

Ⅱ-9 データベース構築におけるデータ収集作業について

新田 裕樹¹
Hiroki Nitta

各務 誠一²
Seiichi Kakumu

満田 広司³
Koji Mituta

【抄録】発足から、現在に至るまでにCORINS（工事实績情報サービス）、TECRIS（測量調査設計業務実績情報サービス）は、数回のソフト及びハードのバージョンアップを重ねてきた。そのことにより、現在では非常に柔軟なデータベースとして整備されている。そのために収集するデータ内容の見直しや、精度の高いデータを収集することを目的として、データ入力時及び登録時のデータチェックなども多数取り入れてきた。データチェックの方法や具体例を挙げるとともに、今後の収集作業も含め、CORINS/TECRISのデータベース構築におけるデータ収集作業を紹介する。

【キーワード】CORINS（Construction Records Information Service）、TECRIS（Technical Consulting Records Information Service）、工事情報、業務情報、入力システム、データベース、情報整備、情報活用

はじめに

財団法人日本建設情報総合センター（JACIC：ジャック）では、昭和60年11月15日設立以来、情報システムの研究開発、建設情報の提供、情報化の啓発・普及等を積極的に推進し、建設分野における建設マネジメント技術の向上を図る等、建設分野の情報化に関する諸事業を展開してきた。

これら各種事業の中から、CORINS（工事实績情報サービス）とTECRIS（測量調査設計業務実績情報サービス）について、データベース構築における情報収集の観点から紹介する。

1. データ収集の方法

CORINS及びTECRISのデータ収集の方法は、ほぼ同じである。国土交通省を始めとする全国の公共発注機関並びに公益民間企業（以下、「発注機関」という。）から、CORINSやTECRISに登録を義務付けられた工事や業務を受注した企業は、登録するための実績データをJACIC独自のデータ作成支援ソフト（以下、「入力システム」という。）を使用して作成する。出来上がったデータをJACICへオンラインによるデータ転送、もしくはフロッピーディスクに格納し郵送等により送付する。JACICでは、専用の登録受付

システムにて登録時にデータチェックを行った上で、データベースに登録する。その結果、全国の発注機関や企業等が活用できる情報が提供されることになる。

（図1）

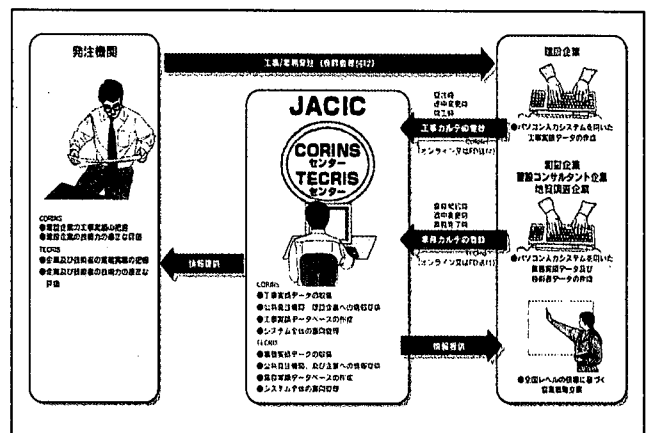


図1. 実績データ収集の流れ

2. 収集情報見直しとシステムのバージョンアップ

次に入力システム及び登録受付システムの変遷について紹介する。

CORINSは平成6年度、TECRISは平成7年度よりサービスを開始した。サービス開始当初から現在に至るまで、データ収集に際しては、一貫して企業にJACIC独自の入力システムを販売もしくは無償配布することで提供してきた。これにより、統一的な形式によるある

1 株式会社 土木情報サービス 03-5114-3191

2 財団法人 日本建設情報総合センター 03-3505-0452

3 財団法人 日本建設情報総合センター 03-3505-0452

程度の精度が保障された実績データの収集が可能となった。

しかし、サービス開始後、実際に運営していく中で、収集情報の見直しを行う必要が出てきた。これは、サービス開始前の検討段階において想定していた情報の活用方法が少しずつ変化し、実態と合わなくなったことや省庁再編等の社会的な変化により、実際の利用者（情報提供先）である発注機関や企業側から収集情報の見直し要望が挙がるようになってきたためである。そのため、JACIC では収集情報について利用者にアンケートを取るなど、収集内容や利用方法について確認を行った結果、これまでに何度かの見直しを行った。

【CORINS の見直し事例】

登録対象工事の拡大

→過去に 2 回実施。1 回目は工事の請負金額 5,000 万円以上から 2,500 万円以上に拡大（平成 9 年度）。2 回目は工事の請負金額 2,500 万円以上から 500 万円以上に拡大（平成 14 年度）。

登録時期の追加

→工事竣工時のみの登録から、工事受注時点や施行途中時点でも登録を行う（施工中工事の把握、監理技術者等の専任配置状況確認のため）

収集項目の追加

→「発注機関情報」の充実（発注機関組織の最新情報を調査し反映させることで、より詳細な情報提供を目的とする）
→「発注機関担当者名」を追加した（企業側で作成した実績データについて、発注機関側の確認者を明記させることにより、精度の低いデータ、不正なデータの登録を防ぐため）
→「交通規制」「近接工事」の有無等を追加した（工事難易度を評価するため）
→数値項目の単位の統一と桁数見直し（入力ミスの軽減。「千円」「千m」「千m²」等を「円」「m」「m²」等に改め、桁数を 3 桁増やした）

【TECRIS の見直し事例】

登録対象業務の拡大

→調査設計業務、地質調査業務に加え、測量業務を登録対象とした（平成 8 年度）。

収集項目の追加

→業務キーワードの充実（新たに要望の多かった 7 千語を加え、1 万 7 千語と充実させた）
→契約形態に「共同設計方式」を追加（旧建設省の通達に基づき、同入札方式が認められたため）
→「発注機関情報」の充実（CORINS に同じ）
→「発注機関担当者名」を追加（CORINS に同じ）
→「技術者 ID」追加（管理技術者や照査技術者等、技術者を一意に特定し、管理するため）

こうした収集情報の見直し結果を受け、JACIC では入力システムを改良し、バージョンアップを行ってきた。CORINS では平成 14 年 10 月に第 5 世代目となる「CORINS 入力システム (Ver. 5.0)」を発売したばかりである。TECRIS についても第 4 世代目となる「TECRIS 入力システム (Ver. 4.0)」を平成 13 年 6 月に発売している。つまり、収集情報の見直しによる入力システムのバージョンアップは、利用者が必要とする情報を網羅した柔軟なデータベースを構築するための重要な作業のひとつであるといえる。（表 1）

そして、その入力システムのバージョンアップに合わせて、登録受付システムもこれら収集情報の見直し等に合わせ、バージョンアップを行っている。

表 1. 入力システムのバージョンアップの変遷

年	CORINS	TECRIS
平成 6 年（1994 年）	Ver. 1（3 月）	
平成 7 年（1995 年）	Ver. 2（4 月）	Ver. 1（4 月）
平成 8 年（1996 年）	Ver. 3（10 月）	Ver. 2（9 月）
平成 11 年（1999 年）		Ver. 3（10 月）
平成 12 年（2000 年）	Ver. 4（10 月）	
平成 13 年（2001 年）		Ver. 4（6 月）
平成 14 年（2002 年）	Ver. 5（10 月）	

3. データベースの精度向上

柔軟なデータベースを構築していくためには、収集情報の見直しのみで実現することは難しいのが現状である。実現するためには、正確な情報を提供できる情報精度の高いデータベースとして整備していく必要がある。したがって JACIC では、データベースの精度を高めるため、次にその手法による作業を行っている。

3-1. 入力システムによるデータチェック

「入力システムによるデータチェック」は先に紹介した企業側で実績データを作成するために使用する入力システム上の話である。入力システムでは、常識的に考えられない内容やシステムの仕様上、想定していない内容で作成したデータを保存しようとした時点で「エラー」と判定する仕組みを持っている。このエラーは、一般的なパソコンソフトなどと同レベルの考え方である。

【入力システムで判定するエラー事例】

- ・工期の日付が開始日と終了日で逆転している
- ・半角英数字しか入らないところにカナや漢字が入力されている
- ・空白（スペース）は入らないところに空白が入っている
- ・必須項目に入力されていない
- ・JIS 第一第二水準以外の文字が入力されている

つまり、入力システムでは JACIC のデータベースに実績データを登録する前の最低限の内容チェックを行っています。

3-2. 登録受付システムによるデータチェック

「登録受付システムによるデータチェック」は、JACIC に送付されてきた実績データについて、登録受付システムで登録する際に、システムによるデータチェックを行い、登録の可否を判定するものである。内容確認の結果は次の3通りに振り分けられる。

【登録受付システムによる判定結果】

- ①データ内容に「問題なし」
- ②データ内容に何らかの不備があるため「エラー」
- ③データ内容に不備の可能性があるため「保留」

①については、システム的には問題がないデータであるため、登録できるとシステムでは判定し、登録作業上も問題はないものとして処理する。

②については、明らかに常識的に考えられない内容やシステムの仕様上、想定していない内容のデータとして、システムでは登録できないと判定する。入力システムのエラーと同じ考え方に基づくもので、基本的には、入力システムで「エラー」としている内容は、

登録受付システムでも「エラー」として処理する。これは、入力システムで作成した実績データを別ソフトで加工してしまったデータや、大手ゼネコンや大手コンサルなど、自社実績を管理する自社システムから入力システムのデータフォーマットに合わせてフロッピーディスクに格納した実績データなど、入力システムを使用しないで作成したデータによる登録を想定したものである。また、登録受付システムでは、入力システムではチェックしていない内容についても「エラー」として判定している。

【登録受付システムでのみ判定するエラー事例】

- ・建設業許可番号／会社コードの間違い
- ・会社名称の間違い
- ・技術者 ID の間違い（TECRIS のみ）
- ・登録番号の間違い

これらのデータチェックは、システムの中で別に比較対象となるデータベースがあって初めて可能となるもので、入力システムでは行っていない登録受付システムの特長でもある。チェックは、比較対象となるデータベースを「正」として照合することで、内容が不一致の場合に「エラー」として処理する。登録受付システムでは、このような比較対象とするデータベースを複数持っていて、それと連動させたチェック行うことで、精度アップの効果を高めている。(図2)

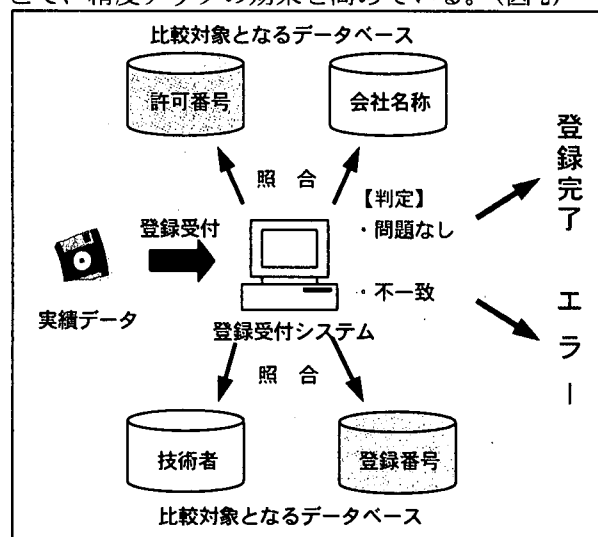


図2. 登録受付システムのみのチェックの仕組み

つまり、登録受付システムによるデータチェックは、

入力システムよりも更に嚴重であるといえる。

③については、「エラー」とは考え方が少し異なるものである。「エラー」が明確にシステムで判定できるのに対して、「保留」はシステムでは内容的に疑わしい可能性があるものの、正しい内容である可能性もあるため、システム上の判定を一旦保留することをいう。この場合、登録を申請してきた企業に内容確認をした上で正否を判定する。企業に内容を確認した結果、問題がなければデータベースに登録することになる。

【登録受付システムで判定する保留事例】

- ・重複登録と思われる実績データの確認（同じ件名で発注機関が工事や業務を発注することがあるため）
- ・発注機関名称がシステム設定内容と異なっている（組織再編等で名称変更となることがあるため）
- ・請負金額 100 億円以上の工事の登録をしようとした（CORINS のみ、100 億円以上の工事は年間に数えるほどしかないため）
- ・技術者情報の不一致による確認（技術者の生年月日や氏名フリカナの間違が多い）

入力システムや登録受付システムによるデータチェックは、運用開始当初から必要に応じて見直しを行ってきた結果、今日では CORINS で約 200 種類、TECRIS で 150 種類以上のチェックを行うまでになっている。

3-3. 登録済み実績データチェック

前述の「入力システムによるデータチェック」「登録受付システムによるデータチェック」の二つは、運営開始当初から現在行っているチェックを確立していたものではなく、年々、チェック内容を増やすなど見直しを行ってきた結果で、今日ではかなり精度の高い実績データを収集するために役立っている。しかし、見直しを行う以前のデータについて精度が低いものが残ってしまっているという問題がある。これを企業に問い合わせを行うことで必要に応じて企業側より修正をしてもらうというのが、「登録済み実績データチェック」である。

手始めとして CORINS では「重複登録」「請負金額（10 億円以上）」について確認作業を行ったところである。一方 TECRIS では、「重複登録」「技術者の情報」について確認作業を行っている。今後もいくつかの項目につ

いて確認作業を行う予定としている。

4. 収集情報の見直しによる検討課題

以上のように CORINS 及び TECRIS に登録される実績データ及び登録済み実績データの精度は高くなってきている状況であり、結果として利用者にとって正確で使いやすい柔軟なデータベースを構築していくことに繋がっている。しかし、その一方で収集情報やデータチェックの見直しを行った結果、特に収集情報が詳細になったことで実績データの作成がしにくく、またチェックの種類が増えたため、エラーや保留になってしまいうデータが増えるなど、簡単に実績データの登録ができないといった問題が出てきている。

【収集情報を見直した結果の問題事例】

・発注機関情報

→当初は機関や事務所単位で管理していたが、Ver. 4 より事務所内の部署等まで管理できるよう詳細な情報（コード表）にした。結果として目的の部署が探しにくくなってしまったことなどにより、誤った内容で発注機関情報を選択してしまう

これらの問題については、これから利用者への確認、エラーや保留となったデータの内容調査・分析など必要な方策立てを行うことも含め、今後の検討課題としている。

おわりに

CORINS 及び TECRIS のデータベース構築におけるデータ収集作業については、収集情報の見直しやデータチェック内容の見直しなどにより、運営開始当初に比べるとその収集精度は格段に上がってきている。しかし、収集情報については、利用者側の目的の変化や社会的な変化により、今後も様々な要求として挙げられてくると思われることや、現在行っているデータチェック作業についても、すべての収集項目に対して完全にチェックを行っているものではない。

JACIC では、更に必要なチェックを導入することも含め、より効果的なデータ収集作業が行えるよう業界団体と連携していく。

CORINS/TECRIS が建設業界を始め、社会に対して広く寄与できることを期待している。