

データ駆動型研究を 推進する解析基盤

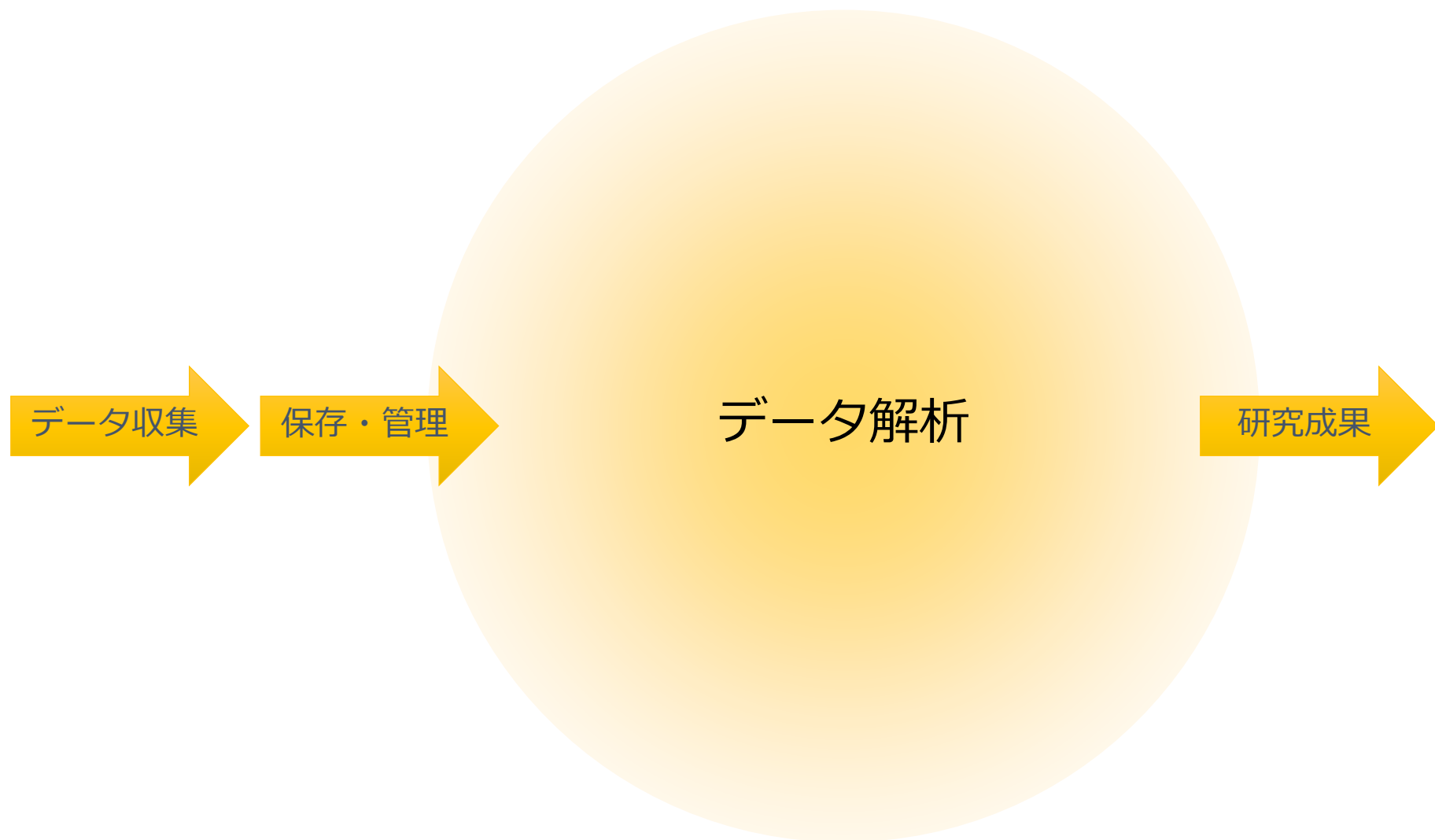
藤原 一毅

オープンサイエンス基盤研究センター

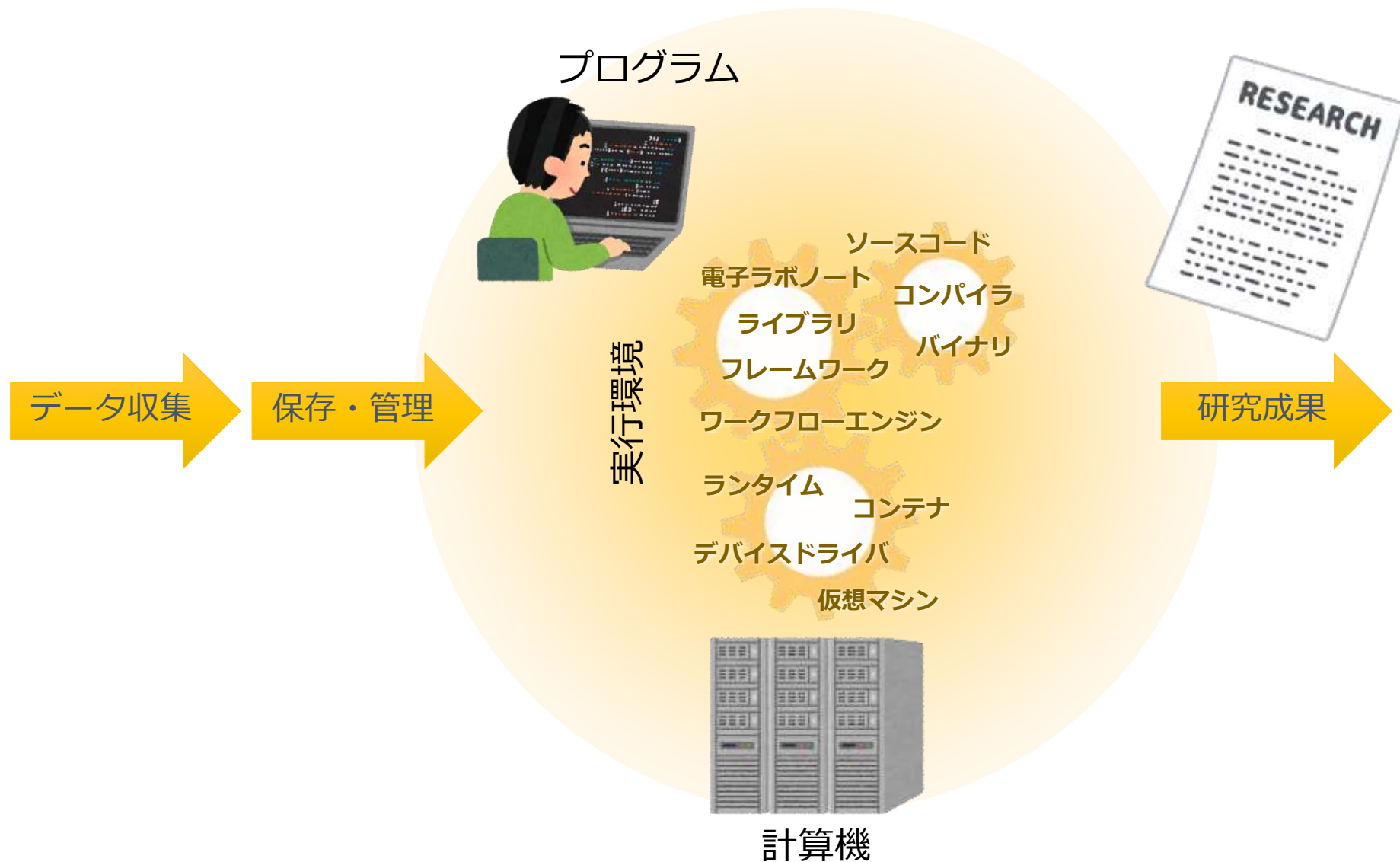
国立情報学研究所

2020/12/11 AXIES 2020年度年次大会
FH1 「2022年から始まる次世代学術研究プラットフォーム」

データ解析 = ?



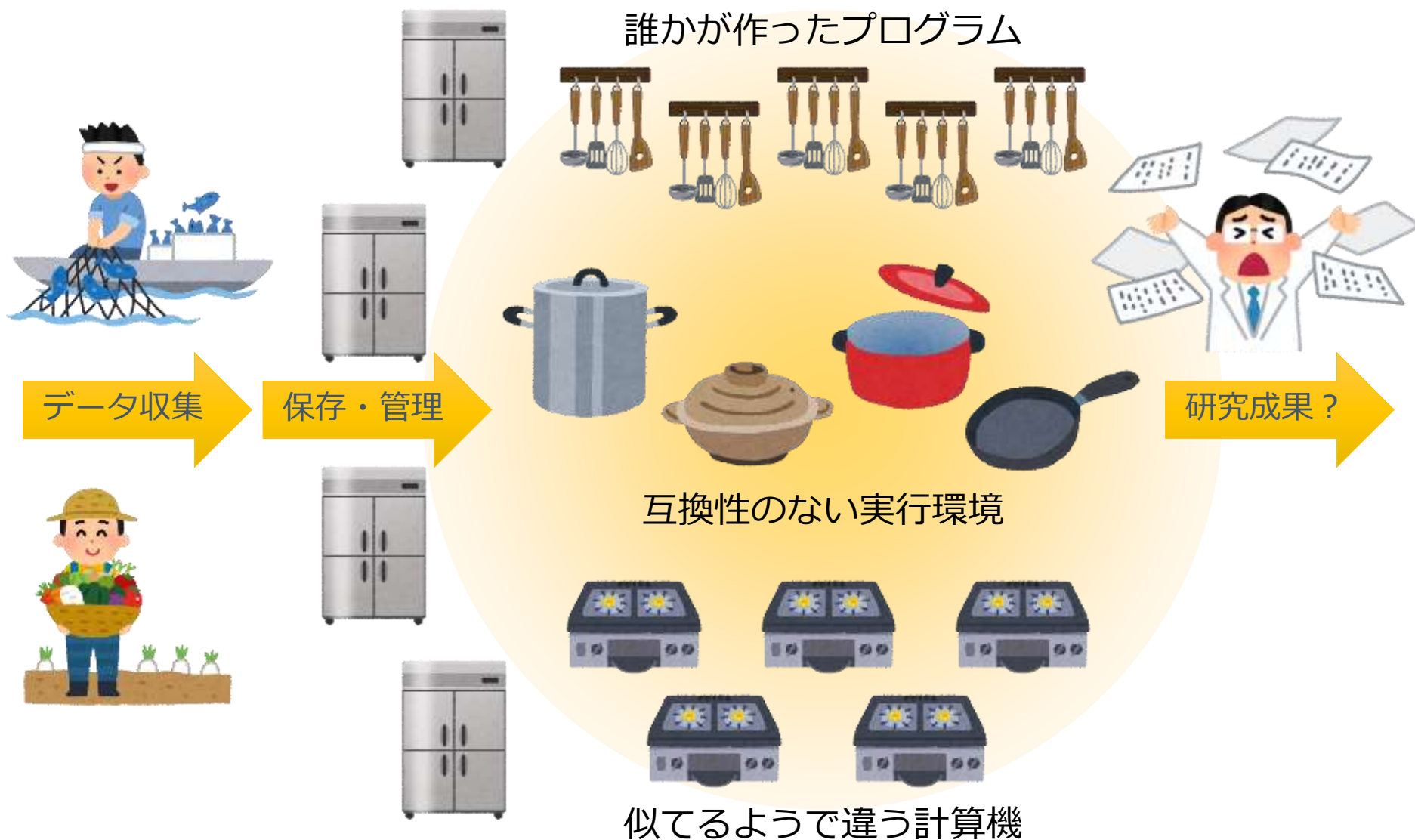
データ解析 = プログラム × 計算機 × 実行環境



データ解析 = プログラム × 計算機 × 実行環境

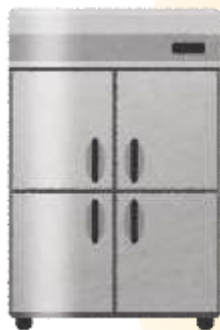


プログラム×計算機×実行環境 = ∞



資源をブローカリングする商社が必要

このデータを解析する
ならこのプログラム！



あなたが使える
一番速い計算機は
これ！



実行環境、作って
おきました！

巨人の肩の
上に立つ！



商社として将来実現したいこと

- 先行研究者が作ったデータとプログラムを発見し、解析内容を再現し、それを出発点として発展的な研究を始められるようにしたい。

論文・データ・プログラムを紐付ける**カタログサービス**

- 任意の計算資源を使って、システムの違いを意識せず、すぐに実行環境を構築してデータ解析を始められるようにしたい。

異なる計算機を同じように利用できる**実行環境構築サービス**

- 分野別に構築された既存のデータ基盤・計算基盤をまたいで、分野横断的なデータ解析をできるようにしたい。

セキュリティとガバナンスを保つ**データ流通サービス**

GakuNin RDMに統合される解析環境

GakuNin RDM マイプロジェクト 検索 Ikki Fujiwara

JH統合 ファイル Wiki メンバー **解析** 設定 証跡管理

このプロジェクトから作成した私の解析環境

[私の解析環境をすべて表示](#)

Server name	URL	Last activity	Launch	Delete
8b9wm-files-osfstorage-python-2vd0f1e4		2 months ago	    	
8b9wm-files-osfstorage-python-i8p14edf		2 months ago	    	
8b9wm-files-osfstorage-python-nyv1dwyo		2 months ago	    	

新しい解析環境

このプロジェクトから作成

外部リポジトリから作成

Python (3.7.3, miniconda 4.6.14) [Change](#)


conda makes this environment a great starting point for installing other languages.
Ubuntu 18.04 Python

Additional Packages ?

Customize the selected environment with any other packages you need. You can also use these package managers to install other package managers, such as for different languages. Packages will be installed on the next capsule run. [Learn more](#).

Package Managers

Packages

apt-get 

build-essential 12.4ubuntu1  git-lfs 2.3.4-1  [+ Add](#)

conda 

Cython 0.29.12  jupyter 1.0.0  matplotlib 3.1.0  networkx 2.1  numpy 1.16.4 
pandas 0.24.2  seaborn 0.8.0  xird 1.2.0  [+ Add](#)

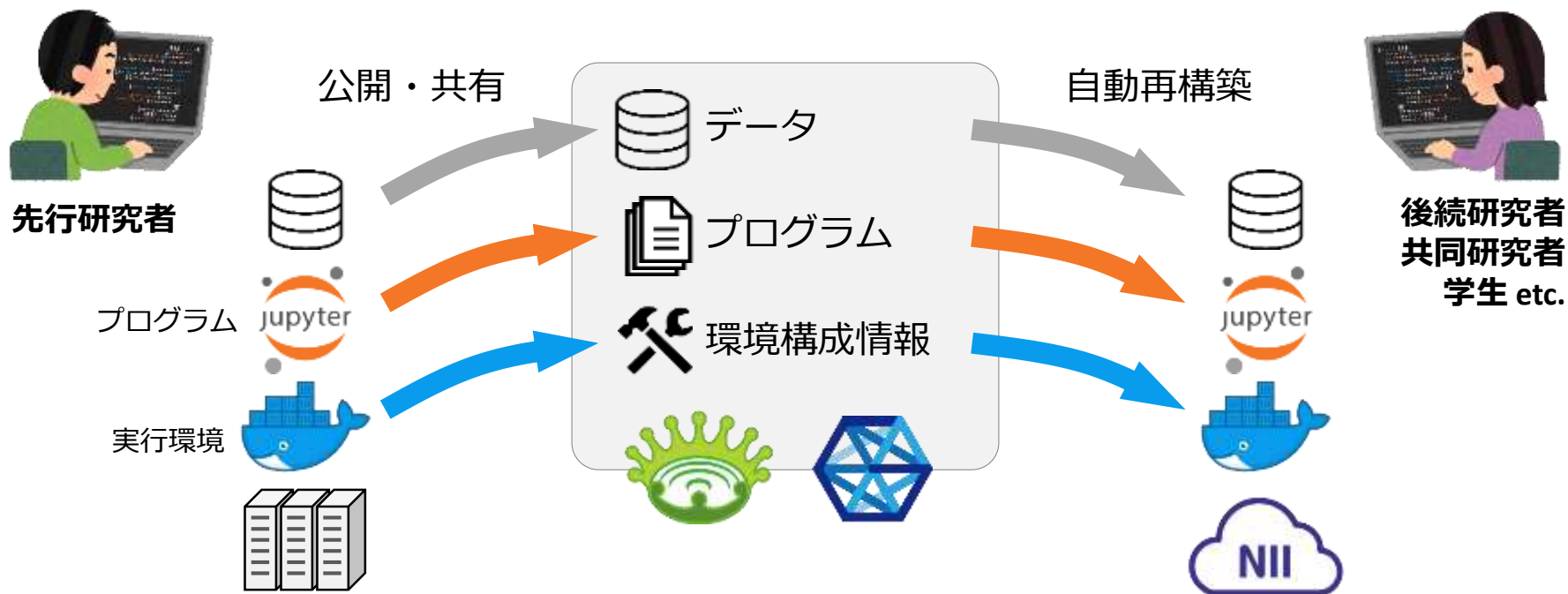
pip 

glypy 0.12.1  importlib 1.0.4  ndex 1.0.11.23  [+ Add](#)

Launch

研究のボタンタッチを簡単に



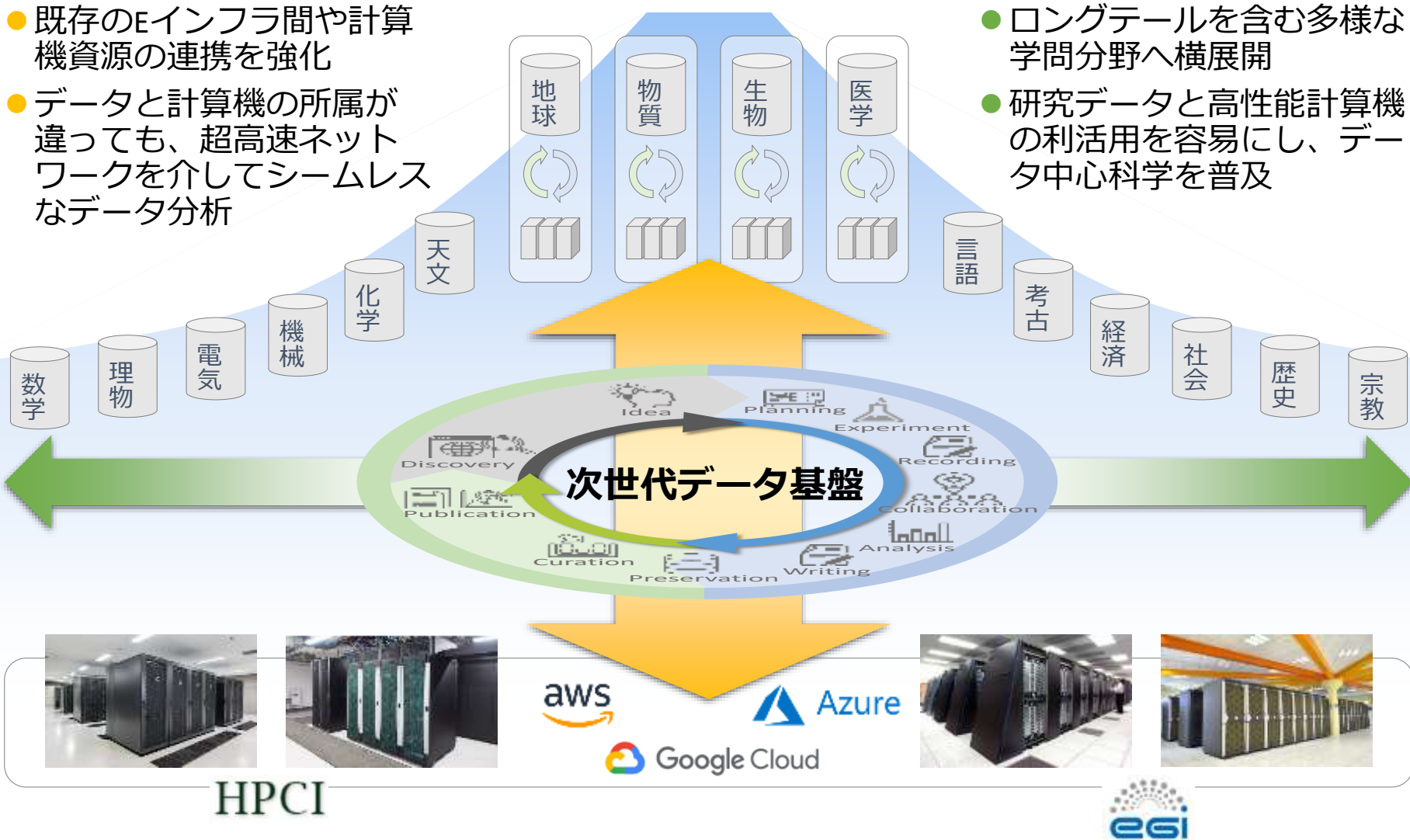
- 先行研究者は、データ・プログラム・環境構成情報をまとめてアップロードする
- 後続研究者は、個人用のコンテナ環境上に実行環境を再構築し、プログラムを実行する
- この**コンテナ再構築作業を自動化**し、誰でも再利用可能にする

研究データと計算機システムをもっと仲良く

- 既存のEインフラ間や計算機資源の連携を強化
- データと計算機の所属が違って、超高速ネットワークを介してシームレスなデータ分析

既存の分野別サービス

- ロングテールを含む多様な学問分野へ横展開
- 研究データと高性能計算機の利活用を容易にし、データ中心科学を普及



オープンサイエンスの環をつなげよう

- 今後の開発について、将来のユーザーとなる皆さまのアイデアをお聞かせください。

