

# The struggle for KOSMOS Ⅲ Data Conversion. (データ移行奮闘記)—This is it!—

たなべ      みのる  
田邊      稔

(メディアセンター本部課長代理)

## 1 プロローグ

最初から“それ”が大変なことはわかっていた。200万件の書誌データ、500万件の所蔵データ、多種多様なデータソースとフォーマット、「負の遺産」と呼ばれる遡及データ群、揺れる仕様、パッケージシステム(以下、Aleph)のデータ構造や機能の理解不足、海外ベンダーとの英語での不慣れなコミュニケーションなど、眼前に立ちはだかる壁はいくつもあった。それだけでなく、データ移行とは神経を使うきめ細い作業である。どう考えたって簡単に行くはずがなかった。しかし、データ移行チーム<sup>註1</sup>のリーダーに私が任命された時、さほど大変さを理解していなかった。これまでデータ移行は数多くこなしてきたし、今回は大半の作業をパッケージベンダーがやってくれるものと信じていたからだ。後に、その予想が見事に打ち砕かれ、未知なる荒海に放り込まれることになろうとは、知る由もなかった。

データ移行チームが、データ・システム・人とどう向き合い、どう戦ったか?の魂の記録。“これ”が“それ”だ!(This is it!)

## 2 データを移行する意味

言うまでもなく、データとはシステム上のコンテンツであり、システム検証に不可欠な存在である。データとシステムはいわば「鍵」と「鍵穴」の関係にある。車にキーを差して回さないと動かないのと同様に、データを流し込まない限り新システムは動き出さない。同時に、データとはシステムを映す鏡でもある。データを見ればシステムの内部構造がわかると言っても過言ではない。KOSMOS IIのコアシステムである“CALIS”とそれを取り巻く分散システム群である“中から”、“KOHEI”、“ちょいす”、“RUI”等、どれもデータ構造がユニークで、良くも悪くもKOSMOS I時代からの歴史を引きずってきた。KOSMOS Iのがつりと階層化されたデータ群が、KOSMOS IIで緩やかに分散し、再度KOSMOS IIIの固定長の世界へと収斂されて行く様は圧巻である。可変長の緩いデータを、バリデートされた固

定長のデータに変換するのは至難の業である。データを何度もゴリゴリと「(篩<sup>ふるい</sup>)」にかけて、新システム仕様の型に合うように磨き上げて行く。いわば、“データのクレンジング”である。

また、データ移行とは過去(歴史)を知り、現状と向き合い、未来へとバトンをつなぐ作業でもある。その意味で、我々データ移行チームは、新旧システムをつなぐ「時代の架け橋」的役割を担ったといえる。

## 3 移行方針の決定と覚悟

ものごとを「選択する」ということは「決める」こと。「決める」とは「覚悟すること」。「覚悟すること」は「決めたことを遂行すべく、潔く前に進む」ことである。覚悟したら、あとは粛々とやるだけだ。Alephというパッケージを選択した時点で我々は覚悟をしなくてはならなかった。しかし、何をどう移行するか?など決めるべきことを決めていなかった。だから、覚悟もできず、なかなか前に進めなかった。

2009年1月末に日吉AVホールにて3日間かけて行われたEx Libris(以下、E社)とのデータ移行のキックオフミーティングは最悪だった。慶應側の移行方針がまったく決っておらず、人によって言うことがバラバラでE社からの質問にその都度フリーズし、まともに答えることができなかった。果ては、内輪揉めまで起こす始末。恥ずかしくて、情けなかった。E社から見たら、日本の大学のお粗末な姿は、いささか滑稽に見えたことだろう。先行き不安な船出となった上に、このミーティングの衝撃から立ち直るのに相当時間がかかった。

E社は、データ移行に限らず、自社の開発フレームの適用を強要してきた。自分たちのシナリオで顧客を型にはめ込み、強引にサービスイン(以下、STP: Switch To Product)へと持ち込む作戦だ。パッケージベンダーなら至極当然のこと。しかし、それを受け容れることは、慶應側に考える隙が与えられないこと(つまり、思考停止状態)を意味する。肝心なことは、E社のペースに吞まれずに、慶應のペースで早め早めに手を打つこと。そのためには、迅速な意

思決定と仕様の明確化が必要だった。

そこで、我々はリスク回避のために「大きな英断」をした。当初、旧→新コードのマッピングを含め、大半のデータ移行をE社に任せるつもりだった。しかし、日本のベンダーでも難しいデータ移行の細かい変換作業をE社に英語で説明し、理解・納得させ、プログラムを開発させるのは得策とは思えなかった。そのため、『E社に引き渡すデータフォーマットの種類をできる限り集約し、コード類をすべて慶應側でマッピングし、E社には右から左へデータをロードしてもらっただけにする』という大胆な方針転換を行った。結果的に、この作戦は我々を成功へと導いたが、膨大なコードマッピング作業には相当手を焼いたことは言うまでもない(図1)。

また、設計段階で工夫した点の1つに、共通仕様

の作成がある。データ移行仕様書について、当初E社製のデータ移行仕様書(Data Conversion Specification・図2)のみで済ませようとしていたが、これでは現場との仕様確認や移行結果の検証が困難と判断し、誰が見ても分かり易く、データの流れや手順が一望できる「データ共通仕様」(図3)を別途作成することにした。具体的には、1)データソースの項目と2)E社に引き渡す項目、そして3)ロード先のAlephテーブル項目の関連が一目でわかるようにした。

なお、技術的には、後戻りのない形でしっかり作ることを考えた。システム屋の間では一度しか使わないプログラムを「ゴミプロ」と呼ぶが、「たかがゴミプロ、されどゴミプロ」。その場限りの「やっつけ仕事」ではなく、STP後の運用に「つながる仕事」にしようと誓った。

CALIS	Type of Patron(CALIS)	Sublibrary 1	Status code1	CALIS	地区コード50だったら	CALIS	地区コード70だったら
1	学部	KEI50	01	1	サブライブラリ理工は02を移行	1	サブライブラリ信濃
2	学部(医学部)	KEI50	01	2	サブライブラリ理工は02を移行	2	サブライブラリ信濃
3	大学院前期博士	KEI50	05	3	サブライブラリ理工は07を移行	3	サブライブラリ信濃
4	大学院後期博士	KEI50	05	4	サブライブラリ理工は07を移行	4	サブライブラリ信濃
5	大学院博士一貫	KEI50	05	5	サブライブラリ理工は07を移行	5	サブライブラリ信濃
6	看護短大生	全地区移行しない		6		6	
7	通卒登録者	KEI50	01	7	サブライブラリ理工は02を移行	7	サブライブラリ信濃
8	夏スク・夜スク	KEI50	01	8	サブライブラリ理工は02を移行	8	サブライブラリ信濃
9	通スク	KEI50	01	9	サブライブラリ理工は02を移行	9	サブライブラリ信濃
10	通信一般	全地区移行しない		10		10	
11	国セ日本語別科	KEI50	01	11	サブライブラリ理工は02を移行	11	サブライブラリ信濃
12	国セ短期	KEI50	01	12	サブライブラリ理工は02を移行	12	サブライブラリ信濃
13	国セ日本教	KEI50	05	13	サブライブラリ理工は07を移行	13	サブライブラリ信濃
14	国セ帰国	KEI50	01	14	サブライブラリ理工は02を移行	14	サブライブラリ信濃
15	生徒(幼稚園)	全地区移行しない		15		15	
16	生徒(中学校)	全地区移行しない		16		16	
17	生徒(高校)	KEI50	移行しない、または99	17	サブライブラリ理工に32を移行	17	

図1. コードマッピング仕様～ PatronStatus サンプル

### 3.3 Oracle Z30 table - CALIS ITEMS

The table below is the complete specification of Aleph Z-file containing the items. The name of the Oracle table is « Z30 ».

Oracle field name	Type	Description	Source field
Z30-REC-KEY			
Z30-DOC-NUMBER	M 9(9)	DESC: System number of the administrative record (ADM) associated to the item. CONV: ADM system number.	Generated by Ex Libris
Z30-ITEM-SEQUENCE	M 9(6) Leading zeroes	DESC: Item sequence number, within one ADM number. CONV: Create consecutive numbers, starting with 000010, adding "1" for each new copy sequence. Begin again from 000010 for each new ADM system number.	Generated by Ex Libris
Z30-BARCODE	M X(30)	DESC: Unique identifier of the item. The Z30-BARCODE can be created automatically based on the definitions of the tab_checksum table of the library's tab directory. Currently two routines are available: - checksum_create_z30_by_seq: creates the barcode based on the last-item-barcode sequence of the Z52 record. - checksum_create_z30_by_doc_n	Unique across all items. Leading zeroes are meaningful. "BOOKID"

図2. Data Conversion Specification サンプル

CALIS所蔵情報M所蔵(CALS)		※引渡し項目は雑誌と同フォーマットとする	
CALISデータ項目	移行時引渡し項目	ALEPHテーブル項目	コンバート手順
所蔵レコードID	Note	Z30-NOTE-INTERNAL	NOTEの先頭に
登録日	移行しない	移行しない	
更新日	update_date	Z30-UPDATE-DATE	
BOOK ID	BookID	Z30-BARCODE	
登録番号	Registration number	Z30-INVENTORY-NUMBER	
書誌レコードID	Bibliographic ID in CALIS	移行しない(書誌で持つ)	複数の書誌がリンクしている場合は最初の
請求記号	Call no.	Z30-CALL-NO-1 Z30-CALL-NO-2	
地区コード	Sublibrary code	Z30-SUB-LIBRARY	z305-sub-library-mapping 確定090323 参照
配架場所コード	Collection	Z30-COLLECTION	z30-Collection-mapping090929 確定参照
受入日	Acquisition date	Z30-ARRIVAL-DATE	
受入区分			
管理区分	Budget information	Z601-Budget-Number Z30-PAGES	新旧予算コードMapping TableのCalis仕様によって、を割り当てる。

図3. データ共通仕様サンプル

#### 4 移行データと“これでもか”のスケジュール

移行対象データは、書誌関連はプリントのみとし、電子リソースは対象外とした<sup>注2</sup>。また、最終的に著者典拠データも加わった。所蔵データは基本全件対象とし、閲覧関連は利用者データ全件と現貸出データ・未収金データのみとし、過去の貸出履歴や予約データなどは移行しないこととした<sup>注3</sup>。詳しくは次の表(図4)を参照されたい。

移行スケジュールについては、当初E社との間で

データ種別	ソース	フォーマット	移行件数
メイン書誌	CALIS	BinaryMARC	1,783,856
中から書誌	中から	BinaryMARC	104,835
和装本・旧分類	Google	BinaryMARC	101,956
著者典拠	メイン 中から	ASF-MARC	571,795 645,104
所蔵	図書 雑誌	CALIS Text	3,597,389 1,658,100
雑誌所蔵範囲	CALIS	Text	77,118
利用者	CALIS	Text	61,103
貸出	CALIS	Text	37,824
未収金	CALIS	Text	297
図書発注	ちよいす	Text	3,837
雑誌発注	KOHEI	Text	16,017
雑誌書誌	KOHEI	Text	491
業者	RUI	Text	2,295
予算	KOHEI	Text	357

図4. 移行対象データと最終本番移行件数(差分除く)

は本番を含めて3回の移行を予定していたが、結果として、全8回もの移行を行う結果となった。理由は大きく3つある。1つは、慶應側の仕様の揺れとE社側の不具合対応遅延のため。次に、著者典拠の移行が加わったこと。そして、本番前に時間測定と手順確認のために、どうしてもシミュレーションを行いたかったことである。

また、運用上図書・雑誌の日々の受入をストップできないため、差分移行を行わざるを得なかった。差分移行は“追加のみ”とし、更新と削除は対応しないこととした。というのも、更新や削除まで欲張って差分移行してトラブった事例は枚挙に暇がないからだ。それに、追加だけであれば、慶應、E社とも通常の新規移行スクリプトが流用できる。

#### ★全8回の移行スケジュール

- 0次(2009年5~7月): サンプルデータ
- 1次(2009年8~10月): フルデータ
- 2次(2009年11~12月): フルデータ
- 3次(2010年1月): フルデータ+典拠
- 4次シミュレーション(2010年2月): フルデータ+典拠
- 5次シミュレーション差分(2010年2月): 差分データ
- 最終本番(2010年3月): フルデータ+典拠
- 最終本番差分(2010年4月): 差分データ+図書発注

0次~2次移行では、仕様の揺れがあり、手探りだった作業手順が、3次ぐらいから安定してきて、データFIXから提出まで1ヵ月ぐらいかかっていた

た作業が、約 2 週間で提出できるまでになった。

また、シミュレーションと最終移行では、E 社のデータコンバージョンセンターがあるドイツ・ハンブルクとの時差（東京-8 時間）を最大限利用できるような時間単位の綿密なスケジュールを作成した。

## 5 牙を剥いた所蔵データと 2GB の壁

上記データ種別の中で最も苦労したのは、巨大な所蔵データのハンドリングである。所蔵データには、Item Status（資料取扱区分）や Item Process Status（資料状態コード）、Enumeration（擬似巻次）などコードマッピングが必要な項目が多数ある。数が少なければ、手作業でも変換できるが、500 万件は尋常ではない。CALIS の所蔵形式から E 社に引き渡すデータを用意するのに約 10 日間を要した（ちなみに、書誌データ 200 万件は同じ MARC フォーマットでも約 3~4 日間）。

データ移行プログラムの開発は、生産性（開発効率）と他の開発者との互換性を考え、大半を MS-Access2007 で開発した。KOSMOS II 時代の資産（Access クエリやモジュールなどの部品群）を最大限活かし、開発時間を短縮したかったからだ。しかし、大きな落とし穴があった。現状 Access の DB（以下、MDB）では、保存できるファイルサイズは最大 2GB までに制限される。また、Access ではデータ更新を行う際に、更新前のトランザクションログを確保するが、すべて MDB 内部に保存されるため、サイズが膨れ上がる。500 万件の所蔵データを一気に更新すると、300MB 程度のファイルが軽く 2GB を超えてしまう。そのため、館別に所蔵データ処理用の MDB を分けてから処理を行うこととした。それでも、三田 1 館で 200 万件を超えるため、MDB の最適化（ファイル圧縮）を繰り返しながら変換を繰り返した。正直この作業は心身ともかなり堪えた。結局、本番移行では時間が足りないと判断し、一部 Perl で処理を行うことにした。

## 6 ぎりぎりまで追い込んだデータ整備

「できる限り、過去のデータ不備を引き継がず、データをきれいにしてから Aleph へ移行しよう！」という基本方針の元に、不備データを抽出し、一括修正できるものは本部システム担当で対応し、個別処理の必要なものは図書・雑誌テクニカル担当

に修正してもらった。また、配置場所一括変更や一括除籍など予め計画されているデータ整備については、2009 年 9 月末までで KOSMOS II 修正依頼を締め切った。

書誌一所蔵リンク切れ整備、合綴本所蔵データ分割、中からリンク整備、全除籍処理、KOHEI 書誌 ID 付与、書誌データ長修正などについては、期限ぎりぎりまで追い込みをかけた。しかし、どうしても 4 次のシミュレーションまでに間に合わないものが発生したため、「とにかく Aleph に全データを移行し、改めて整備計画を立てよう！」と腹を括った。

## 7 E 社の評価と信頼性

全 8 回のデータ移行を通じて E 社対応の総合評価は 65 点といったところだろうか。結果的に及第点は付けられるが、ロード後のインデックス生成を忘れたり、利用者名などの日本語が文字化けしていても気づかなかったり、MARC フォーマットのタグ 780/785 の雑誌変遷リンクの要求仕様をなかなか理解してくれなかつたり、GUI での結果確認を忘れたりとお粗末な面も多かった。中でも 1 次⇒2 次で移行品質が大きく下がったのは誤算だった。結局、2 次&3 次のバグをシミュレーションまで引きずったが、最後はきっちり決めてくれた。データ移行が終わってから聞いた話だが、今回 E 社側はドイツのデータコンバージョンセンターの K 氏がほぼ一人で対応したらしい。孤軍奮闘という点では共感を得たが、いくら移行のエキスパートといっても一人では無理がある。逆にミスを生むリスクも高い。最低でも二人一組でクロスチェックを行う必要がある。これは慶應側にも同じことが言えた。今後の課題としたい。

## 8 E&K 効果～愛すべき未来へ

ハンブルクとの時差もあって、データ移行作業を平日の深夜～朝方や休日に自宅でやるが多かった。睡魔に襲われて寝落ちしたり、とっ散らかってパニックを起こしたり、気がふれそうになる瞬間も多々あった。そんなとき、いつも自分を鼓舞してくれたのが、EXILE<sup>注4</sup>と KREVA<sup>注5</sup>の音楽だった。EXILE のアルバム『愛すべき未来へ』の透明感のあるバラードと KREVA のアルバム『心臓』のエッジの利いたナンバーを交互に iPod でガンガン鳴らし

ながら、孤独で泥臭い作業を乗り切った。正直、彼らの音楽がなければ深夜・休日作業を頑張り通すことはできなかつたろう。彼ら輝かしい“E&K”は、暗中模索する我ら“E社とKeio”を“愛すべき未来”へと導く一筋の光となった。

## 9 エピローグ

兎にも角にも、無事にデータ移行が完了し、STPを迎えられてほっとした。成功したのは、言うまでもなく、データ移行チーム全メンバーと仕様確定や結果検証に関わってくれた全スタッフの努力の賜物に他ならないが、個人的に成功の要因を上げるとすれば、それはリーダーの「大なる勘違い」と「根拠のない自信」かも知れない。“これが自分の最後の仕事(いわば集大成)”という気概と“俺はプロだ!”, “俺がやらなきゃ誰がやる!”という自己暗示。それに「絶対うまくいく!」というあてのない自信。それらが、自分を駆り立て、心折れることなくメンバーと共に最後まで走り抜けることができた。

最後に、これだけは自信を持って言えることがある。『このメンバーだから乗り越えられた!』と。各メンバーがプロ意識を持ち、心を“鬼”にして臨んだからこそ成功に漕ぎ着けたのだと確信している。

## 謝辞

CALISデータ抽出やデータ整備など多大なる協力をいただいた有限会社アユソフト殿、複雑な著者典拠データ作成を快く引き受けてくれた京セラ丸善システムインテグレーション株式会社殿、ちよいす図書発注データをご提供いただいた丸善株式会社殿に深く感謝したい。また、度を越した在宅作業に不満はありながらも傍で支えてくれた家族に、この場を借りて心から感謝したい。

## 注

- 1) データ移行チームは、次期システム開発プロジェクト室の各モジュール担当4名(目録1,受入1,雑誌1,閲覧1)とシステム担当2名(内1名はリーダー田邊)の6名で構成。
- 2) 電子リソースは、情報探索ツールのKOSMOS(Primo)へ直接ロードすることに決定したため。
- 3) 予約については、データFIX後にCALISから貸出中/確保中/移動中の予約リストを作成し、データ移行後~STP直前に閲覧担当者が手でAlephにデータを移植した。
- 4) EXILE(エグザイル)は、日本の音楽(J-POP)とダンスパフォーマンスの融合を目指す14人組のヴォーカル&ダンス・ユニット。
- 5) KREVA(クレバ、本名: 畠山貴志)は、日本の男性ヒップホップMC、トラックメーカー、作詞家、DJ。慶應義塾大学環境情報学部(SFC)卒。

