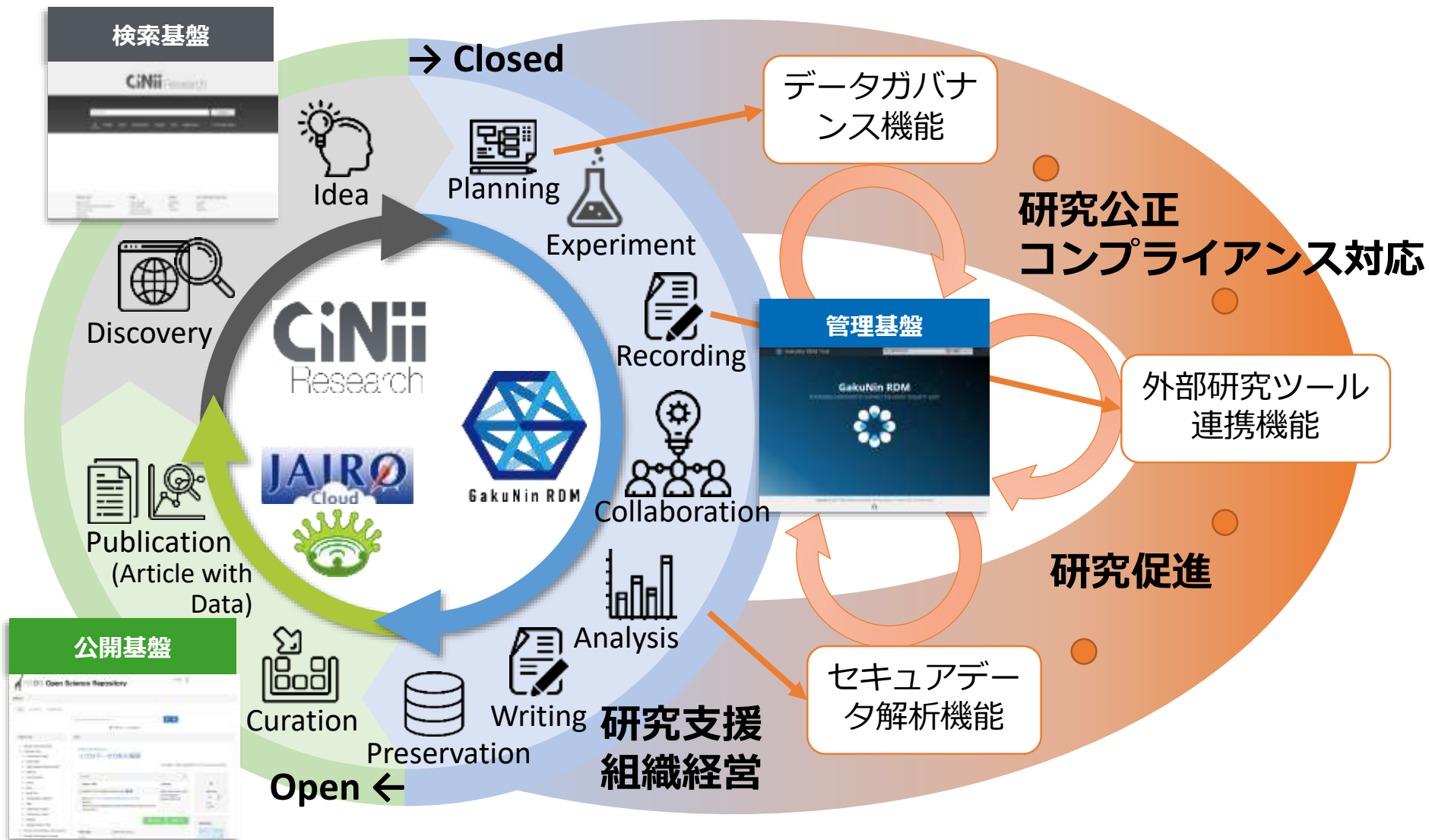


コードの再利用を容易にする データ解析機能

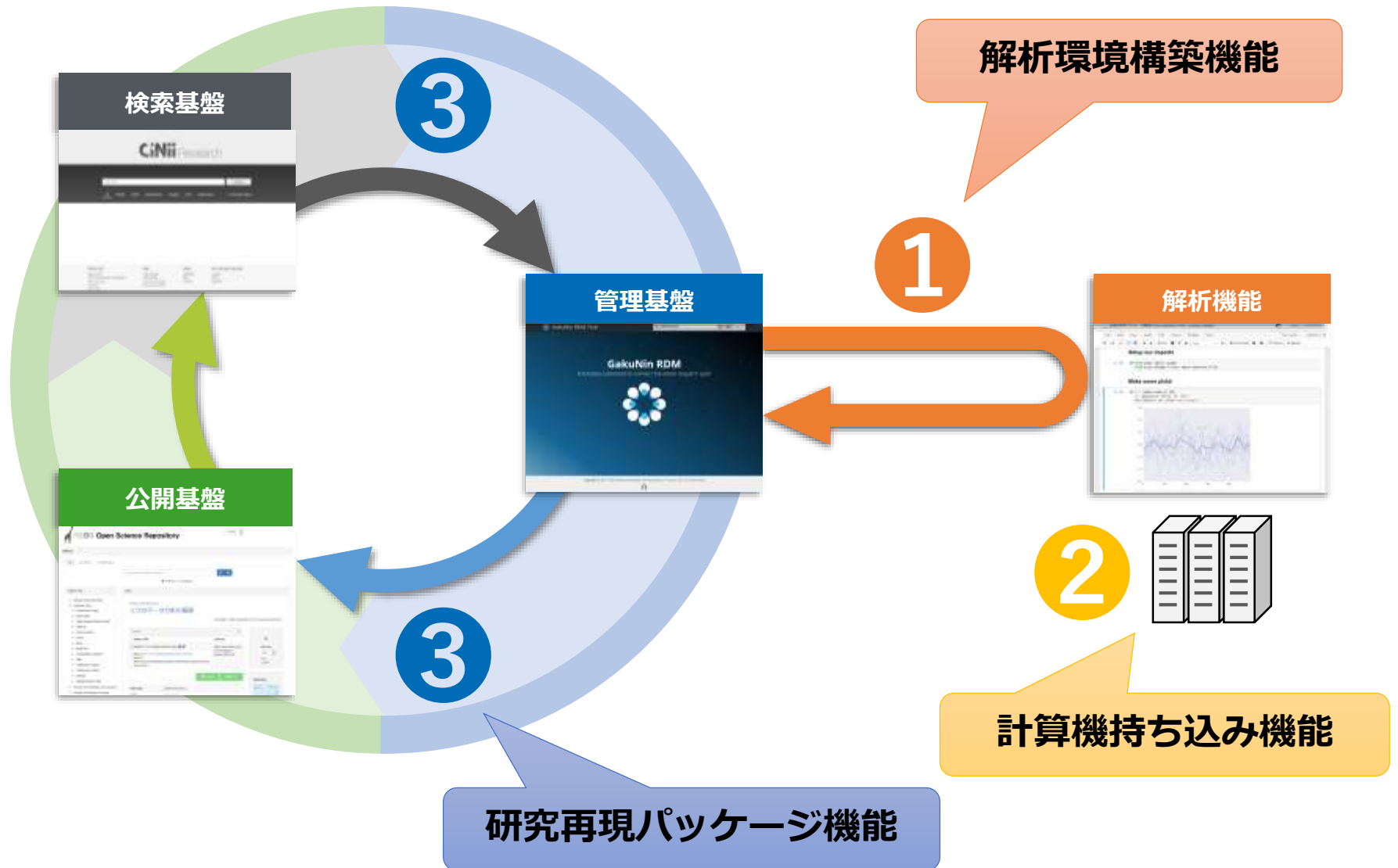
藤原一毅

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

次世代学術研究プラットフォーム



次世代学術研究プラットフォームの データ解析機能



データ解析も GakuNin RDM を中心に

1 解析環境構築機能

準備中

プログラム開発・実行環境を NII のクラウド上に自動構築し、GakuNin RDM を介して共有・再利用する。

2 計算機持ち込み機能

開発中

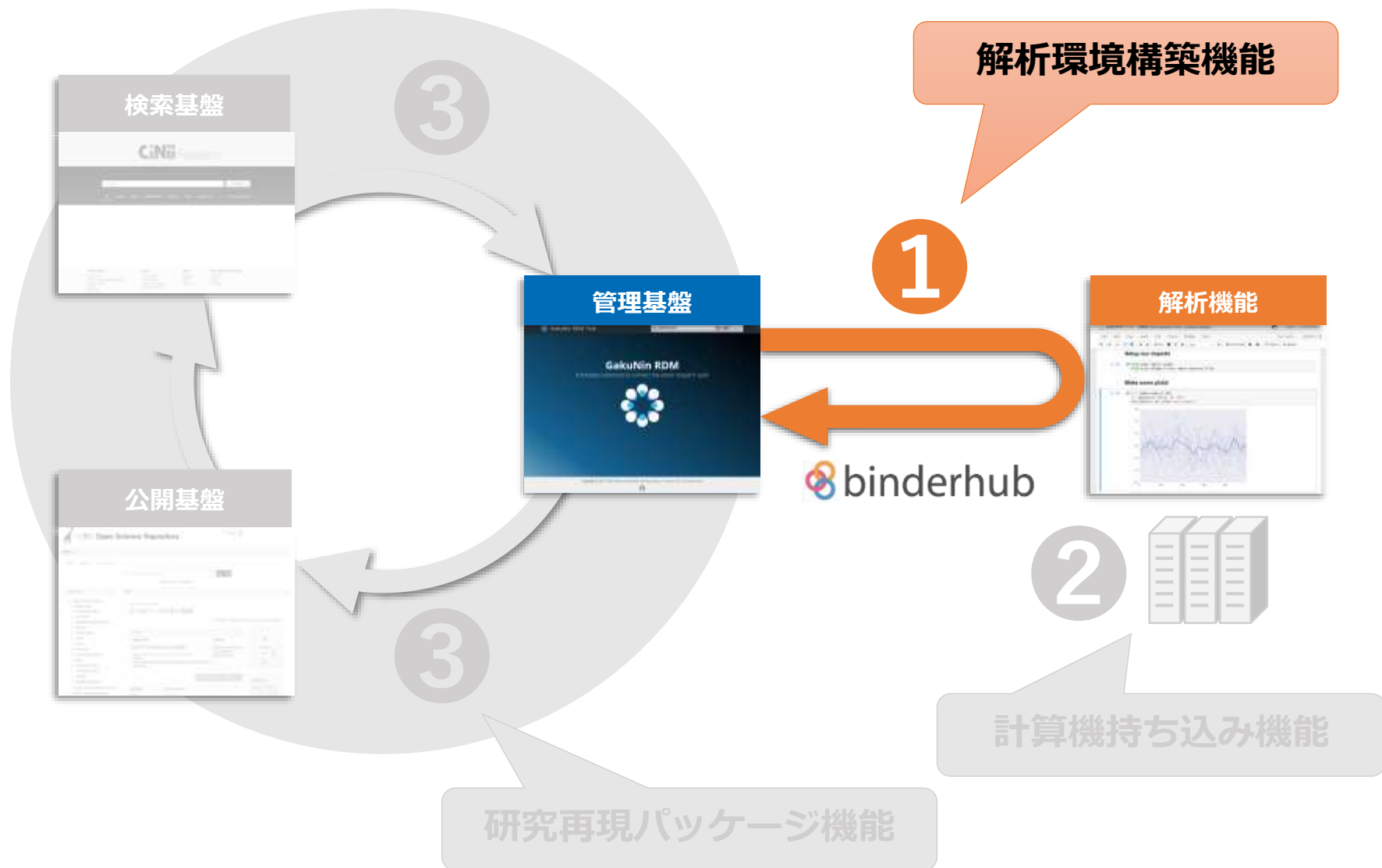
利用者が自身の（所属機関の）計算機を GakuNin RDM に登録し、そこにプログラム開発・実行環境を自動構築する。

3 研究再現パッケージ機能

構想中

論文・データ・プログラムを紐付けた「研究再現パッケージ」を GakuNin RDM で作成し、NII Research Data Cloud の各サービスを通じて共有・公開・検索・再利用する。

1 解析環境構築機能



1 解析環境構築機能

- プログラム開発・実行環境を自動的に構築する
- 環境構成情報は GakuNin RDM に保存され、共有・再利用できる

(利用手順イメージ)

1. GakuNin RDM に追加される「解析」画面を開く
2. ベース環境と追加パッケージを選択し、[Launch] ボタンを押す
 - 現在のところ Jupyter Notebook, JupyterLab, RStudio をサポート
3. プログラム開発・実行環境を含む個人用コンテナが NII のクラウド上に生成され、すぐに利用できる
 - GakuNin RDM の標準ストレージにあるファイルはコンテナにコピーされる
 - GakuNin RDM に接続された外部ストレージはコンテナにマウントされる
4. プログラムを作成・実行する
 - 解析結果ファイルを [result] フォルダに出力する
5. 書き戻しボタンを押す
 - [result] フォルダの中身が GakuNin RDM にコピーされる

解析環境を自動的に構築する

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

解析環境名	URL	最終更新	起動	削除	
vdqrb-osfstorage-6zq3agwc	/user/test001@nii.ac.jp/vdqrb-osfstorage-6zq3agwc/	2021-03-29 01:13 PM	  		
vdqrb-osfstorage-fp0c2a6t	/user/test001@nii.ac.jp/vdqrb-osfstorage-fp0c2a6t/	2021-03-29 01:13 PM	  		

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

① ベース環境を選ぶ

環境

R (4.0.3, RStudio 1.2.5019) ✓

R is a language and environment for statistical computing and graphics. RStudio is an integrated development environment for R.

変更

追加パッケージ

apt-get sl: ✕ +追加conda +追加R (CRAN) +追加R (GitHub) +追加

② 追加パッケージを選ぶ

インストール後に実行するスクリプト

デフォルトストレージ内にpostInstallスクリプトを定義すると、イメージビルド時に任意の処理を実行できます。 [ファイル一覧を開く](#)

③ 起動！

更新

ビルド状況



解析環境を利用する

The screenshot shows a JupyterLab interface with the following elements:

- Header:** "jupyterhub index Last Checkpoint: 2分前 (unsaved changes)" and "Logout Control Panel".
- Menu:** File, Edit, View, Insert, Cell, Kernel, Widgets, Help.
- Toolbar:** Includes icons for file operations, a "Run" button, and a "Code" dropdown menu.
- Code Cell 1:** Titled "Setup our imports", containing the code:

```
In [3]: from numpy import random
from scipy.ndimage.filters import gaussian_filter
```
- Code Cell 2:** Titled "Make some plots!", containing the code:

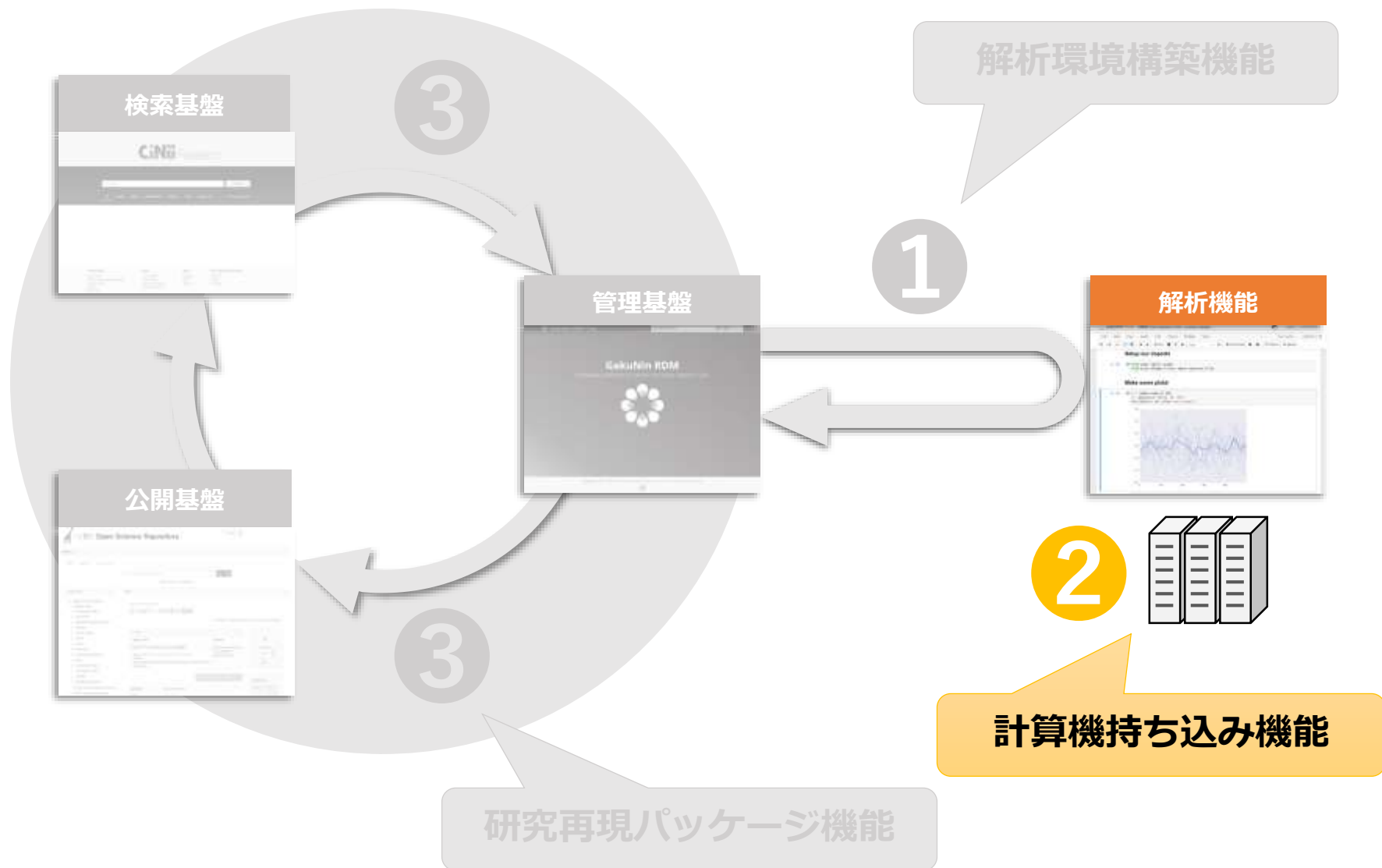
```
In [3]: x = random.randn(10, 500)
x = gaussian_filter(x, [0, 10])
sns.tsplot(x, err_style='unit_traces');
```
- Figure:** A time series plot showing 10 overlapping noisy signals in light blue and their smoothed version in a darker blue. The x-axis ranges from 0 to 500, and the y-axis ranges from -0.6 to 0.6.
- Code Cell 3:** Titled "In [4]:", containing the code:

```
In [4]: mat = random.randn(10, 20)
sns.heatmap(mat);
```
- Figure:** A heatmap visualization of a 10x20 matrix, showing a color gradient from blue to red.

⑤書き戻し

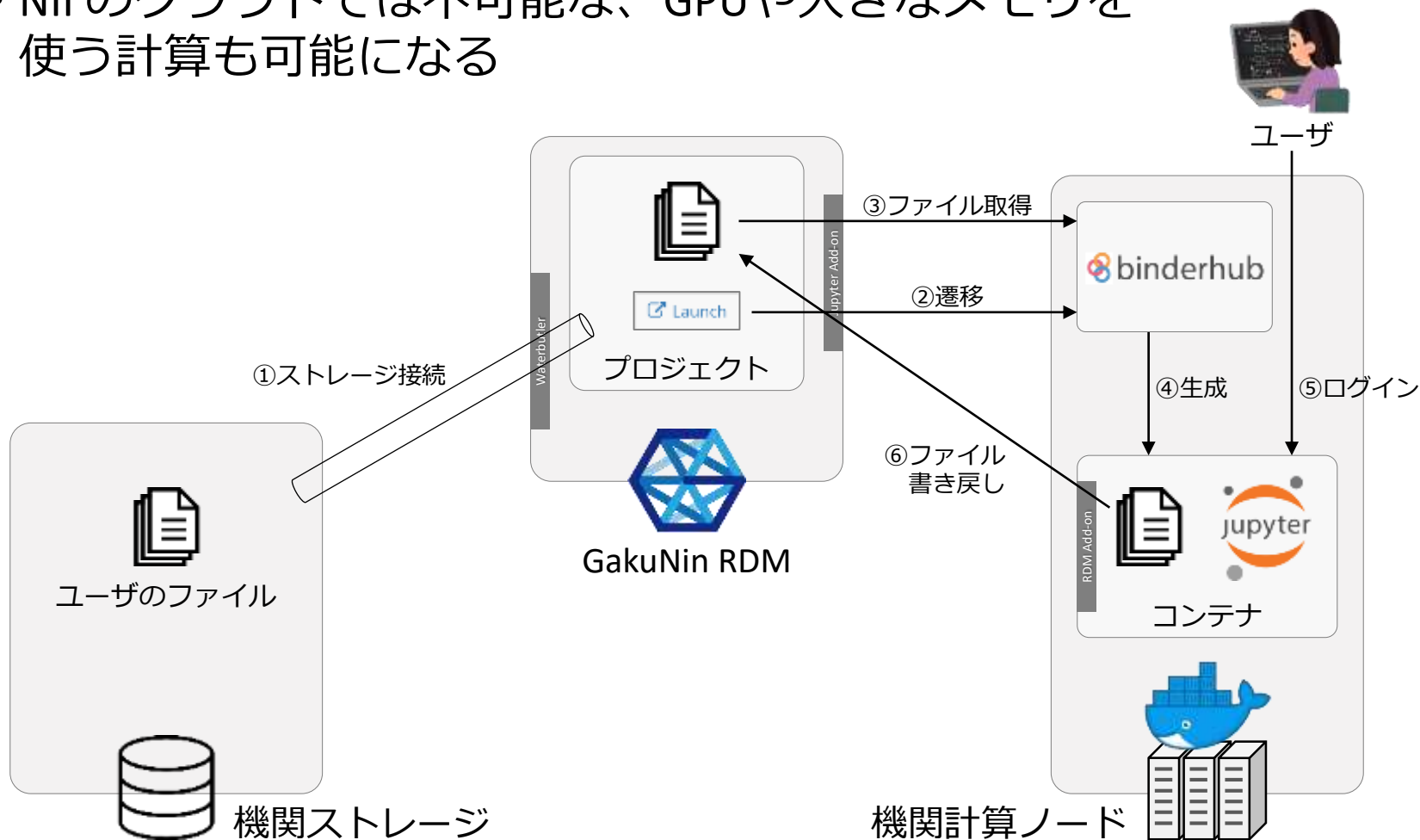
④プログラム編集・実行

② 計算機持ち込み機能



② 計算機持ち込み機能

- ユーザーの（所属機関の）計算ノード上に解析環境を構築する
- NII のクラウドでは不可能な、GPUや大きなメモリを使う計算も可能になる



持ち込んだ計算機に解析環境を構築する

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

解析環境名	URL	最終更新	起動	削除	
vdqrb-osfstorage-6zq3agwc	/user/test001@nii.ac.jp/vdqrb-osfstorage-6zq3agwc/	2021-03-29 01:13 PM			
vdqrb-osfstorage-fp0c2a6t	/user/test001@nii.ac.jp/vdqrb-osfstorage-fp0c2a6t/	2021-03-29 01:13 PM			

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](#)

環境

R (4.0.3, RStudio 1.2.5019) ✓

R is a language and environment for statistical computing and graphics. RStudio is an integrated development environment for R.

変更

追加パッケージ

apt-get sl: ✕

conda

R (CRAN)

R (GitHub)

計算機を選択

● ● 大学スパコン ▼

環境構築先を選ぶ

インストール後に実行するスクリプト

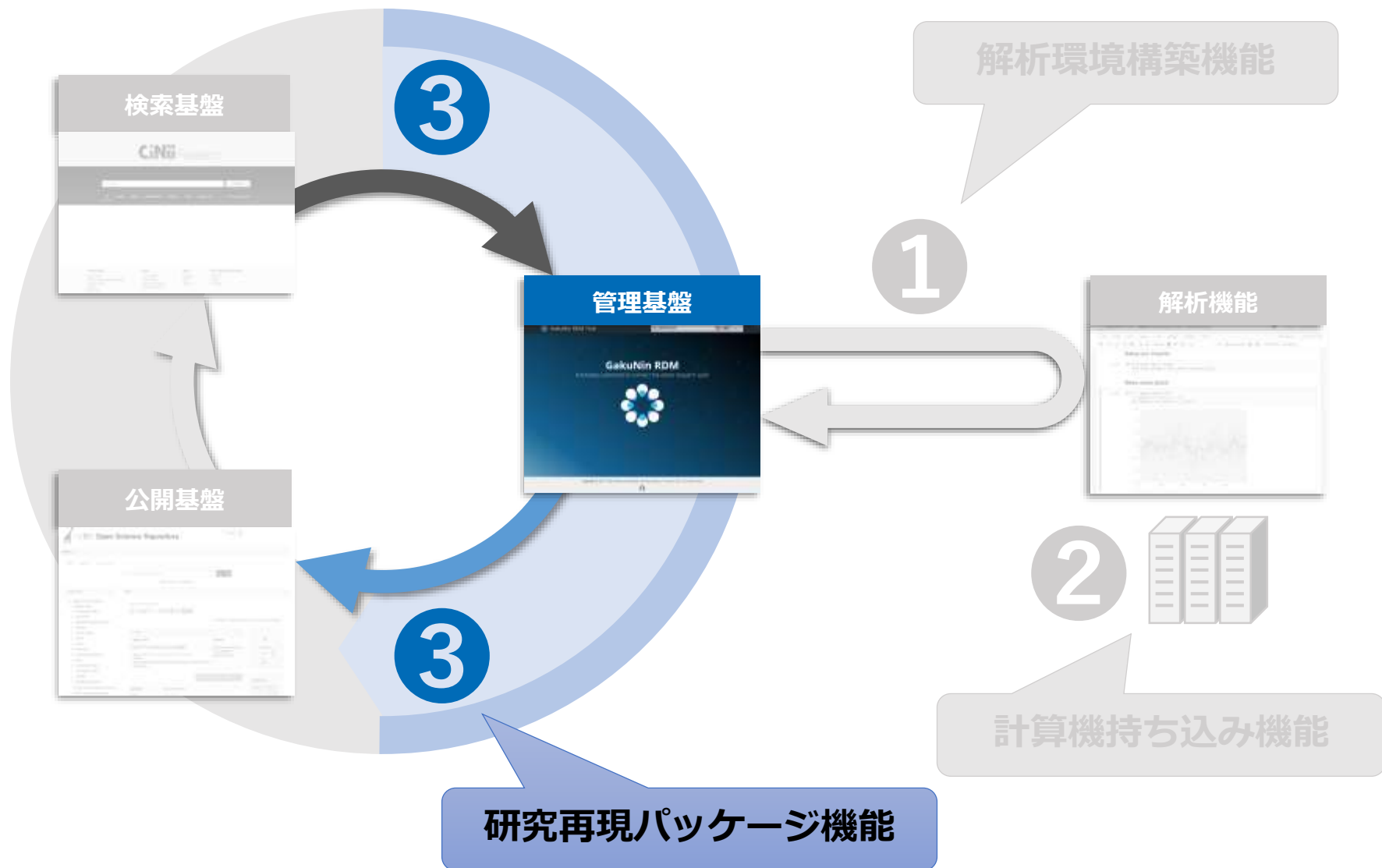
デフォルトストレージ内にpostInstallスクリプトを定義すると、イメージビルド時に任意の処理を実行できます。 [ファイル一覧を開く](#)

更新

ビルド状況



3 研究再現パッケージ機能

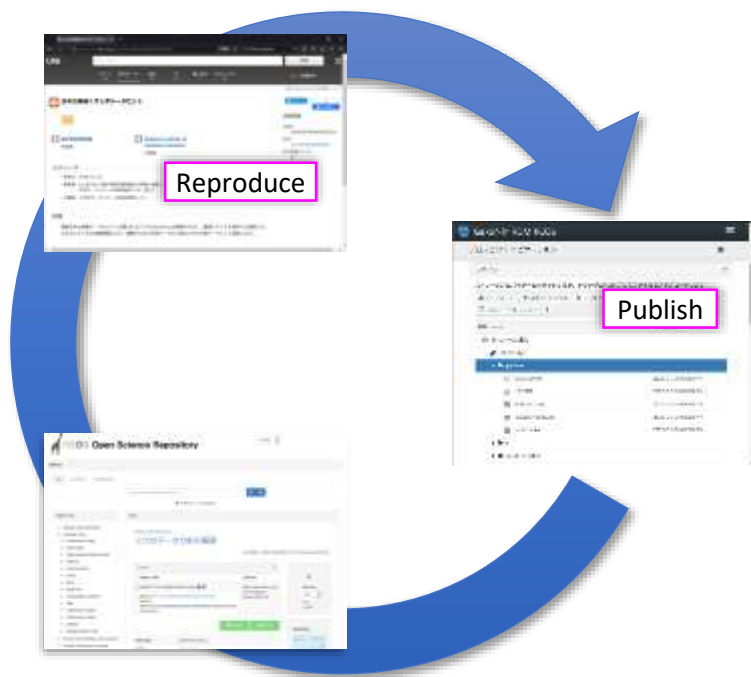


③ 研究再現パッケージ機能

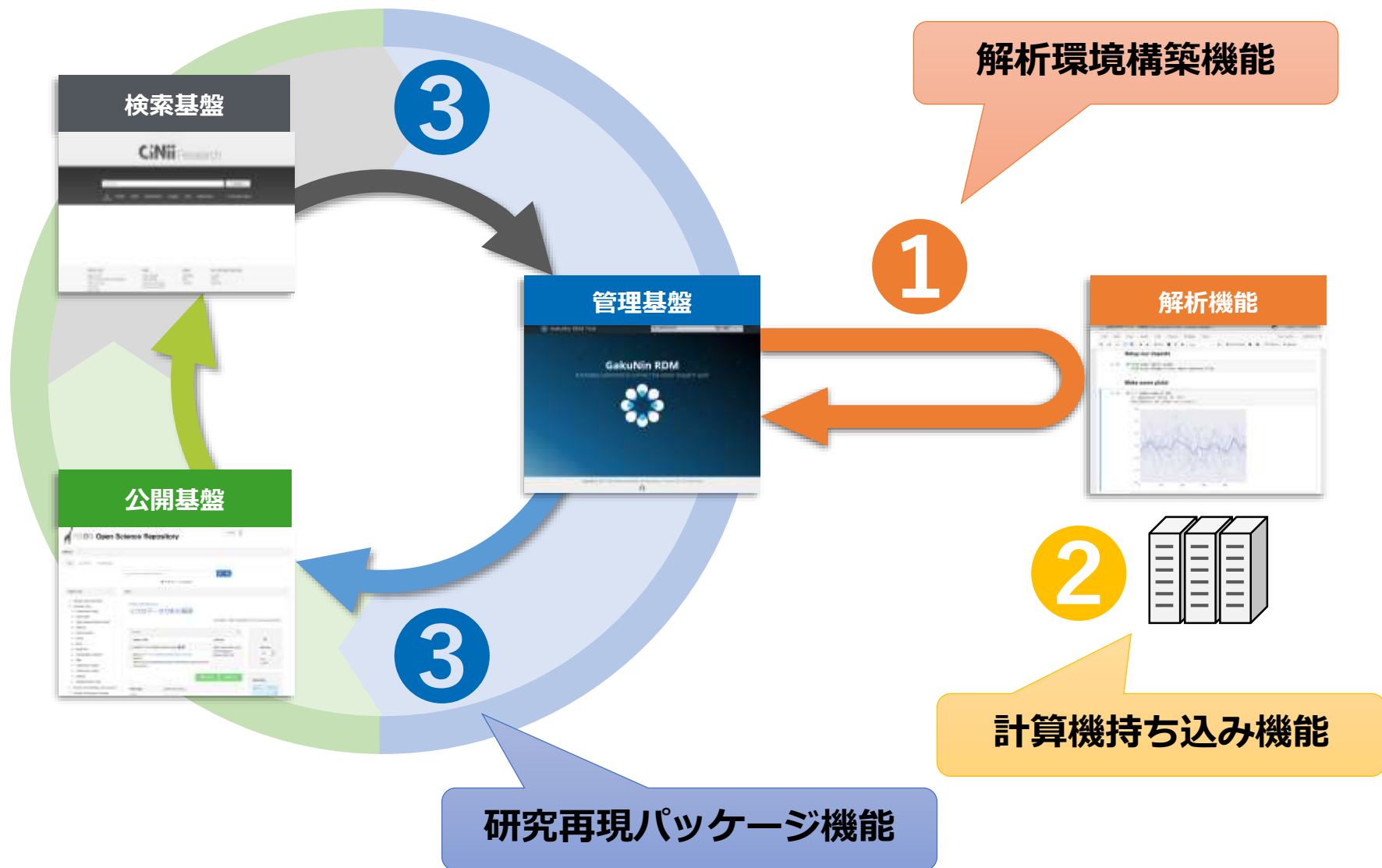
- 論文・データ・プログラムを含む一連の研究成果を、再利用可能なパッケージとして流通させたい
- 先行研究を容易に再現し、発展的な研究をすぐに開始できる

(再現手順イメージ)

1. 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 GakuNin RDM のひとつのプロジェクトに保存する。
2. GakuNin RDM の「パッケージ公開」画面を開き、関連論文などのメタデータを入力し、[Publish] ボタンを押す。
3. パッケージが JAIRO Cloud に登録・公開され、CiNii Research で検索可能となる。
4. 後続研究者は、CiNii Research の検索結果画面で [Reproduce] ボタンを押すと、新しい GakuNin RDM プロジェクトが作られ、パッケージの中身が展開される。



オープンサイエンスの環をつなげよう



ご利用案内

- **0** 解析環境構築機能（BinderHub単体）
 - 試験提供中。詳細は →
- **1** 解析環境構築機能（GakuNin RDM連携）
 - 2022年度より、GakuNin RDM利用者に標準機能として提供予定
- **2** 計算機持ち込み機能
 - パートナー募集中。ご協力いただける方はご連絡ください
- **3** 研究再現パッケージ機能
 - 構想中

RCOS
cs-support@nii.ac.jp