

情報処理の概念

#11 デジタル世界の出版とアーカイブ

Yutaka Yasuda

出版の歴史

- 出版とは何か？
 - 著作物を世に出すこと
- 出版＝著作物の大量複製
 - グーテンベルグ
 - 15世紀なかば、活版による活字印刷

出版の歴史

- 手書きの書物一冊でも出版にはなるが
 - 複製：より多くの読者に伝えるための手段
 - 写経をはじめとした聖書複製の歴史を見よ
 - 死海写本 (B.C.200頃から)
- 世に出す = 大量に出す
 - デジタル世界で複製の概念は一変
 - 容易な複製は容易な出版に直結する (はず)
- 電子出版への流れ

電子出版はどこから始まったか

- 1981 Xerox Star / 1984 Macintosh

WYSIWYG (What You See Is What You Get)

見たままが得られるという操作性

- 1986 LaserWriter

Adobe PostScript の発明

DTP (Desk Top Publishing) の誕生

ワーノックとジョブスの出会い

DTP : DeskTop Publishing

- DTP のポイント

成果物は紙

紙への出力の過程を電子化したもの

- 電子出版過程の一部という視点

成果物が紙ではなくデータ（電子化された情報）であることが究極の電子出版ではないだろうか？

- ではPDFはいかに？

PDF : Portable Document Format

- Adobe Portable Document Format

交換可能な、という意味

文書を作成したアプリケーションに関わりなく、本来のフォント、レイアウト、カラー、グラフィックスをすべて保持できるフォーマット

- 印刷イメージを電子化

ブラウザで見る

紙に印刷する

- これが電子出版だろうか？

Web による情報発信

- 誰もが出版（情報発信）することが可能に
 - 低い参入障壁
 - 資金・設備
 - 技術力（blog などツールの広まり）
- Web出版という「スタイル」の特徴
 - 即時公開
 - 散在（出版社などによる集中管理がない）

Web による情報発信

- Web出版「物」の特徴

 - 断片的（まとめて書ける人は少ない）

 - 流動的（固定されない）

 - 信頼性（査読、保証が無い）

 - 紙への出力を目的としない

- 新しいかたちの情報発信の普及

 - 既存のメディアとの関係は？

意見を下さい

情報が多くのところから大量に発信されるようになったとき、新聞に残される価値は何でしょう？
(新聞社に残される価値はどこでしょう？)

サーチエンジン

- はじめは人間が読み、登録し、並べたリンク集

1994, Yahoo! (from Stanford University)
デビッド・ファイロ、ジェリー・ヤン

- ロボット型の登場

自動的にネットを巡回

プログラムが Web ページを読み、登録

1998, Google
ラリー・ペイジ、セルゲイ・ブリン

機械可読であるということ

- HTMLの機械可読性が活きている

一次情報はまず機械が読むという感覚

- 大量の情報発信がもたらすもの

溢れる情報

人間が振り回されるのはおかしい

今後はニュースなどもまず機械が読む

裏書き、編集の重要性

(新聞社は何を売っているのか?)

出版と蓄積

- 従来：出版と蓄積は分業されていた

出版社：publishing / 図書館：archiving

- 納本制度

国立国会図書館法(S.23) 制定

文化財の蓄積及びその利用に資するため、発行者は、発行の日から30日以内に、最良版の完全なもの1部を国立国会図書館に納入（義務）

- 出版側と蓄積側の連携がなされてきた

Web の蓄積

- 電子出版では出版と蓄積は同義である
 - 出版（公開）と同時に自動的に一次情報として利用され、二次情報が生まれる
 - そこで一次情報にアクセスできなかったら？
- ネットワーク上の出版物が納本制度の対象になっていない
 - 出版側と蓄積側の連携が失われている
- Web の弱点
 - Web は動的（内容が変化する）で、消えることもある
- どのような蓄積の可能性があるか？

WARP : Web Archiving Project

- 国立国会図書館 <http://www.ndl.go.jp/>

「更新・削除されていく」 Web の内容を保存

選定・許諾依頼によって選択的に保存

「インターネットのアーカイブは貴重な情報資源」

Wayback Machine

- Internet Archive <http://www.archive.org/>
- サンフランシスコに本部：非営利団体
- 85 billion pages (June, 2007)

その他の電子アーカイブ

- 過去の著作物から積極的に電子化
- 著作権法の期限外のものから

グーテンベルグ計画

青空文庫

British Library

HUMI Project

グーテンベルグ計画

- <http://promo.net/pg/>
<http://www.gutenberg.org/>
- イリノイ・ベネディクティン大学マイケル・ハートが推進
- 1971年開始
- 2007現在20,000超
- テキスト主体

青空文庫

- <http://www.aozora.gr.jp/>
- 利用に對価を求めない、インターネット電子図書館
- 著作権の切れたもの、自由に出せるものを対象
- テキストとHTML、エキスパンドブックで提供
- 1997年スタート
- ボランティアで入力、校閲
- 6272 本収録（2007.5 現在）

The screenshot shows the Aozora Bunko website in a browser window. The address bar shows <http://www.aozora.gr.jp/>. The page features a navigation menu with links for 'メイン', 'お知らせ', '掲示板', '別館', '資料', '運営', and 'リンク'. A search bar is present with the text 'www.aozora.gr.jp 内を Google で検索' and a '検索' button. A prominent notice states '著作権の保護期間延長に反対します' (Opposition to copyright protection period extension). The main heading reads 'インターネットの電子図書館、青空文庫へようこそ。' (Welcome to the Internet Electronic Library, Aozora Bunko). Below this, it says '初めての方はまず「青空文庫早わかり」をご覧ください。' (For first-time users, please see 'Aozora Bunko Early Understanding'). A blue banner announces '2007年7月7日「青空文庫10歳」記念パーティ開催!' (2007 July 7 'Aozora Bunko 10th Anniversary' commemorative party!). The 'メインエリア' (Main Area) contains several sections: '青空文庫早わかり' (Aozora Bunko Early Understanding), '総合インデックス' (General Index), '公開中 作家別' (Currently Published by Author), '公開中 作品別' (Currently Published by Title), '作業中' (In Progress), and '青空文庫 分野別リスト' (Aozora Bunko List by Field). The '公開中 作家別' section lists authors by their first initials: あ行, か行, さ行, た行, な行, は行, ま行, や行, ら行, わ行, 他. The '公開中 作品別' section lists titles by their first characters: あ, か, さ, た, な, は, ま, や, ら, わ, い, き, し, ち, に, ひ, み, り, を, う, く, す, つ, ぬ, ぶ, む, ゆ, る, ん, え, け, せ, て, ね, へ, め, れ, お, こ, そ, と, の, ほ, も, よ, る, 他. At the bottom, a blue banner contains the text '知る 問く 学ぶに 歓びを' (Joy in knowing, asking, and learning).

HUMI project

- 慶應義塾大学
HUmanities Media Interface
 - 人文科学研究における新しい情報伝達手段の導入
- 1996 開始
- 稀覯書（歴史的価値のある古い書物）のデジタル化
 - グーテンベルク聖書（1455年頃、ドイツ・マインツ）

電子出版の流れ

- DTPからWebへ

出版とは広めることが目的

電子化(digital化)は出版の可能性を大きく変える

- 電子文書は今まさにさまざまな形で立ち上がっている

PDF / Web / 電子図書館 / 電子書籍 / 青空文庫

答えはひとつではない

- 蓄積を忘れない

電子出版では出版と蓄積は同義

動的な形態の情報発信こそ蓄積が重要

Webがもたらしたもの

- 無制限な多くの作家の登場

情報発信形態の自由度

散在する断片的なドキュメント

- 機械可読であることの可能性

サーチエンジンが何を可能にしてくれたか

一次情報として使えるデータを提供するという感覚

- コンピュータが人間の知的な活動を支援できるという事実