



Solving the right problems: **Requirements capture for large-scale, evolving research projects**

Gabrielle M. Schroeder
Newcastle University RSE Team
RSECon24

Embarking on a new RSE project

Goal: Map the course of the project

Image: Tabea Schimpf (Unsplash)

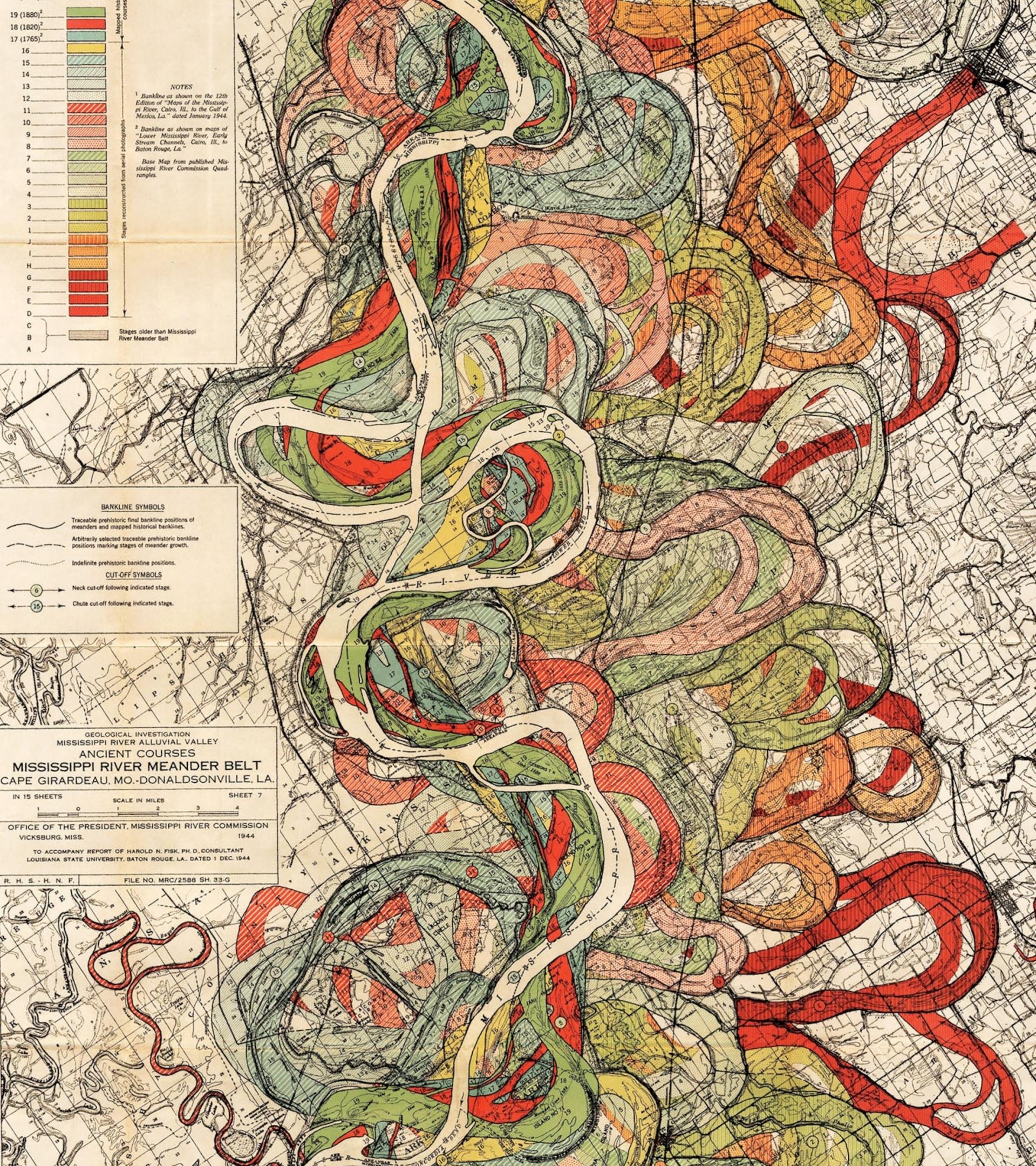


Embarking on a new RSE project

Goal: Map the course of the project

Challenge: Large, evolving research projects

“The nature and origin of the Alluvial Valley of the Lower Mississippi River” by Harold Fisk (1944)



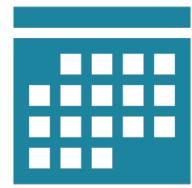
OpenScan

Driving **horizon scanning**
research at the National
Institute for Health and Care
Research **Innovation**
Observatory (NIHRCO)

Image: Clemens van Lay (Unsplash)



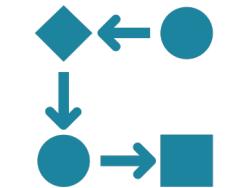
Capturing OpenScan's requirements



Assessment
stage



Regular
meetings



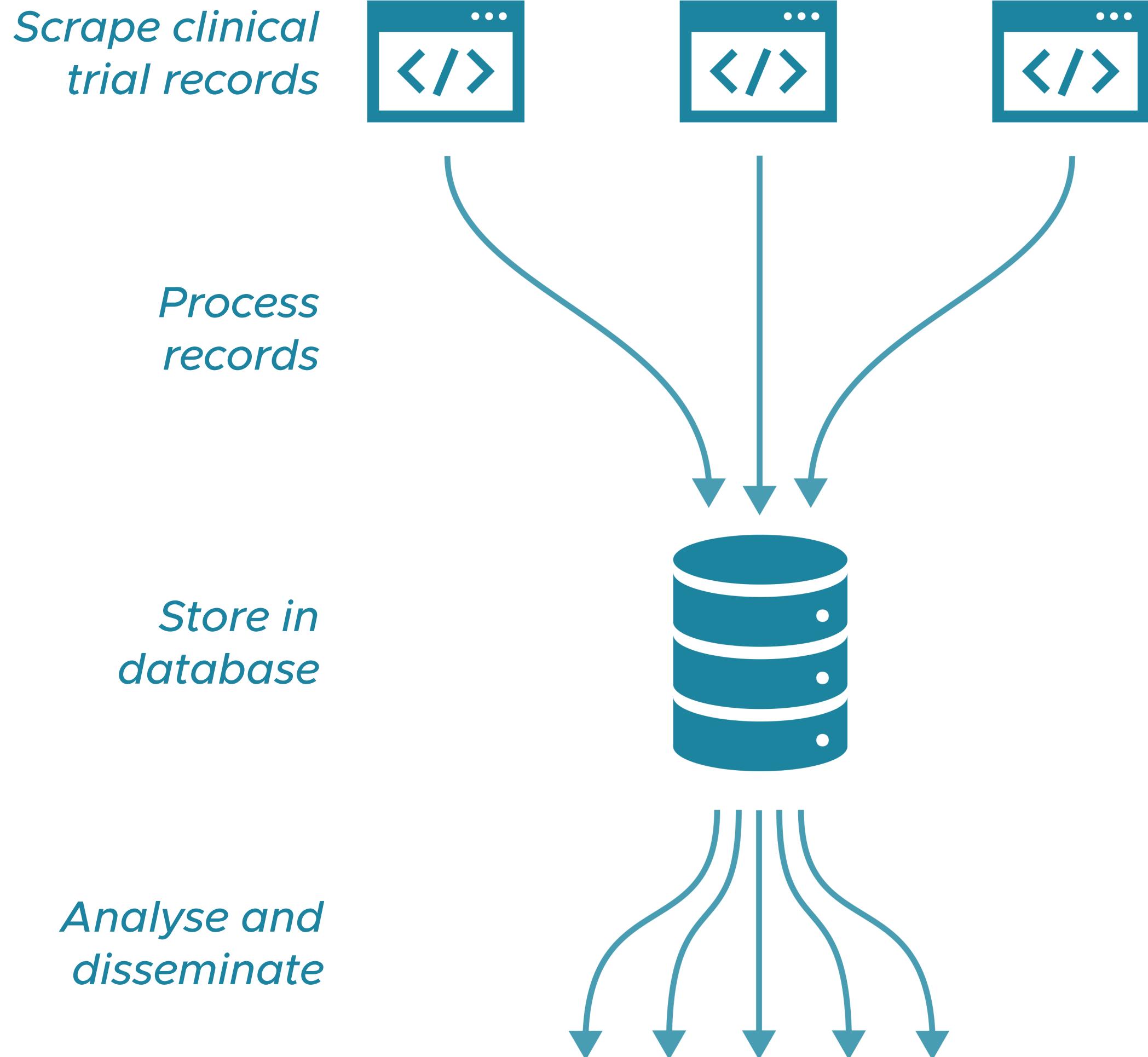
Document
current state



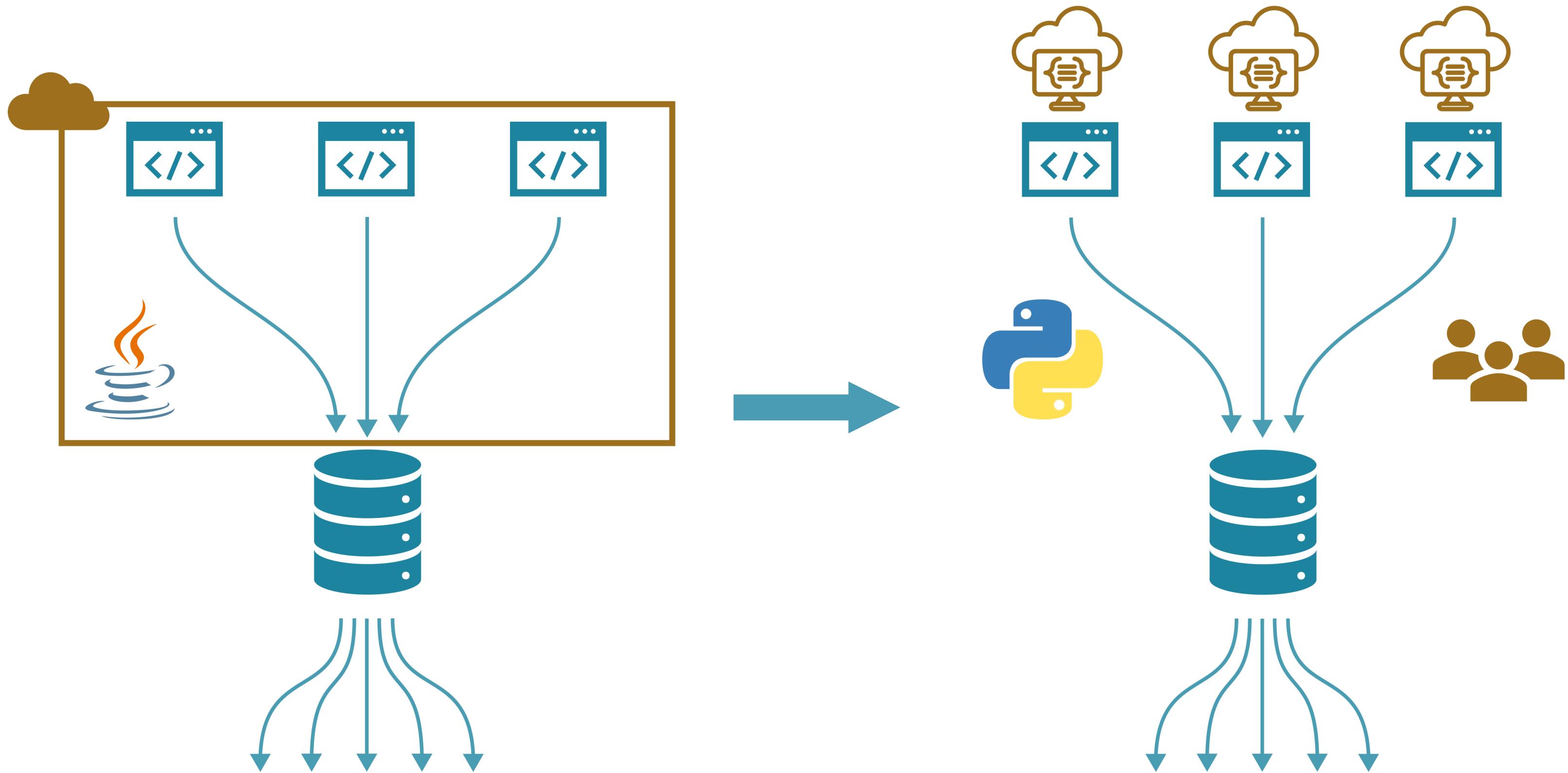
Document
planned work

Example work package

Re-designing the web scraper system



Requested changes



Strategy: User stories

As a **<user>** I want to **<goal>** so that **<benefit>**.

Strategy: User stories

As a **<user>** I want to **<goal>** so that **<benefit>**.

Goals should not include the technical solution:

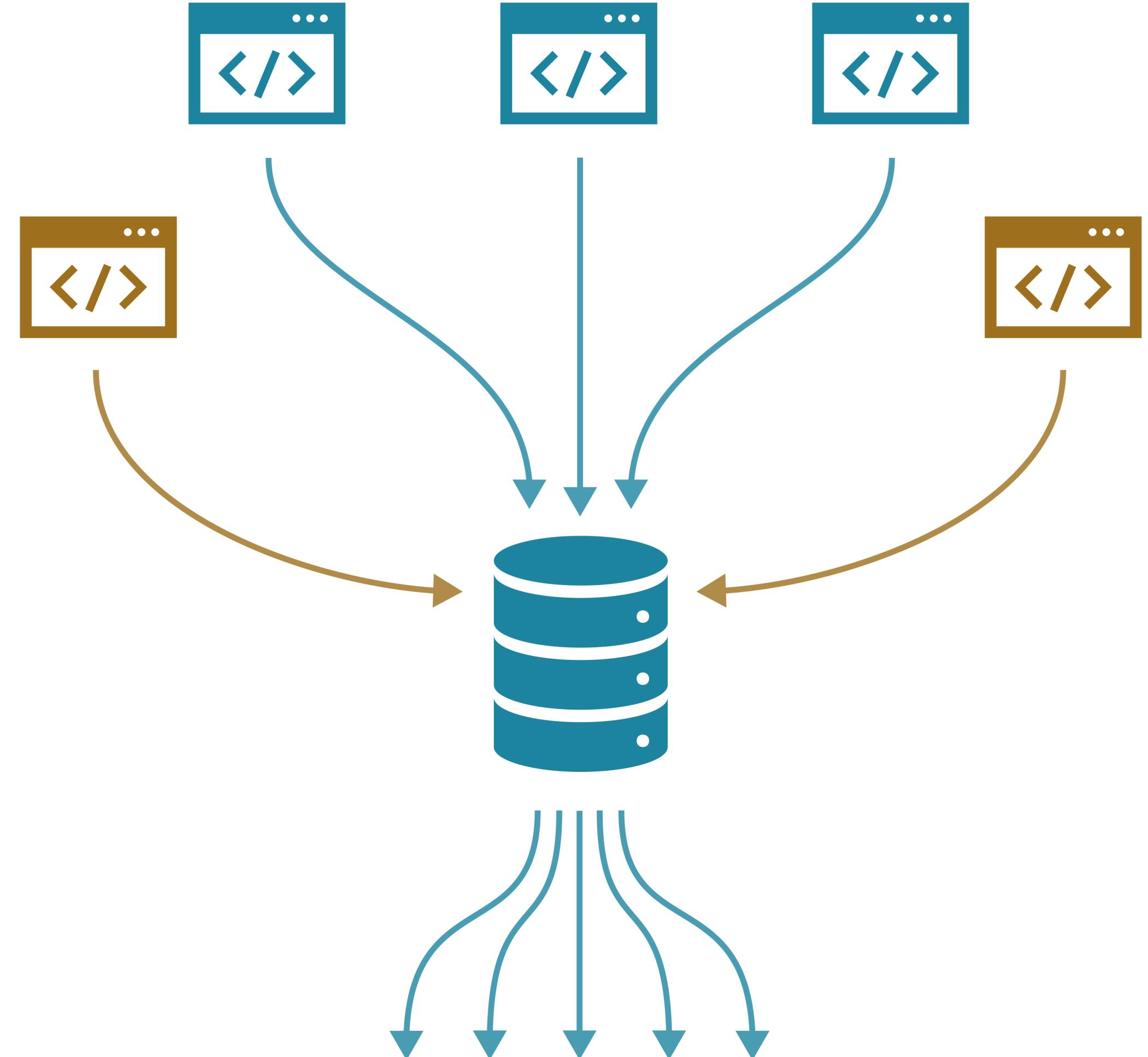
I want to easily train researchers to write web scrapers

rather than

I want to write web scrapers using Python

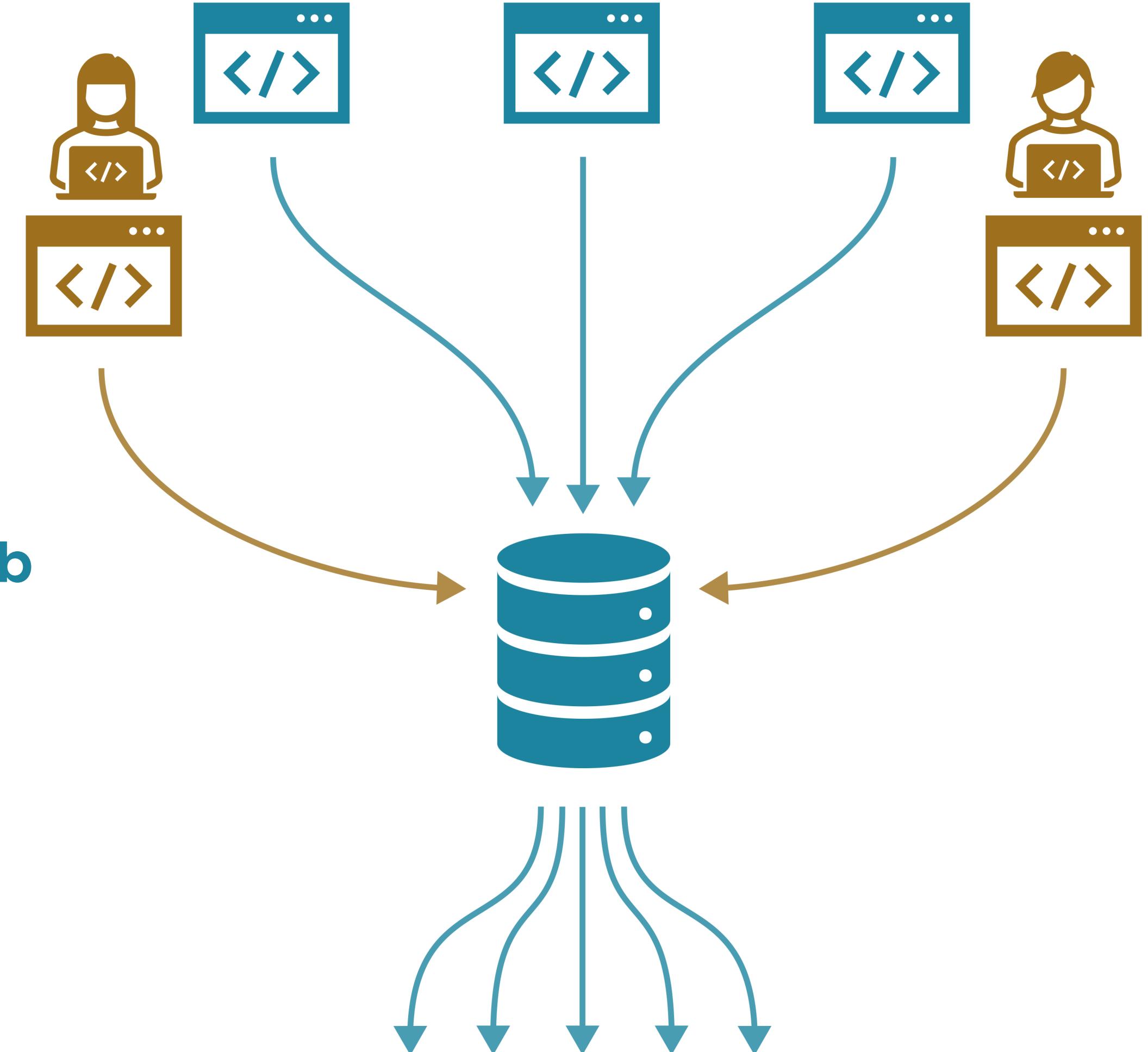
User stories

As an **NIHRI** developer,
I want to **easily add new
web scrapers**
so that **we can expand
OpenScan and develop
similar projects.**



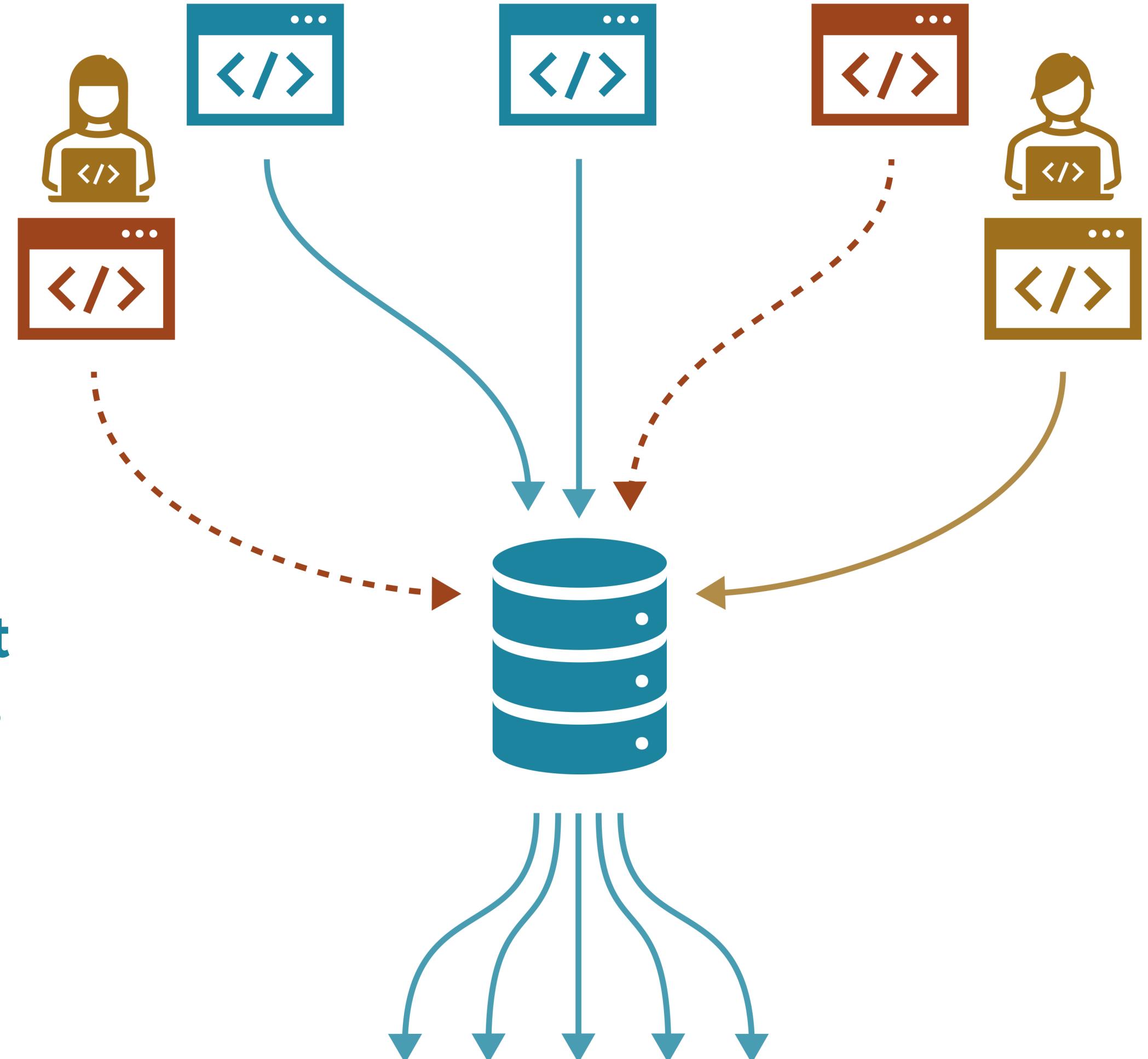
User stories

As an **NIHRIo developer**,
I want to **collaborate on web scrapers**
so that **multiple people can maintain and develop OpenScan**.



User stories

As an **NIHRIo developer**,
I want to **ensure that
different developers do not
break other's web scrapers**
so that **our pipelines are
reliable and robust**.

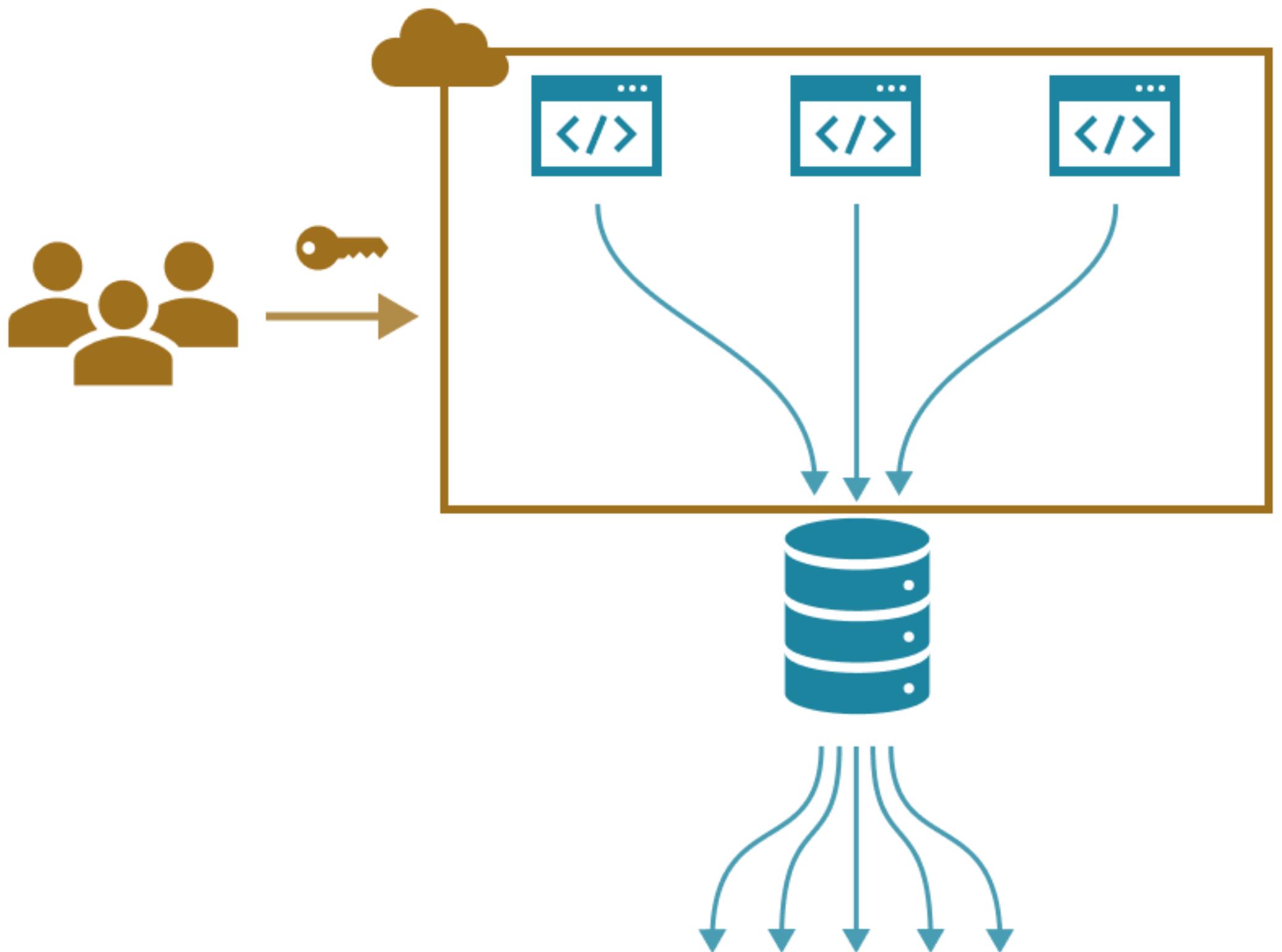


Strategy: Investigate current workflows

Strategy: Investigate current workflows

One login for server!

- Difficult to deploy
- Poor version control
- Fragile



New requirement

Need **easy-to-use, reproducible, and traceable workflows** for developing web scrapers.

Create function Info

Choose one of the following options to create your function.

Author from scratch

Start with a simple Hello World example.

Use a blueprint

Build a Lambda application from sample code and configuration presets for common use cases.

Container image

Select a container image for your function.

Basic information

Function name

Enter a name that describes the purpose of your function.

myFunctionName

Use only letters, numbers, hyphens, or underscores with no spaces.

Runtime Info

Choose the language to use to write your function. Note that the console code editor supports only Node.js, Python, and Ruby.

Python 3.11



Architecture Info

Choose the instruction set architecture you want for your function code.

x86_64

arm64

Permissions Info

By default, Lambda will create an execution role with permissions to upload logs to Amazon CloudWatch Logs. You can customize this default execution role.

▶ Change default execution role

New requirement

Need **easy-to-use, reproducible, and traceable workflows** for adding web scrapers.

Solution:
Infrastructure-as-Code
(in Python)

```
# Serverless function (AWS Lambda) for web scraper
scraper_lambda = lambda_.Function(
    self,
    self.scrape_label + "ScraperLambda",
    function_name = self.scrape_label + 'Scraper',
    runtime = lambda_.Runtime.PYTHON_3_12,
    code = lambda_.Code.from_asset('lambda/lambda_clinicaltrials'),
    handler = 'crawler_lambda.lambda_handler',
    timeout = Duration.minutes(2),
    memory_size = 5308,
    layers = [layer_util, layer_requests, layer_powertools],
    environment = {
        "POWERTOOLS_SERVICE_NAME": self.crawler_label,
        "POWERTOOLS_LOG_LEVEL": "INFO"
    }
)
```

Summary

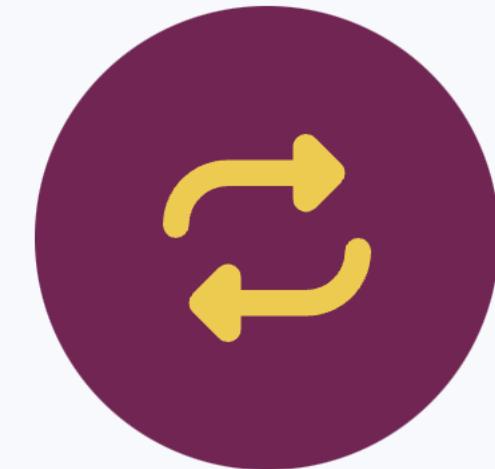
We identified additional requirements for OpenScan's redesign.

Careful requirements capture will help make OpenScan maintainable and reproducible.



Maintainable

Easy to adapt and to correct faults



Reproducible

Enable trust in research

Lessons for RSE projects

- Dedicate time for requirements capture.
- Investigate the current state and pain points.
- Document requirements to formalise your thoughts.
- Plan for long-term needs.



Thank you

Contact

gabrielle.schroeder@ncl.ac.uk

www.linkedin.com/in/gabrielleschroeder



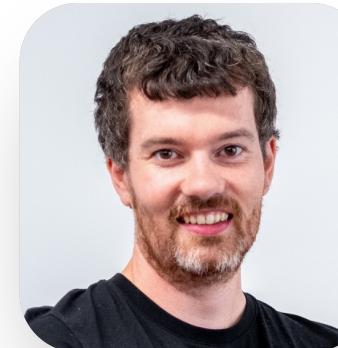
Robin
Wardle



Gabrielle
Schroeder



Kate
Court



Mark
Turner



Chris
Marshall



Saleh
Mohamed



Amey
Vedpathak



Hongbo
Bo



Kieran
McDonough



Karim
Elkobrossy



Teresa
Fortune



Dawn
Craig