NII RCOS
Research Center for Open Science and Data Platform

管理基盤
データガバナンス機能

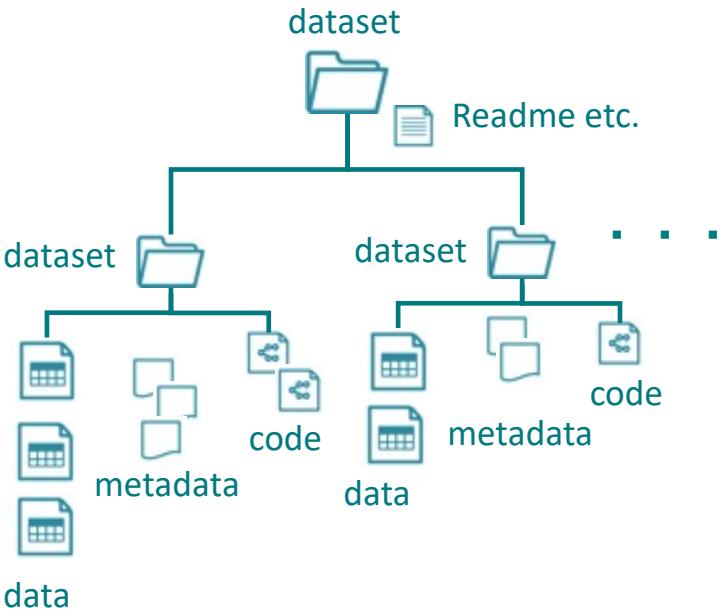
2021-7-7

国立情報学研究所
オープンサイエンス基盤研究センター
横山重俊

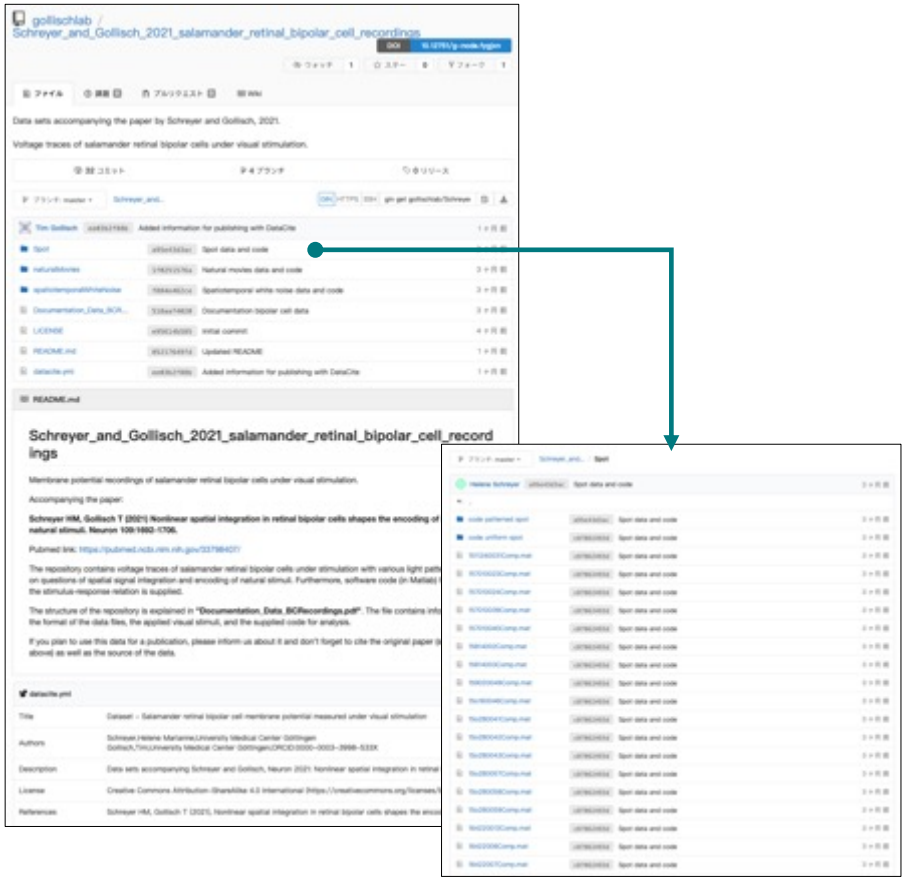
1. データガバナンス機能って？

研究データ

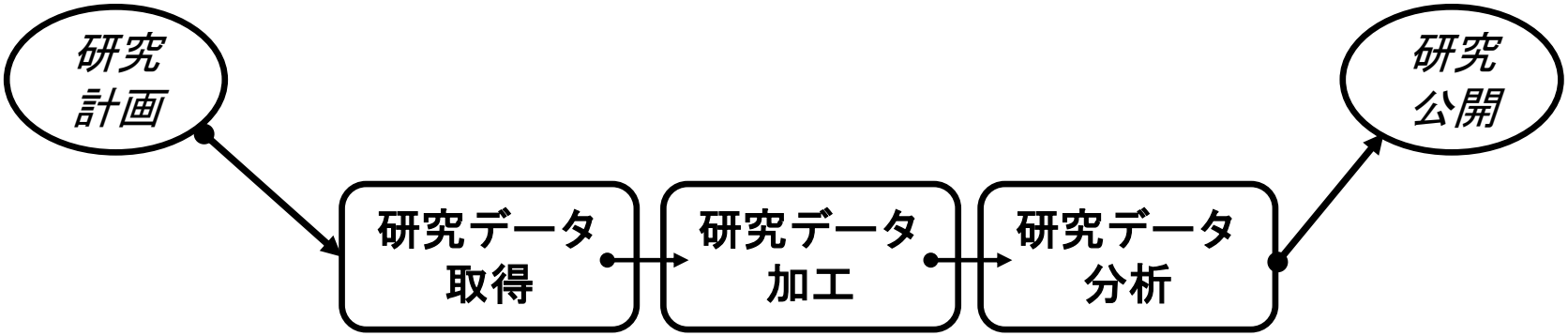
研究データの構成

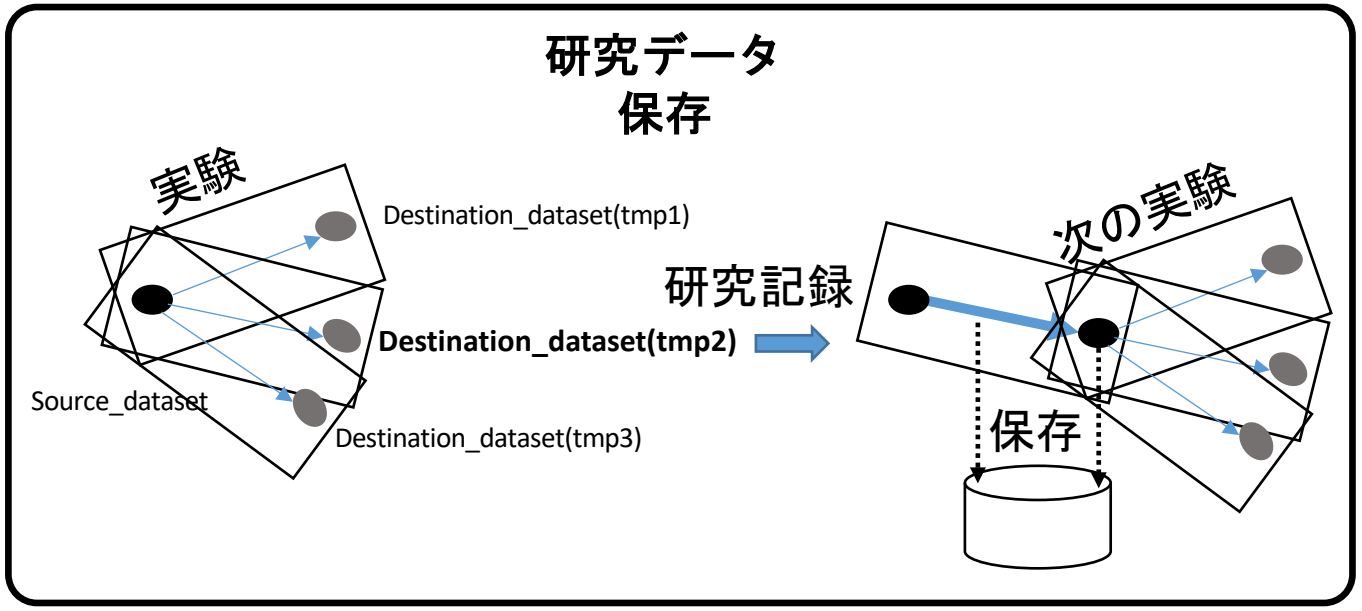


研究データの例



研究データ管理

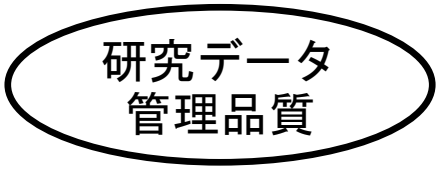

研究
計画
研究データ
取得
研究データ
加工
研究データ
分析
研究
公開


研究データ
保存
実験
Destination_dataset(tmp1)
研究記録
Destination_dataset(tmp2)
Source_dataset
Destination_dataset(tmp3)
次の実験
保存

研究データガバナンス

研究データ管理 と 研究データガバナンスの関係

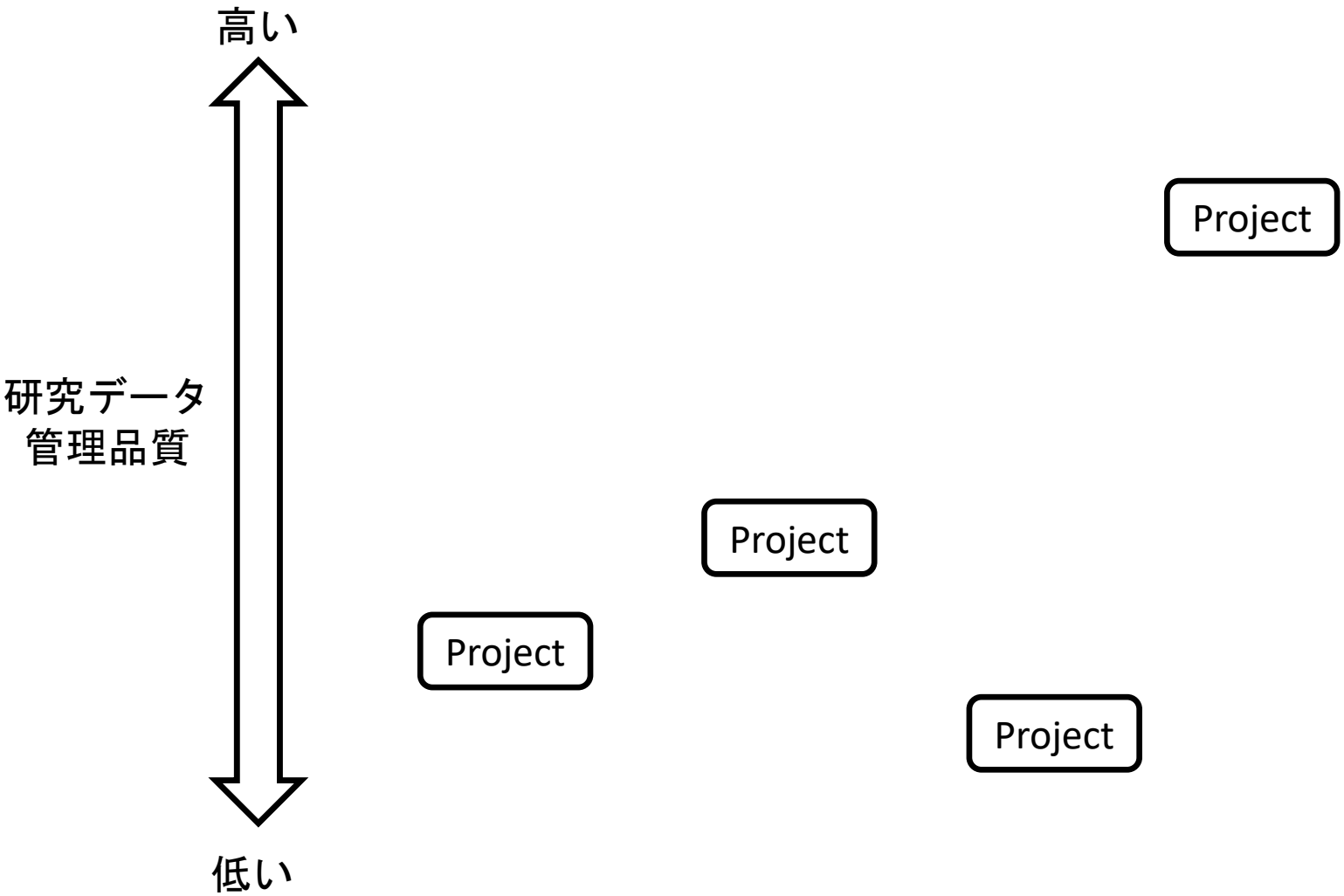
研究データガバナンス　：　研究データ管理品質を守る規律に従う行為



研究データ管理　：　研究データの取得・加工・分析・保存・公開などの行為

仮説：
研究者や研究グループ自身が，研究データガバナンスを，
日頃から行うことが，研究データ管理品質の向上に大きく寄与する.
（ガバナンスは監視ではない）

研究データ管理品質のばらつき



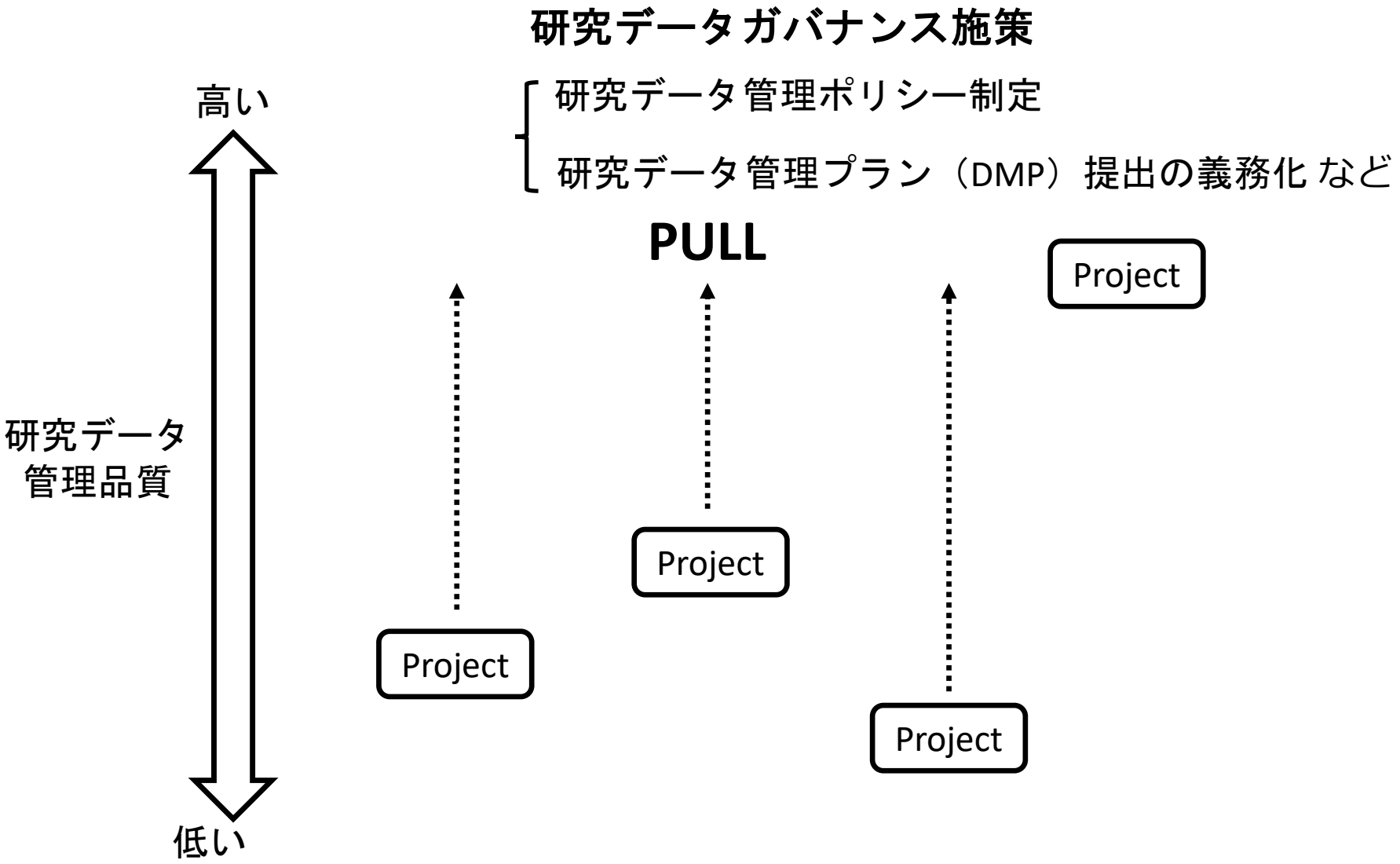
研究データ管理の品質が悪いと

- そもそも必要な時に，どこにあるのか分からない.
- 本当にあるのかすら，分からない.
- ある場所が分かっても， 実際には手に入らない.
- 手に入っても，それをどうやって扱うのか分からない.
- データの由来が不明なので信用できない.
- 保護しなければならないデータを， 漏洩させてしまう.
- 整理が悪くて，とても公開までに手間がかかる.

研究データ管理の品質

1. **FAIR**原則の満足度
   Findable, **A**ccessible, **I**nteroperable, **R**eusable
2. トレーサビリティ
3. 再現性
4. データ保護 など

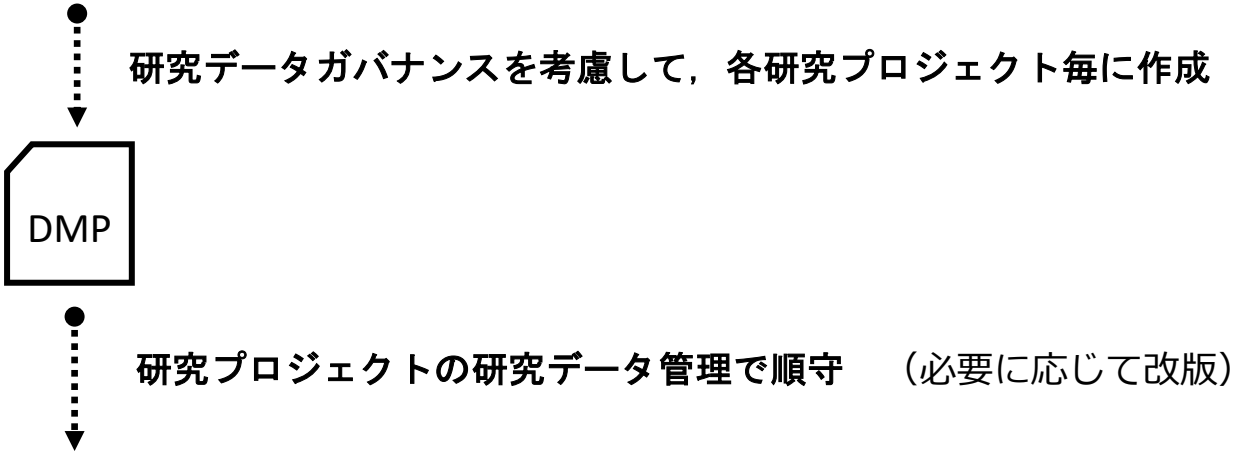
施策の展開による研究データ管理品質向上



データマネジメントプラン (DMP)

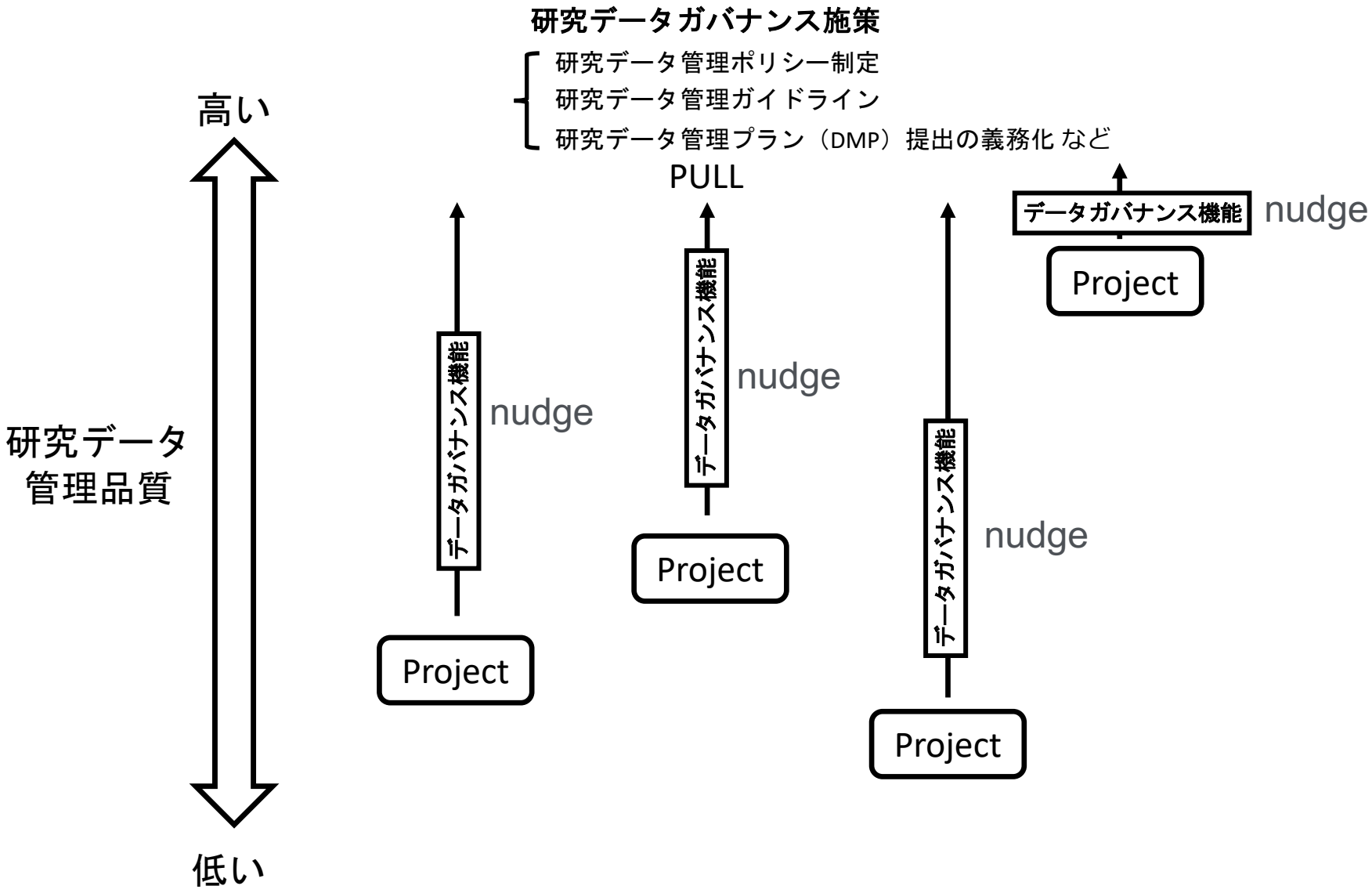
データマネージメントプラン(DMP)は，研究のために収集・作成する研究データの取扱いや整備・保存・公開についての計画を定めた**文書**である.
DMPの記述内容は提出先の助成機関によって異なるけれど，データ管理者，データの説明，データの種類，データ量，公開レベルなどが記述される.

研究データガバナンス　：　研究データ管理品質を守る規律に従う行為



研究データ管理　：　研究データの取得・加工・分析・保存・公開などの行為

データガバナンス機能の役割



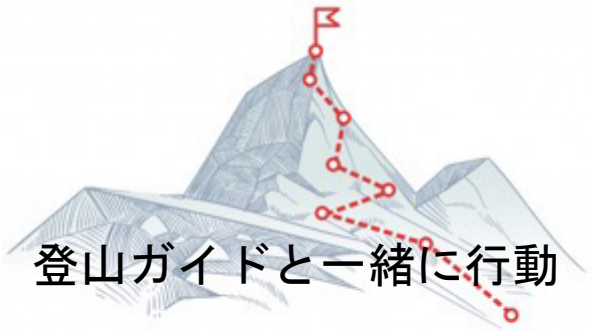
2. データガバナンス機能って，やっぱり何？

データガバナンス機能 概観

計画

登山計画書の提出

実行



登山ガイドと一緒に行動

下山届の提出

登山計画書：DMP

生成

研究計画

思い：maDMPとして表現

生成

登山ガイド：ワークフローツール

研究実行

ワークフローツール
を使って環境構築
→ 環境利用→ 研究遂行

その間，必要に応じて
リルートあり

研究公開

ワークフローツール
を使って研究公開

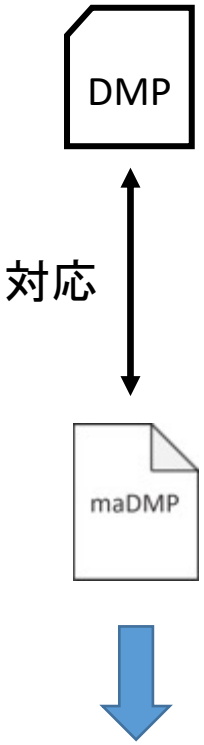
この間必要に応じてDMP更新
版作成

登山と研究

登山	研究
思い	maDMP
登山届	DMP
地図	Datasetのweb
登山ルート	Datasetのweb上のPath
GPS	Datasetのweb上の位置情報把握
登山ガイド	Datasetのweb上の案内人
写真	Dataset
登山記録	研究記録（Datasetのweb上のPath）
安全管理組織	研究管理組織
山小屋など	研究管理サービス
登山情報管理機能	**データガバナンス機能**

https://yamap.com/　https://www.yamareco.com/

maDMP　(machine actionable DMP)

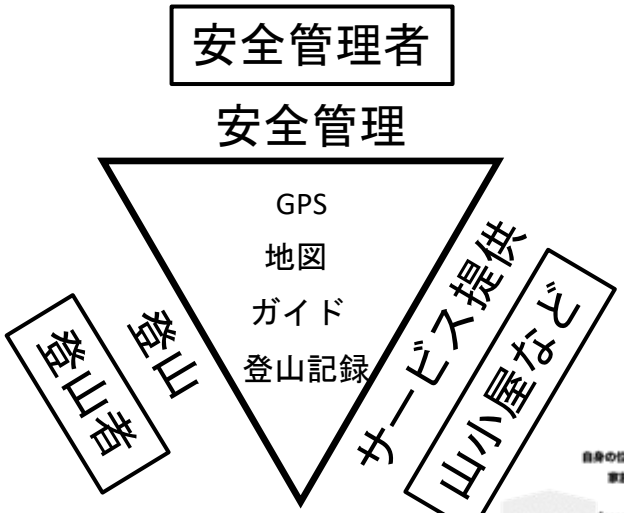


Human Readable な 文書

Machine Readable で Actionable な 情報

データガバナンス機能への入力として使える情報　（∵ データガバナンス機能はMachine）

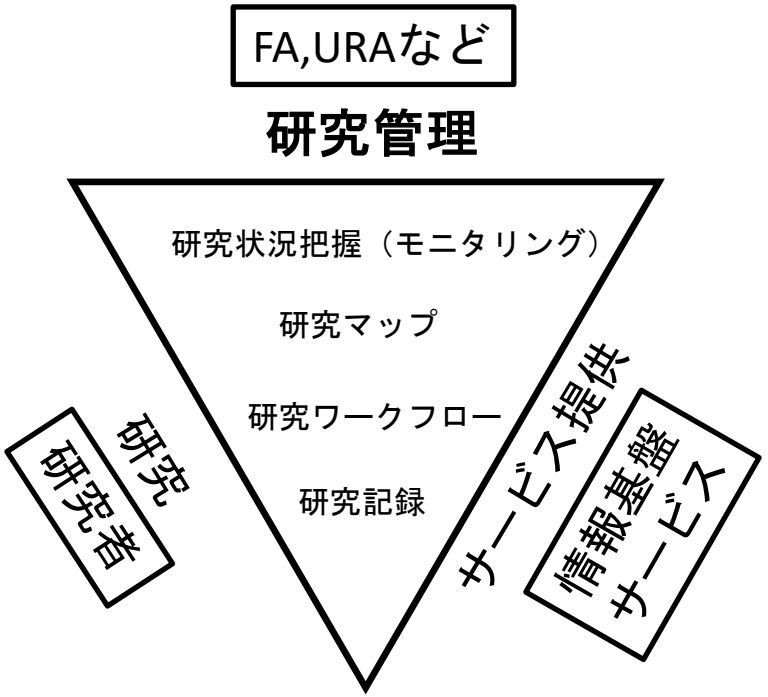
データガバナンス機能の役割



【参考情報】

https://yamap.com/

https://www.yamareco.com/



データガバナンス機能

RDCの中でのデータガバナンス機能

「プレゼンテーション層」

User Interface

「アプリケーション層」

. . .

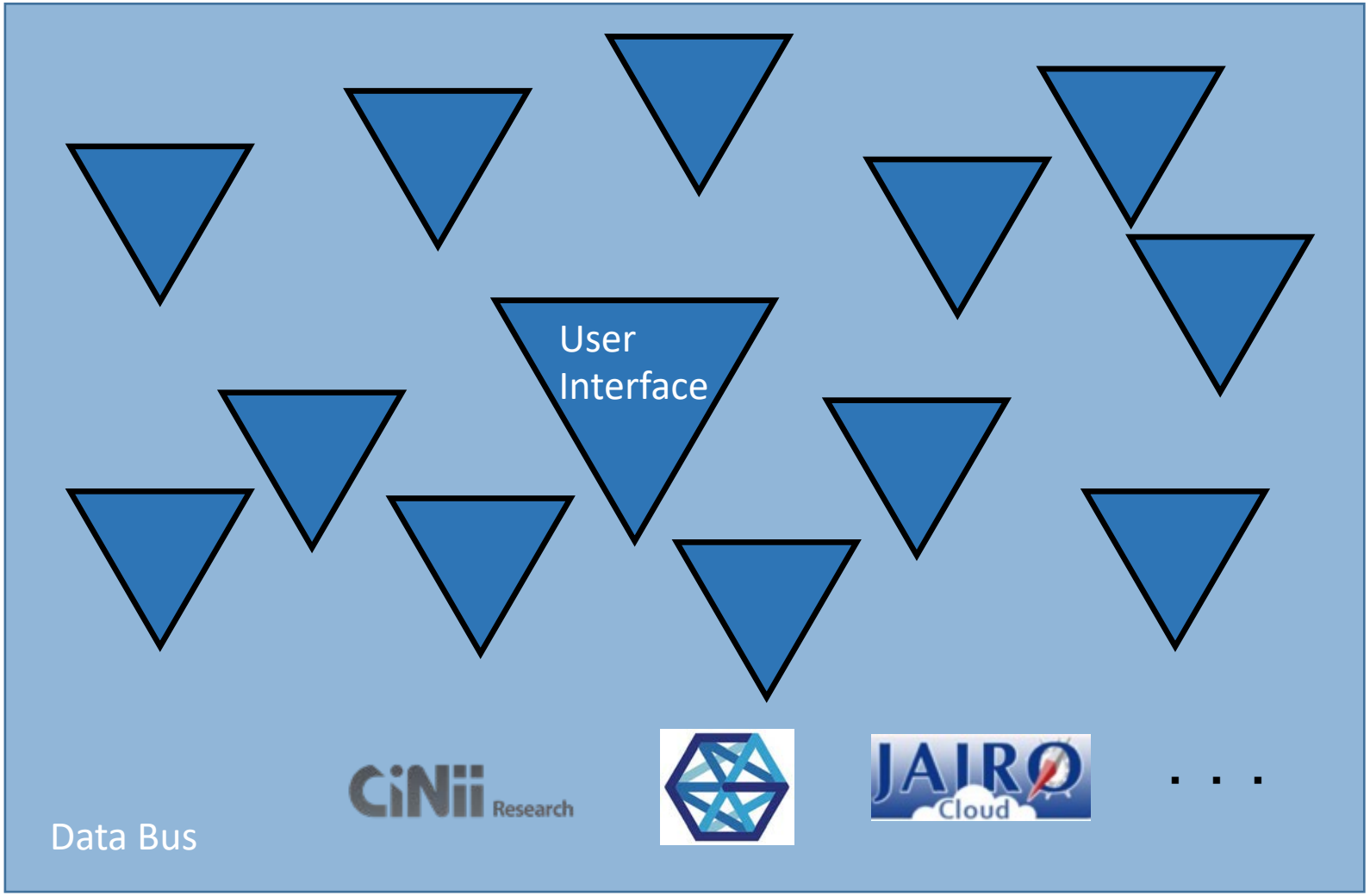
「データ層」

Data bus

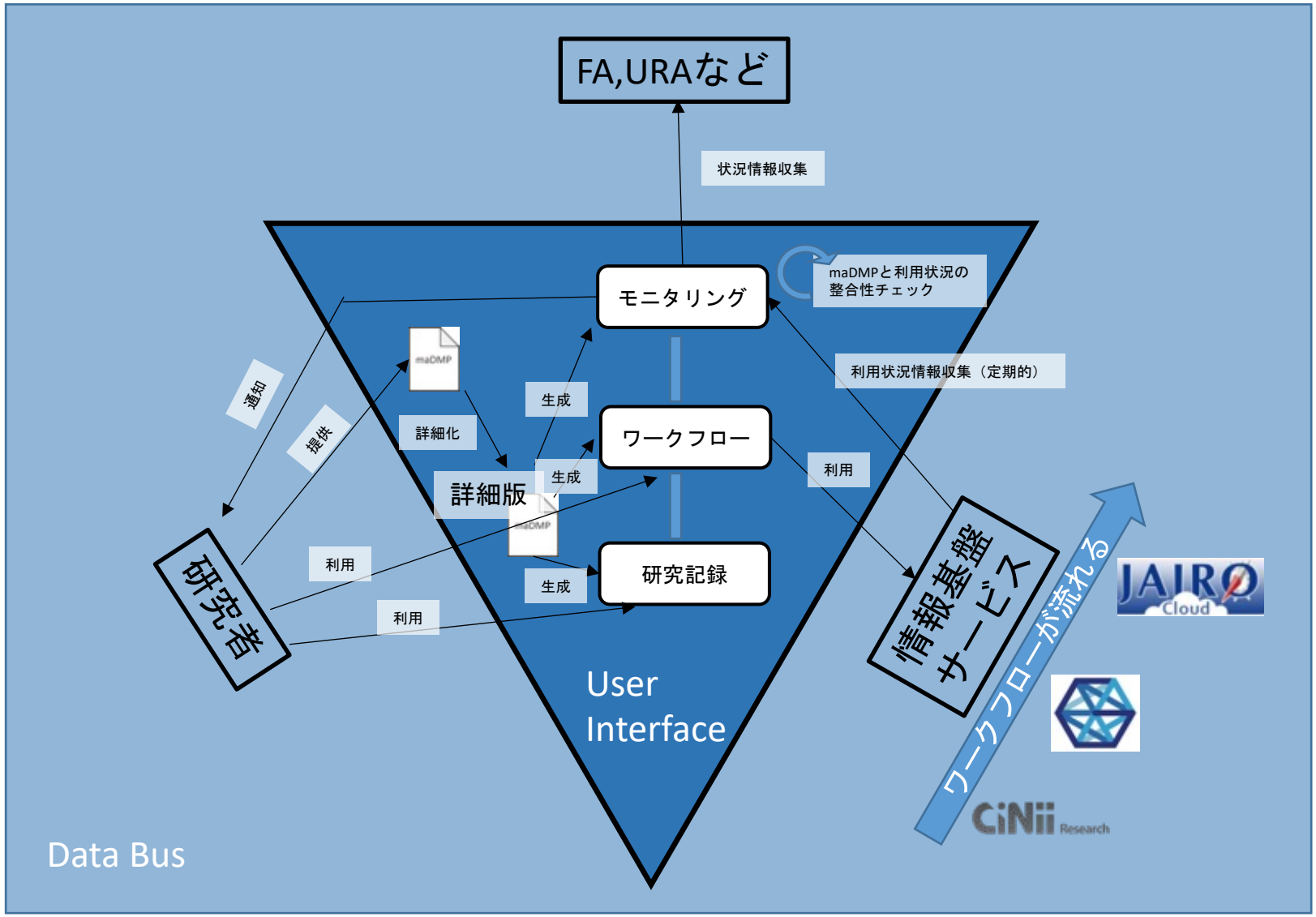
Global Positioning System

データガバナンス機能の動作イメージ

Data busがみんなをつなぐ



データガバナンス機能 内部



3. データガバナンス機能って，本当は何？

ゼロから考える必要はない！

「オープンソースソフトウェアの歴史」に学ぶ　と良いかも.

A tale of two 'opens': intersections between Free and Open Source Software and Open Scholarship https://osf.io/preprints/socarxiv/2kxq8/

	オープンソース
ソースコード カバナンス	オープンソース ポリシー文書 コーディング規約 セキュリティコーディング規約
	プロジェクト推進方針　　**code of conduct** ライセンス
	ガバナンス機能（CI/CD，課題管理など）
ソースコード 管理	管理サービス　（github, gitlab, gitbucket, gogsなど）
	管理技術　（git など）
	コード　データ構造　ドキュメンテーション

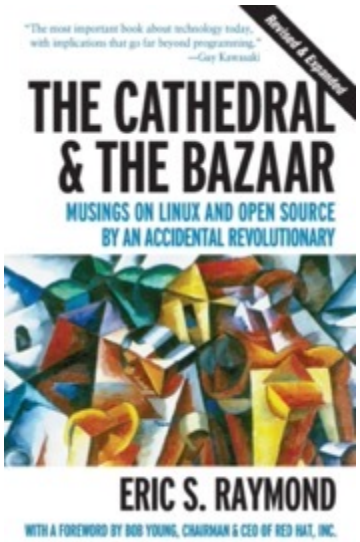
ガバナンスと管理の繋ぎを真似る　と

	オープンデータ
研究データ カバナンス	研究データ管理ポリシー
	DMP/maDMP
	データガバナンス機能 （ワークフロー，モニタリングなど）
研究データ 管理	管理サービス　（GRDM, **gin**, github+S3など）
	管理技術　(git)+(**git-annex**) など
	コード　データ構造　ドキュメンテーション **データ**

gin https://gin.g-node.org/
git-annex https://git-annex.branchable.com/

もっと「オープンソースソフトウェアの歴史」に学ぶ と

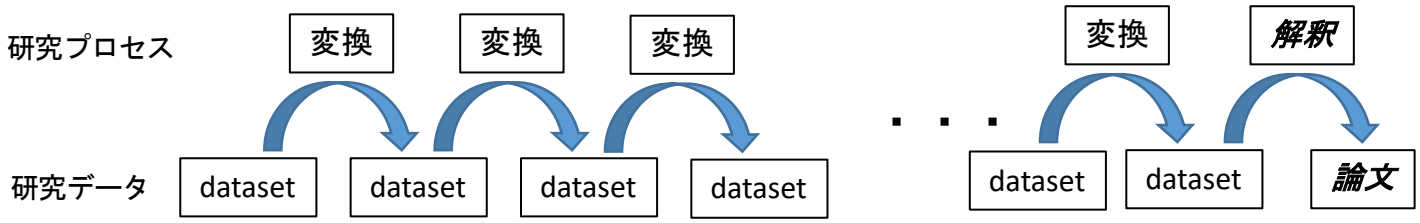
オープンソースソフトウェアの本質は
オープンコード　ではなく　オープンプロセス



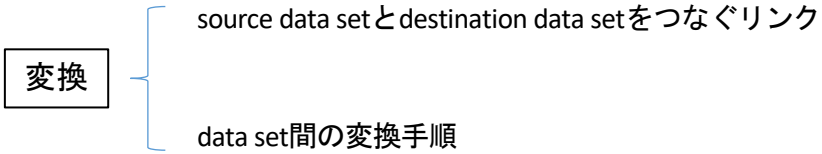
https://en.wikipedia.org/wiki/The_Cathedral_and_the_Bazaar

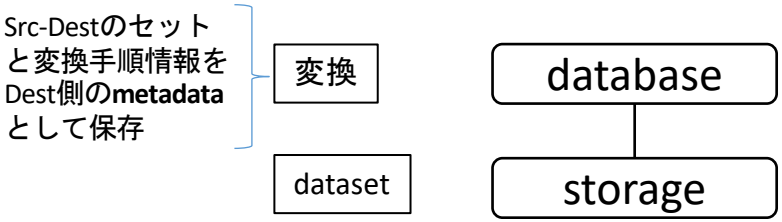
研究データと研究プロセス

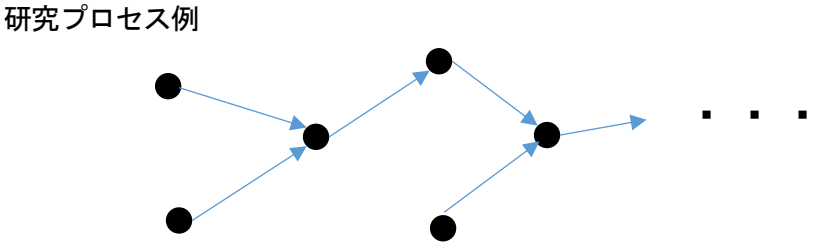
1. 研究プロセスと研究データの関係



2. 研究データと研究成果をつなぐものの実態

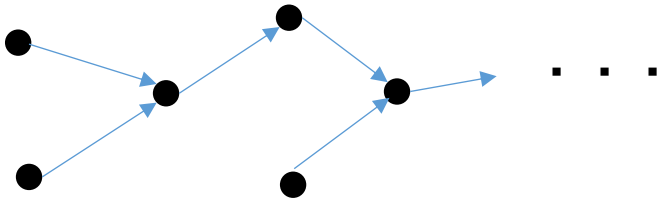




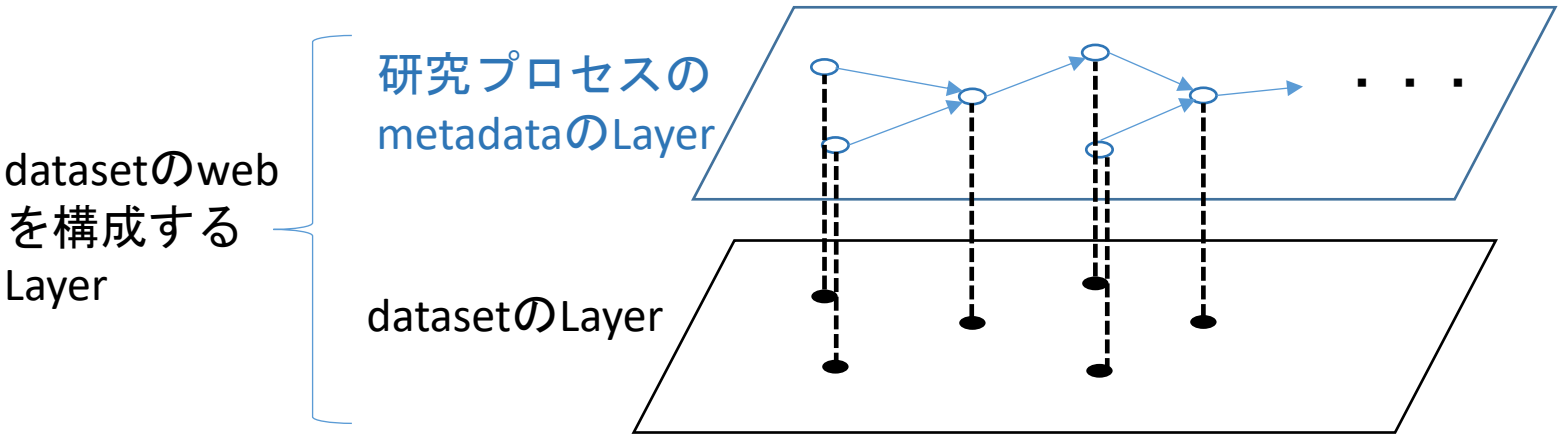


データガバナンスの重要な側面

データガバナンスとしてDatasetのみの管理に目が行き易いけれど，研究プロセスを表現したmetadataでdatasetをリンクしている構造（datasetのweb）を管理することも重要である.



datasetのweb



Thank you