

企業内資料のデジタルアーカイブ化 技術革新に関する企業内資料の調査研究(2) Building digital image archive of the corporate documents A survey of corporate documents concerning technological innovation (2).

下川 雄一¹・山崎 幹泰²・中森 勉³・竺 覚暁⁴

SHIMOKAWA Yuuichi¹・YAMAZAKI Mikihiro²・NAKAMORI Tsutomu³・CHIKU Kakugyo⁴

¹金沢工業大学 博士(工学)・²金沢工業大学 博士(工学)・³金沢工業大学 工修・⁴金沢工業大学 工博

¹Kanazawa Institute of Technology, Dr. Eng・²Kanazawa Institute of Technology, Dr. Eng・³Kanazawa Institute of Technology, M. Eng
・⁴Kanazawa Institute of Technology, Dr. Eng

デジタルアーカイブ、工作機械、画像ファイル、データベース
Digital Archive, Machine Tool, Pictorial Data, Database

はじめに

本研究は、社会や文化のニーズと技術革新との相関関係を知るための基礎資料を蒐集、組織化し、知識基盤として公開することを目指している。ここでの基礎資料とは、製品カタログなどの販促宣伝資料、仕様書・規格・技報・設計図などの技術資料、企画書・開発記録などの社内資料、などを指している。我が国の基幹産業技術企業における、昭和戦後期のこうした資料を蒐集して、デジタル画像化して記録し、さらにこの画像資料に適切な資料組織法、検索システムを開発して、これらを画像データベースに組織し、研究用に公開することを研究の目的としている。平成17年度に構築した画像データベースシステムの概要については既に報告した¹⁾。現在は工

作機械分野の資料蒐集と並行して、入手できた資料のデジタルアーカイブ化(スキャン作業、PDF変換、データベース登録)を進めている段階である。10月末の時点で10社から資料を借用することができており、その中で株式会社池貝(以下、池貝)より借用した資料についてはほぼその作業が完了している。その内容については別の機会で報告することにし、本稿では池貝の資料を例にデジタル画像(PDFファイル)化に関する検討事項(表1)とその対応方法、および今後の課題について報告する。

当該企業資料に含まれる他社資料保存への対応

借用した資料の中では池貝の資料は量も多く、資料内容も多岐に渡る。その中で、他社の製品情報も掲載されている工作機械の製品総合カタログや完全に社外の資料など、他社に関する情報も幾つかの形で存在している。これらについては同時代の技術革新の傾向を表す資料として、全てデジタル画像化することとした。

重複する資料への対応

池貝の資料では、原本とそのコピー、もしくはコピーが複数存在するようなものも存在する。物理的な資料保存においてはコピーを予備として保管する価値は比較的高いと考えられるが、デジタルアーカイブ自体が1つの予備(更にそのバックアップデータも存在する)と捉えた場合、原本のデジタル画像があれば、コピーのデジタル画像の価値はそれほど高くないといえる。但し、池貝の資料では次のような状態の資料も確認されている。

- ・ コピーに手書きのメモ等が記入されている
- ・ 同一資料のコピーが複数存在するが、単独で存

表1 資料のデジタル画像化における検討事項

検討事項	理由
他社資料保存への対応	一頁(又は同一資料)に複数社の製品が含まれる製品カタログ等がある
重複する資料への対応	原本とそのコピーが両方存在する
微細な色情報への対応	入手した資料全てをできる限り原物に近い状態で記録する必要がある
資料形態の記録への対応	文字・画像情報を含まない白紙頁や資料裏面等もできる限り記録する必要がある
グループ化された資料への対応	複数の資料がより大きなレベルでまとまりを持っており、どのレベルで1つのPDFファイルにするか、判断が難しいものがある

在するもの、他の資料と合わせてファイリングされているもの等、保管状態が異なる

このように単なるコピー以上の価値があると判断される場合は、デジタル画像化の対象とした。

微細な色情報への対応

作業開始当初、データ容量を考慮し、モノクロの資料についてはグレースケール方式でデジタル画像化していた。しかし、モノクロのコピー資料に朱書きのメモが存在するようなケースが見つかったため、内容を問わず全ての資料をカラー画像としてデジタル化することとした。

資料形態の記録への対応

資料の中には現時点でそれほど古いとは言えないものも含まれているが、長期的な資料保存媒体としての価値を考慮し、且つデジタル化された画像だけを閲覧するユーザーが原本かコピーか判断するための情報として、表紙・背表紙・裏表紙は勿論、冊子の場合は白紙頁、写真やパネルの場合は資料の裏面の状態も記録することとした。

グループ化された資料への対応

複数の資料が群として存在している場合、どのレベルで1つのPDFファイルとするかは、属性情報の精度やデータベースの検索・参照の効率に大きな影響を及ぼす可能性がある。例えば、池貝の資料では下記のような状態の資料が存在していた。

定期発刊の資料が発刊順に製本（ハードカバー化）アーカイブ化されている。

海外向けの複数の製品パンフレットが一冊にファイリングされている。

様々な資料の原本やコピーが1つのパッケージ（2穴ファイル、封筒など）にまとめられており、そのパッケージに「歴史資料」といった群としての意味情報が明記されているものもある。

本研究では、このように明らかに資料間の連続性や関連性が読み取れるものについては個別の資料としてPDF化することにした（但し、製本表紙には歴史資料的重要性はないので記録しない）。これに対し、このようなケースについては判断が極めて難しい。パッケージ単位で1つのPDFとした場合、資料毎の区別がつきにくく、データベースの属性が不足してしまう。一方、個々の資料毎にPDF化した場合、資料間の関連性や順序、集合体として

の意味情報が読み取れない。両方の方法でそれぞれデータベース化することも考えられるが、利用者が混乱を招く恐れもある。現時点では、このような例を除いて、群としての資料の意味が強く感じられる場合は1つのPDFファイルとしているが、この問題については今後更なる研究の余地がある。

まとめと今後の課題

これまでの作業を通して、多様な資料形態に柔軟に対応できる資料組織法のヒントが得られた。例えば、前出のようなグループ化された資料の場合、個々の資料に関する画像情報とパッケージに関する画像情報がそれぞれ独立したデータベースレコードに対応させ、且つそれらのレコード間で関係性（パッケージと各資料の親子関係、各資料の順序関係など）を記録することが望ましい。その場合、データベースの検索・参照の仕組みにおいても、そのようなデータベース構造に対応した形への改善が求められる。

また、作業の継続にあたって以下2つの属性項目の追加が必要であることが分かってきた。

- ・「資料の形態」：原本、別刷り、コピー、青焼き等、資料の形態を明示する属性
- ・「企業コード」：同一企業から入手された資料であることを示すための属性

「企業コード」と関連する項目として、現在「所在情報」^{注1)}、「製作社(者)」^{注2)}、「管理・所有者情報」^{注3)}といった属性項目を用意しているが、これらは必ずしも企業名称と一致する訳ではない。他社資料保存への対応も含め、同一企業に関係する資料を一意にシンボライズする情報が必要である。

注

- 1) その資料がどこに保管されているか、物理的な場所を表す。
- 2) 製作社(者)の入力に際しては、企業名称の変更も考慮するようにしている。
- 3) 形式的な管理・所有者を特定する情報であり、物理的な保管場所を表すものではない。

参考文献

- 1) 山崎幹泰・下川雄一・中森勉・笠覚暁、技術革新に関する企業内資料の調査研究(1)、特定領域研究「日本の技術革新 - 経験蓄積と知識基盤化 - 」第2回フォーラム報告書、2006年9月