第８回 大学情報・機関調査研究会（２019 年 12 月）
発表要旨

大学 IR 情報の流通における質保証について

森 雅生・大石 哲也（東京工業大学）

１.はじめに

IR への活用を目的とした大学学内の情報収集は，IR 担当者が最初に直面する困難とされている．この課題は，日本に IR が紹介された 2006 年あたりから今日まで，継続的に指摘されているにも関わらず，一般的な解決が未だ見られないようである．原因として，組織論的な要素と技術的な要素の 2 つに分類できるが，後者の技術的な要素の中にも大学が抱える組織論的な問題が見受けられる．この考察では，情報収集における技術的な課題の中に見られる大学組織特有の事情について報告し，その解決策の一つを提示する．

２.背景と注目した課題

Institutional Research が日本に紹介されてから，10 数年が経とうとしている．多くの高等教育機関が，IR 組織を設置し専門人材を配置しつつある．特に，国立大学では教育に関する IR のみならず，研究活動のモニタリング，教員の人員配置や資源の適切な配分に IR を活用しようとしており，その重要性が高まっている．しかし，IR 自身に投資するほどの体力のない大学も多い．そのため，IR はマネジメント・情報処理・統計分析といった多種多様な大量の作業が見込まれる業務であるにも関わらず，十分な人材を確保できないことが大きな課題となっているようである．

IR が最初に直面する難題と言われているデータ収集であるが，IR の情報技術に関する人材不足が困難をさらに深めている． IR は社会学的または組織的な仕事だが，データ収集には情報処理技術が必要である．社会学的背景と情報学のスキルを組み合わせた人的資源を見つけるの簡単ではない．多くの場合，IR 組織に配属されたスタッフは教育学と統計学の専門家であるが，一方で IT スキルを必要とする業務は外部委託されていルことが多い．外注することは，すなわち，業務を言語化した仕様書にすることであるが，言葉にはできないような，機関の文脈にあわせたデータ収集と情報処理は一層困難となろう．

著者らの所属する IR 室は，情報技術に力を入れた IR 機能の構築を行ってきたが，もう一つ，上記のような背景に加え，学内の業務システムを管理運営する場合にも，IT 技術者の不足という問題に直面した．Howard-McLaughlin による IR の情報支援サイクル（図 1）で説明すると，従来の IR における情報収集の課題というのは，”Restructure and Analyze Facts”のフェーズにおける課題である．しかし，IT 技術者の不足により，”Collect and Store Data”のフェーズにおいても，重大な課題が見えた．

学内情報を活用した IR におけるデータ分析の仕事の質は，各部署から出てくる情報の質に依存する．つまり，情報が収集できたとしても，Restructure…のフェーズで再構築できないような情報が来てしまっては，どんなに質の良い分析手法を IR に導入したとしても意味がない．Collect…のフェーズで情報の質が落ちる大きな要因の一つとして，学内の業務システムの更新や改修よって，IR との連携を考慮しないままデータの仕様が変えられてしまうことが挙げられる．

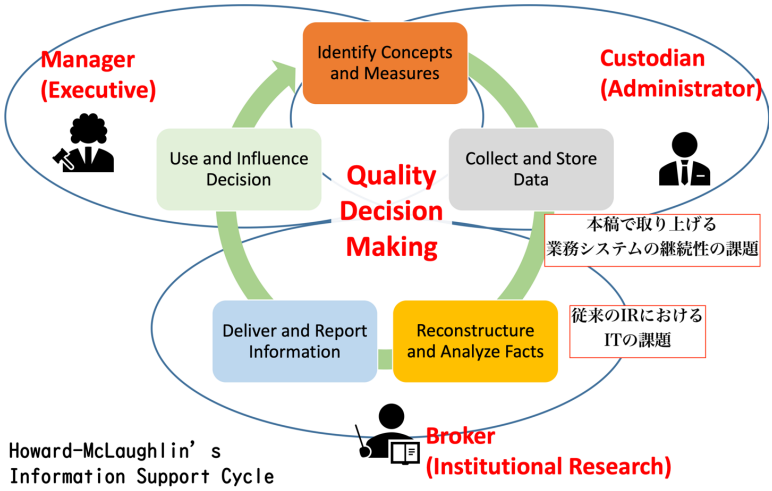


図 1　Howard-McLaughlin の支援サイクル

IR 機能との連携を考慮しつつ業務システムの継続性を担保するには，やはり IT の技能が必要であり，業務システムを事務的に管理運営するだけでは，これを保障することは難しい．この課題を Custodian（管理者・事務局）における IR 情報の質の保証の問題としてとらえ，これを回避・解決するために著者らは 2 つのアイディアを次節以降で示す．

3．FAIR データ原則に基づくデータスキームの検査

IT の専門技能を持たない教職員が業務システムの管理運営を担当する多くの場合，業務の大半を業者に委託するわけであるが，IR との連携を維持するためにどのような方法によって技術的な検査を行えば良いかが不明瞭であることが，まず一つ，解決すべき事柄である．

FAIR Data Principle（図 2）は，科学研究データを効率的に共有するために，一次所有者がデータをアーカイブする際に整備しておくべき事項が記されたガイドラインである．欧州の研究者，図書館員，研究資金提供者のコミュニティである FORCE11 において議論され，2011 年にリリースした．もともと EU は，科学技術政策の一つとして研究データを保存することを掲げている．データの再利用や研究不正防の推進には，FAIR データ原則の活用が欠かせないと言う認識が広まり，EU の大学や研究機関で間も無く採用されることになった．この原則は共有が期待されるデータが，持つべき 4 つ性質をいくつかの細目に分けて掲げられている．検索可能（Findable），アクセス可能（accessible），相互運用可能（interoperable），そして再利用可能（re-usable）であること，の 4 つである．

この原則で検査していることは，データが永続的 ID を持つこと，情報の内容を解説するメタデータがあること，電子媒体・ネットワークが利用でき，権限やセキュリティが適切に設定されているかということである．著者らは，FAIR データ原則を IR 情報の場合に適用できるようにチェックリストを構成した．このリストをシステムの更新・改修時に仕様書などに盛り込むことで，当該システムの IR 連携に必要な要件について，業者に理解を促すことができるようにしている．

FAIR Data Principle	チェック項目（例）
F1.（メタ）データは、グローバルでユニークで永続的な識別子（ID）を持つこと。	1. データテーブルにはPIDが割り当てられているか？
	2. どの識別子がそのデータテーブルのレコードの識別に用いられているか。
	3. データテーブルは識別子のマスターテーブルにアクセスできるのか。
F2. データは、十分なメタデータを持つこと。	1. メタデータを記述している文書はあるか。
	2. メタデータのメンテナンスの責任者は誰か。
F3.（メタ）データは、検索可能なリソースとして、登録あるいはインデックス化されていること。	（検討中）
F4. メタデータは、データ識別子（ID）を明記していること。	（検討中）
A1.（メタ）データは、標準化された通信プロトコルで、識別子（ID）により入手できること。	
A1.1 プロトコルは公開されており、無料で、制限無く実装できること。	（検討中）
A1.2 プロトコルは必要な場合に、認証や権限付与方法を提供できること。	1. 認証の手続きが用意されているなら、認証に関する責任者は誰か。
A2. データが利用不可能になったとしても、メタデータにはアクセスできること。	1. メタデータを記載した文書は機関のどこにあるのか。
I1.（メタ）データは、定型的かつ可達であり、共有された適用可能な言語を使うこと。	1.（メタ）データの記述は形式を持ってなされているかいないか。
	2.（メタ）データの保存先はどこ？
I2.（メタ）データは、FAIR原則に従う語彙を使うこと。	1. 機関はそのメタデータで用いられる用語の定義をしているか。
	2. メタデータで用いられる用語集の管理者はだれか・
I3.（メタ）データは、他の（メタ）データに適した参考情報を含むこと。	1.（メタ）データは必要に応じて引用を行なっているか。
R1.（メタ）データは、多くの正確な関連属性をもとに記述されること。	
R1.1（メタ）データは、明確でアクセス可能なデータ使用ライセンスを持つこと。	1.（メタ）データに関する適切なライセンス管理はなされているか。
R1.2（メタ）データは、詳細な出典を有すること。	1. メタデータはどの業務システムのものであるか追跡できるか。
R1.3（メタ）データは、分野ごとののコミュニティ基準を満たしていること。	（検討中）

図 2　FAIR データ原則

４. データウェアハウス・デザイン・シート

業務システムにおけるデータ構造やスキーム情報（属性情報）を，業務システムの運営担当と IR 担当との間で共通の認識や共有を促すための仕掛けとして，FAIR データ原則を利用したデータテーブルチェック方法を前節で示した．では，データを取得した後の再構築に関わる部分のメタ情報や，データウェアハウスにおける分析方法の定義情報について，どのような共有方法が考えられるのであろうか．

データが持つメタ情報や構造の表現方法，そのメンテナンス方法に関し，IR 情報の基になる業務システムのデータベースと，データウェアハウスにおける分析方法の定義を示した次の 4 つの要素を取り上げる．これらはデータウェアハウスを記述する最低限の項目であり，これらは業務システムの更新や改修に伴う変化において注意すべき要点である．

（１）　メタ情報……業務システムのデータベースが持つテーブルの属性と型情報．
（２）　ER ダイアグラム……テーブル間の関連を示す図．
（３）　スタースキーマ……ER ダイアグラムに基づき分析方法を示した図．
（４）　多次元分析……スタースキーマに基づき多次元キューブの実装法を示した図．

データウェアハウス・デザイン・シートは，業務システム管理者と IR 担当がこの 4 つの要素を共有するツールキットである。上部 2 要素が，業務システムに関するメタ情報であり下部 2 要素はデータウェアハウスに関する定義となっている（図 3)。

５. 最後に

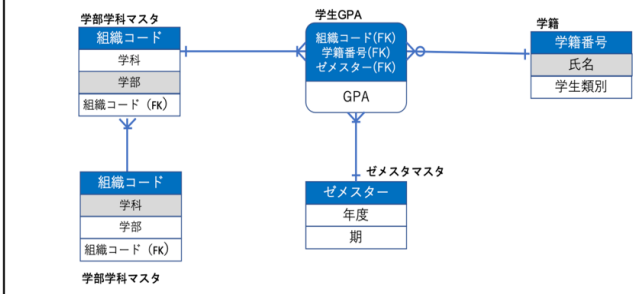
本稿では，　IR 情報の流通，特に管理者から IR への情報フローに関する課題として，スキルを持つ人材や　IR　情報の管理に関する問題点を指摘した．その解決策の一例として，FAIR データ原則に基づくシステムのチェックリストとデータウェアハウス･デザインシートを提案した．

後に残った課題として，これらの提案を IT 人材のいない大学の中でどのように運用していくかと言うことが課題である．一例として，こうした情報の更新管理を，業務システムの更新や改修の際に，業者による納品時の義務としておくことが挙げられる．何れにしても，IT 技術を超えた組織論的な対策が必要となることである．

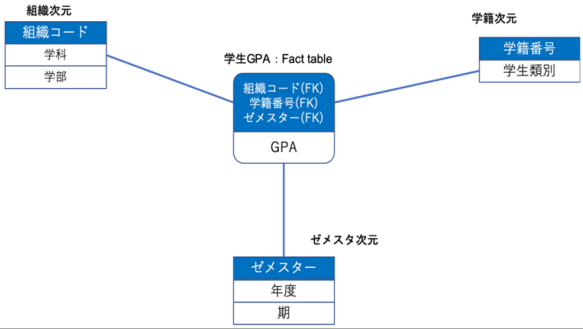
メタ情報

テーブル名	属性名	データ型	Null_OK？	Note
学籍情報	学籍番号（主キー）	文字列	FALSE	
	氏名	文字列	FALSE	
	性別	"女","男"	TRUE	
	学生類別	"一般","編入","留学生"	FALSE	
	生年月日	日付型	TRUE	
	出身高校コード（FK）	文字列	TRUE	
学部学科	組織コード（主キー）	文字列	FALSE	
	学部名	文字列	FALSE	
	学科名	文字列	FALSE	
	下部組織コード（FK）	文字列	TRUE	
	課程コード（FK）	文字列	FALSE	
ゼメスタ	ゼメスタ（FK）	文字列	FALSE	
	年度	整数値	FALSE	
	期	"前期","後期"	FALSE	
学生GPA	ID（主キー）	文字列	FALSE	
	組織コード（FK）	文字列	FALSE	
	学籍番号（FK）	文字列	FALSE	
	ゼメスタ（FK）	文字列	FALSE	
	GPA	実数値（浮動小数点数）	TRUE	

ER図



スタースキーマ



多次元分析

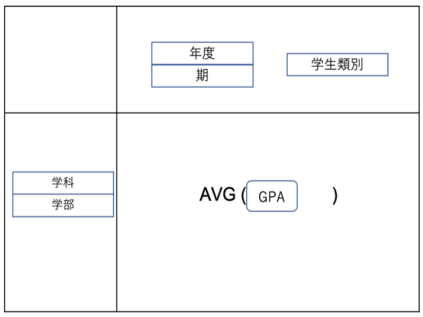


図 3　メタデータおよびデータスキーム記録のフォーマット例

参考文献

[1] Wilkinson, D.M., et al.（2016）, “The FAIR Guiding Principles for scientific data management and stewardship”, *Scientific Data*, vol.3, Article number: 160018.

[2] 鈴木健司，データウェアハウスがわかる本，オーム社出版局，2000年.