



研究データとオープンサイエンスフォーラム  
～RDA東京大会における議論を踏まえた研究データ共有の最新動向  
主催：国立国会図書館／国立情報学研究所／科学技術振興機構

2016

17<sup>th</sup> Mar

# RDAの活動と議論，日本の課題解決への示唆：人材育成と法的枠組み

池内 有為

筑波大学大学院図書館情報メディア研究科

ikeuchi.ui@gmail.com <http://oui-oui.jp>

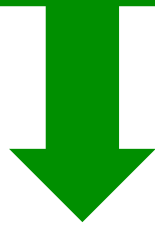


# Contents

1. 自己紹介

2. データサイエンスの人材育成

3. 相互運用のための法的枠組み



1) 日本の課題【内閣府報告書】

2) RDAの活動・議論【報告】

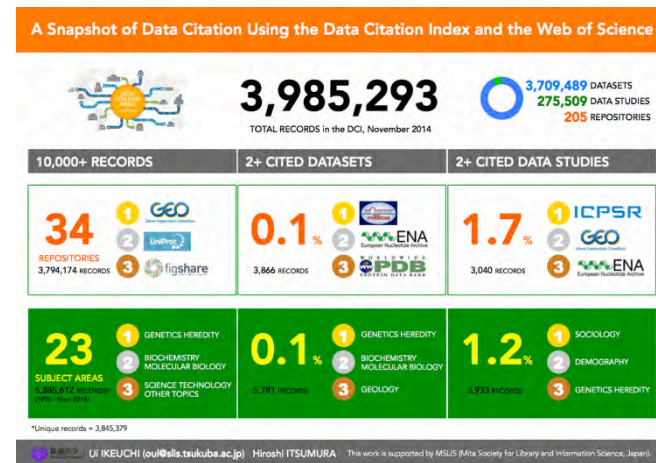
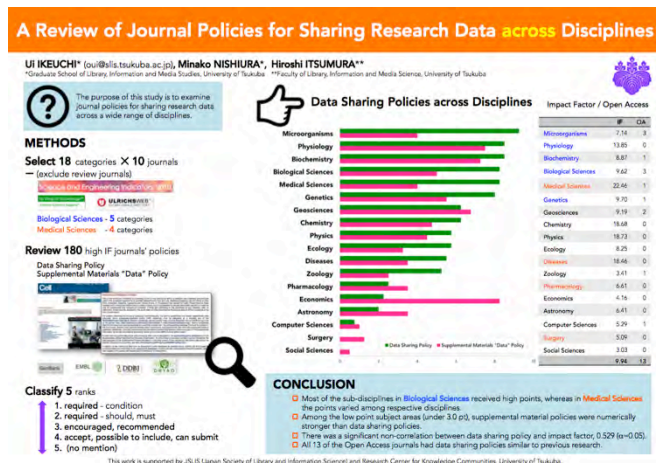
3) 課題解決への示唆【私見】

# 1. 自己紹介



# 研究紹介

- 研究データ共有の実態分析
  - 分野別学術雑誌のポリシーの状況
  - 公開されたデータの再利用の状況
- 研究データ共有と学術情報流通の変化
  - 研究データ共有のインセンティブ
- 海外の大学図書館の研究データ管理
  - インタビュー調査など



## 2. データサイエンスの人材育成

IG Education and Training on handling of research data  
(Working Meeting Session 1)

Joint meeting of IG Education and Training on handling of  
research data, WG RDA/CODATA Summer Schools in Data  
Science and Cloud Computing in the Developing World  
(Working Meeting Session 2)





[https://twitter.com/LM\\_HATII/status/704703958325956608](https://twitter.com/LM_HATII/status/704703958325956608)

# 日本の課題

## ■ オープンサイエンスのためのスキル

“大部分の研究者にはオープンサイエンスを実現するために必要なスキルを開発するための**訓練**、**教育**が必要である。”

## ■ 技術的インフラ・人材育成

“大学・研究機関等においては、技術職員、URA、大学図書館職員等を中心としたデータ管理体制を整備できるように、**データサイエンティスト**や**データキュレーター**などを研究新人材として位置づけられるよう、**包括的な育成システム**を検討し、**推進**することが必要である。”

内閣府「国際的動向を踏まえたオープンサイエンスに関する検討会」報告書  
“オープンサイエンスを推進する際の留意点” p.20-21

<http://www8.cao.go.jp/cstp/sonota/openscience/index.html>

# “Research Data Science” サマースクール

## ■ 背景・課題

- 研究者・サポートスタッフにデータスキルが必要
- 今後10年, データスキルを持つ人材は欧州で500,000人必要 [Barend Mons]
- データを扱い, 管理し, 分析するスキルがなければ公開されたデータが活用されない



全分野を対象とした入門コース+応用コース



# “Research Data Science” サマースクール

## CODATA-RDA Working Group on Research Data Science Summer Schools

The ever-accelerating volume and variety of data being generated is having a huge impact of a wide variety of research disciplines, from the sciences to the humanities: the international, collective ability to create, share and analyse vast quantities of data is having a profound, transformative effect. What can justly be called the 'Data Revolution' offers many opportunities coupled with significant challenges. High among these is the need to develop the necessary professions and skills. There is a recognised need for individuals with the combination of skills necessary to optimise use of the new data sets. Such individuals may have a variety of different titles: Data Scientist, Data Engineer, Data Analyst, Data Visualiser, Data Curator. All of them are essential in making the most of the data generated.



データサイエンティスト, データエンジニア,  
データアナリスト, データビジュアライザ,  
データキュレータ

2016年8月1日～12日  
申込締切：4月18日  
Trieste（イタリア）

### Mission and Objectives

The CODATA-RDA Research Data Science Summer Schools will:

- address recognised need for Research Data Science skills across disciplines;
- follow an recognised and accredited curriculum;
- provide a pathway from a broad introductory course for all researchers (Vanilla) through more advanced and specialised courses (Flavours and Toppings);
- be reproducible: all materials will be online with Open licences;
- be scalable: emphasis will be placed on Training New Teachers (TNT) and building sustainable partnerships.



<http://indico.ictp.it/event/7658/>

<http://www.codata.org/working-groups/research-data-science-summer-schools>

# “Research Data Science” サマースクール

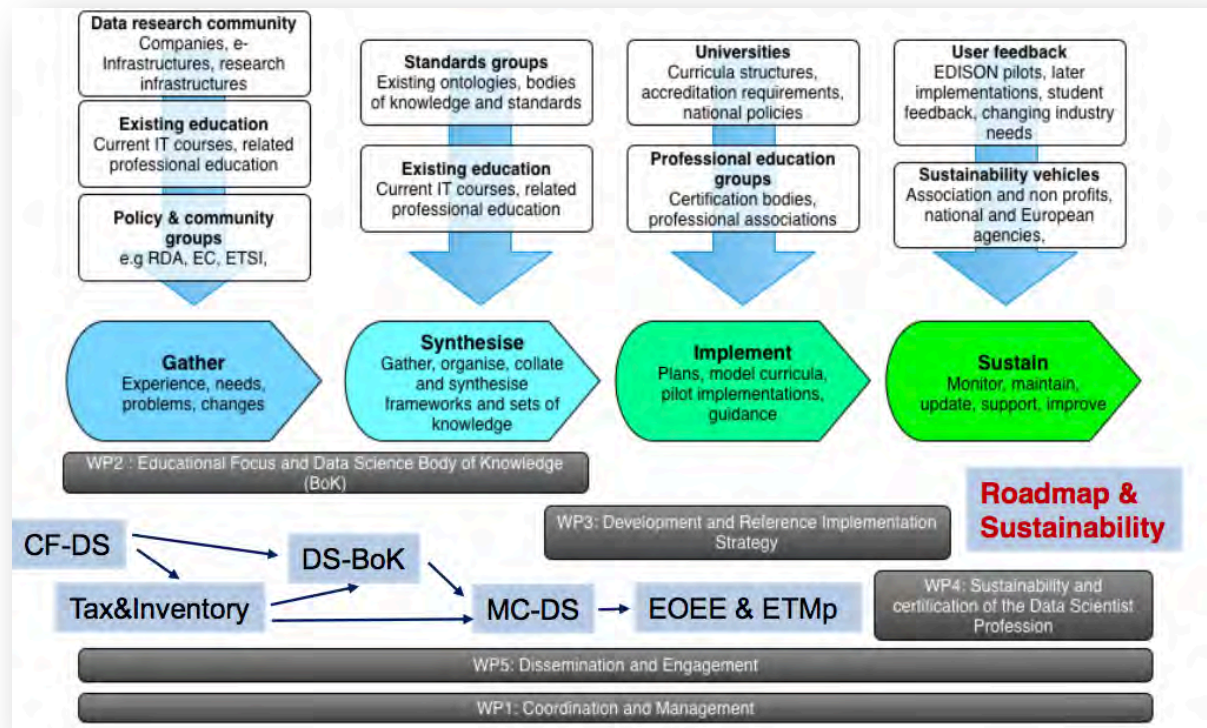
## ■ 入門編カリキュラム（全分野）

- **Open Research Data** : データ出版, ライフサイクル, メタデータとアノテーション
- **Open Science** : 共有・公開のインパクト
- **Data Carpentry** : SQLデータベース入門
- **Software Carpentry** : R, Unix shell, Git
- **Visualisation** : ツール
- **Analysis** : 統計, 機械学習
- **Computational Infrastructure** : クラウドコンピューティング, IaaSクラウド上に仮想マシン構築



## ■ アプローチ

- “データサイエンス”能力のフレームワーク
- コアな知識の特定
- モデルカリキュラムの構築



Horizon 2020  
2015.9から2年間

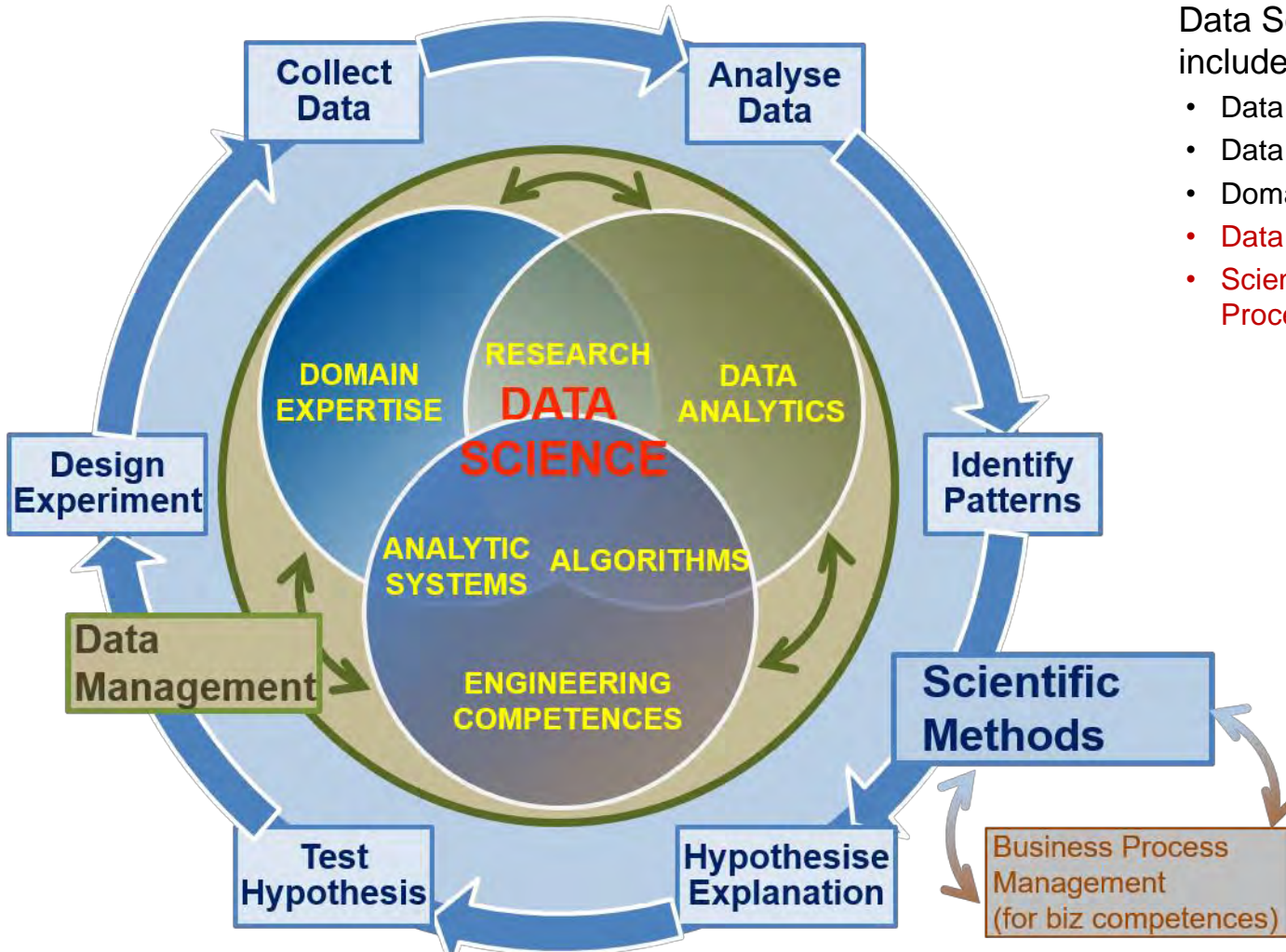
<https://goo.gl/zwbuz4>

# Data Science Competence Groups



- 従来のデータサイエンス
  - データ分析・ビジネス分析・機械学習
  - エンジニアリング・プログラミング
  - 分野の専門知識
  
- EDISON（追加）
  - データ管理, キュレーション, 保存
  - 科学・研究の方法 and/vs. ビジネスプロセス/オペレーション

# Data Science Competence Groups : 研究



Data Science Competence includes 5 areas/groups

- Data Analytics
- Data Science Engineering
- Domain Expertise
- **Data Management**
- **Scientific Methods (or Business Process Management)**

## Scientific Methods

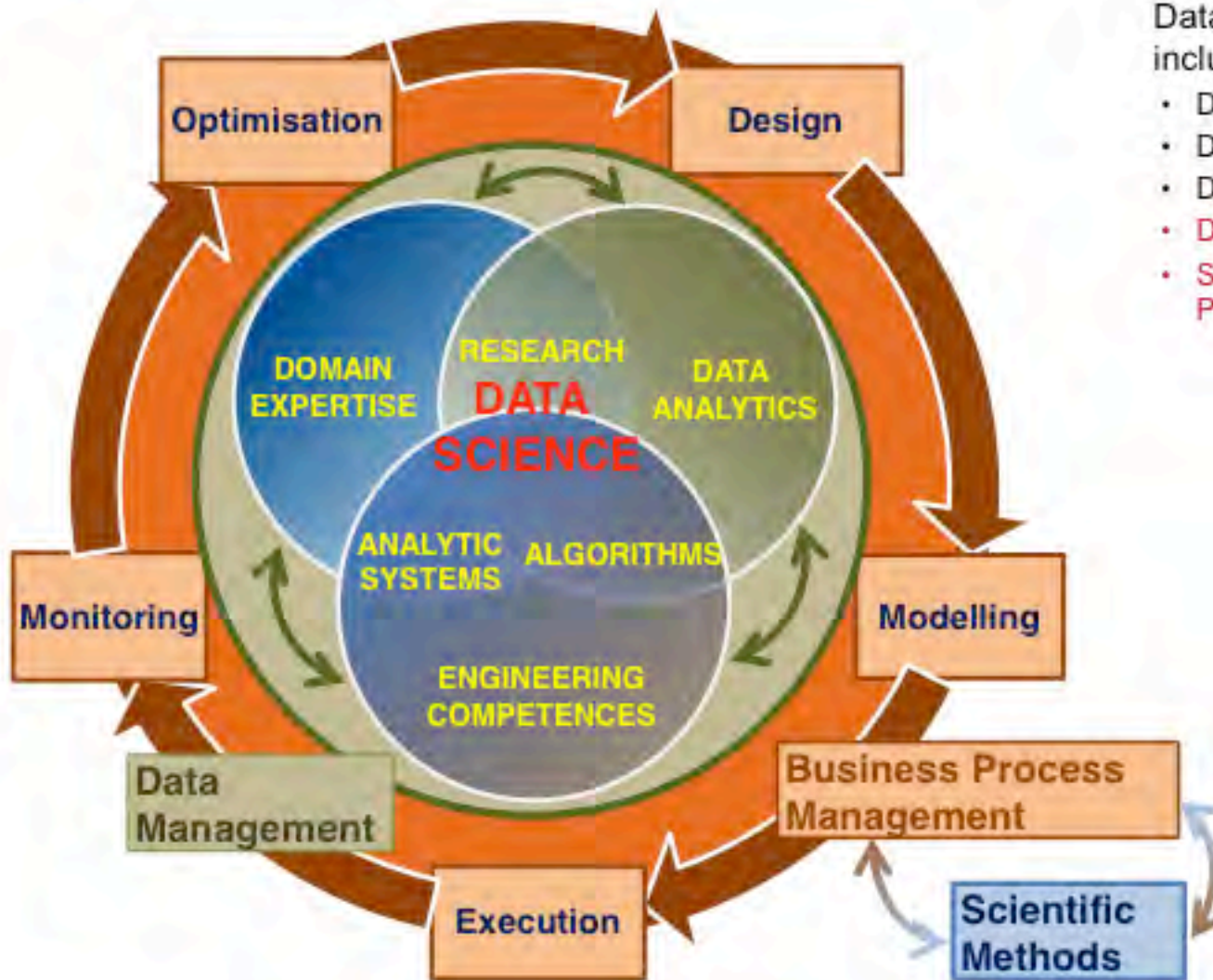
- Design Experiment
- Collect Data
- Analyse Data
- Identify Patterns
- Hypothesise Explanation
- Test Hypothesis

## Business Operations

- Operations Strategy
- Plan
- Design & Deploy
- Monitor & Control
- Improve & Re-design

<https://goo.gl/zwbuz4>

# Data Science Competence Groups : ビジネス



Data Science Competence includes 5 areas/groups

- Data Analytics
- Data Science Engineering
- Domain Expertise
- **Data Management**
- **Scientific Methods (or Business Process Management)**

Scientific Methods

- Design Experiment
- Collect Data
- Analyse Data
- Identify Patterns
- Hypothesise Explanation
- Test Hypothesis

Business Process Operations/Stages

- Design
- Model/Plan
- Deploy & Execute
- Monitor & Control
- Optimise & Re-design

# ポイント



## ■ 連携

- RDA-CODATA, TWAS, EDISON, GEO, DCC, ICTP, CERN, DigCurV,...
- ニーズ調査の結果, モデル・フレームワーク, 既存のコースや教材などを共有
- 連携機関から講師を派遣

## ■ 国際的な標準プログラムの構築を目指す

- クレジット/ア krediyteeshyon
- 中国, ブラジル, インド, 南アフリカ他で実施

# 日本の課題と解決への示唆

## ■ 連携・活用

- (標準化された) カリキュラムの継続的調査
  - ✓ “データサイエンス”にデータキュレーションや管理, 保存といったデータライブラリアンのコアな業務が含まれている
- 参加者からのフィードバック・効果の検討

データライブラリアンや  
キュレータに+

## ■ 日本語教材の開発

- 特定されたニーズ・スキル・フレームワークの活用
- CCライセンスで公開された教材の活用

## ■ キャリアパス (アップ) の検討

- データ管理の論文化 = 共有 + 業績化?

雑誌の例: “*International Journal of Digital Curation*”, “*Digital Scholarship in the Humanities*”

日本のデータライブラリアンや  
キュレータにとって魅力的?



### 3. 相互運用のための法的枠組み ：原則とガイドラインの作成

IG RDA/CODATA Legal Interoperability: Testing the 'Principles' and 'Implementation Guidelines' for the Legal Interoperability of Research Data (Working Meeting Session 5)

<https://rd-alliance.org/ig-rdacodata-legal-interoperability.html>



# 日本の課題

- 論文，研究データの公開・共有化に係る検討  
“共有すべきデータは，**知的財産戦略**を含め，総合的に検討を行う必要”
- 研究データの保存に係る検討  
“公開されたデータや情報の**権利関係（ライセンス）**を明確にし，また…”

内閣府「国際的動向を踏まえたオープンサイエンスに関する検討会」報告書  
“今後の検討課題等” p. 21

<http://www8.cao.go.jp/cstp/sonota/openscience/index.html>

# IG RDA/CODATA Legal Interoperability

## ■ 「原則」と「ガイドライン」を作成中

### 前文

1. 研究データへの合法的アクセスと再利用の促進
2. データの権利と責任の決定
3. 法的利益のバランスを取る
4. 透明性と明確性のある権利の明示
5. 研究データの権利の調和化の推進
6. データの創作者と権利者への適切な帰属の提供

原則（6条）



ガイドライン（8章）

# IG RDA/CODATA Legal Interoperability

- 3-4月：「原則」と「ガイドライン」案を作成
- 4-7月：テスト・レビュー
- 9月：「原則」と「ガイドライン」完成（第8回 RDA総会）



# テスト・レビュー機関（案）

- 研究データコンソーシアム／プログラム／NGO
  - Group on Earth Observations
  - Global Biodiversity Information Facility
  - Consultative Group for International Agricultural Research
  - CERN
  - GenBank
  - COMPARE (infectious disease data)
  - Creative Commons
  - Task Group on Preservation and Access of S&T Data for Development
  - CLARIN (linguistic research)
  - iMarine
  - OpenAIRE

# テスト・レビュー機関（案）

## ■ 政府機関

- 欧州委員会
- OECD
- GRC (Global Research Council)
- G8
- UNESCO
- UNEP
- WHO
- WIPO



# テスト・レビュー機関（案）

## ■ 各国の政府・研究機関

- 米国 (OSTP, NSF, NIST, NIH, DOE, NOAA, NASA, 他?)
- 英国 (RCUK, 他?)
- EU諸国の研究省庁
- 中国科学技術部 (MOST) , 中国科学院 (CAS)
- 日本学術会議
- インド (?)
- 豪州 CSIRO, ANDS
- 南アフリカ NRF, CSIR
- チリ国家科学技術研究委員会 (CONICYT)
- ブラジル(?)



# 日本の課題と解決への示唆

- 日本の法制度との齟齬・問題点の検討→WGへ報告
  - フェアユース, 孤児著作物の規定なし
  
- 日本のガイドラインの検討
  - 1) 内閣府「国際的動向を踏まえたオープンサイエンスに関する検討会」報告書 (p. 18) “関係機関におけるオープンサイエンス実施方針等のあり方”
  
  - 2) 文部科学省 第8期学術情報委員会 (第7回) 「学術情報のオープン化の推進について (審議まとめ案)」 (p. 13) “研究成果の利活用” <http://goo.gl/bTtfKa>

※次頁に本文抜粋



# 参考

## 1) デジタル化された研究データへのアクセス

“研究データは著作物ではないため、**CC0を採用すること**が望ましい。また、データの集積・整理や品質管理などの点で専門的研究者の労力や高度なノウハウが入ったデータベースなど、知的生産物として**著作権が発生する場合にはCC-BYを採用すること**が望ましい。”

## 2) (注記)

“研究データは著作物でないのが通例であるため、(略)**CC0を採用すること**が想定される。一方、研究データ作成への研究者のインセンティブを確保する観点から、利活用の際に**データ作成者の表示を求める**ことが望ましい。(略)方法としては、**CC-BYライセンス**を利用することや、研究コミュニティのルールにより表示することが考えられる。”

# まとめ



## ■ RDAへの参加

- コミュニティに参加し続ける→連携・効率化, 貢献
- 日本の状況を世界に伝える→合意形成に参画

## ■ 人材育成

- “データサイエンス”+“データキュレーション”
- 研究 = 論文執筆 + データ公開が全分野の標準スキルに

## ■ 相互運用のための法的枠組み

- 原則とガイドラインが急速に形成されつつある
- 日本に不都合な点はないか？

# 英国のコンピュータサイエンス専攻 卒業生の失業率が高い

## What should computer science degree students learn?

Conference hears that fundamental skills, rather than specific languages, will serve graduates throughout their career

March 10, 2016

By [David Matthews](#) Twitter: [@DavidMJourno](#)

基本的なスキル (fundamental skills) の  
習得が重要

