

大学における データサイエンスとその教育

安浦寛人

九州大学 理事・副学長



KYUSHU UNIVERSITY

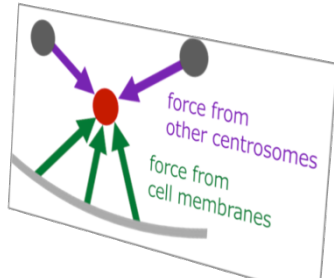


第4の科学的方法論としてのデータ科学

	実世界	サイバー世界
演繹的手法	理論科学	計算科学
帰納的手法	実験科学	データ科学

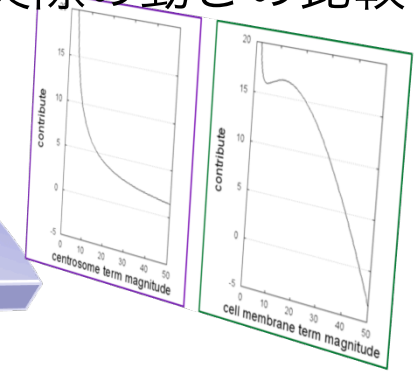


①細胞中心体の動きメカニズムの理論

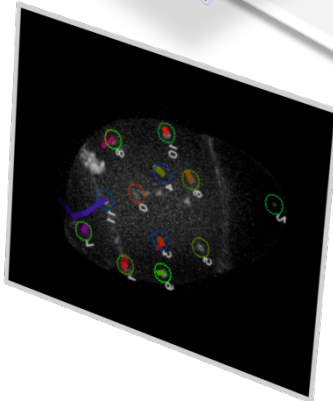


⑤動きメカニズムの理論の新たな展開

④数理モデルに基づくシミュレーションと実際の動きの比較



②実際の動きの画像観察および動きの自動定量化

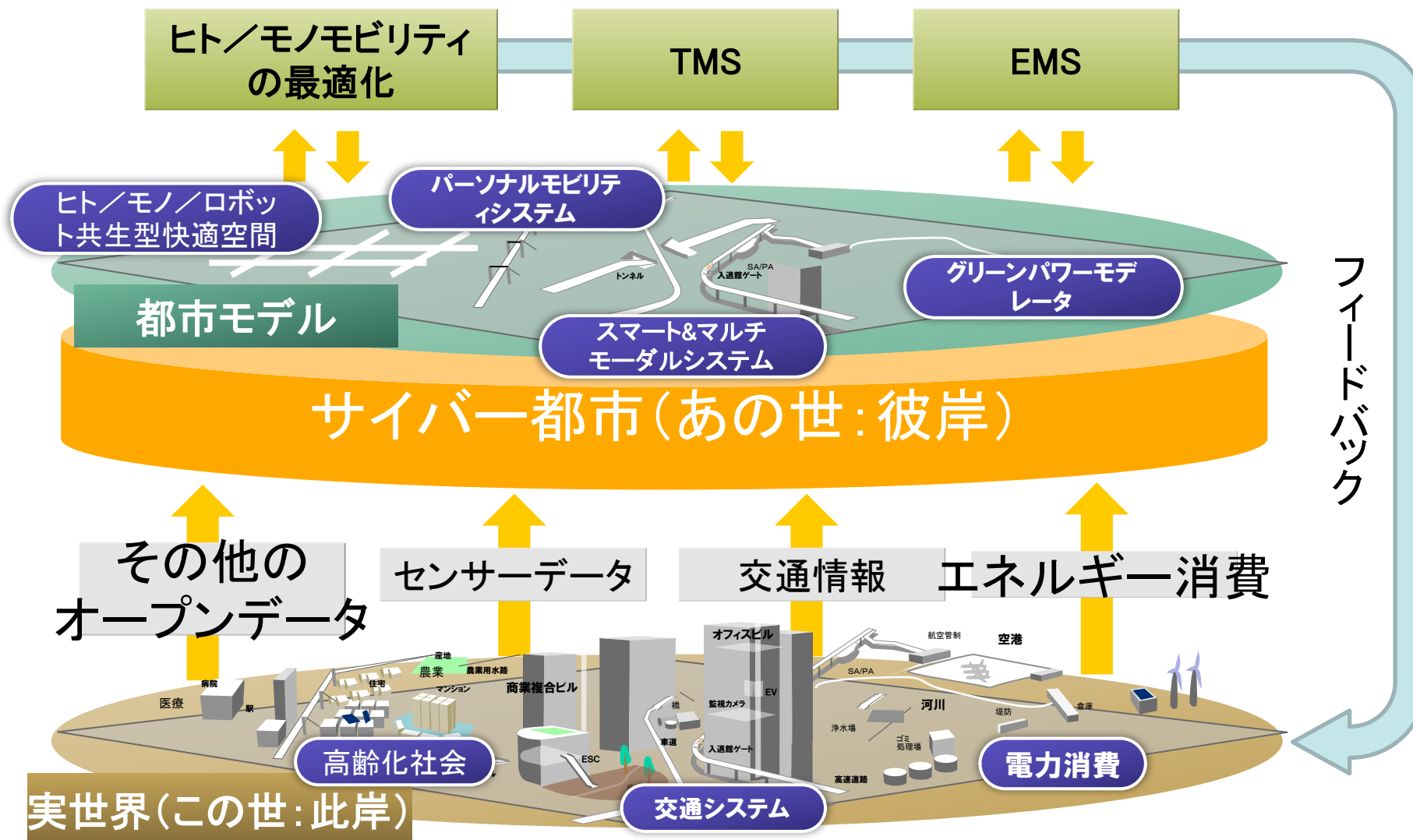


③動きの定量化結果の数理モデル化

$$\Delta P_i = \sum_n \left(a_n \sum_{j \neq i} |D_{ij}|^{\alpha_n} \left(-\frac{D_{ij}}{|D_{ij}|} \right) \right) + \sum_m \left(b_m \sum_k |D_{i,mem_k}|^{\beta_m} \left(-\frac{D_{i,mem_k}}{|D_{i,mem_k}|} \right) \right)$$

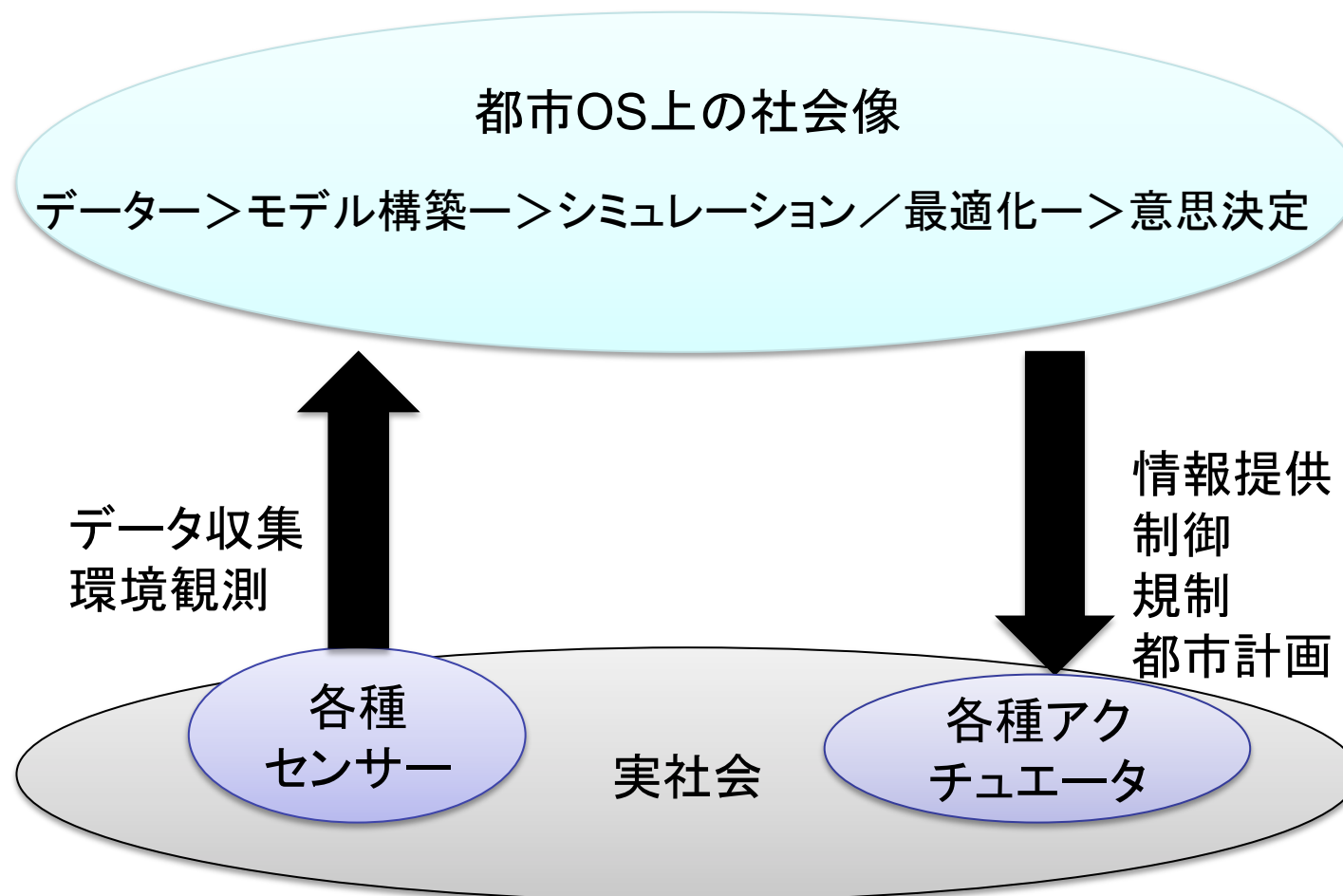


工学でも：都市の設計（都市OS）

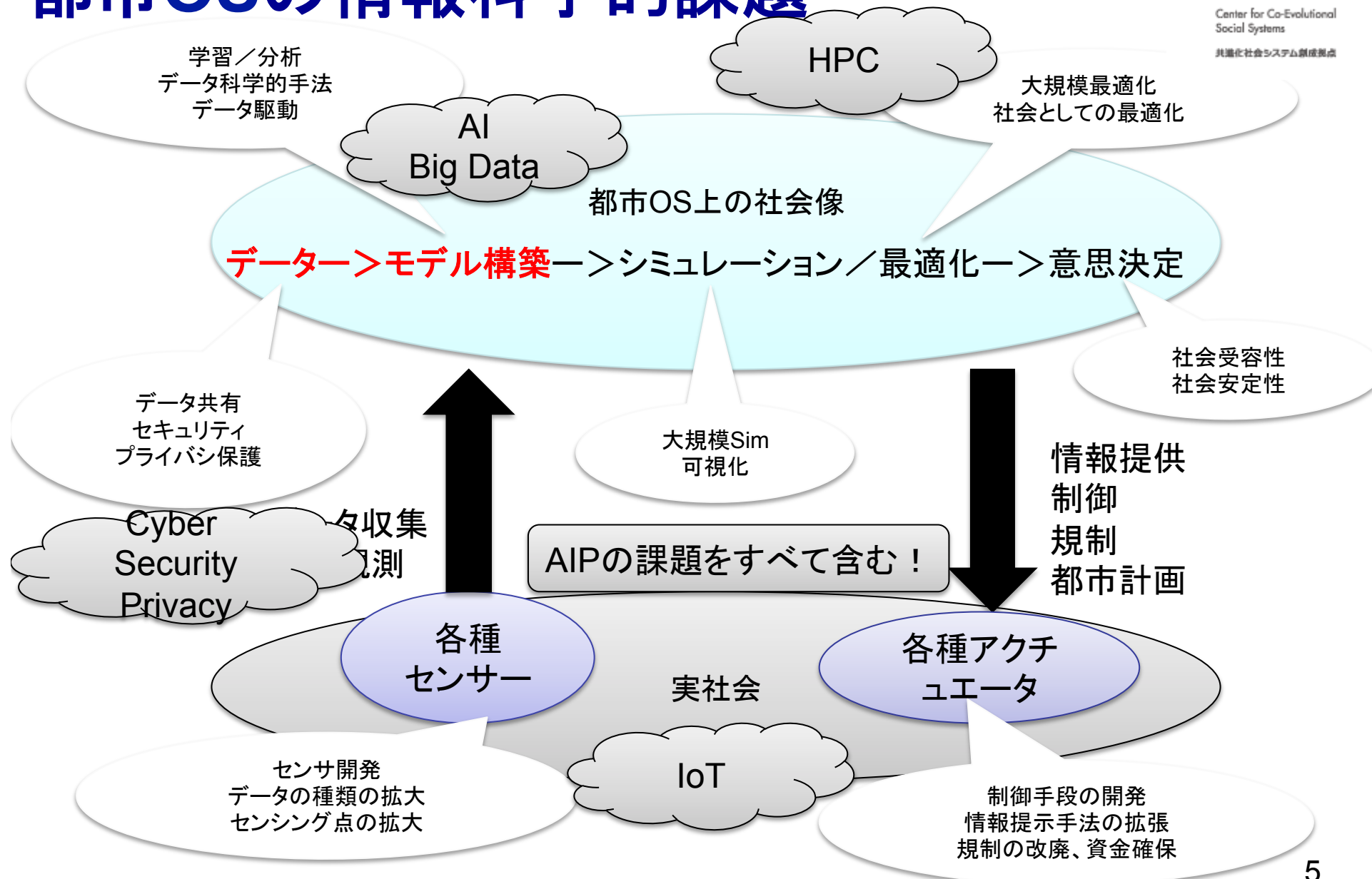




都市OSの概念



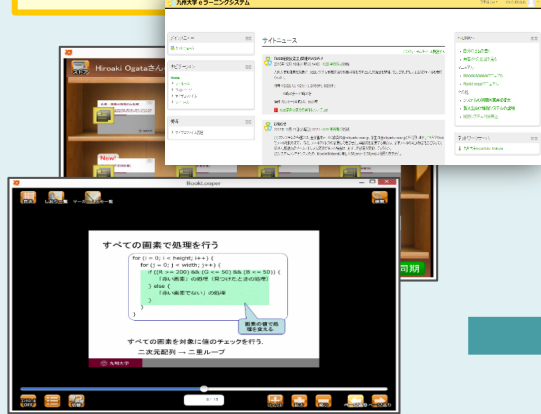
都市OSの情報科学的課題



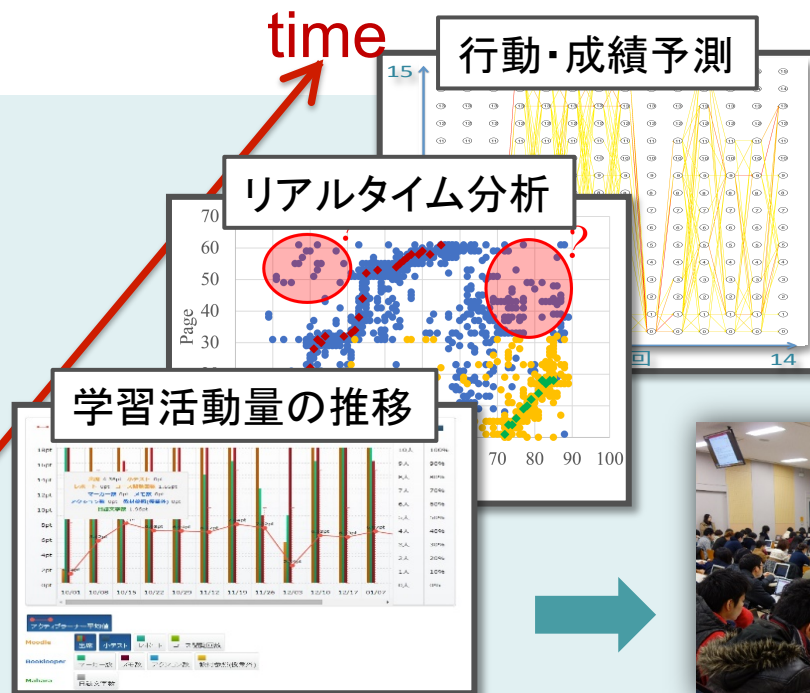


- 社会のあらゆる分野で、情報通信技術に基づくイノベーションが進行しており、この波は「教育」の分野にも当然及ぶ
- 「デジタル教科書」やe-Learning等の教育情報システムを用いて、生涯にわたる学習ログを蓄積し、そのデータの科学的な分析に基づき、教育・学習をサポートする手法の研究・開発が急務
- 「教育データ科学」の学問領域を創設して、デジタル時代の「学び」を解明し、教育の抜本的な改革を目指す

H25年:九州大学PC必携化
H26年:M2Bシステムの導入
現在約3万人、718コースで利用中
1日約**18万件**のログが蓄積



LMS, e-portfolio, e-Book



Learning/teaching analytics tools



Feedback



M2B(みつば)学習支援システム

- ① Moodle(ムードル) : e-Learningシステム
出席管理・レポート管理・掲示板、アンケートなど
- ② Mahara(マハラ) : eポートフォリオシステム
講義日誌の記録による振り返りと情報共有
- ③ BookLooper(ブックルーパー) : デジタル教科書配信システム
講義スライドの共有と配信



1日18万件のログ

利用状況

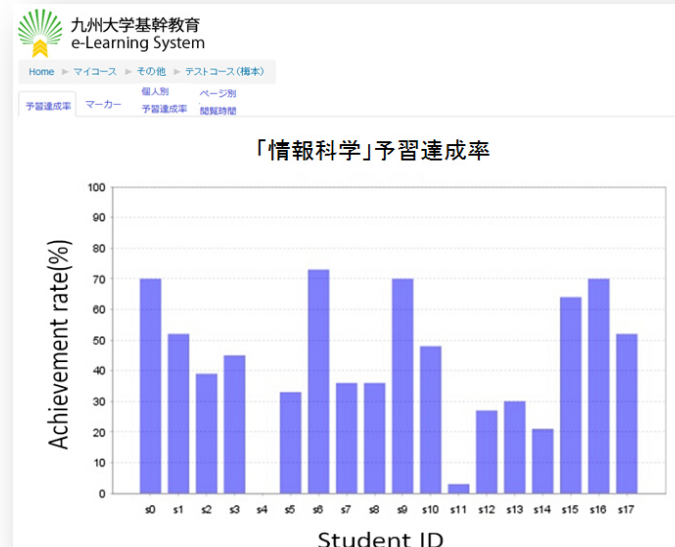
	学生登録数	教員登録数	Moodle	Mahara	BookLooper
2015年度 前期	2,687人	10,490人	206コース	866日誌	132教材
2015年度 後期	19,293人	10,490人	112コース	302日誌	95教材

デジタル教科書の閲覧ログの分析

予習の達成度、マーカー数等



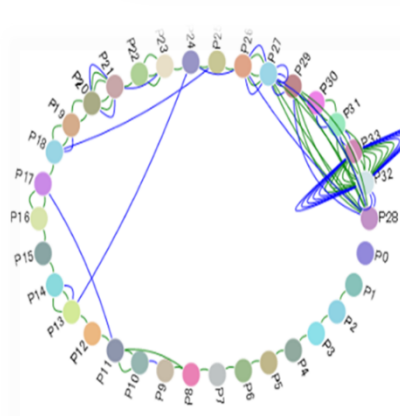
授業内容を適応的に変更



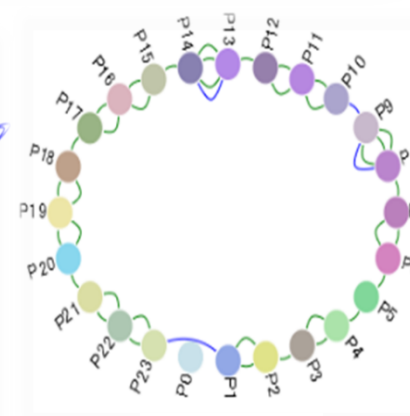
ページの遷移を可視化



閲覧パターンから教材の
改善点を提案



教材A



教材B

教育用ビッグデータを用いた教育・学習の改善

シラバス作成、教材作成、演習問題、
 レポート課題、試験問題等の作成

シラバス確認、参考資料等の事前確認



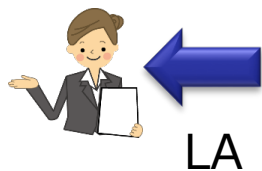
授業期間後の
ワークショップ



ティーチングアドバイザー

- ・ 学習ログ分析
- ・ 教材改善提案

主観や経験に依存しない



LA

教育ビッグデータ

- ・ eポートフォリオ
- ・ eラーニングのログ
- ・ e-Bookの閲覧ログ
- など



LA

ラーニングアドバイザー

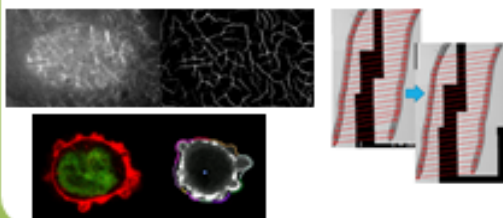
- ・ 学習ログ分析
- ・ 成績予測



データ科学はあらゆる分野に関連する

分野を超えたデータサイエンス研究@九大

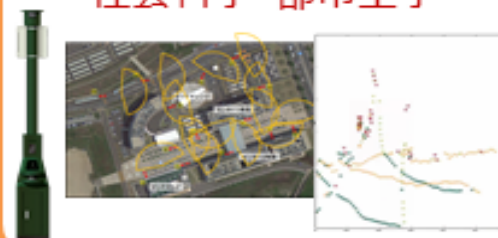
バイオイメージングフォーマティクス



ロボティクス



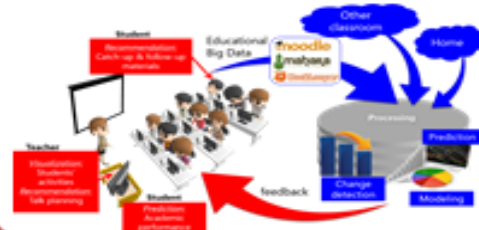
社会科学・都市工学



アグロ・インフォマティクス



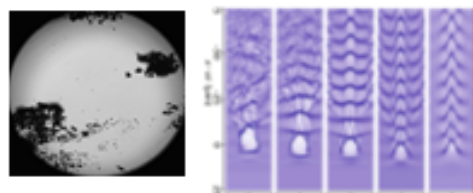
ラーニング・アナリティクス



マス・フォア・インダストリ研究所とのサイバーセキュリティ研究



画像情報学と物理学の連携



情報薬と疾病マネジメント



人文科学





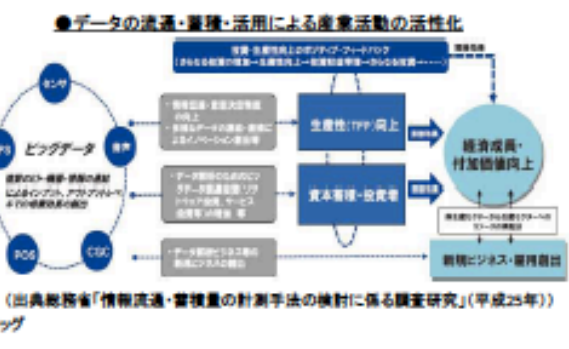
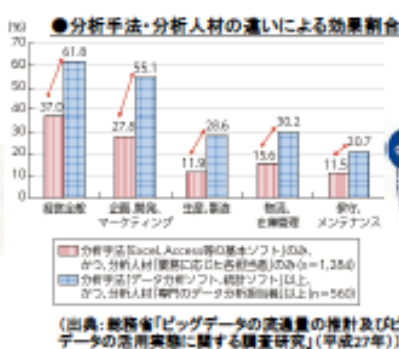
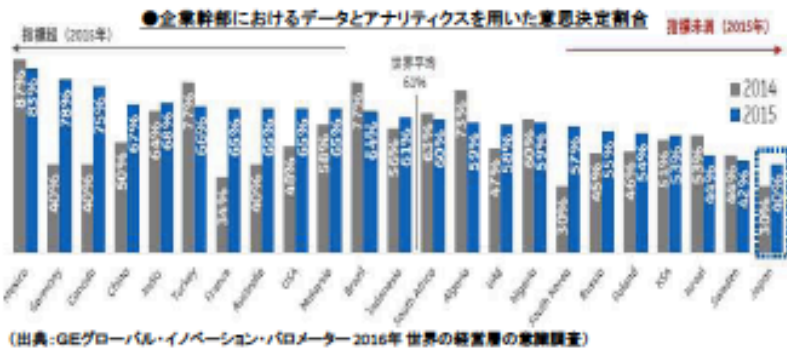
大学の数理・データサイエンスに係る教育強化（概要）

現状

- 膨大なデータが溢れる時代において、諸外国と比較すると企業では意思決定におけるデータとアナリティクスの活用に遅れをとっている状況。
- 世界に先駆けた「超スマート社会」の実現（Society5.0）に向けて、我が国の産業活動を活性化させるために必要な数理・データサイエンスの基礎的素養を持ち、課題解決や価値創出につなげられる人材育成が必要不可欠。

○我が国の企業幹部におけるデータの分析・活用の戦略的価値への認識は、世界の主要国の水準と比べて非常に低い。

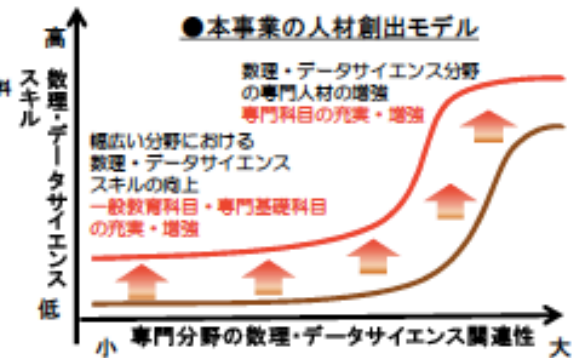
○数理的思考やデータ分析・活用能力を持つ人材が戦略的にデータを扱うことによる経営等への効果は大きい。



専門分野の枠を超えた全学的な数理・データサイエンス教育機能を有するセンターを整備し、専門人材の専門性強化と他分野への応用展開の双方を実現し相乗効果を創出

実現に向けたシナリオ

- ✓ 文系理系を問わず、全学的な数理・データサイエンス教育を実施
- ✓ 医療、金融、法律等の様々な学問分野へ応用展開し、社会的課題解決や新たな価値創出を実現
- ✓ 実践的な教育内容・方法の採用
 - ・企業から提供された実データ等のケース教材の活用
 - ・グループワークを取り入れたPBLや実務家による講義等の実践的な教育方法の採用
- ✓ 標準カリキュラム・教材の作成を実施し、全国の大学へ展開・普及





九州大学におけるデータサイエンス教育と研究

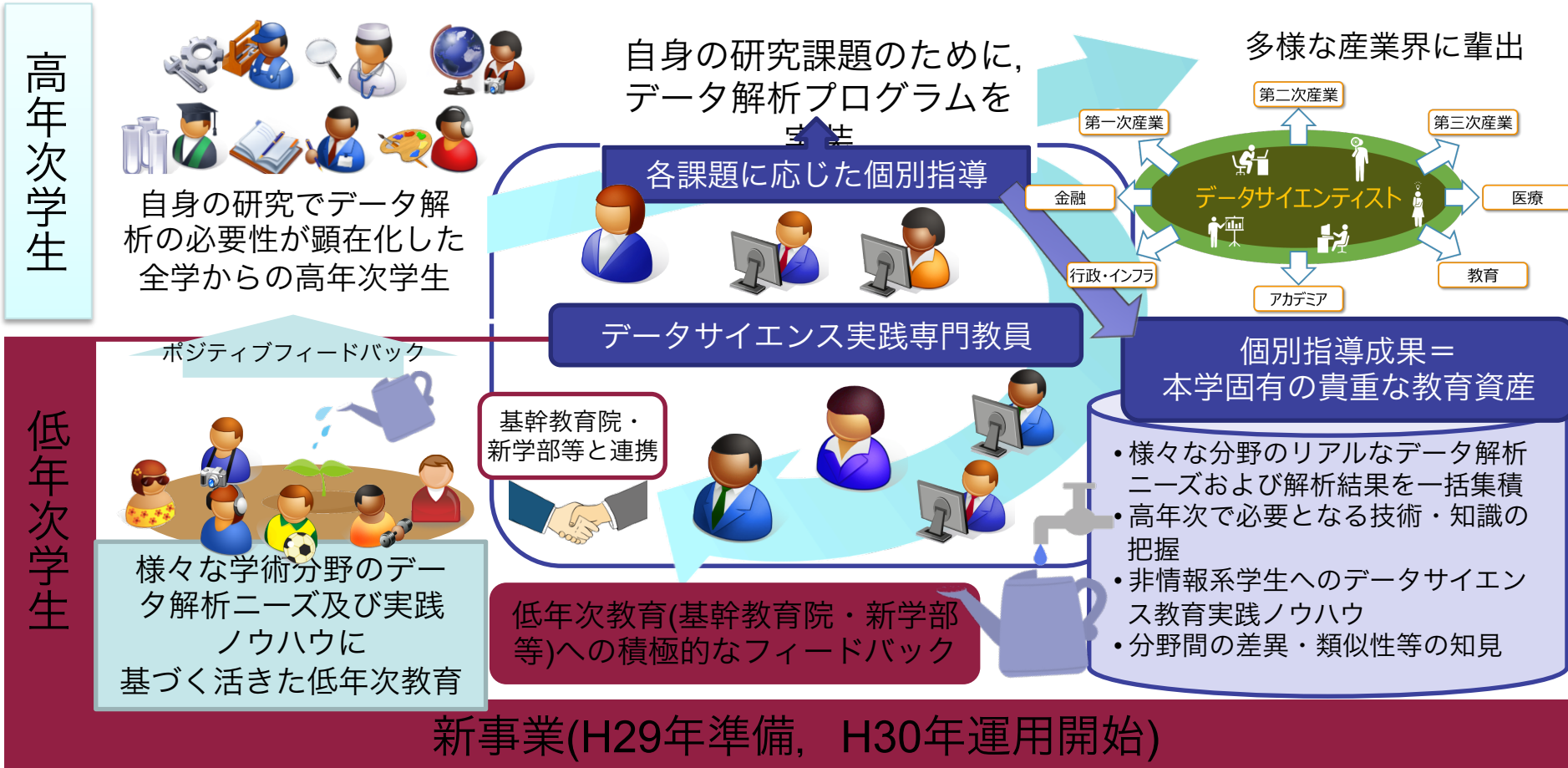
背景

データサイエンス教育の必要性が顕在化
 産業界へのデータサイエンス人材輩出責務
 データサイエンス実践特別講座開講(H29-)
 そこで蓄積されるノウハウは全学で活用されるべき

新事業

データ解析ニーズおよび実践ノウハウを全学的に活用
 基幹教育院・新学部等、低年次教育へフィードバック
 低年次から実践センスを身に着けた学生を育成
 データサイエンス教育のポジティブフィードバックループの完成

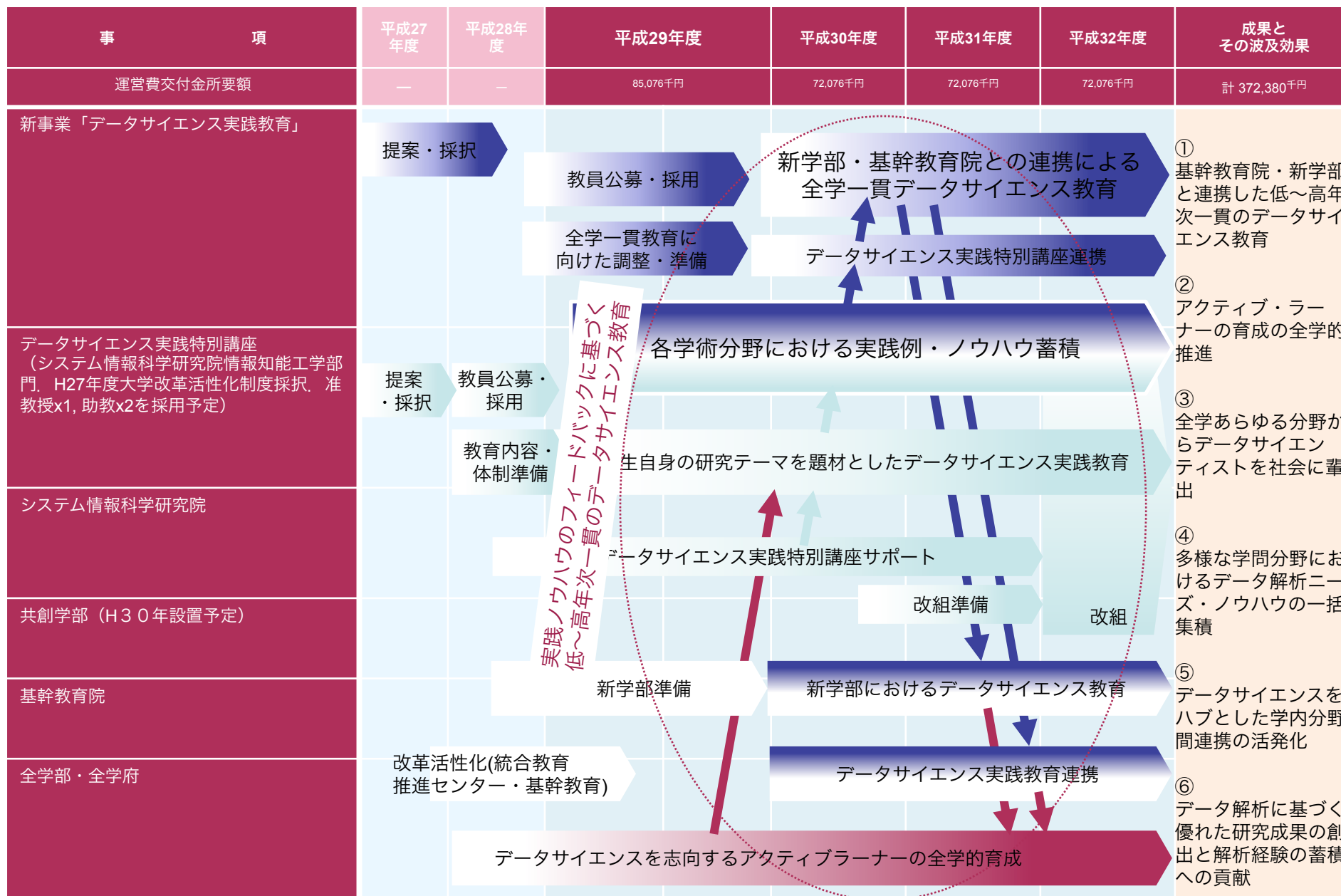
H27年度「大学改革活性化」採択課題(H28準備, H29開講)



「ニーズ及び実践ノウハウのフィードバックに基づく低～高年次一貫型のデータサイエンス教育」の完成



データサイエンス教育のロードマップ（案）

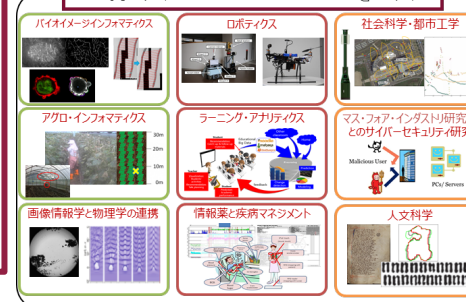




目標

- ・既存科目を再整理し、初年次から大学院まで**階層別の科目をマッピング化** (H30まで)
- ・初年次基幹教育で、**毎年2,500名が数理・情報リテラシーを醸成** (H31まで)
- ・マッピングした科目群から**各専門分野における推奨科目を設定** (H31まで)
- ・高年次基幹教育で、様々な分野に対応した数理・データサイエンス科目を開発 (H32まで)
- ・**数理・データサイエンス教材を開発し、他大学での活用を普及** (H32まで)
- ・新学部 (H30設置) においても学部段階の数理・情報科目をモジュール化 (H33まで)

分野を超えたデータサイエンス研究@九大



データサイエンティストを多様な産業界に輩出



分野を超えた数理・データサイエンス研究

※大学院レベルの実践科目群を社会人科目として解放

学府横断型の「数理・データサイエンス教育プログラム」(副専攻)

例えば、企業との共同教育、異分野学生のPBL実習、海外長期インターンシップなど、より高度な科目モジュールを意欲のある学生に提供

グローバル人材の育成

基幹教育院による全学向け科目モジュール「数理・データサイエンス実践基幹教育Ⅱ」

例えば、文理に応じた、数理・データサイエンスの基礎・応用科目のモジュールを開発し、高年次基幹教育科目として提供

専門分野に関連した知識・技術の涵養

基幹教育院による全学向け科目モジュール「数理・データサイエンス実践基幹教育Ⅰ」

例えば、数学、統計学、情報リテラシーなどの既存基礎科目を整理し、基幹教育科目のモジュールとして低年次学生を対象に広く提供

低年次学生にリテラシーの醸成

大学院

高年次

低年次

各学科
各専攻
専門科目

専門
性の追
求

なぜ、今、データサイエンスか？

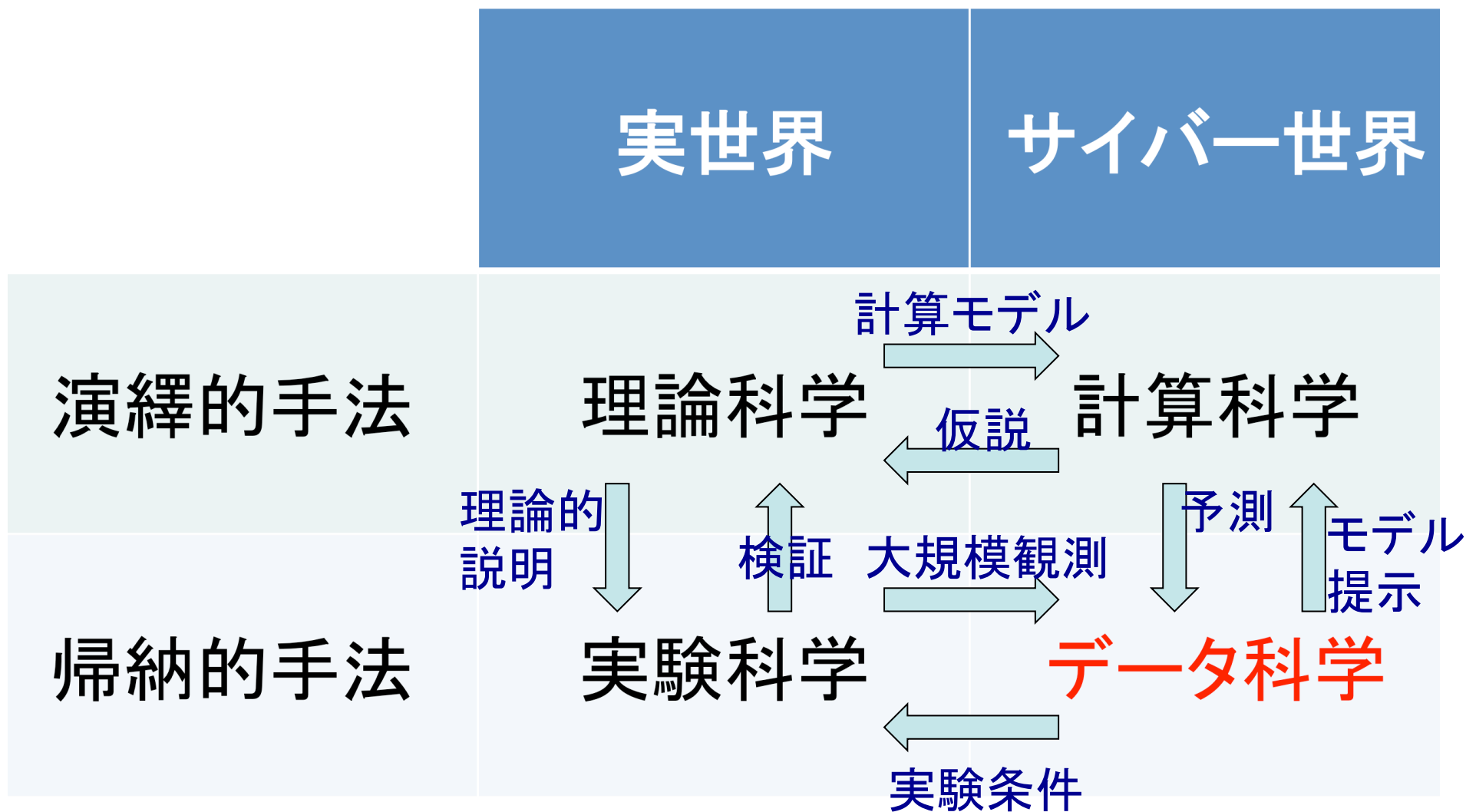
学術的成果の客観的定量化が非常に重要視

- 扱うべきデータ量の爆発的増加
- 解析目的の複雑化・多様化





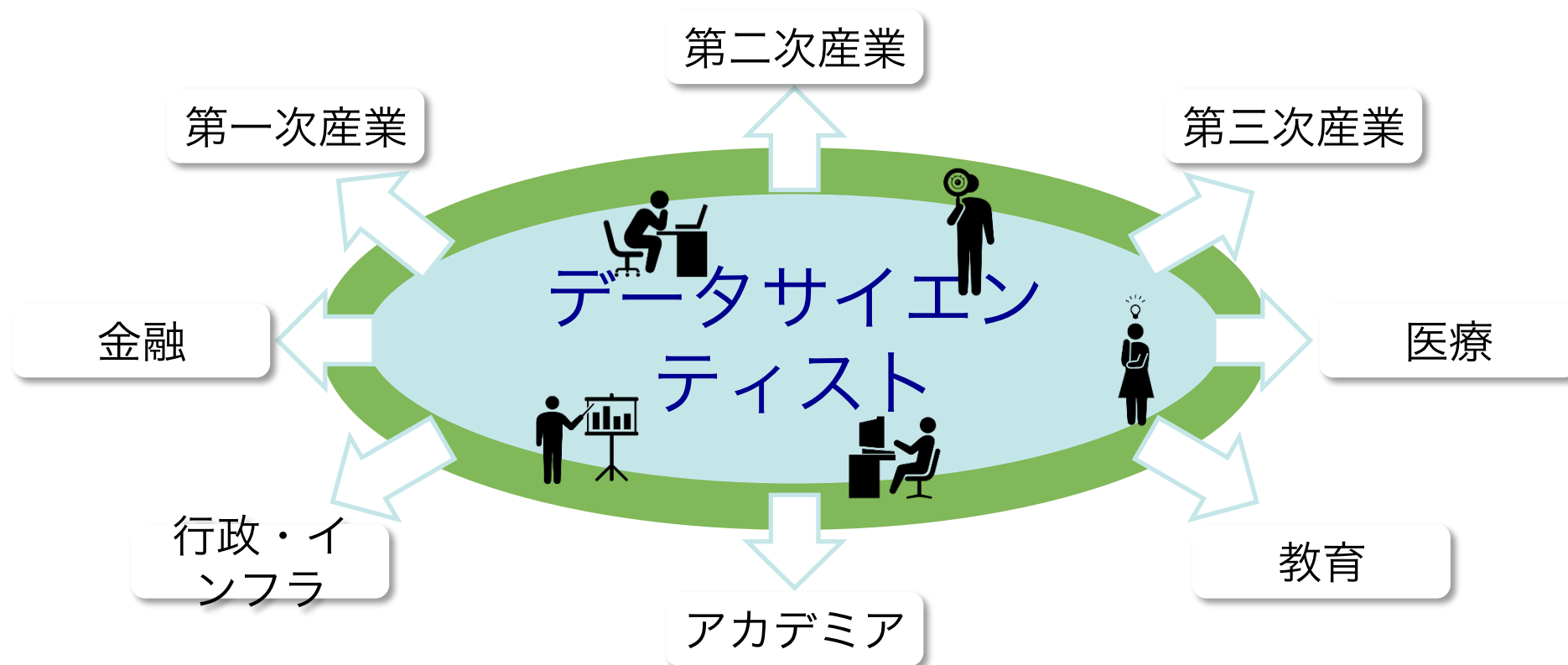
4つの方法論の連携が始まった





社会もデータサイエンティストを求めている

「データサイエンティスト」の輩出が総合大学に対する社会的要請となっている





データサイエンス実践講座が提供する教育

そのために データ解析の基本技術の原理を学ぶ

そのために データ解析の基本技術をどう使うかを学ぶ

そのために データ解析プログラム実装の予行演習

各学生の研究課題のためのデータ解析
プログラムを実装





4年生
大学院生



画像データという、一般的で直感に訴える題材を、積極的に利用



教育効果を高めるために： 分野の壁を越えるための4つの工夫

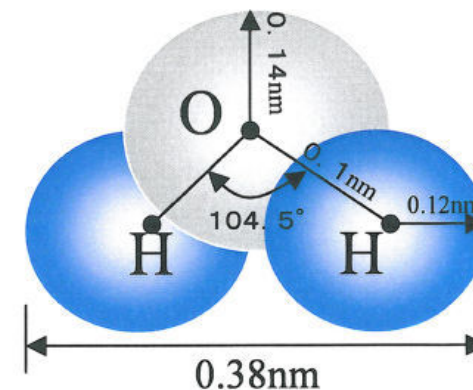
1. 理論教育よりも、データ解析の実践に重点 
2. 画像を練習題材することで直感に訴える 
3. 自分自身の研究内容が題材なので高モチベーション 
4. 専門教員による個別指導の場の提供 

九大発の新しい教育フレームワーク



データサイエンスはあらゆる科学・技術の基礎

- ビッグデータ時代
 - データ量はZetta Byte (10^{21})
 - データに基づく科学的決断
- オープンデータ時代
 - データの公開と共有
 - 人類の知的資産の一部
 - データの著作権
 - プライバシー問題
 - 倫理問題



水の分子構造モデル



太陽と木星の距離778,000,000Km



オープンサイエンスと研究データの管理

オープンサイエンス

- 論文だけではなく研究データもオープンにして、研究の公正性や成果の再利用性を高めようとする、新しいサイエンスの進め方。

研究データを、

- 公開しなければならないのは研究者の責任。
- 保全する環境を整備するのは研究機関の責任。
- 流通させるサポートをするのは図書館の責任。

(ICSU-IAP-ISSC-TWAS working group, Open Data in Big Data World, 2015年12月 より改変)

研究助成団体

- JST: OA(Open Access)を推奨するポリシー → 義務化+研究データについても言及するポリシーへの変更を検討。
 - JSPS: OAなどのポリシーについて検討を開始。
 - AMED: データシェアリングポリシー(義務化)の施行。
- ## 文部科学省、学術会議
- 研究不正対策のために研究資料の10年間の保存を原則。
 - データ・バックアップ用サーバーの提供などインフラ整備は機関の責任。



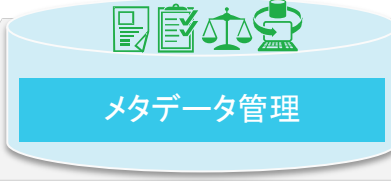
研究データ基盤

提供:NII 山地

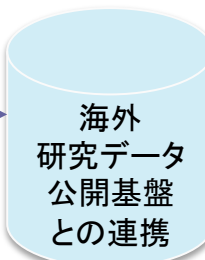
データ検索基盤

- 論文・研究データ関連付け名寄せ機能
- 研究者・研究プロジェクト成果管理機能
- 海外ディスカバリサービス連携機能

CiNii for Data

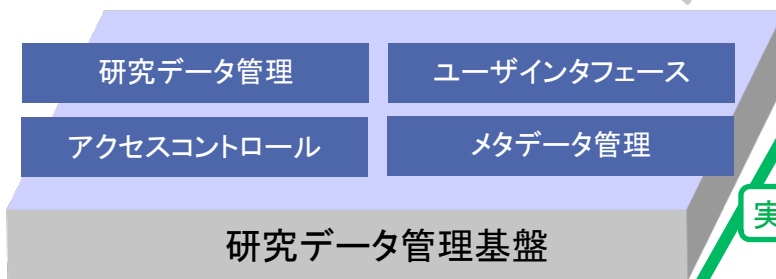
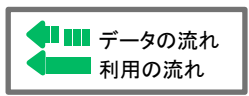


ディスカバリサービス



再利用

メタデータ収集



検索・利用

データ利用者

データ登録者

実験・保存

蓄積

実験データ

成果論文

成果論文

研究データ

機関向け研究データ管理

研究データ用リポジトリ



データ管理基盤

- 学術ネットワークへの高速アクセス機能
- 認証連携対応データ共有機能
- クラウド活用高効率データ保存機能



非公開 共有 公開



長期保存対応ストレージ領域

データ公開基盤

- 研究データ向け高速・簡便セルフアーカイブ機能
- バージョン管理・自動パッケージング機能
- データ提供先に応じたプライバシー情報匿名化機能



for Data₂₃



データサイエンスによる新しい時代

- 人類が経験したことが無い時代
- 教育や研究に新しい発想が必要
- データに基づく新しい科学哲学、倫理、社会の指導原理
- 科学や技術と社会の調和の探求



ご清聴ありがとうございました

九州大学

KYUSHU UNIVERSITY

<http://www.kyushu-u.ac.jp>