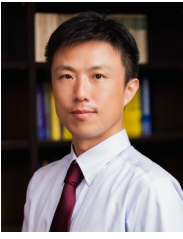


データ記述科学の創出と諸分野への横断的展開（データ記述科学）



領域代表者	京都大学・高等研究院・教授	研究者番号:10432709
	平岡 裕章（ひらおか やすあき）	
研究領域情報	領域番号：22A201 キーワード：データ記述科学、数学、機械学習	研究期間：2022年度～2026年度

なぜこの研究を行おうと思ったのか（研究の背景・目的）

● 研究の全体像

現代の社会にはデータが氾濫しているが、AI技術のブラックボックス問題に見られるように、それらが内包する真の価値を十分に活用できていないと言いき難い。このようなビッグデータを有効利用するには、データの本質となる構造を適切に記述し、その記述言語を用いて現象の背後にある機構に意味を与え理解するプロセスが不可欠である。本領域では、最先端の数学とデータ科学手法を用いてデータの「かたち」と「うごき」に着目した記述子（＝数学言語）を開発することで、この問題を解決する。そのために数学・データ科学・応用の三位一体体制で領域を構成し、既存の学問分野の枠を凌駕する新融合領域「データ記述科学」を創成する。この活動を通じて、データ活用社会の未来実現を飛躍的に加速させることを目指す。

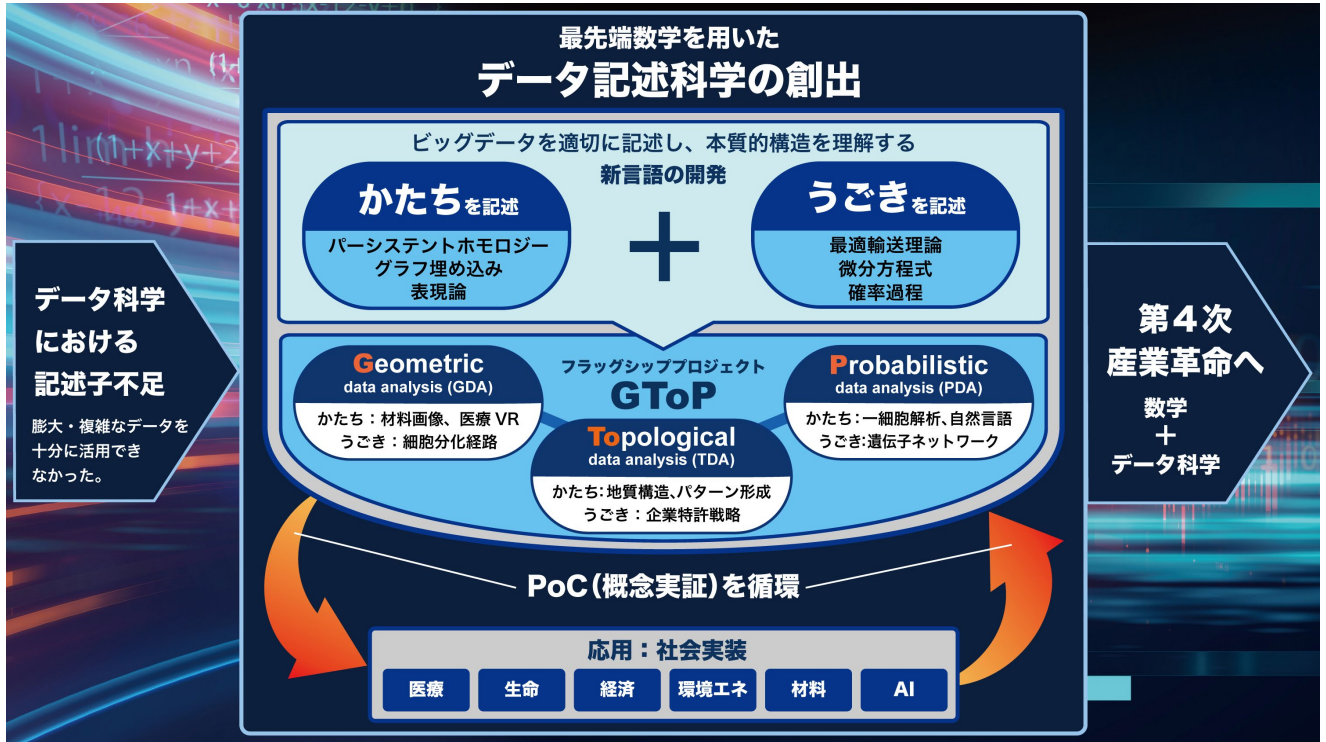


図1 データ記述科学の取り組み

● 研究体制

本領域では3つの研究項目を設けている。研究項目A01ではデータ記述科学の理論基盤構築を目指した数学およびデータ科学研究を、研究項目A02ではデータ記述科学の新分野探索研究を、研究項目A03ではデータ記述科学を材料科学および生命科学に展開する研究をそれぞれ実施する。また理論研究と応用研究を高いレベルで融合させるために、数学・データ科学・応用からなるPOC（Proof of Concept）サイクルを研究の基本ユニットとし、ユニットで設定された共通課題に三者が協働して解決を目指す体制を構築している。

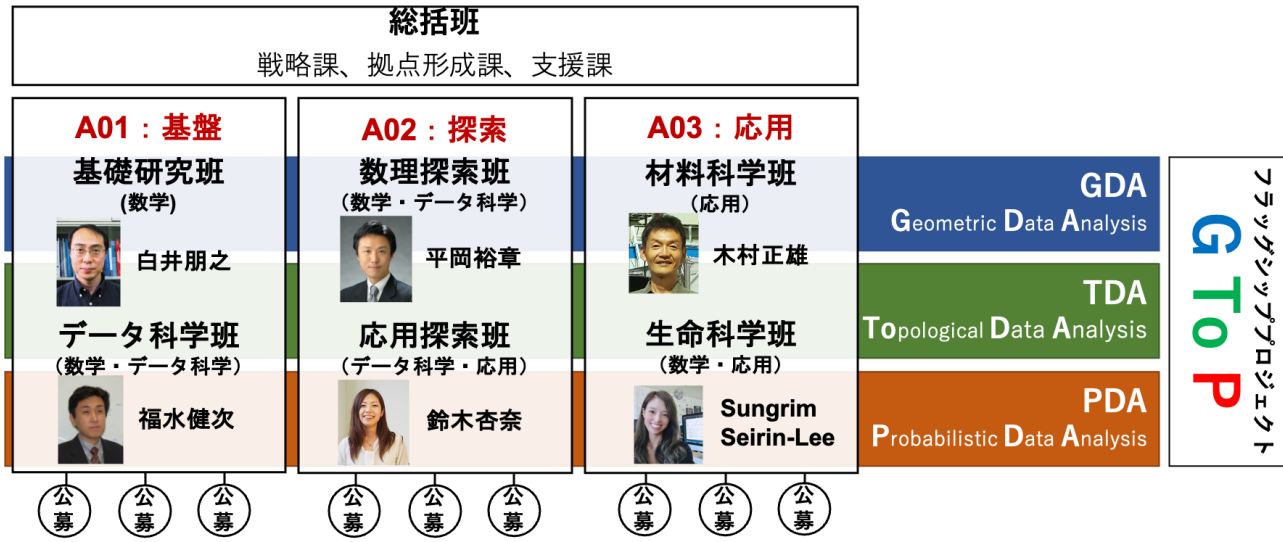


図2 チーム構成

この研究によって何をどこまで明らかにしようとしているのか

●フラッグシップ・プロジェクト「GToP (ジートップ)」

本領域で目指すデータ記述科学の創出に向けて、最も重要な構想がフラッグシップ・プロジェクト「GToP」である。GToPは

- GDA (Geometric data analysis)
- TDA (Topological data analysis)
- PDA (Probabilistic data analysis)

をデータ解析手法の3つの軸とし、本領域の全ての研究課題はこれらの1つ以上に貢献する形で企画される。GToPの基本構想は、データの「かたち」とその「うごき」を幾何学的もしくはトポロジカルに記述し (GDA、TDA)、そのうえでデータのランダム性がここで記述された幾何・トポロジー構造に与える影響を特徴づける (PDA) ことである。このGDA、TDA、PDAの3つの軸に対して「かたち」と「うごき」に関する数学・データ科学・応用の研究テーマを実施し、さらにこれら3つの相互作用が生み出す相乗効果を発揮できる体制を整える。本課題ではこのような多岐にわたる数学分野を包括的にデータ科学と融合させることで、より強固でかつ普遍的なデータ記述法を開発し、その結果幅広い応用展開が可能となる。またこれらの数学理論を用いたデータ記述子は、データの構造を取り出すことが可能であり、従来の統計的特徴量をはるかに超えた能力を有する。

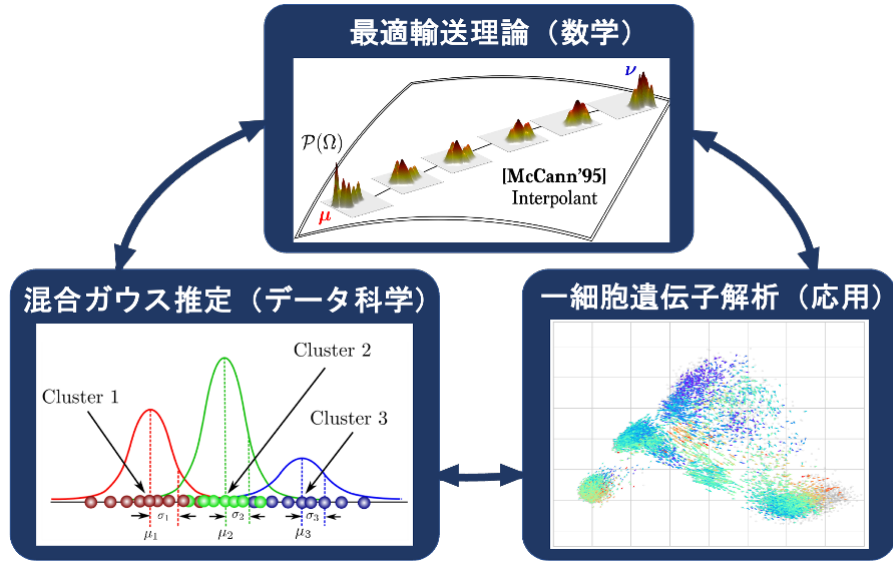


図3 PoCサイクルの例