

京都大学におけるRDM基盤の 構築・運用に関する課題

京都大学 情報環境機構 IT企画室

渥美 紀寿

京都大学



これまでに情報環境機構で提供した 研究用ストレージサービス

- KUMail ストレージサービス (2014/04 –)
 - 学内外でのファイル共有サービス
 - メール認証、パスワード認証によるファイル受け渡し
- ownCloud, Nextcloud (2015/05 –)
 - オンプレクラウドストレージサービス
 - 2017/12 から Nextcloud に移行
- 研究データ 10年保存システムβ (2016/03 – 2017/12)
 - オンプレVM 上の独自Webシステム
 - サービス終了後、保存されているデータは OWCC に移行
- 研究データ保存サービス (2017/11 – 2021/06)
 - 文書管理システム+ 光ディスクストレージ
- Google Workplace (2019/01 –)
 - 教職員グループウェアとして導入
 - Google Drive によるファイル管理
- オブジェクトストレージ (2022/01 –)
 - Amazon S3 + Amazon S3 Glacier Deep Archive



メール添付の削減を目的
2014/01 に試行サービス開始



10年保存用のアーカイブを考慮した
その前段のシステムとして試行導入



10年保存のためにメタデータを付与可能なプ
ロトタイプシステム



データの作成から共有、アーカイブまでのワ
ークフローを備えたシステム



事務用データの保存目的で導入したが
研究用データにも用いられている



様々なシステムから連携して利用可能な
アーカイブシステムとして導入

KUMailストレージサービス (2014/04 -)

オンプレクラウドストレージ (2015/05 –)

- ownCloud (2015/05 – 2017/12) と Nextcloud (2017/12 –)
 - Nextcloud は ownCloud から派生した OSS
- Dropbox と同様のサービス
- PCやスマホ用アプリがあり, 同期可能
- 共同編集可能
- アクセス権の設定が可能
 - 編集, 読取, 再共有など
- 他のユーザとの共有
 - URL, メールアドレスによる共有
 - ユーザ指定による共有
- グループ作成が可能
 - 共有時に対象となるユーザの指定に利用可能

- 10年保存用のアーカイブを考慮したその前段のシステムとして試行導入
- 用途は研究用に限っていない

ストレージ容量: 5.5TB (使用量: 4.0TB)
利用者数: 46,092名 (2022/07/31 時点)
(退職者・卒業生: 20,466名, 学生: 9,282 名を含む)

オンプレの VM 上で稼動しており, VM 環境で使用可能なストレージの都合上, それほど容量を増加することはできない

研究データ10年保存システムβ (2016/03 – 2017/12)

10年保存のためにメタデータを付与可能なプロトタイプシステム

- オンプレ VM 上に独自開発した Web システム
- ファイル1つ (2GB まで) に対して下記メタデータを記録可能
 - タイトル、著者、種別、年月、書誌情報、DOI、メモ
- 全学メールアドレスを用いた共有が可能
- 10GB/年の制限
- メタデータ1件についてファイル1つ
 - ZIP 等で圧縮して登録

研究データ保存サービス (2017/11 – 2021/06)

データの作成から共有、アーカイブまでのワークフローを備えたシステム

- 文書管理システム (OWCC: Oracle WebCenter Content) と光ディスクストレージ (Fujitsu DA700) の構成

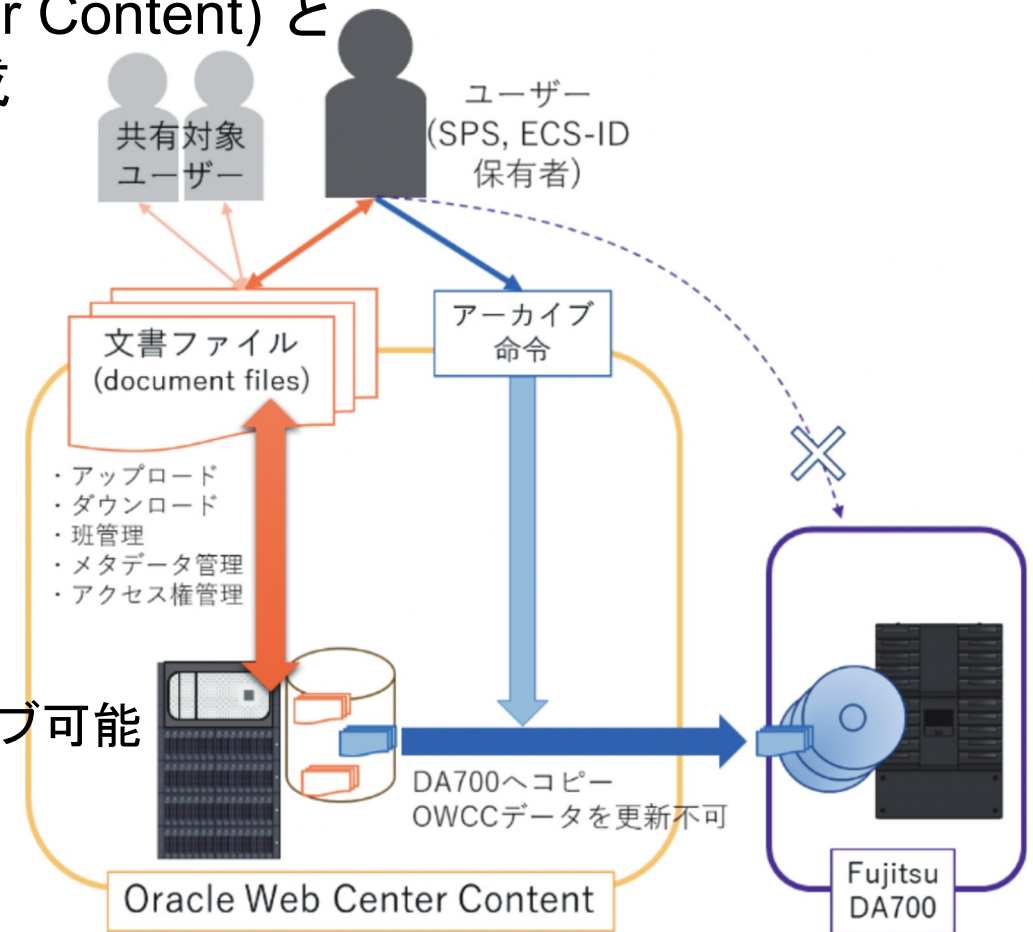
- OWCC

- ツリー構造でフォルダと同様に扱える
- ファイルごとにメタデータ付与
- ユーザを指定した共有, URL による共有が可能
- アクセス権の設定が可能
- バージョン管理可能

- Fujitsu DA700

- OWCC 上でフォルダを指定してダークアーカイブ可能

- 使い勝手が悪かった
 - サービス開始の时期的に部局で検討中でタイミングが悪かった
 - 将来に渡って保証できるサービスでなかった
- などのことが原因であまり利用されなかった



Google Workspace (2019/01 –)

事務用データの保存目的で導入したが
研究用データにも用いられて

- 教職員向けグループウェアとして導入
- 事務関連のファイル管理を想定
- 研究用データの保存場所としても利用されていた
 - 約200名が 50GB/人 以上利用
- 2022/07 以降 100TB に制限される

オブジェクトストレージ (2022/01 -)

様々なシステムから連携して利用可能な
アーカイブシステムとして導入

- アカデミッククラウドシステム (ARCS) で導入
- Amazon S3 + Amazon S3 Glacier Deep Archive
- データ転送量により課金されるが、ARCS では一定量を定額で調達
- 現在は試行期間
1 バケット 容量100GB, 転送量20GiB/月を無料で提供
- 利用方法
 - Amazon S3 に対応したクライアント (WinSCP, CyberDuck など)
 - Nextcloud, GakuNinRDM などの拡張ストレージ

ストレージの利用用途

- 普段の研究活動における研究データ
 - 個人情報を含むなど機密データ (漏洩対策)
 - 他機関のユーザとのデータ共有
 - スパコン等での利用
 - PC との同期
 - 他機関への異動時のデータの移行
- 論文に関連する研究データの10年保存
 - 機密データ (漏洩対策)
 - 長期保存
- オープンサイエンスのための研究データ公開
 - どれくらい参照されているか

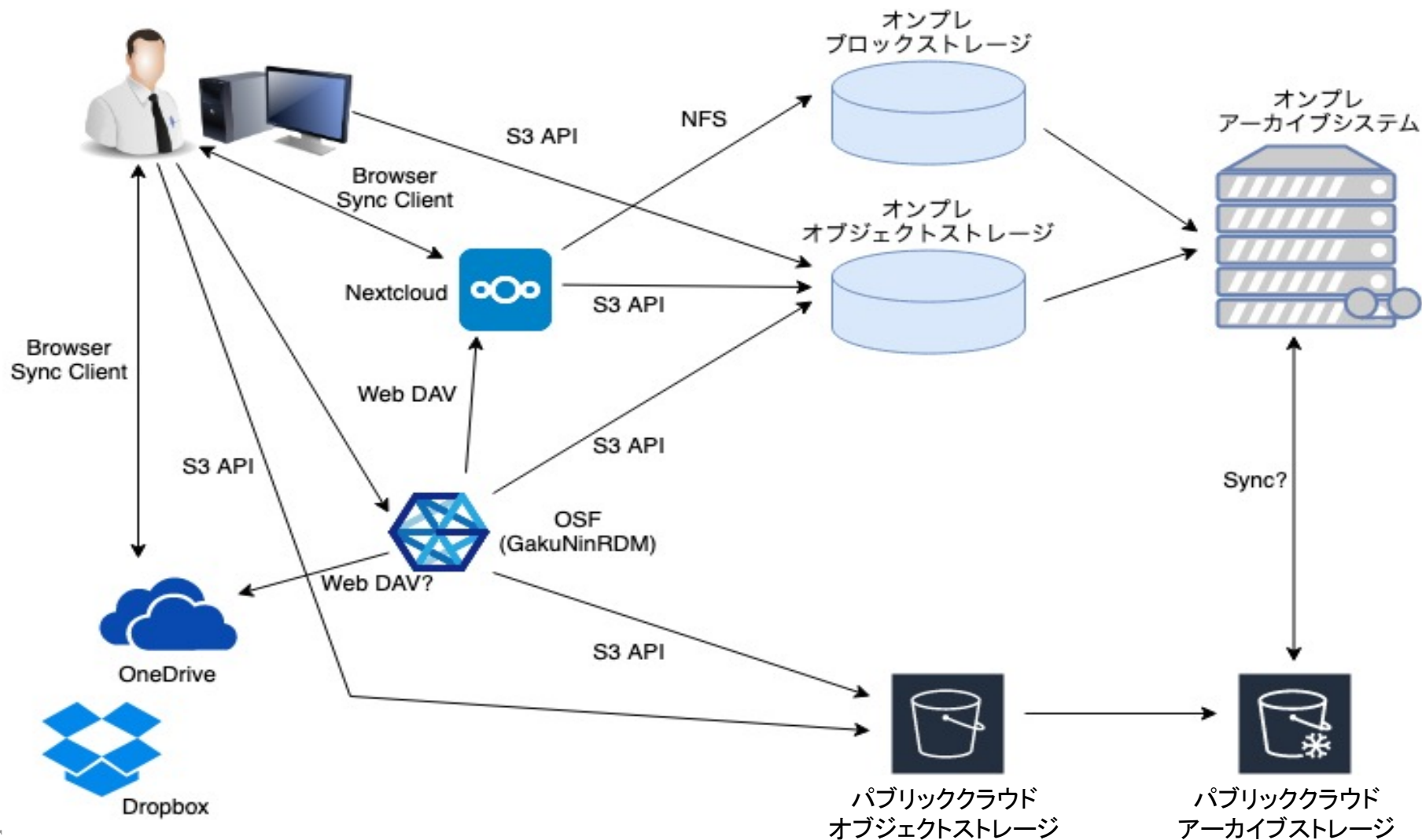
ユーザからの要望

- セキュリティの確保
 - 漏洩のリスク
 - 漏洩した場合の情報収集のしやすさ
(どのデータがどこから流出したか)
 - 個人情報を含むなど機密データの扱い
 - 保存時の注意等
- 10年以上安定して動作するシステム
(容量的にも)
- 多様な環境での利用
 - サーバ等へのマウント等
- ストレージ容量
 - 1ファイル 数百GB
 - 総容量 1TB /人くらいは使いたい
- 他組織への異動後のアクセス
 - 全データの export => NAS or Dropbox
等他のストレージ基盤へ
- 共有の仕方
 - 複雑だとセキュリティ面で不安になる
 - 機密データを含む場合の警告表示等

運用担当からの要望

- ・ 少人数で運用可能
- ・ ユーザ管理のしやすさ
 - ・ 新規ユーザ/離籍者の処理
 - ・ ユーザごとの容量制限
 - ・ 提供バケット, 提供フォルダごとの使用量/通信量の管理
- ・ グループ管理
- ・ AD 連携
 - ・ 他研究機関、企業ユーザの扱い

検討中のストレージ環境



現状の課題

- ・ ユーザの扱うデータ、研究環境の多様性への配慮
- ・ ストレージ基盤の長期間維持
 - ・ データのバックアップ、障害時の迅速な対応
- ・ 継続的な予算確保、費用軽減
 - ・ 増え続けるデータ量、転送量への対応
- ・ 基盤運用の人材確保

多様な環境を考慮しつつ、
低予算、少数での継続的な運用が可能な
ストレージ基盤が必要