



研究データ・学習データ等の管理・利活用

東北大学データ駆動科学・AI教育研究センター
金谷吉成

はじめに

- 社会のデジタル化が急速に進展
- おびただしい量のデータが収集・蓄積
- データを活用することで、新たなイノベーションが創出
- さまざまな社会的課題の解決にも役立つ



- データの適切な管理と利活用が求められる
 - 個人情報やプライバシー情報への配慮
 - 著作権等の知的財産権への配慮

データ利活用関連動画教材

- 研究データ・学習データ等の管理・利活用
- 個人情報、プライバシー情報等の保護
- 個人情報の仮名加工、匿名加工
- 知的財産権の保護
- 著作権の保護
- 授業における著作物の利用
- ChatGPT等の生成AIの概要と教育・学習における利用に関する留意事項

研究データ・学習データ

■ データの重要性

- データは、学術や社会の発展に必須の知の基盤
- 我が国が目指すデジタル社会の実現にとっても不可欠

■ 研究データ

- 研究活動の過程で研究者によって収集・生成された情報
- 研究データを適切に管理することは優れた研究を行う上で必要不可欠

■ 学習データ

- 機械学習やディープラーニング等でAIに与えるデータ
- 研究活動においてAIを扱う場合は学習データも研究データに含まれる
- 学習データには、教師データ・訓練データ・検証データ等を含む

研究データ

- ① 研究素材として収集・生成された一次データ(測定データ、画像情報等)
- ② 一次データ等を分析、処理して生成された情報(加工データや解析データ等)
- ③ 上記データの収集や生成の段階で作成された記録(実験ノート、質問票等)に記載された情報
- ④ 上記のデータを用いて作成された研究成果(論文や講演資料等)に記載された情報
- ⑤ 研究に用いられた有機物等(試料、標本等)に蓄積されている情報
- ⑥ その他研究活動に用いることが予定されている情報及び研究活動に用いられた情報

記録媒体(デジタル・非デジタル)は問わない

研究データの管理・公開に関する留意事項

■法令、契約、大学が定める規程等の遵守

●法令

- 第三者が権利や法的利益を有している場合（個人情報や知的財産にあたる研究データ等）
- 安全保障の観点からその流通が規制されている場合（外国為替及び外国貿易法の輸出規制対象情報等）

●研究契約

- 共同研究契約や外部資金等に基づく研究において締結される契約

●本学が定める規程

- 「東北大学における公正な研究推進のための研究データ等の保存及び管理に関する指針」「東北大学研究データ管理・公開ポリシー」等

●国や国際機関等が定める研究倫理指針

●各研究分野における倫理的要件

研究データの管理・公開に関する留意事項

■オープン・アンド・クローズ戦略に基づくオープンサイエンスの実現

- データの特性から公開すべきもの（オープン）と保護するもの（クローズ）を分別して公開する戦略
 - まずは公開すべきデータと保護すべきデータとを適切に区別する
 - 単に多くの研究データを共有・公開すればよいということではない

■研究データの特性に応じた管理および公開の必要性

- 研究データに含まれる情報は多種多様
- 研究分野の特性や研究データの特質、研究体制等を考慮して、研究データの管理・公開が行われなければならない

研究データの管理

■ 管理の主体＝研究者

- 複数の研究者が共同して研究を実施する場合には、必要に応じ、「研究データ管理責任者」を定めることも考えられる
- 共同研究等の場合は、相手機関等と協議の上、研究者に含めることができる

■ 具体的には……

- 研究前の研究データ管理計画の策定
- 研究中における研究データの収集、生成、解析、加工等の利用行為および保管
- 研究終了後の保存の要否の決定、保存方法等の決定
- 第三者の閲覧や利用の可否、その利用要件等の決定
- 破棄
- etc

研究データの公開

- 研究者は、法令の定められた範囲内およびオープン・アンド・クローズ戦略に反することのない範囲内で、研究データの公開の可否、条件、方法等について決定することができる
 - 学術的意義を有する研究データを社会に公開することは、オープンサイエンスの実現にとって有益
- 研究データの公開にあたっては、管理と同様、関係法令、契約、大学が定める規程、各研究分野において要求される倫理的要件等を遵守しなければならない

法令、契約、規程により非公開とすべきデータ

- 第三者が権利または法的に保護される利益を有しているためにその公開が制限される研究データ
 - 個人情報、プライバシー情報、著作物、不正競争防止法において保護される限定提供データ、秘密管理されたデータ、契約による制限が課せられたデータ
- 外為法に基づき安全保障輸出管理の対象となっている研究データ
- 公開により第三者の利益を害するおそれがある研究データ
- 倫理的要件等から公開に適しない研究データ

上記については非公開とすべき

(ただし、研究遂行上の必要性や公益保護等の特段の事情が認められる場合に、その相手方を必要な範囲に限定した上で公開することはあり得る。)

オープン・アンド・クローズ戦略に基づく公開の可否

■研究データの公開については、戦略的な判断が必要

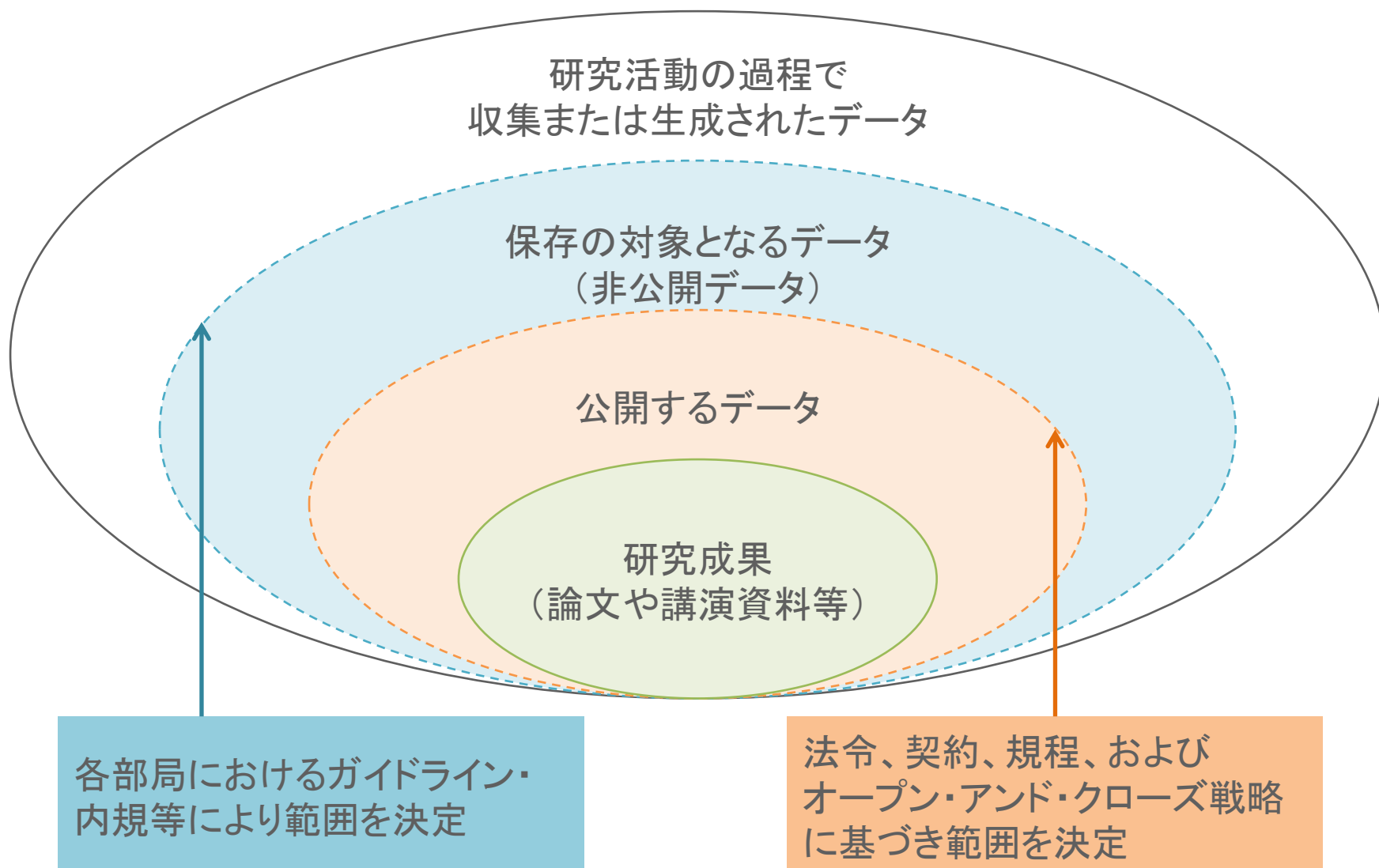
- 研究成果の社会実装やさらなる研究推進のために、知的財産として法的な保護が必要な研究データもある



■公開

- 利用者や利用目的・利用方法等により、さまざまな様態があり得る
 - 利用者を限定せずに利用を認める場合
 - 限定した者のみに利用を認める場合
 - 利用目的等を限定して認める場合
 - etc

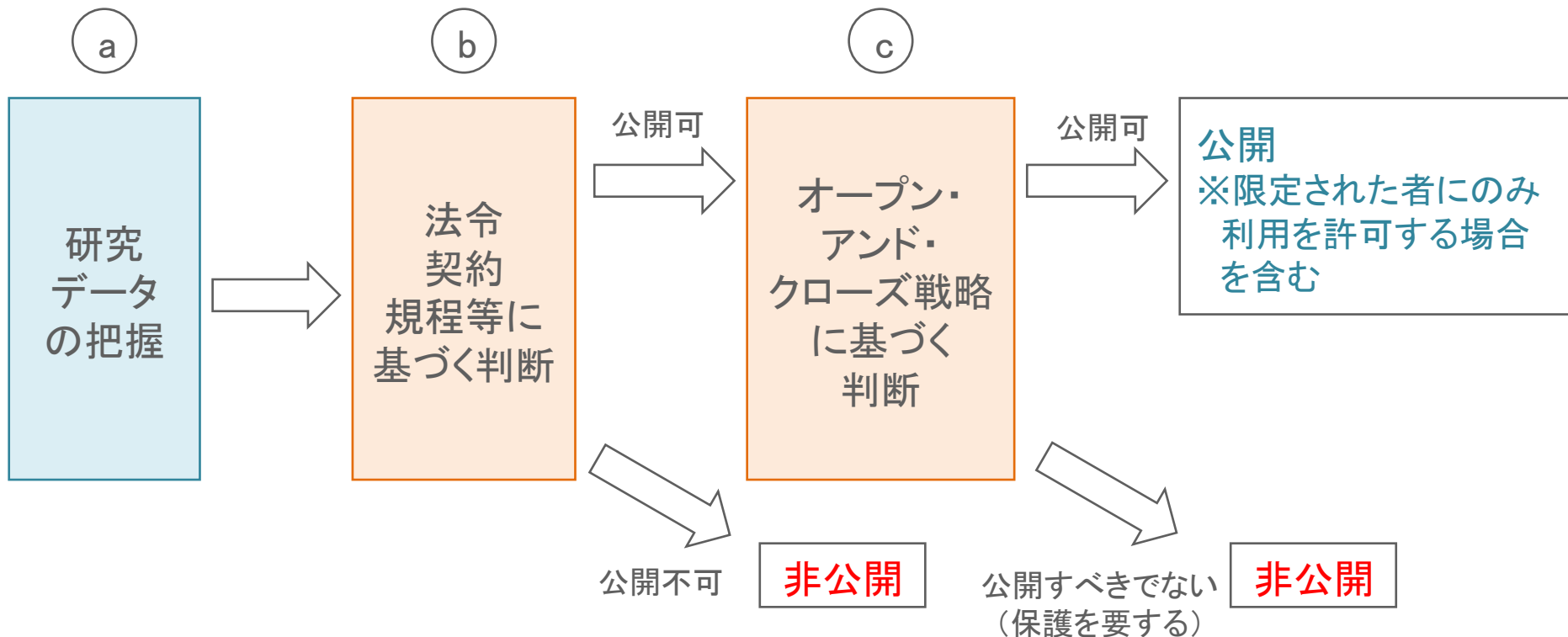
研究データとその区分範囲のイメージ



(出典) 東北大学研究データ管理・公開ポリシーの解説

公開するデータに関する判断の流れ

- a. 対象となる研究データの把握
- b. 法令、契約、規程により非公開とすべきデータの除外
- c. オープン・アンド・クローズ戦略により保護すべきデータ



(出典) 東北大学研究データ管理・公開ポリシーの解説

研究データの公開に関する留意事項

■ 公開する研究データの信頼性の確保

- 公開する研究データには、正確性・完全性・追跡可能性等を担保することが求められる

■ FAIR原則に基づく公開

- Findable(見つけられる)、Accessible(アクセスできる)、Interoperable(相互運用できる)、Reusable(再利用できる)
- 可能な限り「FAIR原則」に則って公開することが望ましい

■ 知的財産の適切な保護

- 知的財産として保護が必要な研究データについては公開してはならない

■ 公開の様態

- 非デジタル形式の研究データについては、公開にあたって、デジタル化や研究データが蓄積された有体物の所在等を示すメタデータによる公開とするなど、可能な限りデジタル化することが望ましい

まとめ

- 研究データの適切な管理・公開が求められる
 - 知の基盤の共有、学術研究の発展を促進
 - 社会への貢献とともに、研究者が、将来においてよりよい研究を行うことに寄与する

参考リンク

- 東北大学における公正な研究推進のための研究データ等の保存及び管理に関する指針

<https://www.bureau.tohoku.ac.jp/kenkyo/fb/files/rules/9.pdf>

- 東北大学研究データ・公開ポリシー

<https://www.tohoku.ac.jp/japanese/2022/01/news20220104-05.html>