

RadonPyプロジェクトの紹介とマテリアルDX推進への提言

吉田 亮

統計数理研究所 先端データサイエンス研究系 教授
マテリアルズインフォマティクス研究推進センター センター長

データ駆動型材料研究の最大の壁：データ資源の不足

オープンデータの不足

文化的問題 短期的解決は困難

コスト

実験・シミュレーション・試料作製・組成解析・構造解析・物性評価に要するコストが高い。

多様性
階層性

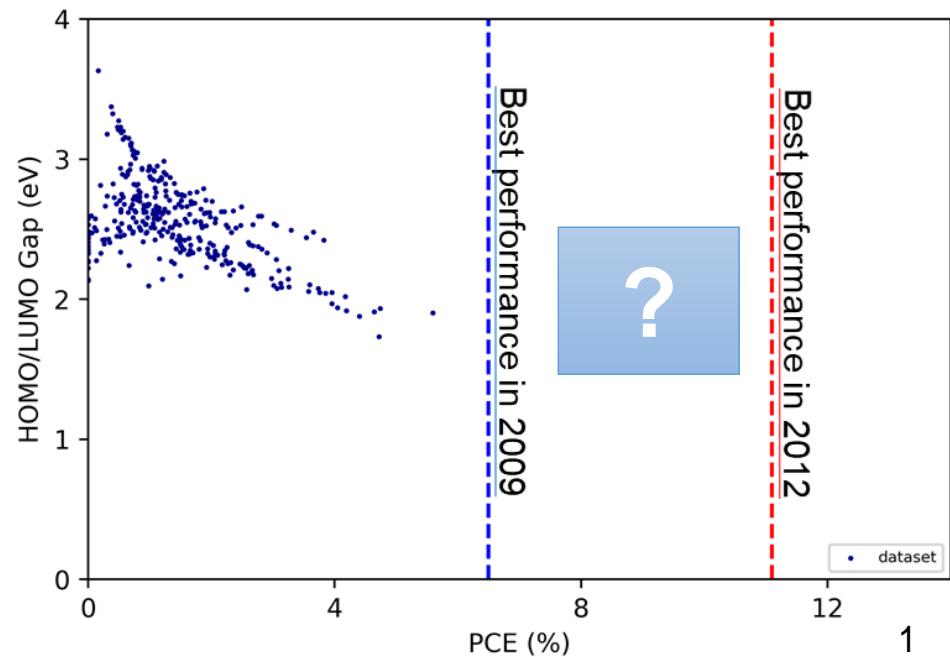
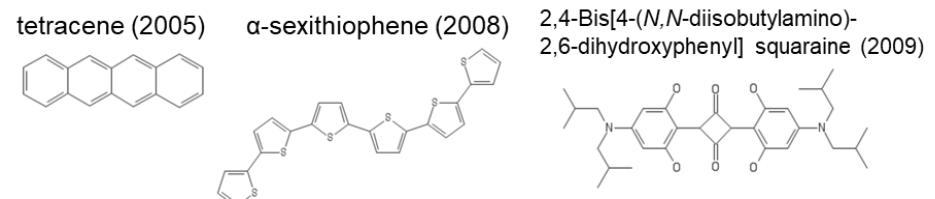
多くの物性がプロセス依存で決まる高次構造に依存するため、物質空間が広大且つ階層的に共通基盤データを構築しにくい。

インセン
ティブ欠如

競合相手に対する情報秘匿の意識が高い。
データを公開するインセンティブが研究者に働きにくい。

革新的材料の周辺にはデータがない

データ科学の内挿的予測の限界

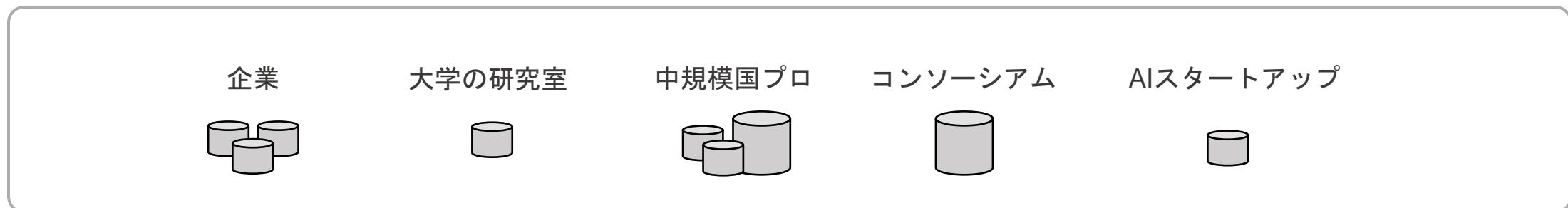


データ駆動型材料研究の在り方とミッション定義

【オープン領域】データ資源の不足を克服するために必要な体系的なコモンデータを整備



【クローズ領域】十分なデータを保有することは困難





RadonPy

RadonPy: 高分子物性計算機実験の全自動化

林慶浩
(ISM)

Hayashi et al., RadonPy: automated physical property calculation using all-atom classical molecular dynamics simulations for polymer informatics. *npj Comput Mater* 8, 222 (2022).

様々な物性や系の計算機実験を全自動化

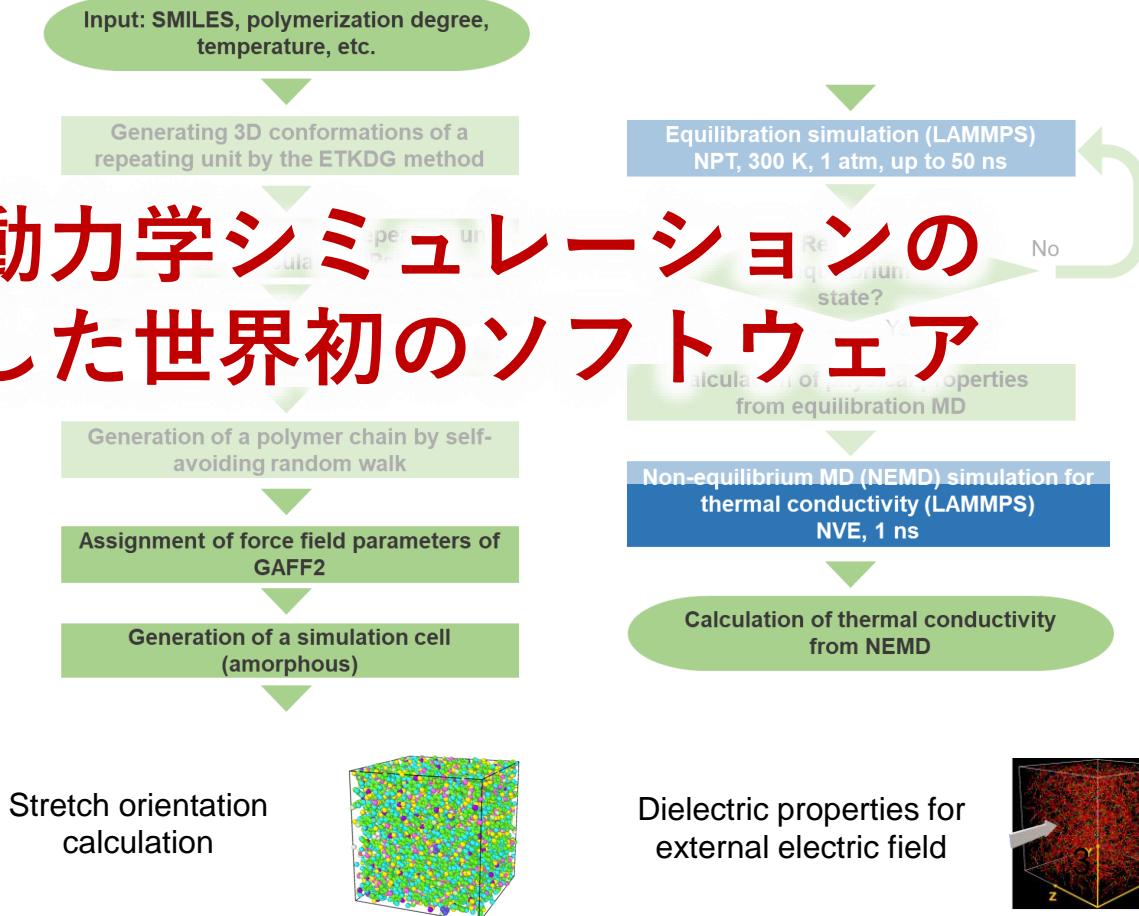
Latest: 34 properties implemented (2024/05/13)

- ◆ Thermal conductivity
- ◆ Thermal diffusivity
- ◆ Density
- ◆ Radius of gyration
- ◆ Specific heat capacity Cp
- ◆ Specific heat capacity Cv
- ◆ Compressibility
- ◆ Isentropic compression ratio
- ◆ Bulk modulus
- ◆ Isentropic bulk modulus
- ◆ Self-diffusion coefficient
- ◆ Thermal expansion coefficient
- ◆ Linear expansion coefficient
- ◆ Dielectric constant (static)
- ◆ Refractive index
- ◆ Solvation free energy
- ◆ Solubility parameters
- ◆ Glass transition temperature
- ◆ Abbe's number
- ◆ Order parameter
- ◆ Dielectric constant
- ◆ Dielectric loss tangent
- ◆ Dipole moment (molar)
- ◆ Thermal conductivity_pe
- ◆ Thermal conductivity_pair
- ◆ Thermal conductivity_bond
- ◆ Thermal conductivity_angle
- ◆ Thermal conductivity_dihed
- ◆ Thermal conductivity_improper
- ◆ Thermal conductivity_kspace
- ◆ Chi parameters

Polymer systems

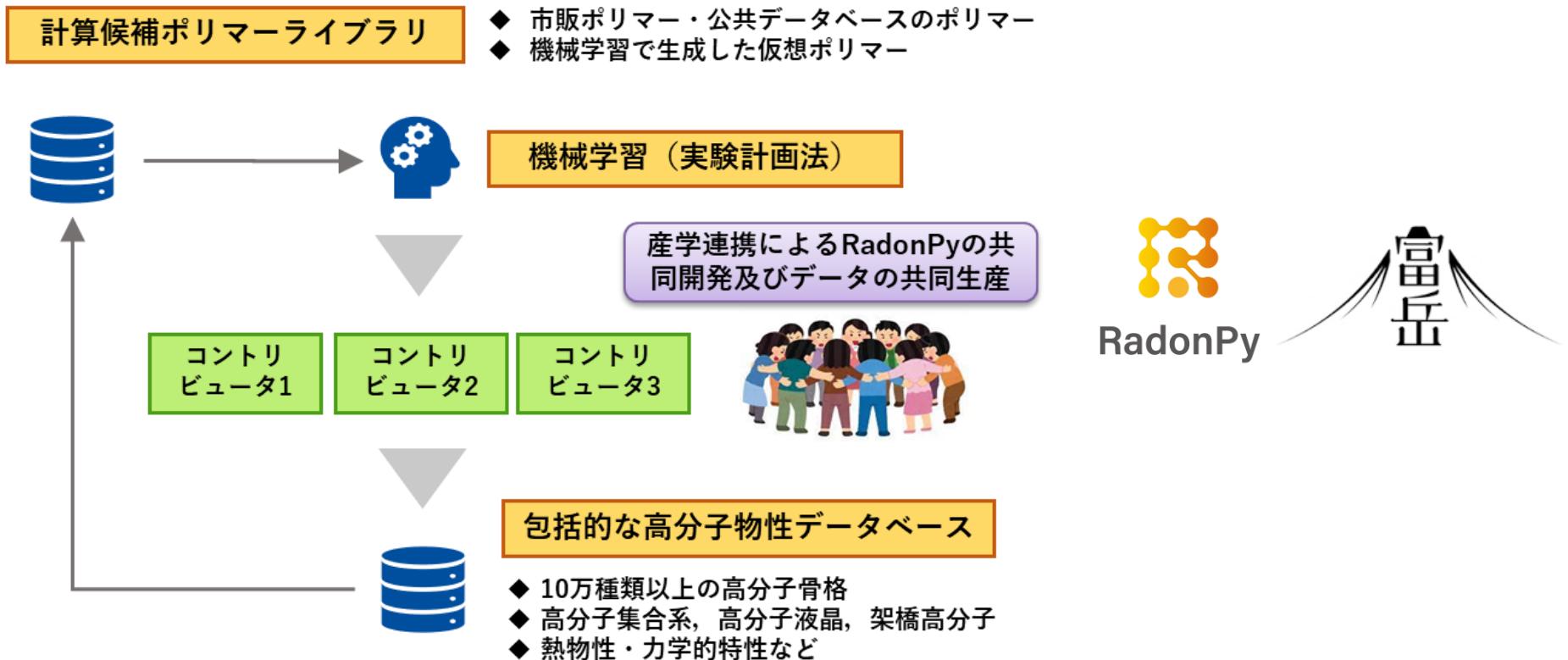
- ◆ Homopolymers
- ◆ Copolymers
- ◆ Cross-linked polymer
- ◆ Amorphous states
- ◆ Stretch orientation
- ◆ Crystalline states
- ◆ Polymer solutions
- ◆ Electric fields

富岳等スパコン上での完全自動計算



世界最大の高分子物性オープンデータベースの产学共創

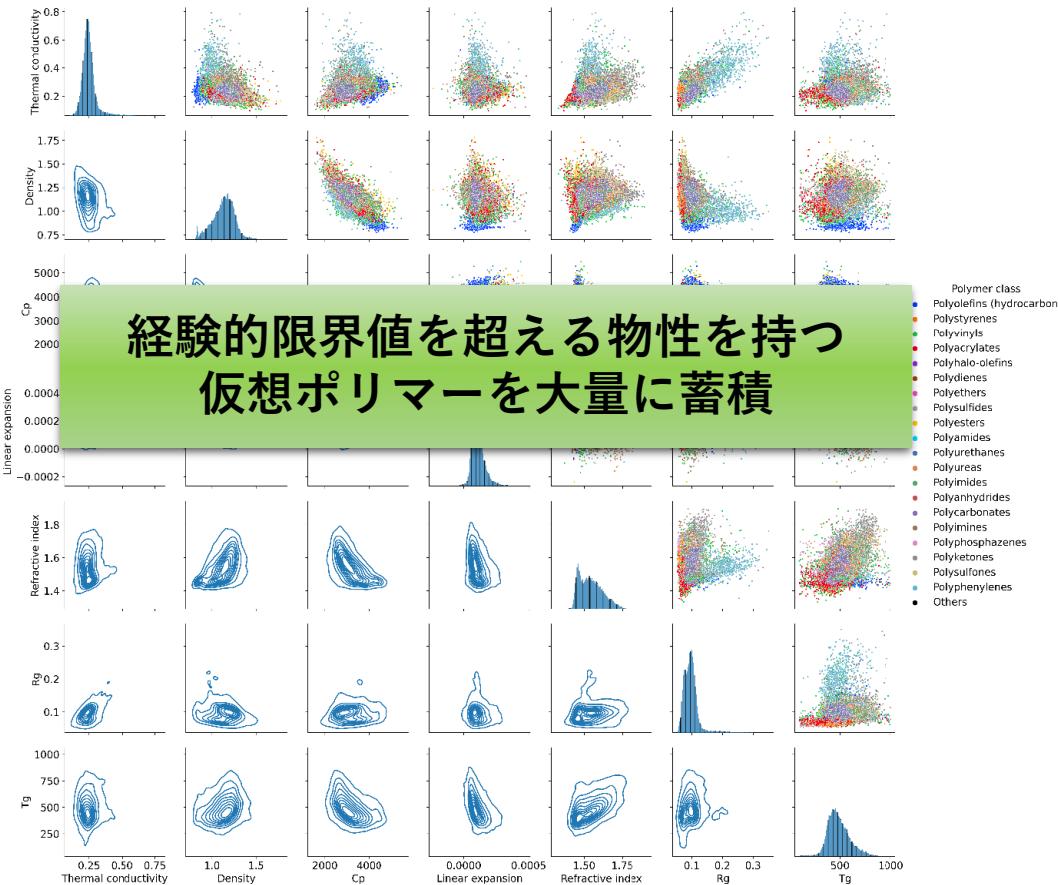
- 产学連携コンソーシアム：統数研・9大学・36企業（239名）+ NIMS PoLyInfo
- 文科省 富岳成果創出加速P「データ駆動型高分子材料研究を変革するデータ基盤創出」（吉田）



計算機実験でマテリアルスペースを埋め尽くす

- 複数物性の同時分布を観測し、パレートフロンティアの位置や構造的特徴を明らかに
- 機械学習のデータ資源としての利用

約8万個の非晶質ポリマーの物性分布



約3万個のセルロース誘導体の物性分布

- 1 Density
- 2 Refractive index
- 3 Rg
- 4 Self-diffusion
- 5 Cp
- 6 Cv
- 7 Compressibility
- 8 Isentropic Compressibility
- 9 Bulk modulus
- 10 Isentropic bulk modulus
- 11 Volume expansion
- 12 linear expansion
- 13 Static dielectric constant
- 14 Thermal conductivity
- 15 Thermal diffusivity
- 16 Tg



RadonPy データベースの活用 : Sim2Real 転移学習

Wu et al. *npj Comput Mater* 5, 66 (2019); Yamada et al. *ACS Cent Sci* 5, 1717-1730 (2019); Ju et al. *Phys Rev Mater* 5, 053801 (2021); Aoki et al. *Macromolecules* 56, 5446–5456 (2023); MInami et al. *NeurIPS* 36 (2023)

Latest: 34 properties implemented (2024/05/13)

- ◆ Thermal conductivity
- ◆ Thermal diffusivity
- ◆ Density
- ◆ Radius of gyration
- ◆ Specific heat capacity Cp
- ◆ Specific heat capacity Cv
- ◆ Compressibility (isothermal)
- ◆ Isentropic compressibility
- ◆ Bulk modulus (isothermal)
- ◆ Isentropic bulk modulus
- ◆ Self-diffusion coefficient
- ◆ Thermal expansion coefficient
- ◆ Linear expansion coefficient
- ◆ Dielectric constant (static)
- ◆ Refractive index
- ◆ Solvation free energy
- ◆ Solubility parameters
- ◆ Glass transition temperature
- ◆ Abbe's number
- ◆ Order parameter
- ◆ Dielectric constant
- ◆ Dielectric loss tangent
- ◆ End-to-end distance
- ◆ Polarizability (monomer)
- ◆ Dipole moment (monomer)
- ◆ Thermal conductivity_ke
- ◆ Thermal conductivity_pe
- ◆ Thermal conductivity_pair
- ◆ Thermal conductivity_bond
- ◆ Thermal conductivity_angle
- ◆ Thermal conductivity_dihed
- ◆ Thermal conductivity_improper
- ◆ Thermal conductivity_kspace
- ◆ Chi parameters

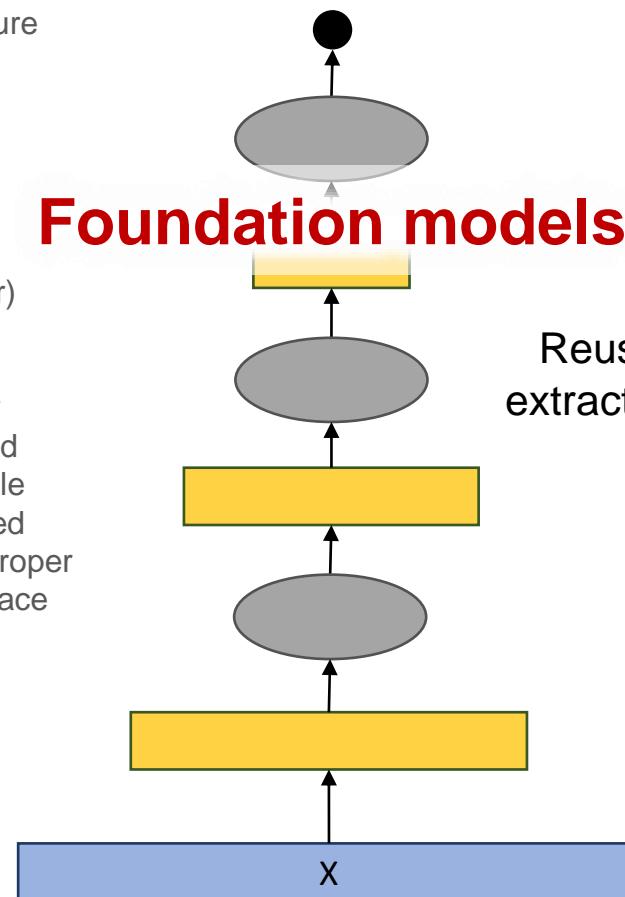
Polymer systems

- ◆ Homopolymers
- ◆ Copolymers
- ◆ Cross-linked polymer
- ◆ Polymer solutions
- ◆ Amorphous states
- ◆ Stretch orientation
- ◆ Crystalline states
- ◆ Electric fields

Source task: RadonPy

Foundation models

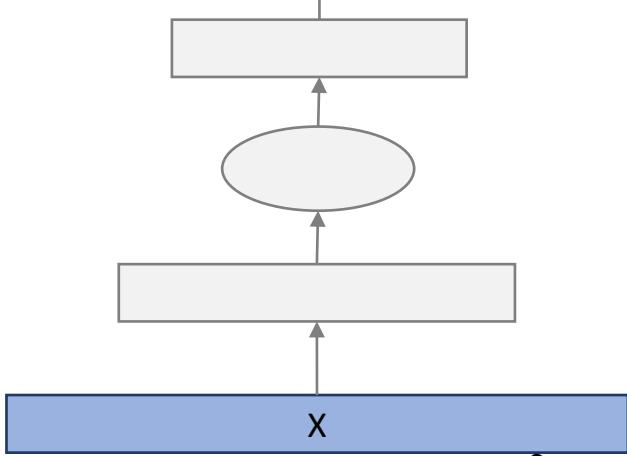
Reuse as a feature extractor or fine-tuning



Target task: real-world systems

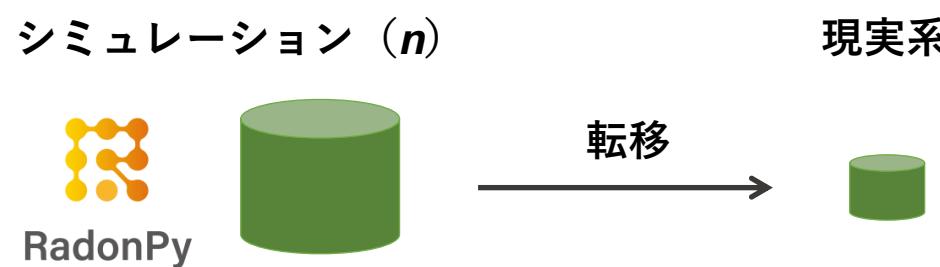
Fine-tuning (limited experimental data)

Reuse as a feature extractor or fine-tuning

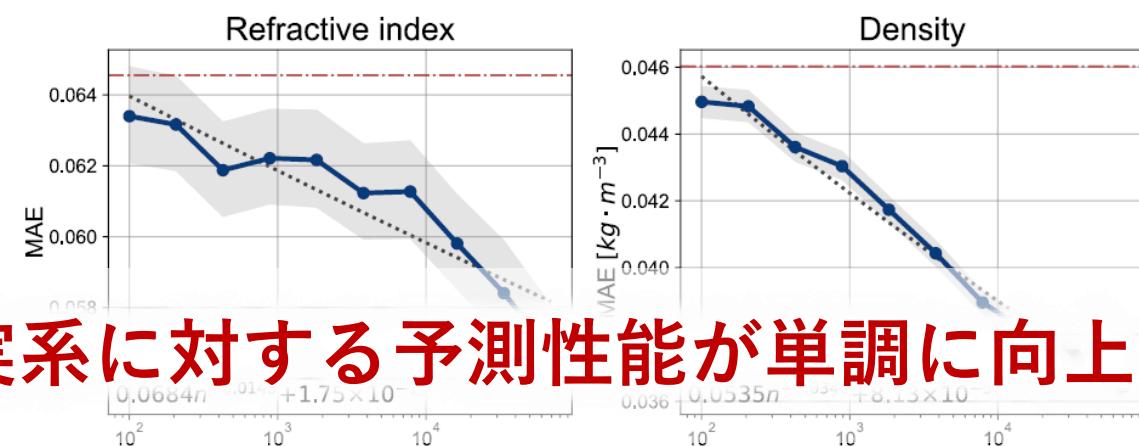


Sim2Real 転移学習のスケーリング則

Mikami and Fukumizu et al., A Scaling Law for Synthetic-to-Real Transfer: How Much Is Your Pre-training Effective?. *ECML PKDD* (2022).



計算データの増加に伴い、転移モデルの実験物性に対する汎化性能がべき乗則に従い改善

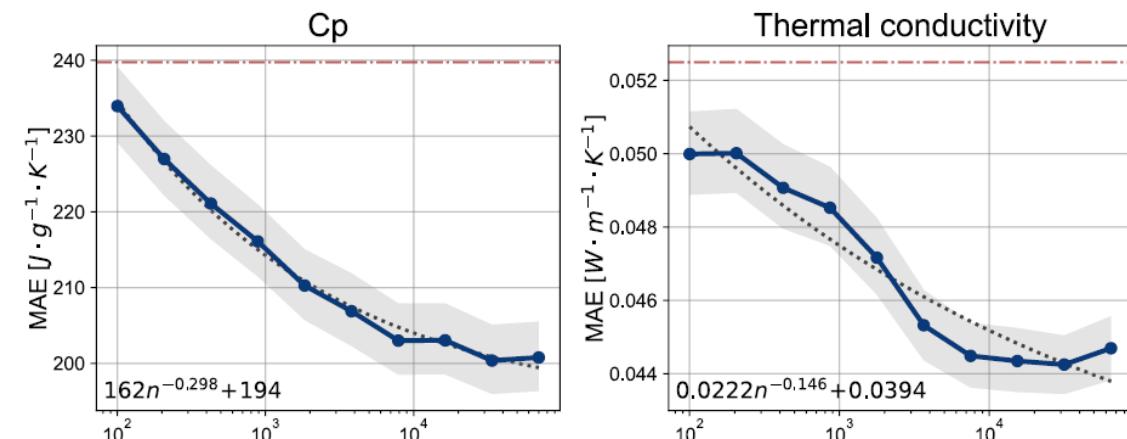


計算データの規模拡大 → 現実系に対する予測性能が単調に向上

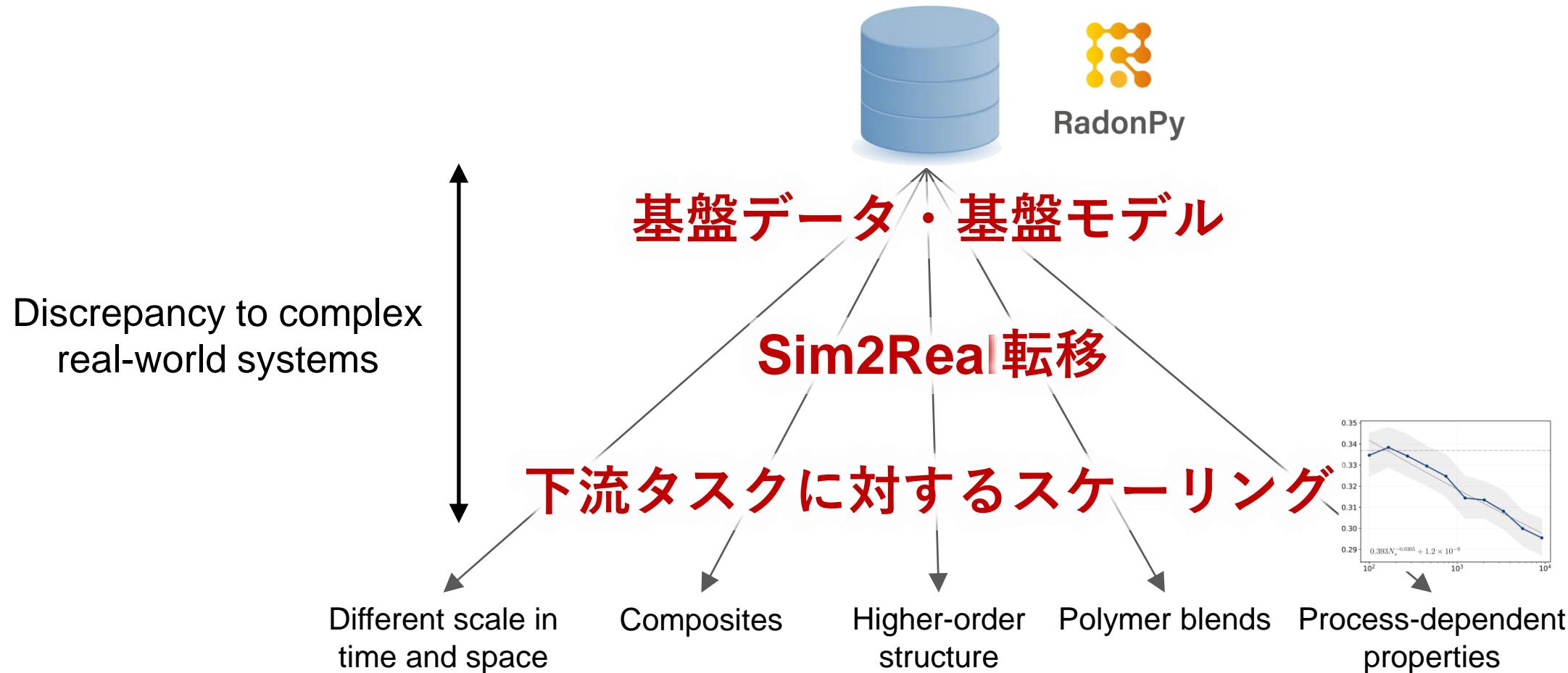
理論予想 $\text{test error} = Dn^{-\alpha} + C$

↑

転移ギャップ (到達可能な汎化性能)
~ 計算データベースの将来価値



戦略目標：スケーラブルに転移可能な基盤データ創出



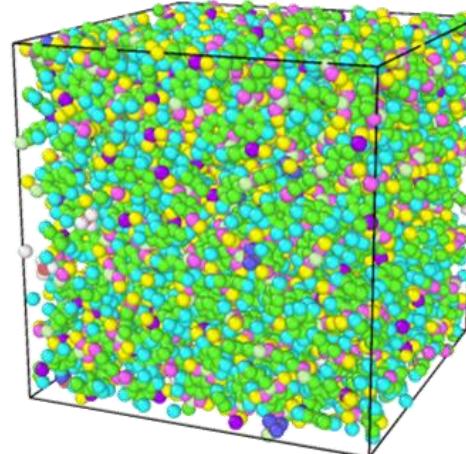
計算世界の拡張とSIM2REAL転移の適用範囲拡大

コンソーシアム型オープンソース開発による加速：計算機実験×機械学習の最高到達点を引き上げる

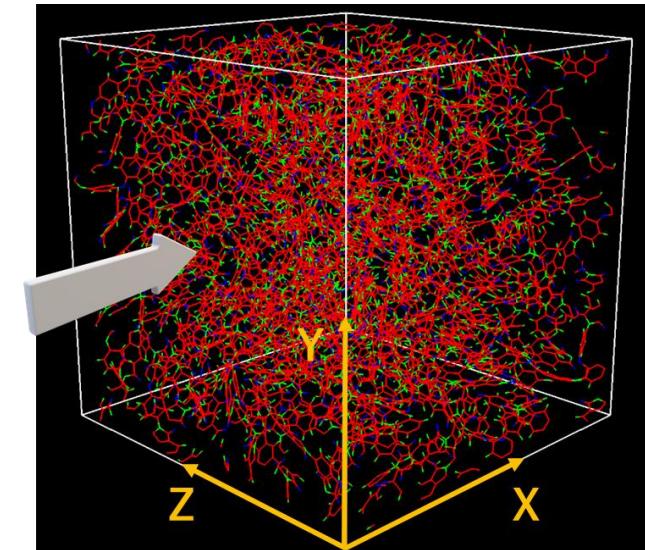
多彩な高度専門人材による開発の加速

- 粗視化MD（ミクロ相分離・複合材特性など）
- ガラス転移温度（杉澤宏樹 三菱ケミカル）
- 配向構造の物性（古屋秀峰・細谷亮平 東工大）
- 溶媒和自由エネルギー（丸山豊 統数研）
- レオロジー物性（企業）
- 量子化学計算（林慶浩 統数研）
- 架橋ポリマー（企業）
- 誘電特性（古屋秀峰・細谷亮平 東工大）
- 共重合体（南條舜 総研大）
- 溶解度パラメータ（企業）
- 物性の温度依存性・組成依存性
- 低分子化合物の材料物性（企業）
- 仮想ポリマー生成（大野充 ダイセル）
- 生分解性ポリマー（篠田恵子 統数研）

延伸配向構造の物性計算



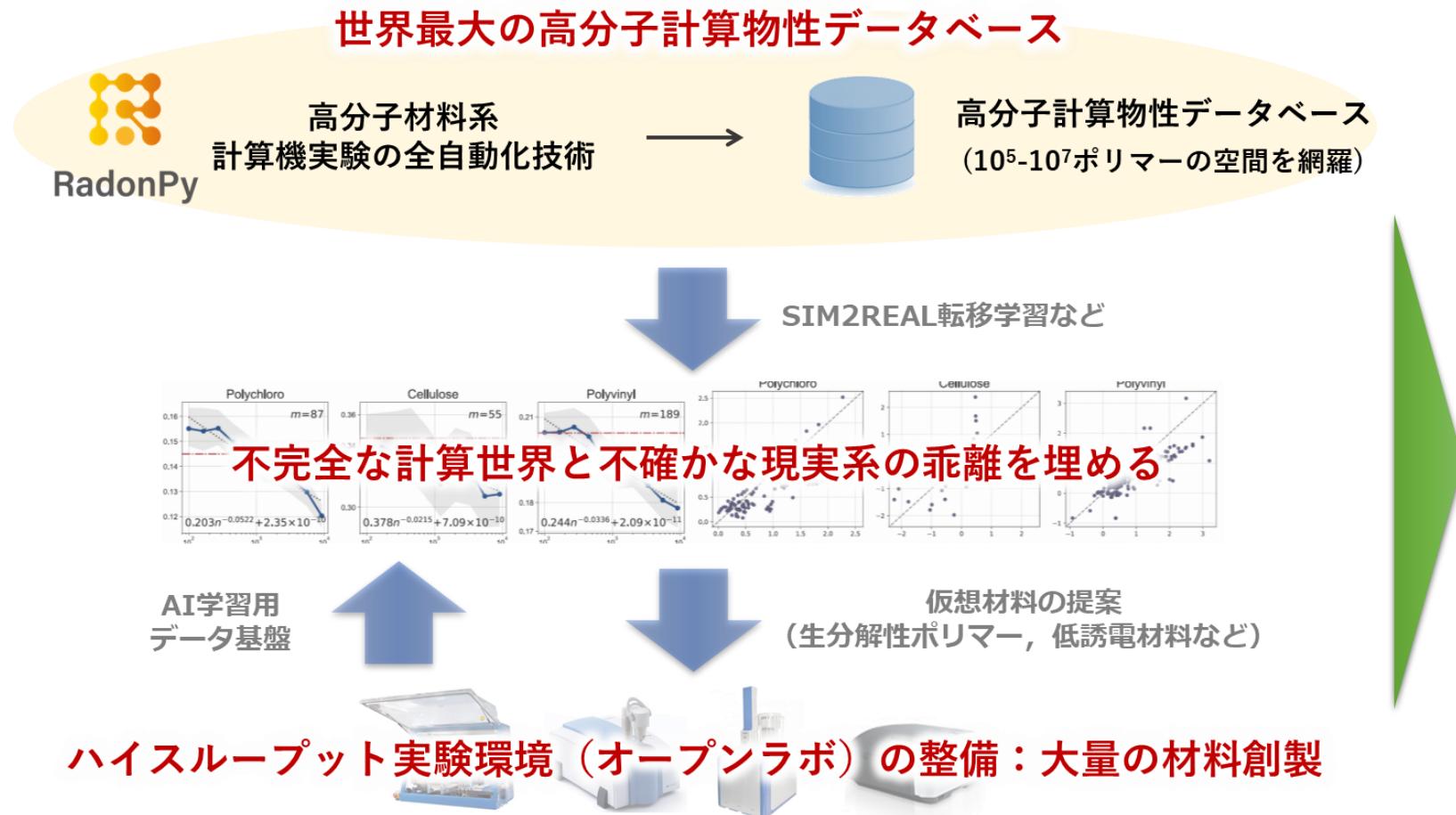
交流電場下の誘電率・誘電正接の計算



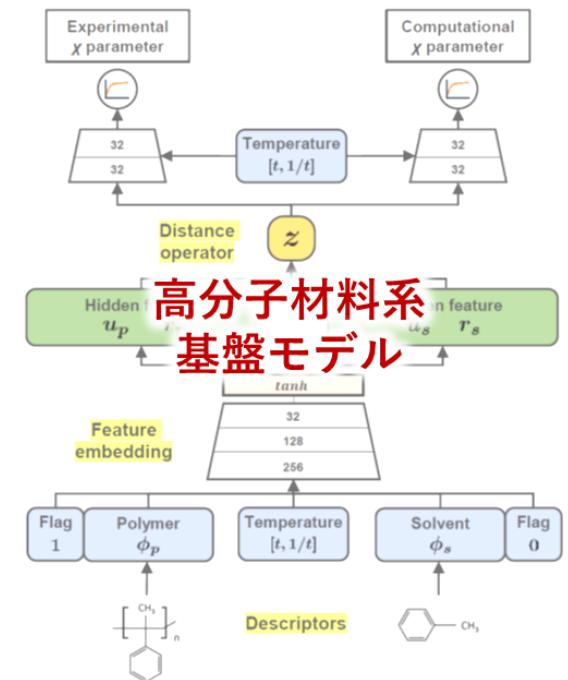
古屋先生・細谷先生@東工大

今後の展開：現実系と計算世界の接続

- RadonPyデータベースは、他の追従を許さないレベルに到達
- 高分子材料系において、大規模計算×大量合成の成功事例をデモンストレーションする。



オープンデータ創出
機械学習分野との合流

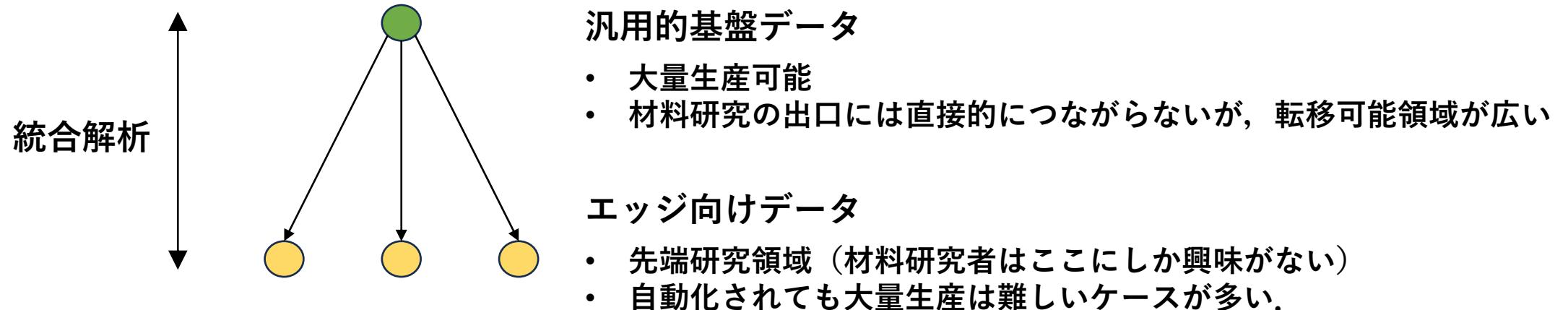


マテリアルDX推進に関する提言

■ 地政学的優位性

- ・ 日本の特異性：RadonPyコンソーシアムのような巨大な产学連携体制が自発的に生まれる環境
- ・ 技術面で劣る点は一切ない。負けているとすれば、リソース、戦略・施策、文化、マインドの問題

■ 階層性・転移性を考慮したデータポートフォリオ構築戦略

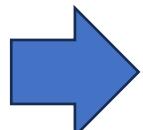


マテリアルDX推進に関する提言

■ 実験自動化の現状と課題

自動実験によるデータの大量生産×AIの成功事例が中々でてこない。

- 研究室に閉じているため、稼働率が低い。
- 運転・維持コストが高い。その内、減価償却費の増加が問題に
- 自動化 자체が目的になっている。
 - エコシステムの構築（データ生産 ⇒ 価値創出 ⇒ 収益 ⇒ 再投資 ⇒ データ生産）
 - オープンラボ・共用施設化
 - 実験費用の低下（例：RNA-seqの受託解析 約5万円/サンプルで可能！）
例：Emerald Cloud Lab – 200種類の実験機器を備えたバイオテック系データ受託生産企業



マテリアルDX推進に関する提言

■ DX推進の問題点

- ・ データ駆動型材料研究では短期的には顕著な成果は生まれない。データを必要としない材料研究者を巻き込んでいいけない。
- ・ リーダーシップをとるもののがデジタル信仰者でなければ、ビジョンやシナリオを描けない。ビジョンがなしに、冬の時代を乗り越えられない。

■ 情報系研究者の参入障壁

- ・ データがないのに、参入してくるはずがない。
- ・ キャリアパスもない。
- ・ オープンデータ創出が必要条件