

エンタープライズコンテンツ管理システムと 光ディスクストレージによるデジタルデータ長期保存の検討

青木 学聡^{1,*}, 梶田 将司¹, 赤坂 浩一², 武田 鋼²

1) 京都大学 情報環境機構

2) 京都大学 企画・情報部

*) aoki.takaaki.6v@kyoto-u.ac.jp

Design of Long-term Data Preservation System using Enterprise Content Management and Optical Disc Archive Storage

Takaaki Aoki¹, Shoji Kajita¹, Hirokazu Akasaka², Hagane Takeda²

1) Institute of Information Infrastructure and Communication, Kyoto University

2) Planning and Information Management Department, Kyoto University

概要: 大学における教育, 研究, 事務管理業務の多くの部分が電子的に行われている現在, デジタルデータを安全に管理することは重要な課題である. 本稿では, データの長期保存に重点を置き, エンタープライズコンテンツ管理システムと光ディスクアーカイブによる長期保存システムの構築について検討する.

1. 背景

現在, 大学における教育, 研究, 事務管理業務に関する情報伝達のほとんどは, 電子メール, ファイル共有, Web システムとデータベース等のように, デジタルデータにより行われている. デジタルデータを適切に管理することは, 効率的, 安全な各業務の遂行の為には必須の事項である. デジタルデータの管理には, 情報セキュリティの維持と情報フローの管理による効率化の視点がある. さらに情報セキュリティにおいては, 機密性に限らず, 可用性や完全性を適切に維持することも重要である.

電子情報を一元的に集約し, 適切なセキュリティモデルの下これらの利活用を図るシステムは, エンタープライズコンテンツ管理 (Enterprise Content Management, ECM) システムと呼ばれる. ECM に一般的に備わる主な機能は以下の通りである.

- 個人, グループロール等によるアクセス権管理
- コンテンツに対するメタ情報の追加

- コンテンツ本文, メタ情報に対する強力な検索機能
- コンテンツの版管理
- ワークフローに沿ったドキュメントの作成, 公開, 保管プロセスの管理.

これらの機能は, 先に述べた情報セキュリティの要件のうち, 機密性については一定の要求水準が達成される. しかし, 法律等で定められるデータの長期保存[1]に要求される可用性, 完全性については, これらを ECM のみの機能で満たすことは以下の理由より困難である.

- 元来, ECM システム並びにこれが使用する HDD そのものは長期間のデータ保存を保証しない. システムの冗長化, バックアップといった副次的な手法でデータを保護することになる.
- 要求される保存期間に比べ, ECM システムの寿命が短い為, システムの更新, あるいはシステムの変更毎にデータ移行作業, そして移行内容の正当性の確認という労力が発生する.

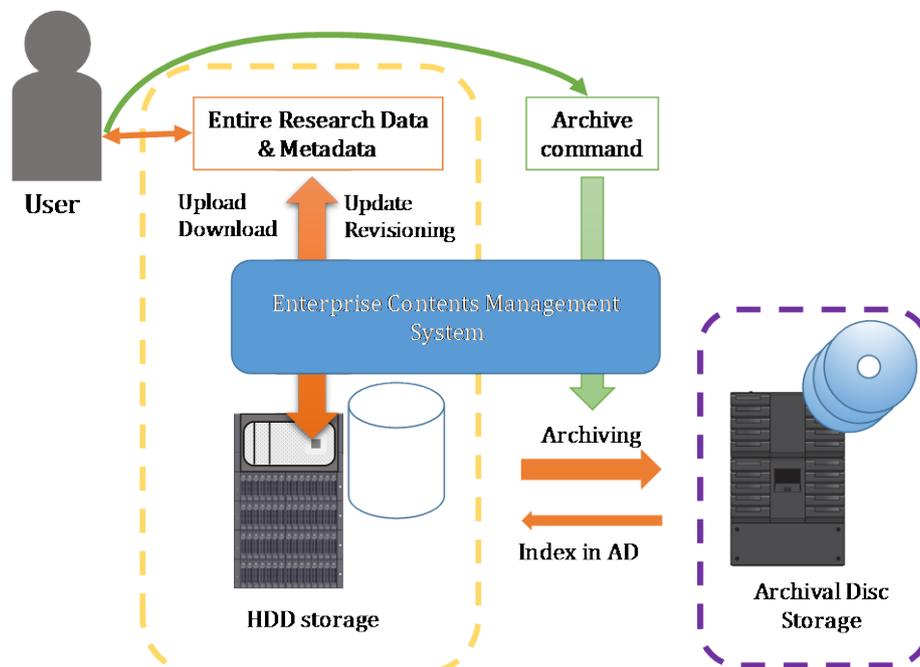


図 1: ECM システムと光ディスクストレージによるアーカイブシステムの模式図

すなわち、データの長期保存については、長期間保存が保証されたメディア上に、ECM システムとは独立した古典的なファイルシステムを構成し、汎用的なデータフォーマットでもってデータを保存する事が望ましいと考える。

京都大学では、2016 年度 12 月に更改される汎用コンピュータシステムにおいて、ECM システムと光ディスクアーカイブシステムが導入されることとなった。そこでこれらを連携させるデータアーカイビングシステムについて検討を始めた。

2. データアーカイブシステムの検討

検討中のアーカイブシステム模式図を図 1 に示す。今回、ECM として Oracle Web Center Content (OWCC)、光ディスクストレージシステムとして、FUJITSU Storage ETERNUS DA700 を導入する。DA700 の光ディスクは Blue-ray の上位企画である、Archival Disc を内蔵する。これは 1 枚当たりの容量が 300Gbyte の Write Once Read Many (WORM) メディアである。

OWCCは先に述べたように、アップロードし

たコンテンツに対し、アクセスコントロールによる他者との共有、版管理、メタデータの記載といった操作が標準で実施できる。このOWCC上に「アーカイブストレージへのコピー」コマンド実施できるよう UI を追加する。このコマンドはコンテンツ管理者が任意のタイミングで実行できるようにする。「アーカイブストレージへのコピー」の際に実施される手続きは以下の通りである。

1. アーカイブストレージの書き込みは、書き込むデータをアーカイブストレージマネージャーにコピーした後実施される。したがって、アーカイビング作業は対象となるコンテンツ ID をキューに登録し、バッチ的に処理することとなる。キューに登録状態のアーカイブ処理は、別管理画面より取り消しが可能とする。
2. バッチプログラムは登録されたコンテンツを順次アーカイブストレージに書き込む、コンテンツがフォルダ構造を持つ場合、配下のコンテンツもアーカイブ対象とする。コンテンツと共に付随するメタデータ、アクセス権情報等を JSON 等構

造化されたテキスト情報として合わせて保存する。

3. アーカイブ元の OWCC 上のコンテンツには、アーカイブストレージへの参照番号を記載する。参照番号は蓄積することができ、簡易的ではあるが、アーカイブストレージ上においても版管理を可能とする。またコンテンツの状態を「アーカイブストレージコピー済み」とし、書き込み禁止属性を付与する。これにより、OWCC 上コンテンツとアーカイブストレージとの内容は一致した状態を保つ。もし利用者がデータを更新したい場合、書き込み禁止属性を解除する手順が必要にある。このことで、頻繁にアーカイブを作成する事への抑止効果を見込んでいる。
4. 前項で説明した通り、各コンテンツの最新のアーカイブ内容は OWCC に反映されているが、それより古いアーカイブ内容は一般ユーザーは参照できない。このことから当該アーカイブストレージの内容はダークアーカイブ的な性質を持つ。もし利用者が過去の版のアーカイブデータの閲覧を求めた場合、管理者は OWCC に記録された参照番号を元に、データを提供する。この際、利用希望者が適切なアクセス権を保有していることを確認することが非常に重要となる。目視に頼らないならば、アーカイブコンテンツ作成時に記録されたアクセス権リストを自動的に適用するファイル共有システムを構築する、といった対応が必要である。

3. まとめ

現在、2016年12月のシステム更新に向け、上記の仕様を実装した光ディスクストレージへのアーカイブシステムを構築中である。これにより各種法人文書、そして特に最近取りざたされるようになった、研究データ[2]の長期保存に

利用されるものと考えられる。また OWCC が元来エンタープライズコンテンツ管理システムとして提供する、情報の集約機能、適切な共有・共同作業環境を最大限活用することで、京都大学における情報セキュリティおよび情報利活用の向上が進むことを期待したい。

参考文献

- [1] 京都大学における法人文書の管理に関する規程
http://www.kyoto-u.ac.jp/uni_int/kitei/reiki_honbun/w002RG00000080.html
- [2] 京都大学のける研究データ保存について
http://www.kyoto-u.ac.jp/ja/research/ethic/research_guide/documents/201602_data_preservation.pdf