクセスを可能とするセキュリティールーム(端末)	

	接続ネットワーク関連	安全な P to P 接続機能、暗号通信、認証機能	情報漏えい防止
		ネットワーク上のデータ改ざん防止・検知	改ざん防止
		複数サイトを跨いだデータの来歴管理(リネージュ) ● 転送データのハッシュ確認など	リネージュ
	媒体の外部保管	媒体の投入／排出／マウント要求(人的動作との連携) 媒体の入出庫管理(RFID 利用など) DC 入り口への余熱室設置(結露防止)／温湿度環境差オペレータ管理	コスト削減
コスト削減	データリダクション	性能的にロスレスな圧縮機能など(LTO は対応) 重複排除機能	
	媒体利用効率向上	磁気テープ上のデータコロケーション	
省エネルギー	DC スペース／空調などのコスト	機器の発熱量だけでなく、容量観点でのスペースと空調コストも考慮	省エネルギー／省電力
	ドライブの電源制御による消費電力削減	オンライン(HDD/SSD)領域で MAID(Massive Array of Idle Disks)機能や電源制御などの取組み	
		光ディスク/磁気テープ・ドライブの未使用時の電源断	
ファシリティ	大型ライブラリ設置に関する考慮	荷重、外寸、消費電力、発熱量、空調能力 サーバ搭載ラックとは特性が大きく異なる 筐体の高さなどが空気の流れに影響するなど。場合によっては仕切りや空調設備の分離などが必要となる。	
	保守スペース	ベンダー保守機器は必要な保守スペースの確保要 特に大型ライブラリ装置	
	媒体・投入／排出経路	媒体の搬入・搬出経路確保	
	物理セキュリティ	媒体への人的アクセスを制限する鍵付き檻など	情報漏えい防止
その他の付帯サービスなど	媒体輸送業者	機密保持契約など	
	産業廃棄物処理業者	機器／媒体の廃棄および再資源化方式 廃棄証明書発行や機密保持契約など	
	媒体復旧	水没した光ディスクマガジンの復旧サービス (分解→洗浄→乾燥→組立→データ復旧)	データ保証 (災害時)

維持体制	ソフトウェア開発 コミュニティ運営	階層制御機能などの開発者コミュニティの維持	長期利用
	ハードウェアの ブロック化 (サイジング/設計)	増設などの基本単位の設計と更新 ● ブロック化されたハードウェアの追加でスケールアウト ● ハードウェアの更新に合わせた最適化(サイジング)を継続 ● 基本性能情報の提示(容量、時間当たりの転送量など)	拡張性 (容易な増設) 導入の容易性

一般社団法人 産業競争力懇談会（COCN）

〒100-0011 東京都千代田区内幸町 2－2－1

日本プレスセンタービル 4 階

Tel : 03-5510-6931 Fax : 03-5510-6932

E-mail : jimukyoku@cocn.jp

URL : <http://www.cocn.jp/>

事務局長 中塚隆雄