

デジタルデータ長期安定保存に迫る危機

(独) 科学技術振興機構 (JST)

河村誠一郎、鈴木慶、永野智己、馬場寿夫、江連三香、佐藤勝昭

JST 研究開発戦略センター (CRDS) では、昨年来、長期にわたって安定にデータを保存できるメモリについて、ニーズ側、シーズ側、それぞれの有識者の協力を得て調査分析を進めてきた。このほど、これらを踏まえて戦略提言「デジタルデータの長期安定保存のための新規メモリ・システムの開発」をまとめ、関係省庁や関係業界などに呼びかけている。この小文では、この提言を出すに至った背景、具体的な研究開発の課題と推進方法、および、新規メモリ・システムの社会経済的効果について紹介する。

1. デジタルデータの爆発的増加にストレージが追いつかない

2012年に全世界で作成・複製されるデータ量は2.8 ZB(ゼタ 10^{21} バイト)、2020年には40 ZBに増加すると言われているが、図1に示すように、2007年時点で、すでに製造されるストレージ機器の容量を上回っているのである。

これにともなって、長期にわたって保存しなければならない価値あるデータも増大しているが、データ保存にはコストがかかるため、多くのデータが消失しているのが現実である。

2. 「ザ・デジタル・ジレンマ」—ハリウッドが突きつけたデジタルコンテンツ保存の問題

現在では、古い書物、美術品、映像、動画のほとんどがデジタル化されているが、現存するストレージ媒体で、100年を超える長期にわたって安定にデータを保存できるのだろうか？この問題に最も早く気づき、警鐘を鳴らしたのが、ハリウッドであった。ハリウッドの映画芸術科学アカデミーの科学技術評議会は2007年「ザ・デジタル・ジレンマ」という報告書を刊行した。この報告書によれば、ほとんどの映画がデジタル制作され、デジタル配信されるようになったにも関わらず、フィルムと同等の長い寿命特性をもつデジタルアーカイブ・マスターフォーマットあるいは処理方法が存在しないと結論し、その上で「映画産業はデジタルアーカイブのテクノロジーとソリューションについて統一見解を出すようにすべき」と業界の真剣な取り組みを促した。

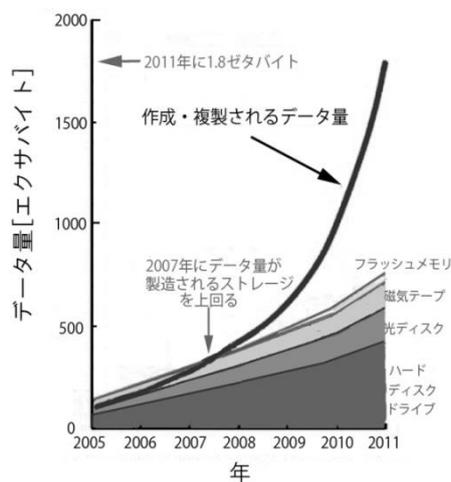


図1. 今日起きている情報爆発 (IDC 白書 2008 年版をもとに作成)

3. たった24年前のデータが読めない—火星探査機バイキングの教訓

NASA は、1975年火星探査機バイキングを火星に送った。デジタルデータは、磁気テープに記録され保管庫で保管されていた。その24年後、南カリフォルニア大学のJミラーがデータをチェックしようとしたところ、磁気テープに保存されたビット

データのフォーマットが失われてしまったため、データを復元できないという事件が起こった。デジタルデータの意味解読には、データの構造を記述するためのデータが必要であることを思い知らされた事例である。これをきっかけに、米国では開放型アーカイブ情報システム(OAIS)が提唱され、ISOの標準規格となっている。

4. ファイストスの円盤とロゼッタストーン of 教訓

ファイストスの円盤は、紀元前1600年以前に作られた厚さ2.1cm、直径16cmの粘土製の円盤で、裏表に「絵文字」が螺旋状に記されている。両面を通じて45種類の「文字」が使われ、全部で241の文字が使われている。20世紀の初頭に発見され、多くの方が解読を試みたが、同じ文字が記された粘土版が他に見つからないことなどのために未だに読み解くことができていない。「文字」自身は3000年もの長期保存に耐えたのに、「意味理解」の方法が伝わっていないため、情報が伝わらない例とされる。

これに対して、ロゼッタストーンは、紀元前196年にメンフィスで出された勅令が石に刻まれた碑文であるが、古代エジプト語に加え、現代でも使われているギリシア文字でも記述されていたおかげで完全に解読された。

これら2つの考古学的遺物の物語から読み取れることは、1000年の長期保存に耐えるメモリが技術的に出来たとしても、意味理解の方法まで伝承されないかぎり、無意味だということである。

5. 10年後にデジタルデータが読めるか？

デジタル情報持続性の危機は、決して海の向こうの話ではなく、身近に起きている。10年前にワープロ専用機で作ったフロッピーディスクの文書を読み出そうとしても、今やワープロ専用機などどこにも見当たらないし、フロッピーディスクを読めるパソコンさえ市販していない。以前録画したビデオテープも再生する装置がない。いま私たちは、デジカメやスマホで写真や動画を撮ってメモ리카ードに記録しているが、これから10年後にそのメモ리카ードが読める装置はあるのだろうか。

6. マイグレーションに膨大なコストとデータ喪失の危険性

ハードディスクのデータは、放っておくと10年ほどで消えてしまう。ほとんどの企業では、マイグレーションといって、サーバーのハードディスクを定期的に更新して、古いデータを新しいシステム上に移行する作業をしている。マイグレーションには膨大な時間と費用がかかる。デジタル化は導入時だけでなく、維持にも相応のコストがかかるということがほとんど認識されていない。

現在主流のマイグレーションが、今後いつまで持続可能であるかについて、技術を担ってきた産業界からでさえ疑問や不安の声が上がり始めている。マイグレーションは、コンピュータの世代が替わるたびに新しい技術体系のシステムに移行していくため、膨大なコストを負担し続けなければならない宿命にある。さらに、障害やエラー等が起きて一度移管が途切れてしまうと、そのデータは永久に失われてしまいかねない危険性もはらんでいる。

7. 国会図書館では長期保存はブルーレイで

国立国会図書館では、平成13年以来、所蔵の紙媒体の出版物のすべてをデジタル化する作業を行っており、平成21年度の補正予算で大きく進展したという。デジタル化資料の保存用画像は、一定の温湿度に保たれた書庫で保存しており、過去にはDVD、現在はブルーレイディスク(BD)に記録しており、総データ量は5PB(ペタ10¹⁵)

バイト)に達するという。光ディスクの耐久性に加え、再生機器や再生技術が長期間担保されるかが課題とされている。

8. NHK では LTO テープに記録して定期的にマイグレーション

NHK では、アナログビデオ時代の 1 インチテープ全部を 10 年もかけて LTO 規格のデジタルテープに変換した。これを長期保存するために、マイグレーションを前提に 6-7 年かけてファイルベースに変換しようとしているが、マイグレーションのコストは膨大。現在の技術では、高品質化に対応できないことや、メタデータを付与する手間など多くの問題を抱えている。

9. なぜここまで放置されてきたのか

以上のようにデジタルデータの長期保存には危機が迫っているにもかかわらず、これまで学术界は分析・認識してきたとは言えず、警鐘や助言・対策等を社会に対して十分に発信してこなかった。なぜだろうか。

今日まで半導体技術が急速な進歩と成長を遂げてきたのは、十分大きな市場が存在し世界中でその市場に大型の投資をしてきたことがある。一方、超長期メモリに関しては、投資からその便益を回収するまでに極めて長時間を要するという本質的な違いがある。現在のデジタル情報社会を支えている電子技術は、“今の仕事”を“処理”する技術として発達してきたが、将来にわたるデータ保存技術への対応は不十分、あるいは無いに等しかった。その背景には、デジタル技術が生まれてまだ半世紀程しか経っていないということもあるだろう。従って、デジタルデータは国や産業界が意志をもって残さなければ、確実に残すことが難しい。

10. 超長期保存メモリ・システムの研究開発課題

超長期保存メモリ・システムを実現するためには、図 2 に示すような材料、製造プロセス、デバイス、回路・設計、システム、情報系、商品ビジネス系といった技術階層毎の研究開発課題を、同時並行的に解決する必要がある。

(1) 材料系の開発課題

材料系に関する研究開発課題としては、長期間安定な 2 つの状態を持つメモリ材料の開発が必要である。使用する材料は長期的に安定で、かつ 2 値 (1 と 0) のデータ記憶に対応する二つの物理的・化学的な状態が超長期間に亘って、ほとんど変化しないような材料が求められる。

(2) デバイス・製造プロセスの開発課題

高密度・高集積化可能なデバイス・媒体技術の開発、高信頼性プロセス技術の開発、高耐久性パッケージ技術の開発などが課題として考えられる。大容量のデジタルデータを長期に保存するためには、超長期保存メモリデバイスとしては、これまでとは桁違いの高密度化・高集積化が必要であり、これを可能とする材料、メモリ構造、新原理メモリなどの研究開発を進めることが重要である。

(3) 回路・設計技術関連の開発課題

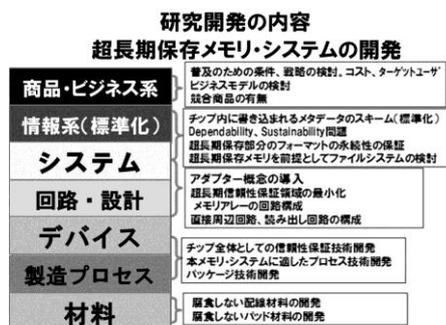


図 2 研究開発の内容

高速・低消費電力な書き込み／読み出し回路の開発、改ざんの防止技術の開発などが課題となる。大容量のデータを微小な領域に高速に書込んだり読み出したりする必要があり、安定した書き込み／読み出しを実現するためには、メモリセル間の干渉を抑制することや、高感度のセンス回路の開発が求められる。

(4) システム関連の開発課題—アダプタ開発の必要性

一般的なデータ保存メモリとの互換・共存技術の開発、オンラインアクセス可能な超長期保存メモリを前提としたファイルシステムの構築が重要な課題となる。

書き込み／読み出しデバイスを、長期的に安定して安価に供給するためには、超長期保存メモリアダプタと外部システムとのインターフェース、そして長期保障される一般的メモリの外部システムとのインターフェースの互換性、共通化が大事である。

(5) 情報系の開発課題—メタデータ・エミュレーション

保存したデータの読み出しと意味理解を保障するためには、超長期保存メモリに、どのようなデータがどのような形（フォーマット）で書き込まれているかなど、意味あるデータとして読み取るための情報（メタデータ）を事前に決めておく必要がある。

フォーマットが後年になっても通常的に使われているかどうかは不明なため、データの読み取りを永続的に保障するか、あるいはフォーマットを認識して、オリジナルのアプリケーションやOSの振る舞いを真似るエミュレーション機能（もとのソフトウェアからの命令を新たなプラットフォームで実行するための技術）により、データを復元して読み出せるようにしておかなければならない。

(6) 国際標準も視野に

超長期保存メモリ・システムの開発・導入にあたっては、まず、十分に考え抜いてシステム設計を進める必要がある。そうでないと、システムを作ってはみたが50年を経ずしてデータが消え去ってしまうということになりかねない。最初の段階から、ハードとソフトにまたがる十分に考え抜いたシステム設計を行い、国際標準を見据えた主要セクター間でのコンセンサス形成を行うことが肝要である。

11. 超長期の保存メモリ・システムの社会的・経済的効果

デジタルデータの長期安定保存が可能になったら、どのような変化が起こるのか。仮に超長期の保存メモリ・システムが実現すれば、データ保存の上で信頼性向上と保存コストの低減を同時にもたらすことになる。また、経済的な恩恵だけでなく、人類の歴史・文化資産、科学的知識の継承にも大きな恩恵を与えることになる。

(1) 科学・技術

宇宙・天文分野では、今後全世界で生成される天文データは年間200PB以上という推計があるが、天体や気象データには同じ状態が一度しかないという性質上、データの保存は重要で、それによって長期にわたるデータの解析が可能になる。同様に、地球科学分野でも先の東日本大震災で認識されたように、土壌や地質学、海洋科学のデータを千年単位で継承・蓄積することで、後世に貴重な情報を残すことができる。

(2) 文化・文明

文明の発展は情報の蓄積と活用の結果であり、これまでの長い歴史における人類のあくなき探究心が現在の文明社会を形成しているとも言える。将来にわたり、わが国が成立し続けるためには、日本の歴史を確実に残して未来へ継承していく必要がある。

膨大なデータが初めからデジタルの状態で作成される現在、社会の持続的発展を維持していくためには、データの長期保存は必須の課題となっている。

(3) 経済的効果

超長期保存メモリ・システムが広く普及するための価格を、現状のメモリ価格を考慮して16GB=400円程度とし、保存のための媒体市場とその1~2倍程度のアダプタ市場が存在すると仮定する。そうすると、2020年40ZBに達すると見込まれるデジタルデータのうち、長期保存の対象となるものが仮に1%とした場合、その保存には媒体で10兆円、アダプタを加えると20~30兆円の市場規模が想定される。

12. 推進には人文科学も含め幅広い分野のサポートが必要

現状では、デジタルデータの長期保存は、強いニーズを持つ国会図書館、放送業界、出版・印刷業界など個別に取り組みが行われているに過ぎない。しかし、長期保存のニーズは行政組織、企業、医療機関、個人レベルまで幅広く存在すると考えられ、100年、1000年という超長期に亘る保存は、企業の通常の存続期間を超えるため、初めは国が積極的に関与すべきである。

また超長期に亘ると、我々の文化自体が大きく変容している可能性もあり、将来に亘って意味理解まで含めた情報の保存を実現するためには、言語学、記号論、文化人類学など、人文社会系の研究者も含めた幅広い視点からの議論が欠かせない。研究開発の推進には、技術関係のみならず、社会・文化さらにビジネス関連など様々な分野の有識者を集めて研究会などを開催し、この分野のコミュニティを広げねばならない。

13. まとめ

映画や文化遺産などの重要なデジタルデータを、長期に亘って安全に保存し、遠い将来でも確実に読み出したいという強い要望があるが、現状では担保されておらず、火星探査機バイキングの例のようにデータ喪失の危機が迫りつつある。

これまで我々は、微細化技術の進展により保存媒体の大容量化を進め、マイグレーションによりデータを移し替えることで、デジタルデータの長期保存問題を先送りしてきたが、そろそろ正面から取り組まないと間に合わない。そういう時期がもう来ている。デジタルデータの超長期保存が過大なコストをかけずにできれば、人類の歴史・文化資産や科学的知識の継承に大きな恩恵をもたらすだけでなく、20~30兆円規模の大きな経済的効果も望めるであろう。

この小文は、JST 研究開発戦略センターが公表した戦略プロポーザル「デジタルデータの長期安定保存のための新規メモリ・システムの開発/12SP07」を要約したものである。<http://crds.jst.go.jp/singh/wp-content/uploads/12sp07.pdf> からダウンロードできる。

連絡責任者 佐藤勝昭

(独) 科学技術振興機構 研究開発戦略センター
〒102-0076 東京都千代田区五番町7 K's 五番町
E-mail: Katsuaki.sato@nifty.com

(2013年4月24日受理)