

SS研 「オブジェクトストレージ座談会」 次世代大規模ストレージの要件 ～データ活用促進の視点で～

2021年3月23日

富士通株式会社 インフラストラクチャシステム事業本部

データシステム事業部 SDSソリューション部

エキスパート 荒木純隆

目次

- オブジェクトストレージの利用シーン
- 大量データの長期保存に関するユーザー調査から見た課題
- 富士通の製品情報
- 参考文献他

オブジェクトストレージの利用シーン

- データの収集、蓄積、処理、分析・活用の各フェーズで利用される
- Active ArchiveからCold Archive(長期保管)
- バックアップストレージ

■ オブジェクトストレージの市場状況

- クラウドサービス 各社にお任せします
- オンプレミス製品 40%以上のCAGR(調査会社の情報)

- 国内の市場規模は海外に比べると非常に小さい
 - データ活用が進んでいない 特に二次利用がなされない(蓄積しない)
inB (特定のコミュニティ内)でのデータ利用が中心
 - データマネージメントソフトウェアの市場調査でも欧米で80%以上でAPACは20%未満だがそのほとんどは中国

国内のデータ蓄積・活用の取組み状況

産

- 大量データを保有することの価値を見出せていない(活用できていない)
 - ・ データはオンライン利用のみ(一次利用)
 - ・ データは捨てられている → 新たな「知」の発見はない
 - ・ データ活用による収入がなければ蓄積に投資されない
- 外部からはデータをもらいたいが、自社のデータは出したくない

学

- エビデンス保存、オープンサイエンティスに向けたデータ公開を進めている
 - ・ 人文系などではデジタル化されていない資料が大量にあり、研究者の退官などにより資料の散逸が起きている(財産であるはずの「知」の保存がなされていない)

官

- 一部はオープンデータとして提供
 - ・ 省庁・自治体ごと対応がバラバラ、提供形態がPDFで活用できないケースが散見
 - ・ 公開することで責任問題になることを避けたい
 - ・ 自治体では今後財政問題の顕在化が懸念 → デジタルデータの消失
- そもそもデジタル化が遅れている

アーカイブとして想定される利用シーン

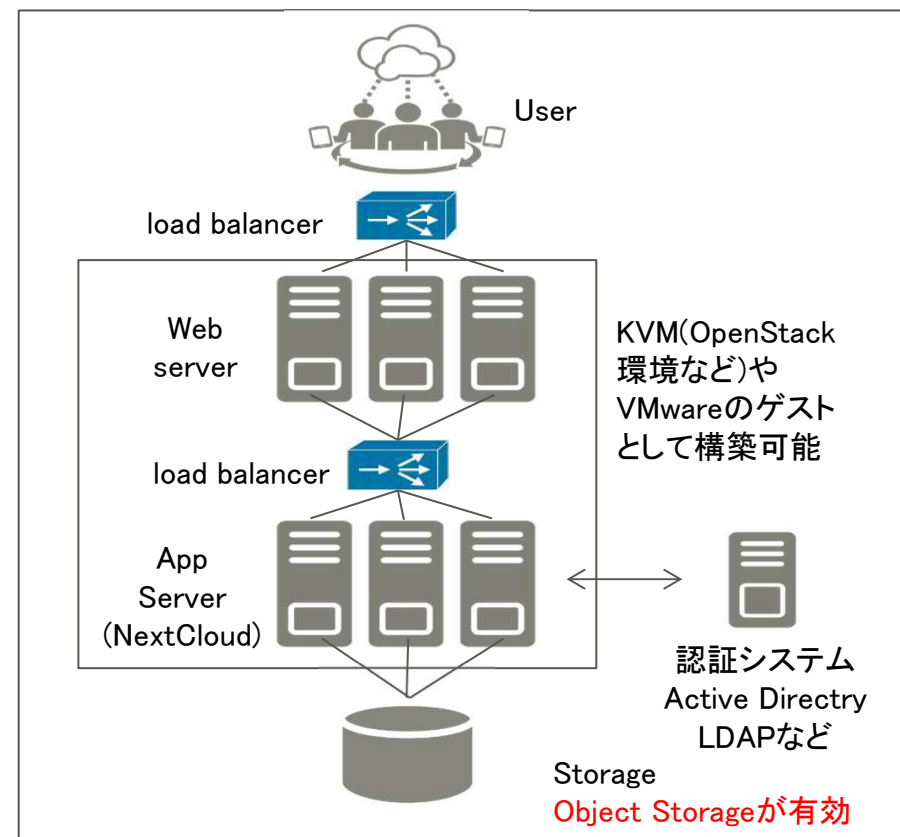


ドキュメント管理		イメージ保管		ヘルスケア	
業種	製造、建設	業種	金融(銀行、保険)	業種	医療、臨床検査
データ種	図面、文書、CAD	データ種	紙文書(伝票、契約書、本人確認書)	データ種	医療画像、ゲノム
AI(Deep Learning)		IoT/Big Data/Analytics		コンピュータフォレンジックス	
業種	全業種	業種	流通、製造	業種	全業種
データ種	学習データ(画像、音声、テキスト)	データ種	取引ログ、センサー、各種アクセスログ	データ種	メール、各種ログ、パケット
映像/画像		研究データ		デジタルアーカイブス	
業種	メディア、報道、CG製作、監視カメラ(全業種)	業種	大学、研究機関、製薬、気象	業種	国/自治体(博物館、図書館、美術館、公文書)
データ種	映像/画像	データ種	実験/観測データ	データ種	文書、画像

NextCloudによるドキュメント管理/共有システム



- 大学・研究機関での利用実績多 (旧ownCloudを含め)
 - Dropbox Likeなサービスをオンプレでも実現
- コンテンツの格納ストレージとしてオブジェクトストレージを利用
- 富士通サーバ ISV/IHV技術情報
 - FUJITSU Storage ETERNUS CD10000を使用したオンプレミス型オンラインストレージサーバソフトウェア「ownCloud」の性能検証結果報告(2015年11月)
 - <https://www.fujitsu.com/jp/products/computing/servers/partner/rov/owncloud/owncloud-1.html>

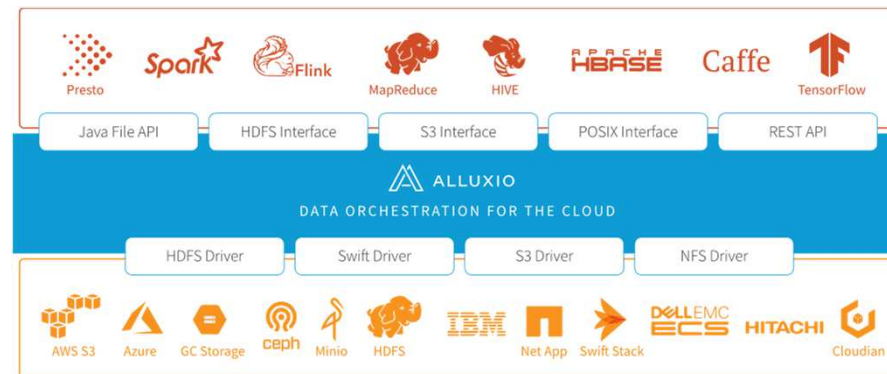


メモリからオブジェクトストレージまでの階層化

■ Hadoop系アプリケーションおよびAIプラットフォームから利用可能

- アプリからの透過的なアクセスを可能に
 - ・ 特定のアプリケーションでは、性能影響をほとんど受けずにコスト削減が可能
- DRAM/Persistent Memory(PM)からHDFS/Object Storageまで

<https://www.alluxio.io/>



大量データの長期保存に関する ユーザー調査から見た課題

- 産業競争力懇談会 2019年度 推進テーマ
「データ連携を支える次世代データ蓄積基盤」 最終報告書
<http://www.cocn.jp/report/bbe468c86d71f75acf4af70e3d8406fbd54a1215.pdf>

目標：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することでデータ駆動社会を支え、SDGs（特にCO₂削減）にも貢献しながら、データ連携の加速と拡大を図ることを目指す。

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

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	データ安全性	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	データ安全性	コスト	性能
企業の会計情報など	法に基づく	時間確保可	データ安全性	コスト	
医療・手術映像	20年	時間確保可	データ安全性	保存環境の継続性	秘匿性
研究データ	10年	時間確保可	データ安全性	保存環境の継続性	コスト
製薬	法に基づく	時間確保可	データ安全性		
地球・宇宙・自然現象観測データ	永年	時間確保可	データ安全性	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	データ安全性	性能
国防、国家機密	永年	時間確保可	秘匿性	データ安全性	
ゲノム・生体データ	永年	時間確保可	秘匿性	データ安全性	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	データ安全性

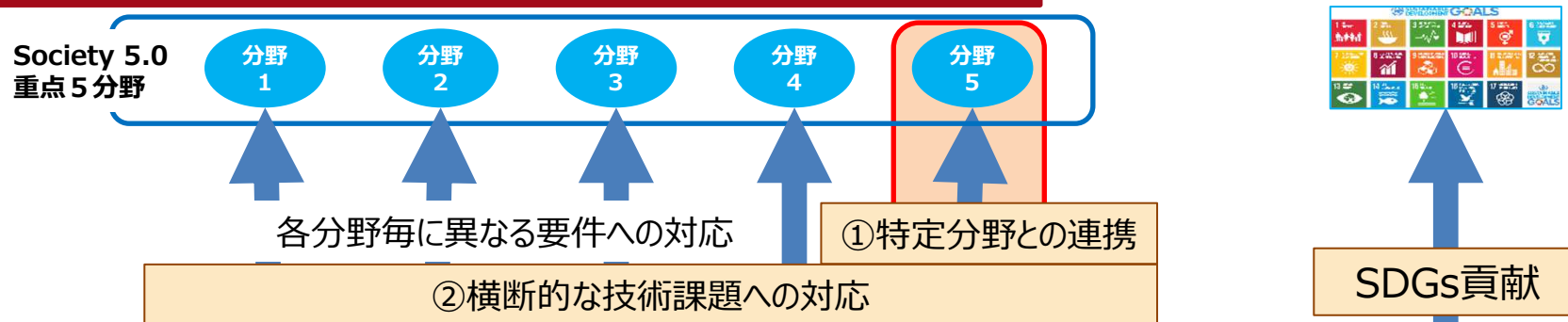
調査から見えた2大課題



ユーザーニーズ	課題と解決手段に関する考察	対応
データ 保全性	データ保全手法の明示と普及啓発(システム化)	階層型ストレージ内の複数種のメディアへの自動コピー(バックアップ) ・改ざん防止(WORMやオブジェクトストレージの機能でも可能) ・データ復元機能 定期的なメディア上のデータチェック
	①” 3 Copies, 2 Different Media Types, 1 Copy Offsite” などのデータ保全手法	
	②機器の故障や災害の想定によるデータ保全手法	
	③ランサムウェアなどへの対応	
保存 環境の 継続性	④データのヘルスチェック	長期保存の観点で、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 ①機器および記録メディアのマイグレーション(物理的老朽化対応) ②データフォーマットやビューアーなどの継続性(論理的老朽化対応) 必要に応じてフォーマット変換 ・OAIS(The Reference Model for an Open Archival Information System) ISO 14721:2012 https://www.iso.org/standard/57284.html ・FITS https://fits.gsfc.nasa.gov/
	①機器および記録メディアのマイグレーション(物理的老朽化対応)	
	②データフォーマットやビューアーなどの継続性(論理的老朽化対応) 必要に応じてフォーマット変換 ・OAIS(The Reference Model for an Open Archival Information System) ISO 14721:2012 https://www.iso.org/standard/57284.html ・FITS https://fits.gsfc.nasa.gov/	

プロジェクトの出口

活動の場：協議会の機能と活動概要



次世代データ蓄積・連携協議会

協議会の機能



階層型ストレージを用いたデータ蓄積基盤構築を推進し、データ連携の加速と拡大を支えることで産業競争力の強化を図る。

協議会の主な活動

- ① 特定分野の活動と連携し、具体的な課題を検証。成果を他分野へ横展開する。
- ② 横断的な技術課題やガイドライン化などに継続的に取り組む。

- データ保全・長期保存に関するガイドラインの作成
 - 既存のガイドラインやJIS/ISOなどの内容整理
 - ユーザ視点でのデータ保全・長期保存の手法について分かり易い記述
 - 階層型ストレージの活用の記述
 - 記録メディアごとの留意事項
- 階層型ストレージの基本設計
 - 機能要件まとめ
 - システムの機能ブロック
- 適用領域拡大

富士通の製品情報

分散ファイルシステム型ストレージ

SUSE Enterprise Storage

- オブジェクトストア基盤である「Ceph」を全面採用
- ハードウェア、ソフトウェアの組合せで提供
 - 富士通専門技術者によるハードウェア・ソフトウェア保守サポート

投資効果

最小構成で導入しても必要な時に拡張ができるので**初期費用を抑制**

柔軟な拡張

オンラインでストレージノード単位の増設、容量と性能を**リニアに拡張**

自動調整

内部でデータ配置のバランスを**自動調整**し、ホットスポット発生を防止

高可用性

複数ノード障害でも安全にデータを保護できる**高い堅牢性**

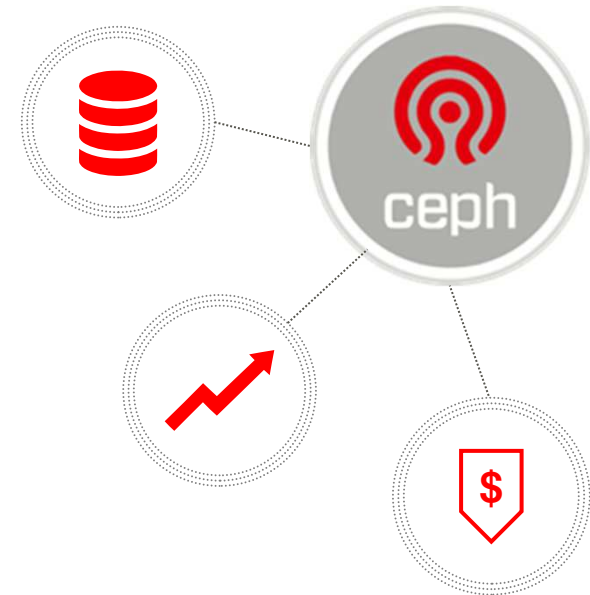
新陳代謝

新・旧ノードを混在させることで**無停止のデータ移行**を実現

主な機能

■ ストレージ機能 (Cephが提供する機能)

- **データ保護** (レプリケーション、イレージャーコーディング)
- **自動データ再配置** (セルフヒーリング)
- **マルチアクセス** (オブジェクト、ファイル、ブロック)
- **RESTful API** (Amazon S3 / OpenStack Swift互換API)
- **OpenStack連携** (Cinder, Glance, Swift, Nova)
- **マルチテナンシー** (ユーザー統計、クォータ、クラスタ分割)
- **ジオレプリケーション** (マルチサイト)



■ 付加機能 (富士通が提供する機能)

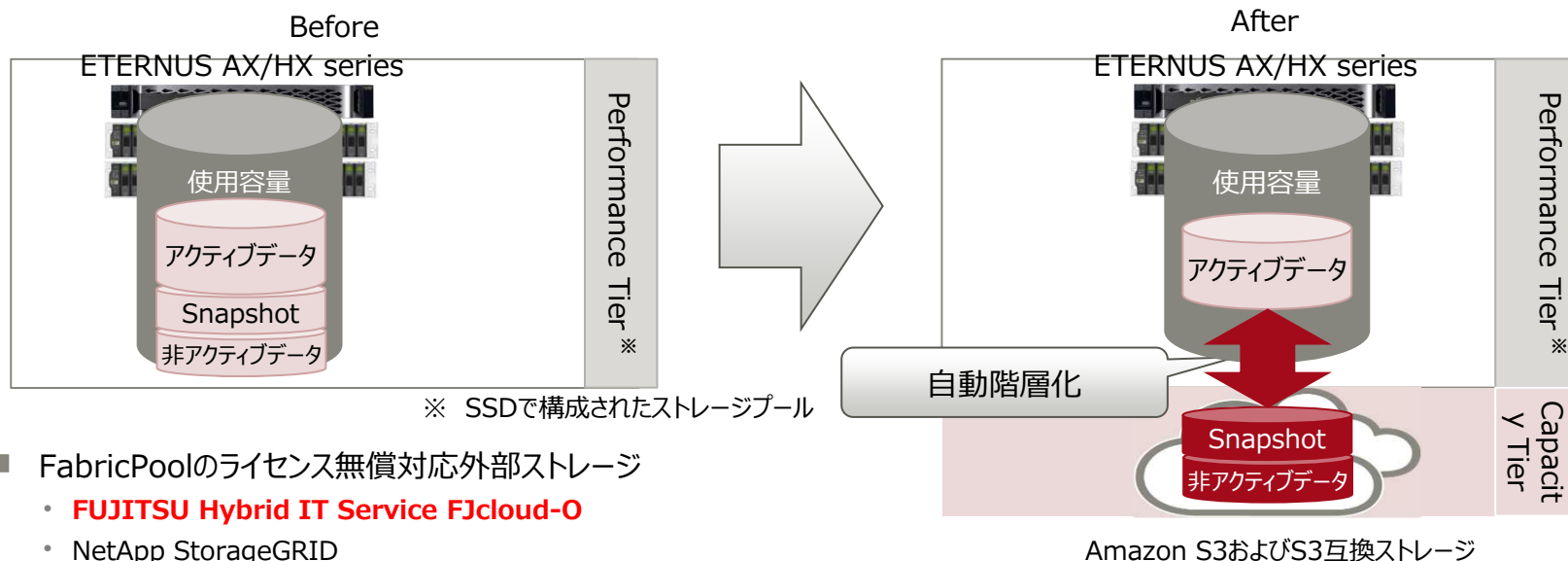
- **リモート通報** (障害予兆自動検知と異常情報のメール通報)

ETERNUS AX/HX series FabricPool



■ ETERNUS AX/HXと外部ストレージを連携、1つのデータプールに見せ、さらにアクセス頻度に応じてデータブロックを自動で移動する自動階層化に対応

- アクセスの少ないデータをコストの低いストレージ領域へ階層化することにより、データ保管コストを削減
- データブロックがクラウド上に移動しても、ユーザーやアプリケーションの見え方は移動前の状態と変更なし
- 重複排除や圧縮といった容量削減効果は保持されるため、クラウドの使用容量も低減



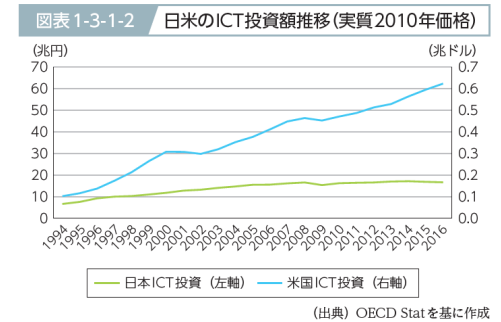
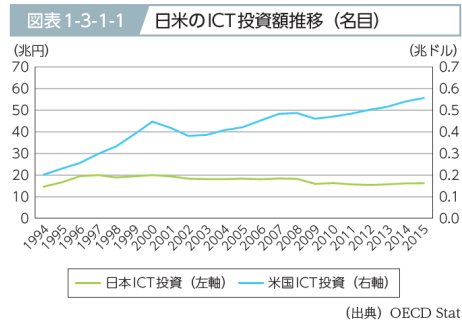
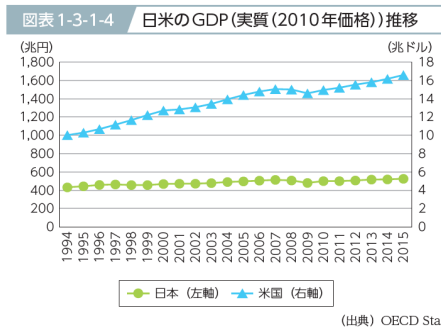
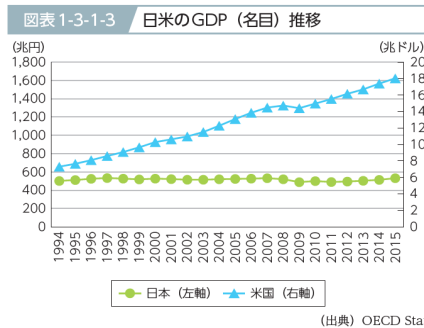
- FabricPoolのライセンス無償対応外部ストレージ
 - **FUJITSU Hybrid IT Service FJcloud-O**
 - NetApp StorageGRID

参考情報

日米のGDPとICT投資額の比較

- 米国のGDPが伸びていることに対し日本は横ばい
- GDP同様、ICT投資額も類似の傾向
 - GDPを増やさなければ、IT投資も増えない

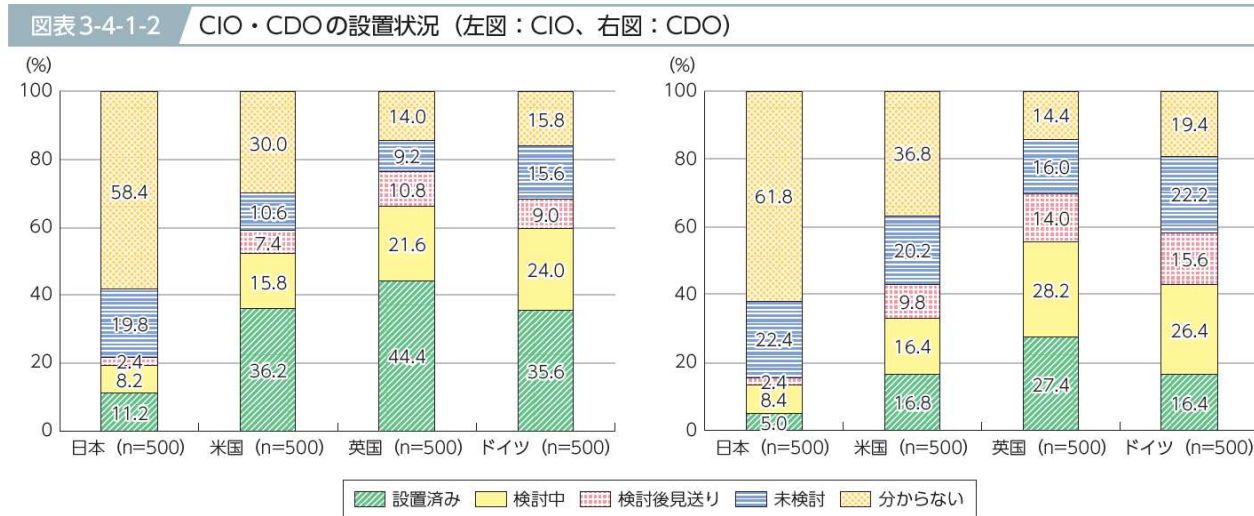
出典:平成30年 情報通信白書



未知のデータ(知)との出会いの場を作り、本当の意味での
データによる産業革命(イノベーション)が必要！！

各国のCIO・CDOの設置状況

- 日本は人的・組織的なICT・データに関するガバナンスが弱い
 - 国内ではCDOやデータスチュワードなどを配置している組織は少ない
 - データマネジメントに関する、理論的・実践的理解が不足している
 - データ利活用の素地ができていない



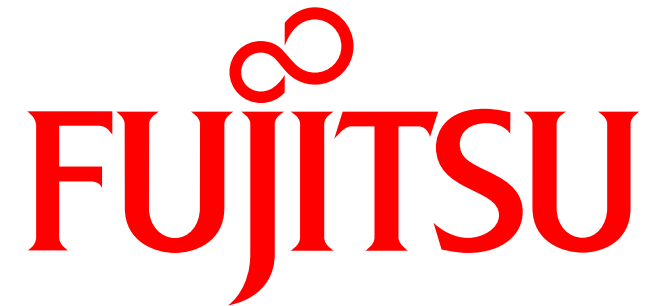
(出典) 総務省「ICTによるイノベーションと新たなエコノミー形成に関する調査研究」(平成30年)

参考文献

- SNIA : Storage Security Data Protection Technical White Paper March 2018 (原文)
<https://www.snia.org/education/whitepapers/security-wp>
- SNIA : ストレージセキュリティ:データ保護 テクニカル・ホワイトペーパー 2018年3月(邦訳)
https://www.snia-j.org/cmm/images/2019/06/Data-Protection_j.pdf
- JIS Z 6017:2013 電子化文書の長期保存方法
https://webdesk.jsa.or.jp/books/W11M0090/index/?bunsyo_id=JIS+Z+6017%3A2013
- 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
https://webdesk.jsa.or.jp/books/W11M0090/index/?bunsyo_id=JIS+Z+6019%3A2018
- 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>
「長期保存用光ディスクを用いたアーカイブガイドライン」
- 内閣府 「デジタルアーカイブのための長期保存ガイドライン」2020
https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/pdf/guideline2020.pdf
- 大学ICT推進協議会 (AXIES) 「学術機関における研究データ管理に関する提言」
<https://axies.jp/files/report/publications/proposal/rdm-recommendation.pdf>

関連団体

- デジタルアーカイブ学会(JSDA)
 - <http://digitalarchivejapan.org/>
- デジタルアーカイブ推進コンソーシアム(DAPCON)
 - <https://dapcon.jp/>
- 産業競争力懇談会(COCON)
 - <http://www.cocn.jp/>
- 次世代データ蓄積連携協議会
 - IDEMA Japan(日本HDD協会) <https://www.idema.gr.jp/>
 - <https://www.idema.gr.jp/committee/cat7/>
- データマネジメント協会(DAMA) 日本支部
 - <https://www.dama-japan.org/>
 - DMBOK2



shaping tomorrow with you