

研究データ管理と 日本の大学

青木 学聡
京都大学情報環境機構

<http://researchmap.jp/takaakiaoki/>
<http://orcid.org/0000-0002-5926-4903>

簡単な自己紹介

- (2000.3~) 各種研究員/産学官連携助手等
 - ナノ加工・製造プロセスの計算機シミュレーション
- (2007.2~) 京都大学工学研究科附属情報センター
兼 電子工学専攻 講師
 - 研究科レベルの情報インフラ整備
 - 情報セキュリティポリシー実装
 - 論文データベース
 - 入試・学部成績データ分析基盤
 - 研究データ保存システムの検討, プロトタイプ(2015.12~)
- (2016.3~) 情報環境機構
兼 学術情報メディアセンター 准教授
 - 研究支援部門として, 全学的なICTの設計・導入・運用
 - スーパーコンピュータシステム (2016.10 更新, 稼働)
 - 汎用コンピュータシステム(2016.12 更新, 稼働)
 - データセンター, ハウジングサービス
 - 全学レベルの研究データマネジメントにも関与を始める

「研究データ管理」の多義性

- 「研究公正とコンプライアンス」のため
- 「オープン(データ)サイエンス促進」のため
- 「学術領域の発展と社会貢献」のため

「研究データ管理」の多義性

- 「研究公正とコンプライアンス」のため
- 「オープン(データ)サイエンス促進」のため
- 「学術領域の発展と社会貢献」のため

特色ある学術データベース(1)

- 貴重資料画像(図書館機構)
 - <http://edb.kulib.kyoto-u.ac.jp/exhibit/>
- 研究資源アーカイブ(博物館等)
 - <http://www.rra.museum.kyoto-u.ac.jp/>
- KEGG: Kyoto Encyclopedia of Genes and Genomes (化学研究所)
 - www.genome.jp/kegg
- World Data Center for Geomagnetism, Kyoto (理学研究科)
 - <http://wdc.kugi.kyoto-u.ac.jp/index.html>
- 地域研究統合DB
 - <http://www.cias.kyoto-u.ac.jp/database/>
- Etc. etc...

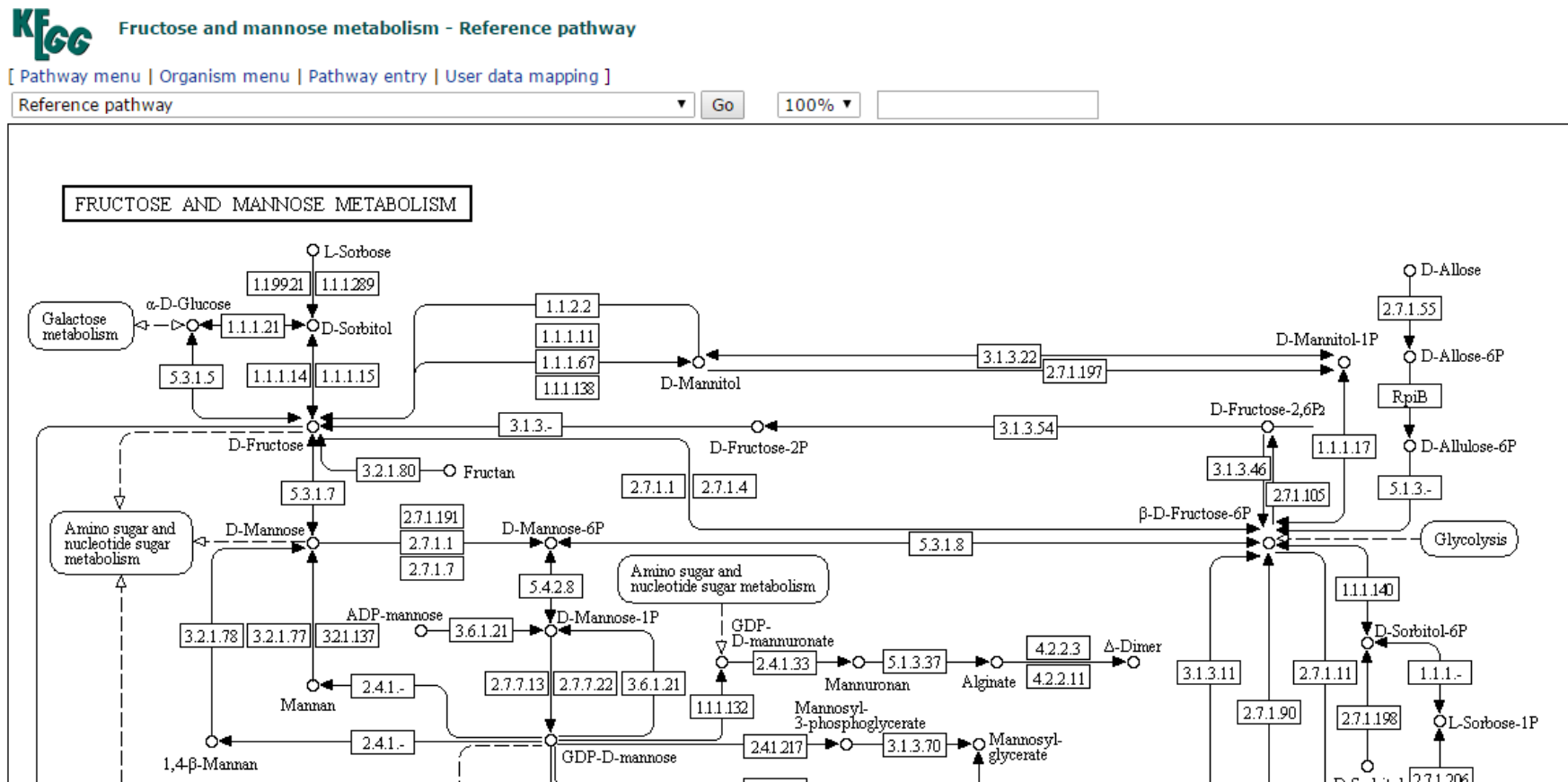
特色ある学術データベース(2)

- 貴重資料画像(図書館機構)
<http://edb.kulib.kyoto-u.ac.jp/exhibit/>



特色ある学術データベース(3)

- KEGG: Kyoto Encyclopedia of Genes and Genomes
 - www.genome.jp/kegg
 - バイオインフォマティクス情報のキュレーションとDB化
 - DBと共にHPCリソースも提供



「学術領域の発展と社会貢献」 のための研究データ管理の視点

- 主にコミュニティレベルで多大な労力, 時間をかけて醸成
- 自発的な取り組みであるため, 高い専門性とデータ利活用を実現
- 持続的な発展が大きな課題
- 大学側から見れば、全容を把握できていない状況
 - 大学としての「コンテンツ発信力」が今後さらに重要視されるのか?
 - 現行の運営からすれば
 - 苦しいから何とかサポートがほしい
 - 余計な手出しは無用
と相反する思いがあるのでは・・・

「研究データ管理」の多義性

- 「研究公正とコンプライアンス」のため
- 「オープン(データ)サイエンス促進」のため
- 「学術領域の発展と社会貢献」のため
 - 主にコミュニティレベルで多大な労力, 時間をかけて醸成
 - 自発的な取り組みであるため, 高い専門性とデータ利活用を実現
 - 持続的な発展が大きな課題

研究公正のためのデータ保存

京都大学での対応

- 2014(H26)年度
 - 8月: 文部科学省「研究活動における不正行為への対応等に関するガイドライン」
 - 2月: 「京都大学における公正な研究活動の推進等に関する規程(※)」
 - 教職員, 部局, 全学の役割, 責任の確認. 研究データ一定保存義務
 - 3月: 日本学術会議「科学研究における健全性の向上について」
- 2015年度(H27)年度
 - 7月: 「(※)の研究データの保存、開示等について定める件」
年度内に管理方針, データ保存計画を, 部局毎に制定
- 2016年度(H28)年度
 - 部局ごとに決定した管理方針, データ保存計画を元に行動中

研究データ保存システムの構築

- 情報環境機構, 工学研究科等で2015年度末より仮運用
 - 電子化されたデータを安心して保存できる仕組みが必要との判断
 - 既に運用, 開発経験のあるWebフレームワークや, CMSを改良しデータ保存システムを構築

ResearchData (beta) Takaaki Aoki Logout

Logged in. Last login: 2016/08/18 15:43:52

My ResearchData

Title	Year / Month	Filename	Filesize	
Molecular Dynamics Simulations Study of Nano Particle Migration by Cluster Impact	2016/02	SMMIB2015-submit-v1.zip	3.4 MB	Detail

Total filesize: 3.4 MB

+ Add

(情報環境機構のシステムの例)

ResearchData Upload

(*) required

Research Data File(*)
[ファイルを選択] 選択してください

Title(*)
Molecular Dynamics Simulations Study of Nano Particle Migration by Cluster Impact

Authors
Takaaki AOKI, Toshio SEKI, Jiro MATSUO

Type
Paper

Year / Month
2016 / 2

Bibliography

Digital Object Identifier (DOI)
http://dx.doi.org/10.1016/j.surface.2016.04.053

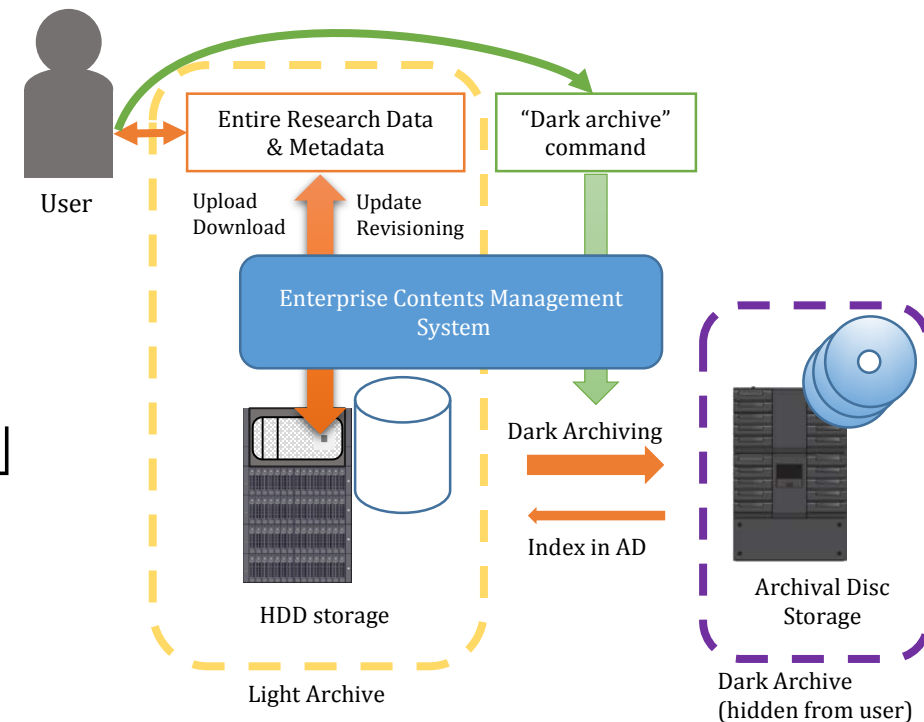
Memo

Share with: (mail addresses of KUMail or KUMCI)
self: mailto:Ta@kyoto-u.ac.jp
memo: jiro.Ta@kyoto-u.ac.jp

Save Cancel

ダークアーカイブ型データ保存システム

- 2016.12 の汎用コンピュータ更新に合わせ, 「情報ライフサイクル管理」のソリューションを導入
- 研究データに限らず, 多くの電子的文書の長期保存を可能に
- 大容量光ディスクを用いたアーカイブシステム (500Tbyteからスタート)
- エンタープライズコンテンツマネジメントシステム(ECM)上で操作
- 研究データ管理の観点からすればあくまでも「研究者・組織の保険」の扱い, 「オープンデータ」とは異なる意味合い



「研究公正とコンプライアンス」 のための研究データ管理の視点

- データの長期保存
- ルール(policy)が先行, 現実的な実施手順(procedure)の策定と実施(deployment)が後追い
- Procedure は部局毎に制定, 大学全体として提供するシステムとスペックが合致するか, これからすり合わせ.
- 残念ながら「ほとんどが死蔵されるデータ」?
 - 研究者個人, 限られた研究グループ内で保持するバックアップ最終手段
 - 最悪の場合(研究公正にかかわるインシデント)への備え

「研究データ管理」の多義性

- 「研究公正とコンプライアンス」のため
 - データの長期保存
 - ルール(policy)が先行, 現実的な実施手順(procedure)の策定と実施(deployment)が後追い
- 「オープン(データ)サイエンス促進」のため
- 「学術領域の発展と社会貢献」のため
 - 主にコミュニティレベルで多大な労力, 時間をかけて醸成
 - 自発的な取り組みであるため, 高い専門性とデータ利活用を実現
 - 持続的な発展が大きな課題

「研究のオープン化」：英国の例

- ほとんどの研究資金配分機関(Research Council)が
 - 論文のオープン化
 - 論文データのオープン化を義務化
- Data Curation Center のwebより
<http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>

● Full Coverage ● Partial Coverage ○ No Coverage

	Research Funders	Policy Coverage		Policy Stipulations					Support Provided			
		Published outputs	Data	Time limits	Data plan	Access/sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
Arts and Humanities	AHRC	●	●	●	●	●	●	○	●	○	○	○
Biotech. & Biological	BBSRC	●	●	●	●	●	●	●	●	●	○	●
Cancer Research	CRUK	●	●	●	●	●	●	●	○	●	○	○
Eng. & Phys.	EPSRC	●	●	●	○	●	●	●	○	○	○	●
Economics & Social	ESRC	●	●	●	●	●	●	●	●	●	●	○
Medical Research	MRC	●	●	●	●	●	●	○	○	●	○	○
Natural Environment	NERC	●	●	●	●	●	●	●	●	●	●	○
Science & Tech. Facilities	STFC	●	●	●	●	●	●	●	○	●	○	○
	Wellcome Trust	●	●	●	●	●	●	●	●	●	○	●

影響が大きいのはEPSRCの要請
 EPSRCは自身のデータレポジトリ,
 データセンターを持たない

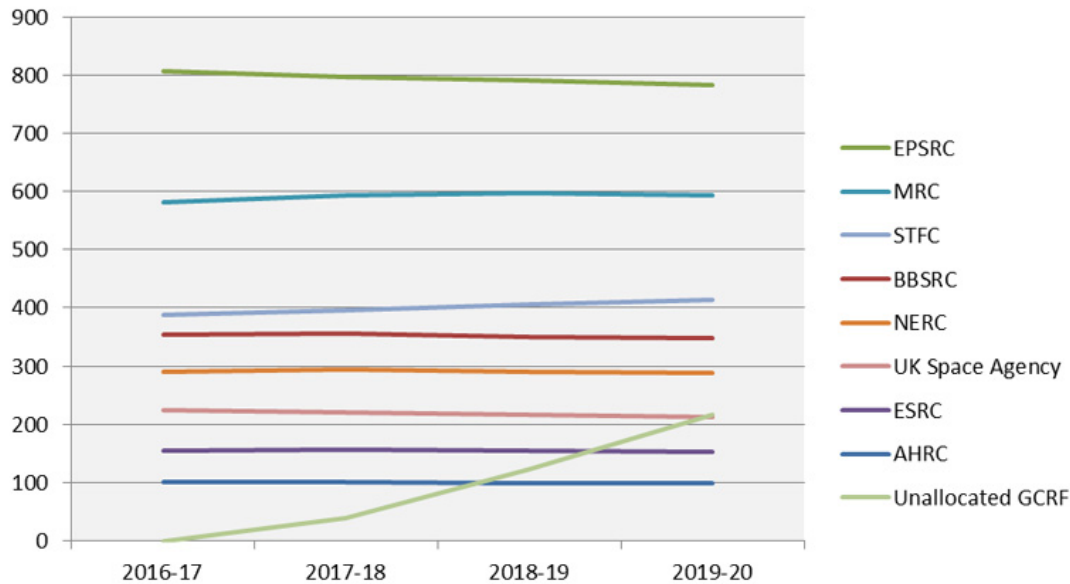


- ・コミュニティが整備する
 - ・出版社が整備する
 - ・機関が整備する
- 等何らかの方法で「データを accessible にしないとイケない」

UK Budget Timeline for Research Councils

<https://www.timeshighereducation.com/news/uk-research-council-budgets-2019-20-revealed>

英国での科学研究予算規模, RC担当分で2,700百万GBP ≈ 3,400億円
(科研費が約2,300億円)



		2016-17	2017-18	2018-19*	2019-20*
AHRC	Arts and Humanities	101	101	99	98
BBSRC	Biotech. & Biological	353	356	350	347
EPSRC	Eng. & Phys.	807	796	790	783
ESRC	Economics & Social	155	157	154	153
MRC	Medical Research	581	594	597	594
NERC	Natural Environment	291	294	290	288
STFC	Science & Tech. Facilities	388	396	406	414
Unallocated GCRF		0	38	122	216
Hefce		1,695	1,716	1,730	1,745

Research council and Global Challenges Research Fund allocations, £ million

Hefce: higher education funding council for England,

大学運営費交付金に相当する資金を配分する機関, ここではそのうちの「研究交付金」の部分

http://www.mext.go.jp/b_menu/shingi/chousa/koutou/062/gijiroku/_icsFiles/afieldfile/2014/11/18/1353540_2.pdf

Edinburgh University の取り組み (1)

- <http://www.ed.ac.uk/information-services/research-support/data-management>
- 先ほどのDCCの母体
- 研究データ管理のベストプラクティスの1つ
- 2014年の図書館総合展フォーラムでの Stuart Lewis 氏の発表 (対訳付き)を始め, 多くの事例紹介, 調査報告あり.
<http://id.nii.ac.jp/1280/00000019/>
- 簡単にどのようなICTサービスを提供しているかを改めて整理

Edinburgh University の取り組み (2)

- “A guide to the Research Data Service” (<http://edin.ac/1Y5k8xf>)
- 研究ステージ毎に利用できるICTを簡潔に紹介
1ページ1機能, 16ページ
 - Before
 - During
 - After
 - (Training & Support)



BEFORE

create a data
management
plan



DURING

working with
data



AFTER

share and
archive your
data



TRAINING
&
SUPPORT

A guide to the Research Data Service (1)

- Before
 - DMPonline (<https://dmponline.dcc.ac.uk/>)による研究データ管理プラン作成支援
 - Finding & analysing data:
 - データ検索に関するコンサルティング
 - オンラインデータ解析サーバー(<http://stats.datalib.edina.ac.uk/sda/>) の紹介
- During
 - ネットワークストレージ
 - DataStore: SMB, NFS等で利用するネットワークドライブ.
教員, 大学院生に0.5Tbyte/人を無償で提供.
 - DataSync: Dropbox のようなクラウド型ファイル共有サービス.
DataStoreの一部(20Gbyte)を切り出して利用.
クライアントPC間のデータ同期, グループ間のファイル共有で利用
 - バージョン管理: ソフトウェア開発用に subversion レポジトリ

A guide to the Research Data Service (2)

- After

- PUREによるデータ保存: 研究者総覧と機関レポジトリを合わせたような仕組み. <http://www.research.ed.ac.uk/portal/> から
 - 人
 - 研究プロジェクト
 - 研究成果(主にOA論文)
 - その他研究活動(受賞等)等をメタデータと共に登録, コンテンツを(利用者の選択により)公開
- DataShare (<http://datashare.is.ed.ac.uk/>): 機関データリポジトリ. PUREよりもオープンサイエンスを志向した構造
- DataVault: 主にPUREの登録データを(メタデータごと)長期保存する

A guide to the Research Data Service (3)

- Training

- MANTRA(<http://datalib.edina.ac.uk/mantra/>):

- 研究者(学生含む)向けのデータマネジメントe-learningテキスト

- MOOC

- <https://www.coursera.org/learn/data-management> にてオンライン学習コースを開講

特に, 若手研究者, 大学院生に対する情報リテラシ教育として普及に注力

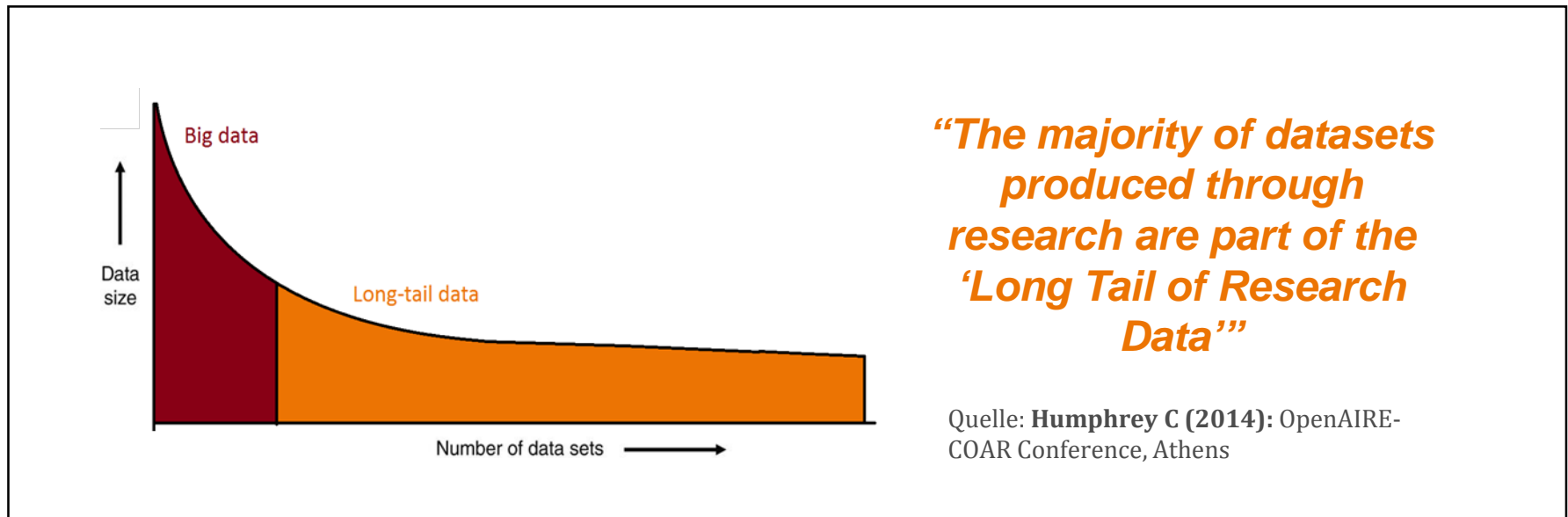
「研究データ管理」の多義性

(「オープン(データ)サイエンス」の視点)

- 「研究公正とコンプライアンス」のため
 - データの長期保存
 - ルール(policy)が先行, 現実的な実施手順(procedure)の策定と実施 (deployment)が後追い
- 「オープン(データ)サイエンス」促進のため
 - 現在「研究データ管理」といえば, おそらくこの部分が中心
 - ファンドの要請に従い, データ管理プランを作成, 実施
 - 十分な研究インフラの提供(大容量ストレージ, 機関データレポジトリ)
機関の体力勝負
- 「学術領域の発展と社会貢献」のため
 - 主にコミュニティレベルで多大な労力, 時間をかけて醸成
 - 自発的な取り組みであるため, 高い専門性とデータ利活用を実現
 - 持続的な発展が大きな課題

「全学的=ロングテール」

- 「オープンデータ化」が必須事項となった場合・・・
 - 全学=(全教職員, 学生)に等しい基盤システムの導入
 - 可能な限りメンテナンスフリー, 汎用的な仕組みが低コストで必要



From “Establishing a generic Research Data Repository: The RADAR Service”, by Dr. Angelina Kraft, TIB at DI4r (<https://www.digitalinfrastructures.eu/content/establishing-generic-research-data-repository-radar-service>)

ドイツの事例: RADAR project (1)

信頼性の高い研究データアーカイブの共同開発

Goal: Establish an interdisciplinary repository for

- archival of (digital) research data as a generic service
- trustworthy preservation & traceable publication

Focus: „**Long Tail**“ – Repository for specialized research disciplines, addition to big data archives

Duration: September 2013 – August 2016,
project funded by German Research Foundation

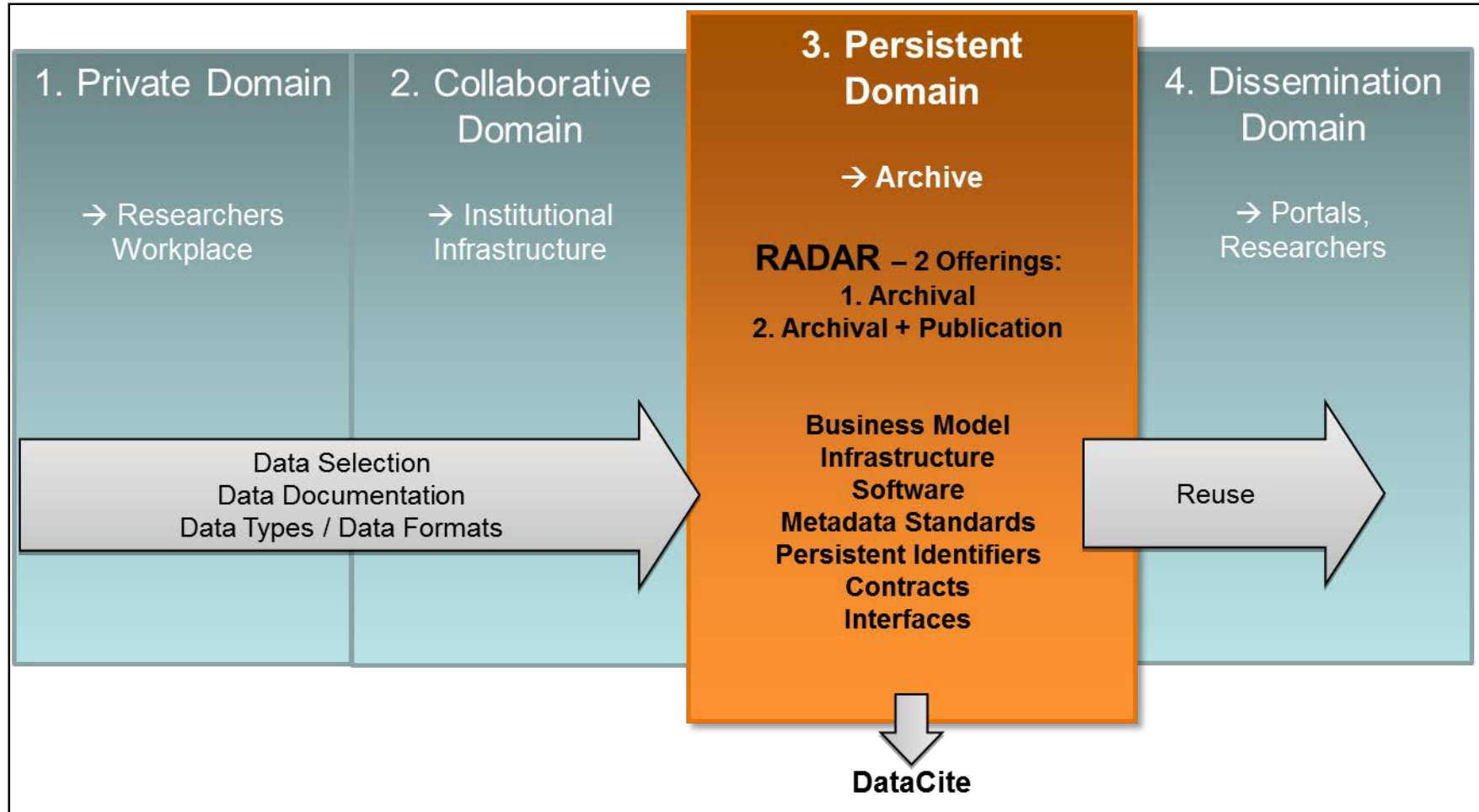
System: Provided by FIZ Karlsruhe, KIT and TIB – Coming soon... in 2016

Info: <http://www.radar-projekt.org>

Service: <http://www.radar-service.eu>

ドイツの事例: RADAR project (2)

研究データライフサイクル上の位置づけ



ドイツの事例: RADAR project (3)

ビジネスモデル

Services:

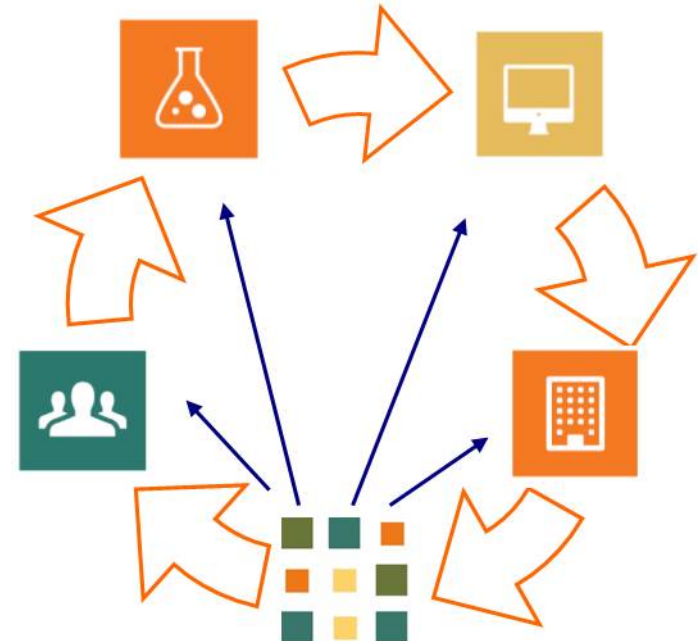
- Basic service: **Archival Storage** 
- Extended service: **Data Publication** 

Furthermore:

- Data Life Cycle support
- **REST API** for clients (customizable)
- Interoperability & **cross-linking** of published datasets via API: DataCite, ORCID & others
- Optional **Peer-Review support**
- Statistics on **downstream data use**

Prices:

- 500 € annual fee + 0,39 € GB data volume per year (net price)



京都大学では何を整備するか? (青写真)

- University of Edinburgh + α のものを目指すのであれば・・・
 - 研究者, 大学院生が自由に利用できるクラウドストレージ
 - アクティブな研究での学内外とのシームレスなデータ共有と協働環境
 - 2 or 3種類の機関データレポジトリ
 - 教員の裁量で自由にデータを公開できるレポジトリ
 - 非常に雑多な内容であるが, 永続的なIDが付与され, 「いつでも」「だれでも」参照できる仕組み
 - OA論文の延長線上としての, 整理されたオープンデータレポジトリ
 - 高度な専門性, ユーザビリティが要求される学術レポジトリ
 - 既存の学術レポジトリとの連携・漸進的な統合
 - 研究コミュニティの核となる先進的なデータレポジトリの構築・提供
 - データキュレーション, ビジュアライゼーション, アナリシス技能の飛躍的な拡大
 - 効率的なコンテンツ配信プラットフォーム(動画, 静止画, 地理情報・・・)
 - アーカイブストレージ, 教員総覧との連携
 - カバレッジ, 評価方法の模索
- コスト, 組織構成等は未考慮, 当然課題は多い

全学研究データマネジメントへの取り組み(1)

「多様な学術研究活動を育む全学研究データマネジメント環境構築事業」(学術情報メディアセンターからの全学経費要求)

参加者:

- 情報環境機構 ■■■
- 研究推進部研究推進課 ■
- 学術研究支援室 ■■
- 附属図書館 ■■
- 総合博物館 ■
- 学術データレポジトリ運用者 ■
- 他研究者有志 ■

「コンプライアンス」

来るべき
「オープンデータ化」

「学術領域の発展と社会貢献」

ミッション:

大学における学術研究のライフサイクルに沿った研究データの蓄積・共有・公開および長期保管に必要な研究データマネジメント環境に求められる要求要件のとりまとめ

全学研究データマネジメントへの取り組み(2)

活動内容

- 京都大学オープンサイエンスデータプロジェクトを組織化, 定期会合
- 学内調査
 - 研究者アンケートを実施 (11月頃)
- 海外調査
 - 米国 (8月下旬, 10月下旬)
 - 欧州 (9月下旬)
- 国内調査
 - 研究助成団体
 - 日本学術振興会 (JSPS), 科学技術振興機構 (JST), 情報通信研究機構 (NICT) など
 - 大学の枠を超えたあり方
 - 国立情報学研究所 (NII) など

まとめに代えて

- 京都大学における研究データ管理の現状を
 - 「コンプライアンス」
 - 「オープンデータ化」
 - 「学術領域の発展と社会貢献」に分類し, それぞれの要件を検討.
- いずれは, 「オープンデータ」という大きな概念として, ポリシー, 手順, システム等が集約されていくと考え, 準備を進める.
- オープンデータ義務化は大学全体のICTインフラの底上げに直結. 各組織が十分なサポート体制を構築できるかが課題

(宣伝) 大学ICT推進協議会 (AXIES) 年次大会で 研究データマネジメントのセッションを開催

- 日時: 2016年12月14, 15, 16日
(本セッションは16日13:30-15:00を予定)
- 会場: 国立京都国際会議場
- エジンバラ大学図書館 Dominic Tate 氏の招待講演
「Research Data Management in Europe, UK and the University of Edinburgh」
- 「京都大学における全学研究データマネジメント環境構築事業」中間報告
- 他 EDUCAUSE CEO, John O'Brien 氏による基調講演など
詳細は <http://axis.jp/ja/conf/conf2016> にて

謝辞

資料作成に当たり

- 京都大学「多様な学術研究活動を育む全学研究データマネジメント環境構築事業」の支援を受けました。
- 国立情報学研究所の山地一禎 准教授, 船守美穂 准教授には米国, 欧州の研究データ管理の現況についてご指導, 助言をいただきました。
- 快く調査に応じてくださった Univ. of Edinburgh の
 - Kevin Ashley 氏
 - Dominic Tate 氏
 - Tony Mathys 氏
 - Robin Rice 氏
 - Anthony Weir 氏に感謝いたします。