

# 教学IRデータを研究に用いるための 必要事項の整理

(2020年2月29日)

首都大学東京 大学教育センター  
教学IR推進室 教学IRデータ管理・分析部門  
松田岳士



# Outline



1. 教学IRと研究  
用語の整理・研究発表への壁

2. 情報収集時の課題  
根拠・匿名性・オプトアウト

3. データ分析・発表時の課題  
研究者の資格・学内チェック体制

各課題への首都大の対応



# IR(Institutional Research)とは？

## ■定義(日本国内の例)

- 大学経営の質の向上と内部質保証のツール(高田ほか 2012)
- 個別大学内の様々な情報を収集して、数値化・可視化し、評価指標として管理し、その分析結果を教育・研究、学生支援、経営等に活用すること(国公立4大学IRネットワーク 2009)
- 高等教育機関における計画立案や意思決定に有効な諸情報の集約、分析および報告を行う組織的機能(藤原ほか 2009)

(参考)文部科学省:教育情報の活用・公表・教学ガバナンス

高等教育の諸課題のうち、1)大学経営上の課題、2)大学教育上の課題、3)評価や情報公表への対応に共通する要素が「情報」であり、大学が必要な情報を活用していくことにより、大学改革の推進を期待(義本 2011)



# 教学IRとは？

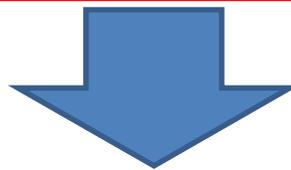
- 教育・学修(学習)に関するデータを調査・分析の対象とするIR
  - 日本独特の概念(分野): 直接該当する英語なし
  - なぜ教学分野のIRが先行・注目されているのか
    - 表向きの理由: 教学データ分析に基づく改革サイクル確立・社会への説明責任が求められている  
例: 中教審答申(平成24年8月)「大学教育の質的転換」  
改革のサイクル確立のためのエビデンスの明確化  
内部質保証を重視  
データ分析、テスト理論、学修評価等の知見を有する専門家育成
    - 本音では: 他分野より導入しやすい(組織的・財政的に)、大学ポートレート・認証評価・エンrollmentマネジメントに使える→要求が広範・曖昧



# Learning Analytics

## ■定義

- 学習や学習が起こる環境を理解し最適化することを目的として、学習者やコンテキストに関するデータを測定、収集、分析、そして報告すること (SoLAの定義: Siemens & Baker 2012)
- 教育コミュニティの行動を評価するために蓄積されたデータの収集、分析、応用 (Larusson & White 2014)



特徴1. 分析対象のデータが学習者に限定されない

→ 多様化した教育データ分析が先行

特徴2. 収集から報告までのプロセスを含む

→ 測定・分析だけでなく、実践への応用まで



# EDM

- Educational Data Mining
- LAとの5つの相違点
  - ✓ 人間の決定支援が重要、発見の自動化は手段⇔自動化が重要
  - ✓ システム全体の理解⇔構成要素の簡素化と個々の要素の関係解析
  - ✓ セマンティック・ウェブが起源⇔教育ソフトと学習者モデリングが起源
  - ✓ 情報提供と教授者・学習者の関係強化⇔適応の自動化
  - ✓ ネットワーク分析・学習者の成功予測⇔クラスター分析・可視化

Table 1: A brief comparison of the two fields

	LAK	EDM
Discovery	Leveraging human judgement is key ; automated discovery is a tool to accomplish this goal	Automated discovery is key; leveraging human judgment is a tool to accomplish this goal
Reduction & Holism	Stronger emphasis on understanding systems as wholes, in their full complexity	Stronger emphasis on reducing to components and analyzing individual components and relationships between them
Origins	LAK has stronger origins in semantic web, "intelligent curriculum," outcome prediction, and systemic interventions	EDM has strong origins in educational software and student modeling, with a significant community in predicting course outcomes
Adaptation & Personalization	Greater focus on informing and empowering instructors and learners	Greater focus on automated adaptation (e.g. by the computer with no human in the loop)
Techniques & Methods	Social network analysis, sentiment analysis, influence analytics, discourse analysis, learner success prediction, concept analysis, sensemaking models	Classification, clustering Bayesian modeling, relationship mining, discovery with models, visualization

(出典) Siemens & Baker 2012

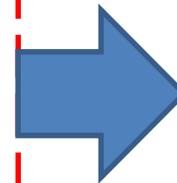


# 教育ビッグデータの特徴

## ■ (購買行動やサービス受領を対象とした) 通常のビッグデータとの相違点

一般に教育ビッグデータは

- 1) 列は短く、行が長い
- 2) 匿名性が低い
- 3) データの意味解釈が容易
- 4) 因果関係は複雑  
→ 少なくとも非線形構造
- 5) 多様化・細分化が進行中  
→ 収集可能範囲が日々拡大



- ✓ 個々のデータの意味は理解可能
- ✓ データ同士の関係は不明
- ✓ 相関しか分からないことも多い
- ✓ 可視化方法を工夫する必要



# 使用できるデータ例

☆個人属性  
学習者(保護者)  
学習支援者  
講師

◎プレアンケート  
受講動機・期待  
学習傾向  
学習環境

時系列(時間推移)

◇LMSログ  
アクセス・操作  
履歴

○学習成果  
成績

△コミュニケーションデータ  
学習者⇔  
教育提供者  
学習者同士

位置データ・生体データ

●ポストアンケート  
結果  
満足度  
自己効力感  
学習経験(態度など)

※行動変容  
パフォーマンス  
態度



# 技術的・物理的に研究可能≠研究可能



## ■ 教学IR研究の促進を阻む壁(例)

- ✓ 情報公開の文化がない
  - そもそもどこにどのようなデータがあるのか分からない
- ✓ どのように許諾を取っているのか分からない
  - 研究目的で収集していないおそれ
- ✓ データの所有権が不明
  - どのデータを誰の許可を得て、誰が研究できるのか
- ✓ 個人情報流出への懸念
- ✓ ひとつの研究機関内に限定
  - 複数大学間でデータを比較できない

課題と対応策を整理



# 収集方法と利用許諾

	収集方法例	許諾方法	課題
調査	アンケート	都度書面	不同意(含氏名 不記入等)・ 無回答の場合
	インタビュー	都度対面・ 書面	
成績	テスト	一括入学時	記憶・証拠なし
	eポートフォリオ	使用開始時 システム上	不同意が不可能
活動	システムログ(LMS等)		
	履修関係	一括入学時	記憶・証拠なし



# データ収集時のチェックポイント

- ✓ポイント1: 研究に用いてもよい理由(根拠)の確認
- ✓ポイント2: 匿名性の程度
- ✓ポイント3: オプトアウトへの対応可能性

## 具体例:

- 入学から卒業までの成績データに関して入学前後に一括で利用許諾の同意を取る大学  
→学生は自分自身の成績が研究されていると認識しているのか?
- 大学で用いられるLMSの利用規約に「研究の対象となることがある」と書かれている  
→そのLMSを使わざるを得ない立場の学生には、同意しないことが実質的に不可能



# ポイント1：法的根拠

## ■ 東京都個人情報保護条例 第十条

第十条 実施機関は、保有個人情報を取り扱う事務の目的を超えた保有個人情報の当該実施機関内における利用(以下「目的外利用」という。)をしてはならない。ただし、次の各号のいずれかに該当する場合は、この限りでない。

- 一 本人の同意があるとき。
- 二 法令等に定めがあるとき。
- 三 出版、報道等により公にされているとき。
- 四 個人の生命、身体又は財産の安全を守るため、緊急かつやむを得ないと認められるとき。
- 五 専ら学術研究又は統計の作成のために利用する場合で、本人の権利利益を不当に侵害するおそれがないと認められるとき。
- 六 同一実施機関内で利用する場合で、事務に必要な限度で利用し、かつ、利用することに相当な理由があると認められるとき。





# データ棚卸

## 所管・所在

- どこにどんな種類のデータがあるか

## 形式・可塑性

- 管理状況、フォーマット、何年前から

## 収集時の許諾等

- 使用目的、許諾の取り方

WGを作り  
1年間かけて確認





# 分析・発表時のチェックポイント

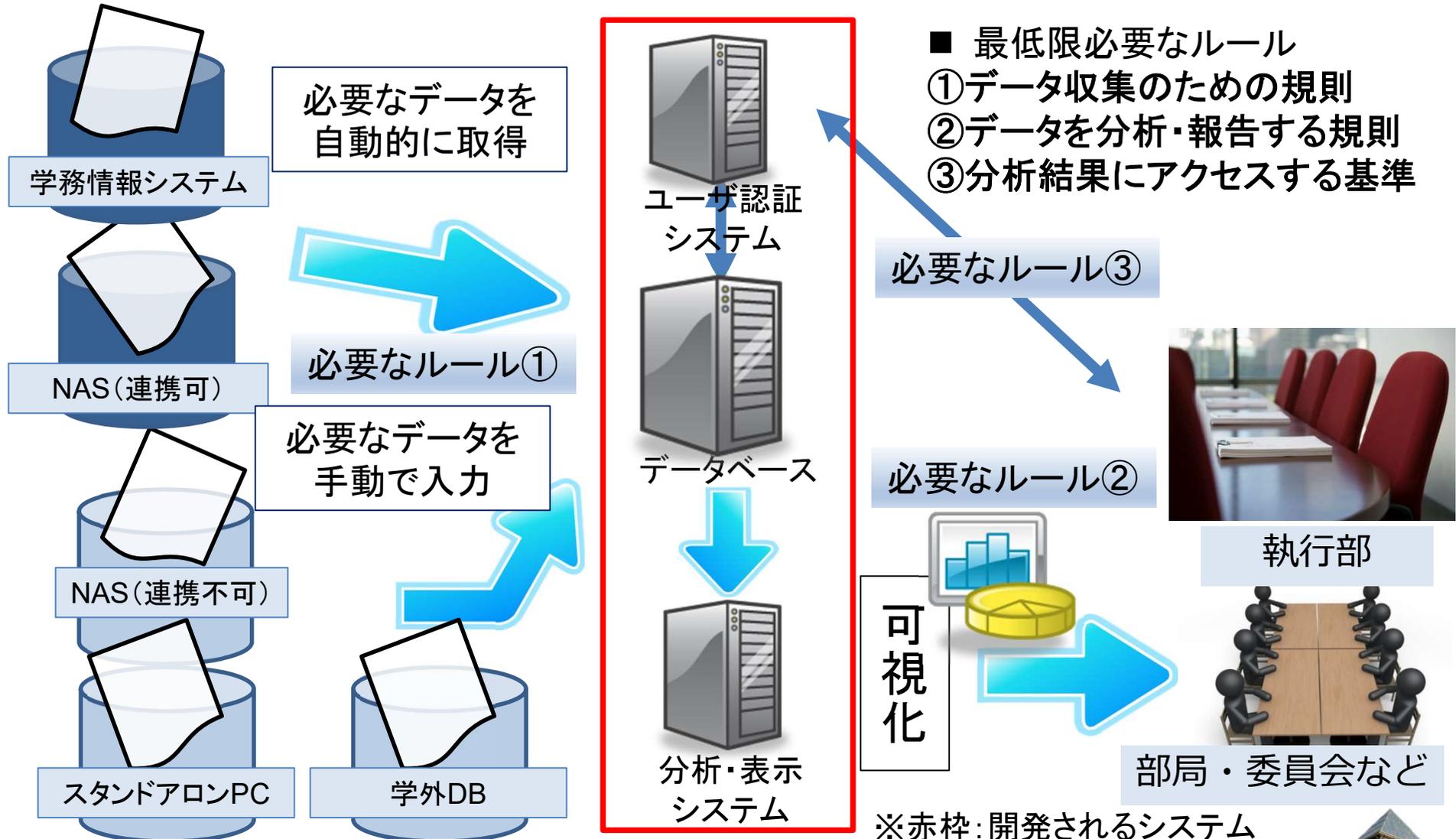
✓ポイント4:なぜその研究者が研究してよいのか

✓ポイント5:研究成果発表までに学内でどのようなチェックを経るのか

- スライド10にあげたデータはほとんどの場合、(研究者ではなく)大学が収集している
  - 研究者はあくまで二次使用者
  - IR担当教員しかIRデータを用いた研究ができない状況はよくない
- 研究倫理審査委員会のチェックだけでよいのか
  - 基本的にデータ提供者と研究者の間の取り決めやアカウントビリティを研究倫理の観点からチェックする組織



# システム開発に応じたルール作り



# 必要なルール（例）

※実際には、ふたつ以上の目的が含まれるルールが作られることが多い

目的	規則・ルール例	主な具体的内容例
①データのスムーズな収集	教学IRデータマネジメント規定	IR情報の定義、IR業務の目的、部局へのIR情報提供プロセスなど
	教学IR室規則	IR室の目的、構成員、室長・室員の権能、IR室運営機関（委員会）
	教学IRシステム利用要項	システムが扱うデータの範囲・秘匿レベル、利用者の定義（下記③を含む）
②データの整理・分析業務遂行	教学IR室業務規定	IR室の業務範囲、利用権限、倫理規定
	教学IRデータ取り扱い要項	IRデータの保護管理規定（個人情報・プライバシー情報との整合）
	教学IRシステム運用内規	データ管理責任者、分析指揮権限、セキュリティレベル、運用体制
③分析結果へのアクセス範囲規定	秘密データ指定ガイドライン	データの種類、秘匿レベル指定方法
	教学IRシステム利用規定	利用者の範囲、運用管理、禁止事項



# アクセス制限の考え方（例）

- データの共有（活用可能性）を優先させたレベル設定  
→ユーザはシステム上でアクセス申請手続き可能
- 注意点
  - ✓ レベル2: アクセス許可手続き（申請先と判断基準）
  - ✓ レベル2と3の判別基準
  - ✓ どのような情報があるか（リスト）は全員に公開

レベル3: 許可された者のみアクセス  
個人情報・機密性の高い情報

レベル2: 申請してアクセス  
従来申請が必要なレベル

レベル1: 無条件アクセス  
学内公開可能レベル



## それでも残る課題

- ✓ 個人情報保護 × 情報公開
  - 「何人以下だと」個人が特定されるか
- ✓ 匿名 × 実名
  - 匿名でデータを取りたがる場合
- ✓ 組織を超えた情報互換 × ルール・組織不在
  - 同業者団体（専門職の横のつながり）

- 本発表で整理したポイントに基づくガイドライン作成
- データからリサーチクエスチョンを導く方法の提案

