

Training Information Professionals for the Emerging Data Ecosystem

Cragin, Melissa

Midwest Big Data Hub at National Center for Supercomputing Applications (NCSA), University of
Illinois at Urbana-Champaign : Executive Director

<https://hdl.handle.net/2324/1794499>

出版情報：九州大学ライブラリーサイエンス専攻シンポジウム「オープンデータとデジタルヒューマニ
ティーズ」, pp.1-, 2017-01-18. 九州大学統合新領域学府ライブラリーサイエンス専攻

バージョン：

権利関係：



Training Information Professionals for the Emerging Data Ecosystem

Presented by: Melissa Cragin

-Slides are selected from original presentation

データ・エコシステムのための 情報専門職の養成

Presented by: Melissa Cragin

* このスライドの日本語訳は、シンポジウムの配布資料として作成したものです。
専門家による訳ではないため、間違いがある可能性があります。

Overview

- ❖ Scientific and scholarly practice in the iSchool curriculum
- ❖ Data as a focal area across the curriculum
 - ❖ Biological Information Specialist (2006)
 - ❖ DCEP (2007)
 - ❖ DCEP-H (2008)
 - ❖ SODA (2012)
 - ❖ Masters degree in Information Management
- ❖ Development of the Data Curation Education Program (DCEP)
 - ❖ Components of the DCEP
- ❖ The new Masters in Information Management
- ❖ Preparing information professionals for the evolving data ecosystem

概要

- ❖ iSchoolのカリキュラムにおける科学的・学術的実践
- ❖ カリキュラム全体にわたる重要な領域としてのデータ
 - ❖ 生体情報の専門家 (2006)
 - ❖ DCEP (2007)
 - ❖ DCEP-H (2008)
 - ❖ SODA (2012)
 - ❖ 情報管理の修士号
- ❖ データキュレーション教育プログラム開発 (DCEP)
 - ❖ DCEPの構成
- ❖ 情報管理分野の新たな修士課程
- ❖ データエコシステムを発展させるため情報学の専門家の養成

Scientific and scholarly production a focus within iSchool research & curriculum

- ❖ History of social computing and communication of scientific and technical information
- ❖ Scientific and “information work” practices
 - ❖ Lab to institution
- ❖ Understanding and assessing the shift toward a digital information landscape
- ❖ Direct engagement with domain researchers
 - ❖ Informatics
 - ❖ Representation of scientific information

iSchoolの研究及びカリキュラムが焦点とする科学的・学術的成果

- ❖ ソーシャルコンピューティングと科学技術情報の伝達の歴史
- ❖ 科学的で情報学的作業を取り入れた演習
 - ❖ 研究室から組織へ
- ❖ デジタル情報環境へシフトする状況の理解・評価
- ❖ 各領域研究者との直接的関与
 - ❖ 情報学
 - ❖ 科学情報の表現

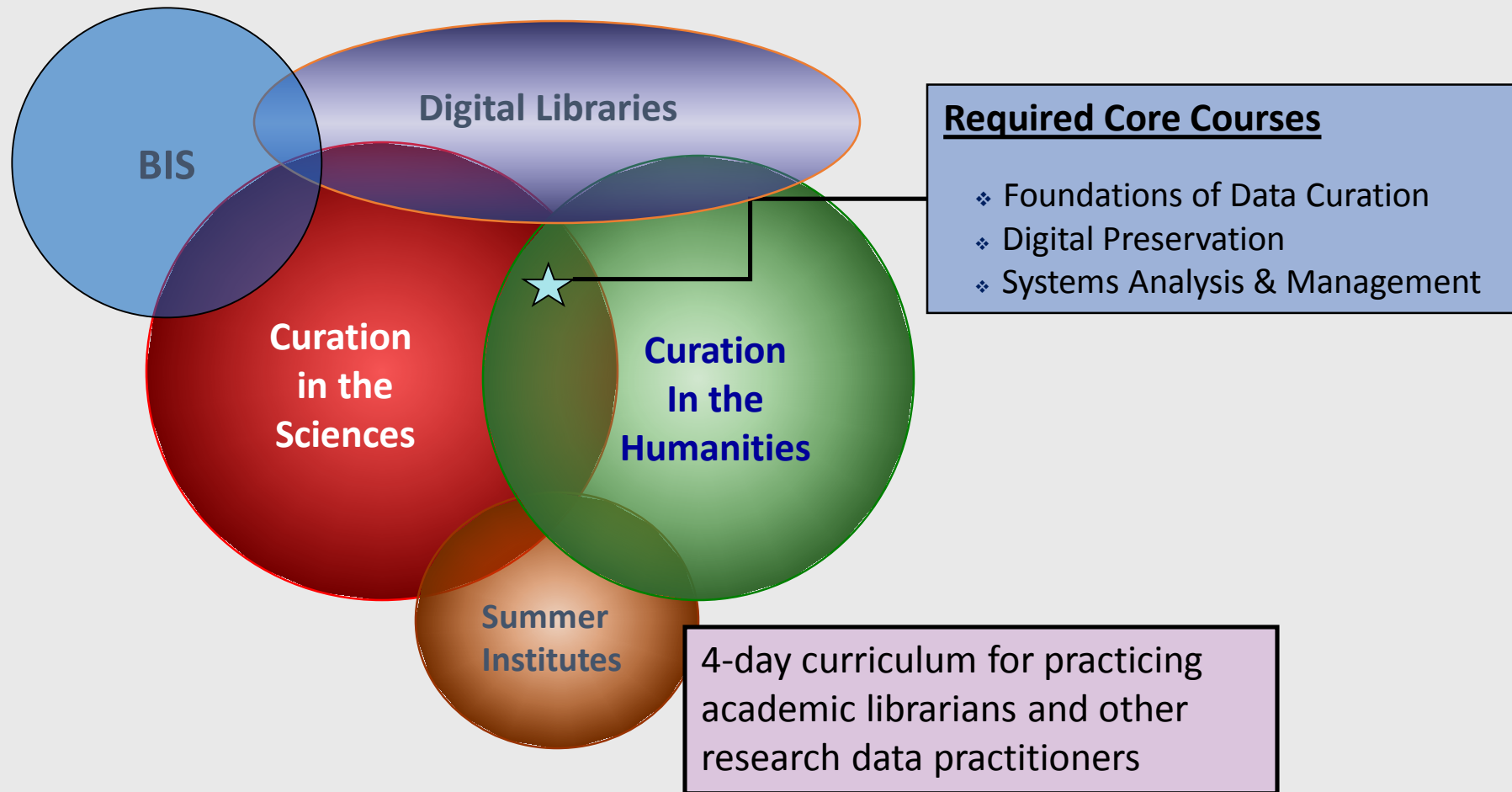
Data as a focal area across the curriculum

- ❖ Biological Information Specialist (2006)
 - ❖ Part of a campus-wide masters in Biological Informatics
- ❖ DCEP (2007)
- ❖ DCEP-H (2008)
- ❖ SODA (2012)
- ❖ Masters degree in Information Management
 - ❖ Above specializations absorbed into this new degree program

カリキュラム全体にわたる重要な領域としてのデータ

- ❖ 生物情報の専門家 (2006)
 - ❖ 学内の生物情報学の修士課程のひとつ
- ❖ DCEP (2007)
- ❖ DCEP-H (2008)
- ❖ SODA (2012)
- ❖ 情報管理の修士号
 - ❖ 上記の専門課程はこの新修士課程へ吸収された

Building a workforce through professional education

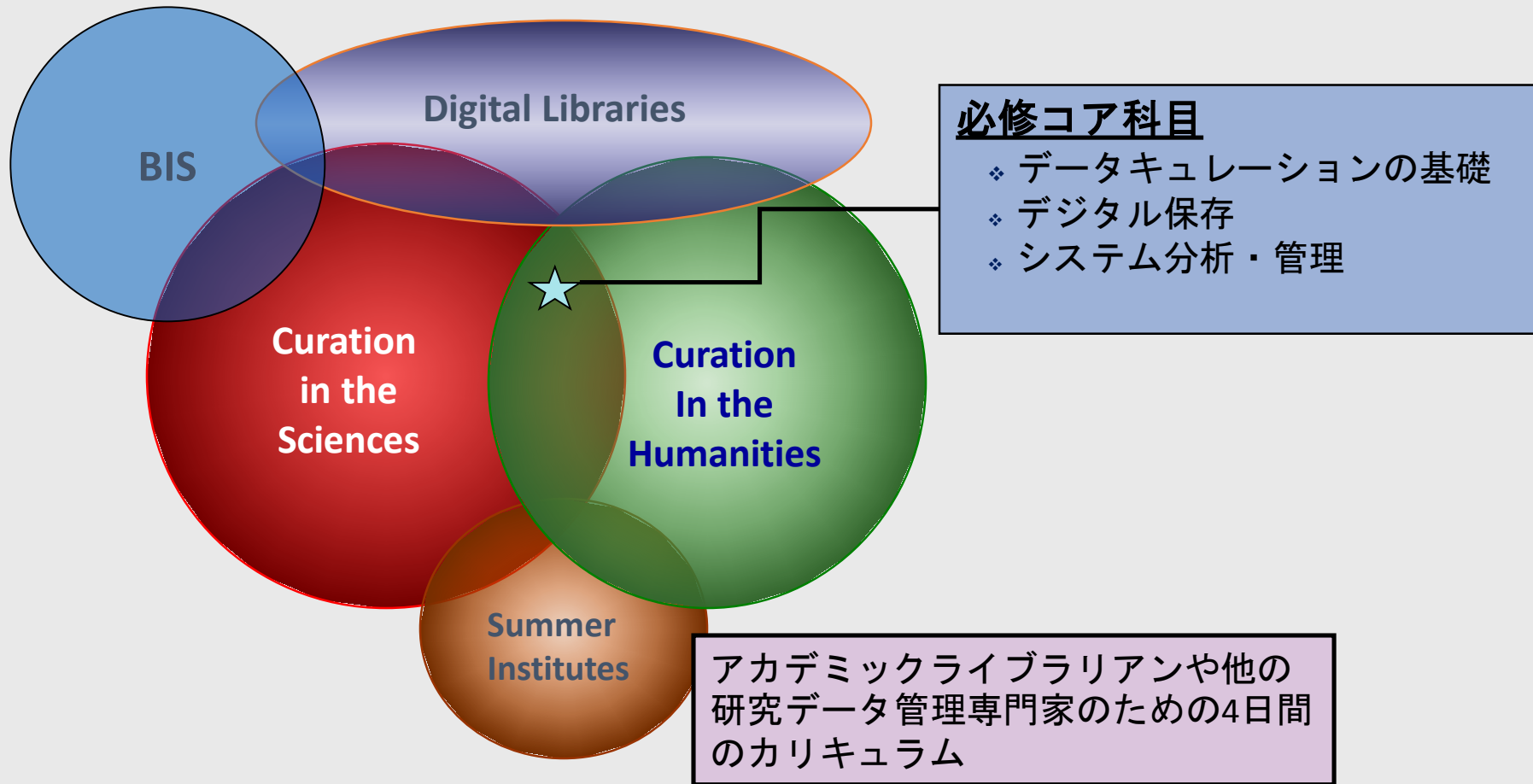


SODA:

Specialization in Socio-technical Data Analytics

MS/IM: Masters in Information Management

専門教育を通じたワークフォースの形成



SODA:

Specialization in Socio-technical Data Analytics

MS/IM: Masters in Information Management

Biological Information Specialist (BIS)

Began as part of campus-wide bioinformatics program

Campus core requirement in biology, CS, & bioinformatics

Others concentrations in CS, crop sciences, animal sciences, chemical & biomolecular eng.

GSLIS only department not focused on computational molecular biology, but biological informatics broadly construed:

Knowledge base: user communities, interoperability, digital preservation, data modeling, ontology development, digital aggregation, information architecture

- Emphasis on data curation and supporting integrative science.

Requires expertise in LIS and research domain.

Biological Information Specialist (BIS)

生体情報専門家

学内の生体情報学プログラムの一部として開始
生物学、コンピュータ科学、生体情報学における必修プログラム

コンピュータ科学、作物学、動物学、化学、生体分子工学における中心プログラム

GSLIS (図書館情報学研究科)は、分子生物学だけに焦点をあてた専攻ではなく、広い意味での生体情報学にも焦点をあてている

知識ベース: ユーザコミュニティ、相互運用性、デジタル保存、データモデリング、オントロジーの開発、デジタル統合、情報基盤

- データキュレーションと統合科学の支援を強調

図書館情報学と研究領域における専門性を要求。

The Data Curation Education Program

Grants:

1. Data Curation Education Program (DCEP) - IMLS/LB, 2006 - Heidorn, PI
2. Extending Data Curation to the Humanities (DCEP+) - IMLS/LB - 2008, Renear, PI



- ❖ Specialization within MSLIS
 - ❖ distance option
- ❖ Foundation in digital data collection & management, representation, preservation, archiving, standards, policy.
- ❖ Emphasis on enabling data discovery and retrieval, maintaining quality, adding value, and providing for re-use over time.

データキュレーション教育プログラム

助成金:

1. Data Curation Education Program (DCEP) - IMLS/LB, 2006 - Heidorn, PI
2. Extending Data Curation to the Humanities (DCEP+) - IMLS/LB - 2008, Renear, PI



- ❖ MSLIS(LISにおける修士号)における専門課程
 - ❖ 遠隔講義のオプション
- ❖ デジタルデータの収集・管理、表現、統合、保存、アーカイブ、標準化、政策の基礎
- ❖ データ発見・検索、質の維持、価値の付与、再利用を可能にすることを強調

Defining Data Curation

The active and on-going management of (research) data through its lifecycle of interest and usefulness to scholarship, science, and education.

Activities

- ❖ enable data discovery and retrieval
- ❖ maintain data quality
- ❖ add value
- ❖ provide for re-use over time
- ❖ archiving
- ❖ preservation

Tasks

- ❖ appraisal and selection
- ❖ representation
- ❖ authentication
- ❖ data integrity
- ❖ maintaining links
- ❖ format conversions

データキュレーションの定義

研究、科学、教育に有用な学術研究データを、
そのライフサイクルを通して継続的に管理すること

活動

- ❖ データの発見性・検索性向上
- ❖ データの質の維持
- ❖ 価値の付与
- ❖ 再利用
- ❖ アーカイブ
- ❖ 保存

責務

- ❖ 評価・選別
- ❖ 表現
- ❖ 信頼性の維持
- ❖ データの完全性維持
- ❖ リンクの保持
- ❖ フォーマットの互換性

Content of core courses

Foundations of Data Curation

- ❖ Digital data and scholarly communication
- ❖ Lifecycles
- ❖ Workflows; Data Re-use & Value
- ❖ Collections
- ❖ Metadata, standards & protocols
- ❖ Infrastructures & Repositories
- ❖ Selection and Appraisal
- ❖ Archiving & Preservation
- ❖ Intellectual Property & Legal Issues
- ❖ Policy & Cooperative Alignments

Kinds of assignments:

Comparative analysis of different metadata standards

Critiques of data management plans

Digital Preservation

- ❖ Archival Theory & Diplomatics
- ❖ OAIS Reference Model
- ❖ Data Formats
- ❖ Digital Archival Objects
- ❖ Preservation Strategies:
- ❖ Emulation vs. Migration
- ❖ Authenticity, Integrity & Trust
- ❖ Evaluation & Value
- ❖ Digital Preservation & The Law

Kinds of assignments:

Planning Grant Application

Trusted Repository Assessment

コアコースの内容

データキュレーションの基礎

- ❖ デジタルデータと学術コミュニケー
- ❖ ライフサイクル
- ❖ ワークフロー; データの再利用と価値
- ❖ コレクション
- ❖ メタデータ, 標準化とプロトコル
- ❖ 基盤とリポジトリ
- ❖ 選別と評価
- ❖ アーカイブと保存
- ❖ 知的財産と法律の問題
- ❖ 政策と協力連携

課題:

異なるメタデータスタンダードの比較分
データ管理計画の批評

デジタル保存

- ❖ アーカイブの理論と古文書学
- ❖ OAIS参照モデル
- ❖ データフォーマット
- ❖ デジタルアーカイブ
- ❖ 保存戦略
- ❖ エミュレーションvs.マイグレーション
- ❖ 真正性、完全性、信頼性
- ❖ 評価と価値
- ❖ デジタル保存と法律

課題:

助成金申請の計画
信頼性の高いリポジトリの評価

Skills for data curation

- ❖ Knowledge of scholarly communication processes and how research works
- ❖ Domain knowledge (or access to it!)
- ❖ Ability to talk to domain experts, programmers, and technologists
- ❖ Systems analysis
- ❖ Ability to track and assess emerging technology
- ❖ Metadata (incl. disciplinary standards)
- ❖ Understanding of how databases work
- ❖ Technical and programming expertise

データキュレーションのためのスキル

- ❖ 学術コミュニケーション過程 と 研究の行われ方に関する知識
- ❖ 領域に関する知識 (またはそれへのアクセス!)
- ❖ 領域の専門家、プログラマー、技術者との会話能力
- ❖ システム分析
- ❖ 新たな技術を収集、評価する能力
- ❖ メタデータ (専門分野のスタンダードを含む)
- ❖ データベースの動作・機能に関する理解
- ❖ 技術とプログラミングの専門的知識

Outcomes of the DCEP & DCEP-H

- ❖ 121 students through the specialization (through 12/2016)
- ❖ Professional “In-service” development
 - ❖ Summer Institutes (1 general; 2 science; 1 humanities)
- ❖ New collaborations
 - ❖ Examples on following slide
- ❖ New research and training grants
 - ❖ Data Curation Profiles Project (Purdue University Libraries, D. Scott Brandt, PI, IMLS NLG 2007-2009)
 - ❖ Data Curation Education in Research Centers (DCERC)
- ❖ Many papers, talks, posters
- ❖ DH Curation Guide: a community resource guide to data curation in the digital humanities – a best practice guide



DCEP と DCEP-Hの成果

- ❖ 専門課程を履修した学生121名 (2016年12月まで)
- ❖ 専門職の“現職” 研修
 - ❖ 夏期講座 (総合系1; 理系2; 人文系1)
- ❖ 新たな連携
 - ❖ 次スライド参照
- ❖ 新たな研究と研修のための助成金
 - ❖ Data Curation Profiles Project (Purdue University Libraries, D. Scott Brandt, PI, IMLS NLG 2007-2009)
 - ❖ Data Curation Education in Research Centers (DCERC)
- ❖ 多数の論文、講演、ポスター
- ❖ デジタルヒューマニティーズキュレーションガイド: デジタルヒューマニティーズにおけるデータキュレーションのための分野別リソースガイド – ベストプラクティスガイド



Summer Institute on Data Curation:

Extending the DC Curriculum to Practicing LIS Professionals

1st Summer Institute on Data Curation

focus on scientific data topics:

- ❖ Digital data
- ❖ Data integrity & authenticity
- ❖ Appraisal and selection
- ❖ Preparation for ingest
- ❖ Digital preservation standards
- ❖ Day-to-day preservation work
- ❖ Repository architectures

2nd Summer Institute: Humanities Data Curation

focus on managing textual data topics:

- ❖ metadata
- ❖ XML/TEI text encoding
- ❖ format and encoding management
- ❖ institutional repository systems
- ❖ digital preservation
- ❖ management of versions and provenance

データキュレーションの夏期講座:

図書館情報学実務者のための発展的なデータキュレーションカリキュラム

第1回データキュレーション夏期講座

主題は自然科学データ:

- ❖ デジタルデータ
- ❖ データの完全性と真正性
- ❖ 評価と選別
- ❖ 受入準備
- ❖ デジタル保存標準
- ❖ 保存の日常業務
- ❖ リポジトリ基盤

第2回夏期講座:

人文科学データキュレーション

主題はテキストデータの維持:

- ❖ メタデータ
- ❖ XML/TEIテキストエンコード
- ❖ 書式とエンコードの管理
- ❖ 機関リポジトリシステム
- ❖ デジタル保存
- ❖ バージョンと来歴の管理

Examples of partnerships with research & data centers



❖ Science

- ❖ BIRN (Biomedical Informatics Research Network)
- ❖ Smithsonian Libraries, Biodiversity Heritage Library
- ❖ U.S. Geological Survey
- ❖ Marine Biological Laboratory
- ❖ Missouri Botanical Garden
- ❖ Field Museum of Natural History
- ❖ US Army ERDC-CERL
- ❖ Snow and Ice Data Center
- ❖ Johns Hopkins Libraries

❖ Humanities

- ❖ Perseus Project
- ❖ OCLC
- ❖ Women Writers Project, Brown University
- ❖ Unit for Digital Documentation, University of Oslo
- ❖ IATH, University of Virginia
- ❖ Center for Computing in the Humanities, Kings College



リサーチ&データセンターとの協力例



❖ 自然科学

- ❖ BIRN (Biomedical Informatics Research Network)
- ❖ スミソニアン協会図書館, 生物多様性遺産図書館
- ❖ アメリカ地質調査所
- ❖ ウッズホール海洋生物学研究所
- ❖ ミズーリ植物園
- ❖ フィールド自然史博物館
- ❖ US Army ERDC-CERL
- ❖ Snow and Ice Data Center
- ❖ Johns Hopkins Libraries

❖ 人文科学

- ❖ ペルセウスプロジェクト
- ❖ OCLC
- ❖ Women Writers Project, ブラウン大学
- ❖ Unit for Digital Documentation, オスロ大学
- ❖ IATH, バージニア大学
- ❖ Center for Computing in the Humanities, キングス大学



Socio-technical Data Analytics (SODA)

- ❖ Socio-technical data analytics combines both
 - ❖ **technical** (databases, social networking, data mining, and text mining) and
 - ❖ **social** (economic, ethical, policy, and political) aspects of data analytics.
- ❖ Foundational courses:
 - ❖ LIS452 Foundations of Information Processing in Library and Information Science*
 - ❖ LIS590AG Evidence-based Discovery or LIS542 Data, Statistical Models, and Information
- ❖ Students are required to have a hands-on experience (either a practicum or an internship)

Socio-technical Data Analytics (SODA)

- ❖ Socio-technical data analyticsは以下の2つの組合せ
 - ❖ **技術** (データベース、ソーシャルネットワーク、データマイニング、テキストマイニング)
 - と
 - ❖ **社会** (経済、倫理、政策、政治) のデータ分析の側面
- ❖ **基礎コース:**
 - ❖ LIS452 図書館情報学における情報処理の基礎*
 - ❖ LIS590AG 根拠に基づく発見 または LIS542 データ、統計モデル、情報
- ❖ 受講者は実地体験を有すること (実務またはインターンシップ)

New courses developed for the data-oriented programs

❖ BIS

- ❖ Literature-based discovery
- ❖ Data mining

❖ DCEP

- ❖ Foundations of Data Curation
- ❖ Digital Preservation

❖ DCEP-H

- ❖ Integration of Humanities into the DCEP curriculum
- ❖ Text Encoding Initiative (TEI) workshops

❖ SODA

- ❖ Evidence Based Discovery
- ❖ Data, Statistical Models, and Information

データ指向プログラム用に開発された新しいコース

❖ BIS

- ❖ 文献に基づく発見
- ❖ データマイニング

❖ DCEP

- ❖ データキュレーションの基礎
- ❖ デジタル保存

❖ DCEP-H

- ❖ 人文科学のDCEPカリキュラムへの統合
- ❖ Text Encoding Initiative (TEI) ワークショップ

❖ SODA

- ❖ 証拠に基づく発見
- ❖ データ、統計モデル、および、情報

Master of Science in Information Management

- ❖ Educating more students with a data focus
 - ❖ No longer declaring a specialization in masters degree programs
- ❖ Pathways of professional focus
 - ❖ Data science and analytics
 - ❖ Privacy, trust, security, and ethics
 - ❖ Information architecture and design
 - ❖ Knowledge management and information consulting
- ❖ Required courses
 - ❖ Information Modeling
 - ❖ Data, Statistical Models and Information
 - ❖ Sociotechnical Information Systems

情報管理の修士課程

- ❖ データに焦点を当てたより多くの学生の教育
 - ❖ 修士課程での専門課程をもう設定していない
- ❖ 専門性の特化
 - ❖ データサイエンスと分析
 - ❖ プライバシー、信頼性、セキュリティ、および、倫理
 - ❖ 情報基盤とデザイン
 - ❖ ナレッジマネジメントと情報コンサルティング
- ❖ 必須科目
 - ❖ 情報モデリング
 - ❖ データ、統計モデル、および、情報
 - ❖ 社会技術的な情報システム

Challenges in the data ecosystem

- ❖ Open research questions
 - ❖ what are data? (aligning ontologies)
 - ❖ what do we keep?
 - ❖ annotation systems (quality; error/bias; application parameters)
 - ❖ impact of cloud services?
- ❖ Varying views of important concepts
 - ❖ data integration
 - ❖ big data
- ❖ Costs and business models
 - ❖ sustainability (preservation and continuity of access)
- ❖ Need for advocacy from many sectors
 - ❖ success cases and best practice guidelines

データエコシステムにおける課題

❖ 未解決な研究課題

- ❖ データとは何か？（オントロジー間の対応）
- ❖ 何を維持するのか？
- ❖ アノテーションシステム（品質、エラー/バイアス、実行パラメータ）
- ❖ クラウドサービスの影響は？

❖ 重要な概念の観点の変化

- ❖ データ統合
- ❖ ビッグデータ

❖ コストとビジネスモデル

- ❖ 持続可能性（維持とアクセスの継続性）

❖ 様々な部門からのアドボカシーの必要性

- ❖ 成功例とベストプラクティスのガイドライン

On-going practical concerns

Technical issues:

- ❖ large-scale storage (*incl. costs for back-up!*)
- ❖ input/output (*I/O – moving data around*)
- ❖ collections of data sets (*formal descriptions and representation*)
- ❖ software sustainability (*keeping the right code for the long term*)
- ❖ data citation and “linking”

Social and Policy issues:

- ❖ need for governance structures (*relationships among repository organizations*)
- ❖ distributed stewardship
- ❖ structural support for interdisciplinary research collaborations

進行中の実用上の懸念

技術的論点

- ❖ 大規模なストレージ(バックアップ費用を含む)
- ❖ 入出力(I/O - データの移行)
- ❖ データセットのコレクション(正式な記述と表現)
- ❖ ソフトウェアの持続可能性(長期的な適切コードの維持)
- ❖ データの引用と“リンク”

社会的、政策的論点

- ❖ ガバナンス構造の必要性(リポジトリ組織間の関係性)
- ❖ 分散スチュワードシップ(監督と報告の責務)
- ❖ 学際的な研究協力のための構造的支援

It's complicated...

- ❖ Gathering meaningful data on scholarly communities is difficult.
- ❖ Scholars and scientists carry out iterative processes of physical, computational, and intellectual activity.
 - ❖ These tasks are complex, may seem mundane or idiosyncratic, but result in a range of immediate and long-term outcomes.
- ❖ Socio-ecology of science
 - ❖ Networks of interaction
 - ❖ Layers of policy
 - ❖ Grant-driven production structures
- ❖ Socio-ecology of R+D, and data-enabled solutions
 - ❖ Grant-driven production structures for academia
 - ❖ But what's needed are networks of interaction
 - ❖ Layers of policy
 - ❖ Very different institutional norms and practices

It's complicated...

- ❖ 学術コミュニティにおける意味のあるデータを収集することは難しい。
- ❖ 人文学者と科学者は、物理的、計算的、および知的活動の反復プロセスを行う。
 - ❖ これらのタスクは複雑であり、平凡にみえたり、特異にみえたりするかもしれないが、その場の、あるいは長期にわたる成果によってなされる。
- ❖ 科学のソシオエコロジー
 - ❖ 相互作用のネットワーク
 - ❖ 政策の層
 - ❖ 助成金主導の生産構造
- ❖ R + Dのソシオエコロジー、データに対応したソリューション
 - ❖ 助成金主導のアカデミアのための生産体制
 - ❖ しかし、必要なのは相互作用のネットワークである
 - ❖ 政策の層
 - ❖ 非常に異なる制度上の規範と実践

Where do libraries fit?

Verbality	Refers to the great proportion of data being “unstructured text-like artifacts”
Verbosity	What’s the rate of redundancy across data stores?
Versatility	How useful are the data to new uses?
Viscosity	How easy is it to leverage the versatility?
Visibility	What are the access controls around it?

ライブラリーはどこに適用するか？

Verbality	“非構造化テキストのような人工物”であるデータの割合が高いことを示す。
Verbosity 冗長性	蓄積されたデータ全体の冗長性はどの程度か？
Versatility 汎用性	新しい用途へのデータの有用性は？
Viscosity	汎用性を活用するのはどの程度簡単か？
Visibility 可視性	その周辺のアクセスコントロールは何か？

The decadal view:

Something(s) to think about

- ❖ How do data management, public access, and preservation intersect?
- ❖ How do we understand variation in the role(s) of science “drivers”? (e.g. discovery; replicability and reproducibility; etc.)
- ❖ How do Academia-Industry-Government partnerships evolve going forward?
- ❖ How do we move to a culture of sharing?
- ❖ How do we understand the differences in “value propositions” for various stakeholders?
- ❖ What “sustainability models” will be viable and durable?

長期的観点: 考えるべきこと(等)

- ❖ データ管理、パブリックアクセス、および保存はどのように交差するのか？
- ❖ 科学の「ドライバー」の役割における変化をどのように理解するのか？（例えば、発見;複製可能性および再現性;等）
- ❖ 産学官連携はどのように進展しているのか？
- ❖ 共有するという文化へどのように移行するのか？
- ❖ さまざまなステークホルダーの“価値命題”の違いをどのように理解するのか？
- ❖ どのような“持続可能性モデル”が実行可能で耐久性があるのか？

Thank you

cragin@illinois.edu

This work is funded in part by IMLS, grant award # RE-05-06-0036-06.

ありがとう

cragin@illinois.edu

<訳>

ライブラリーサイエンス専攻

萩原泰子

福島香奈

水谷江美子

・

石田栄美

富浦洋一