

Harvester Automation for Metadata Search Web Application

Part One: Methodologies by Terry Yuan, SIParCS Intern
Mentors: Nathan Hook, Saquib Aziz Khan, Christy Grant, Eric Nienhouse

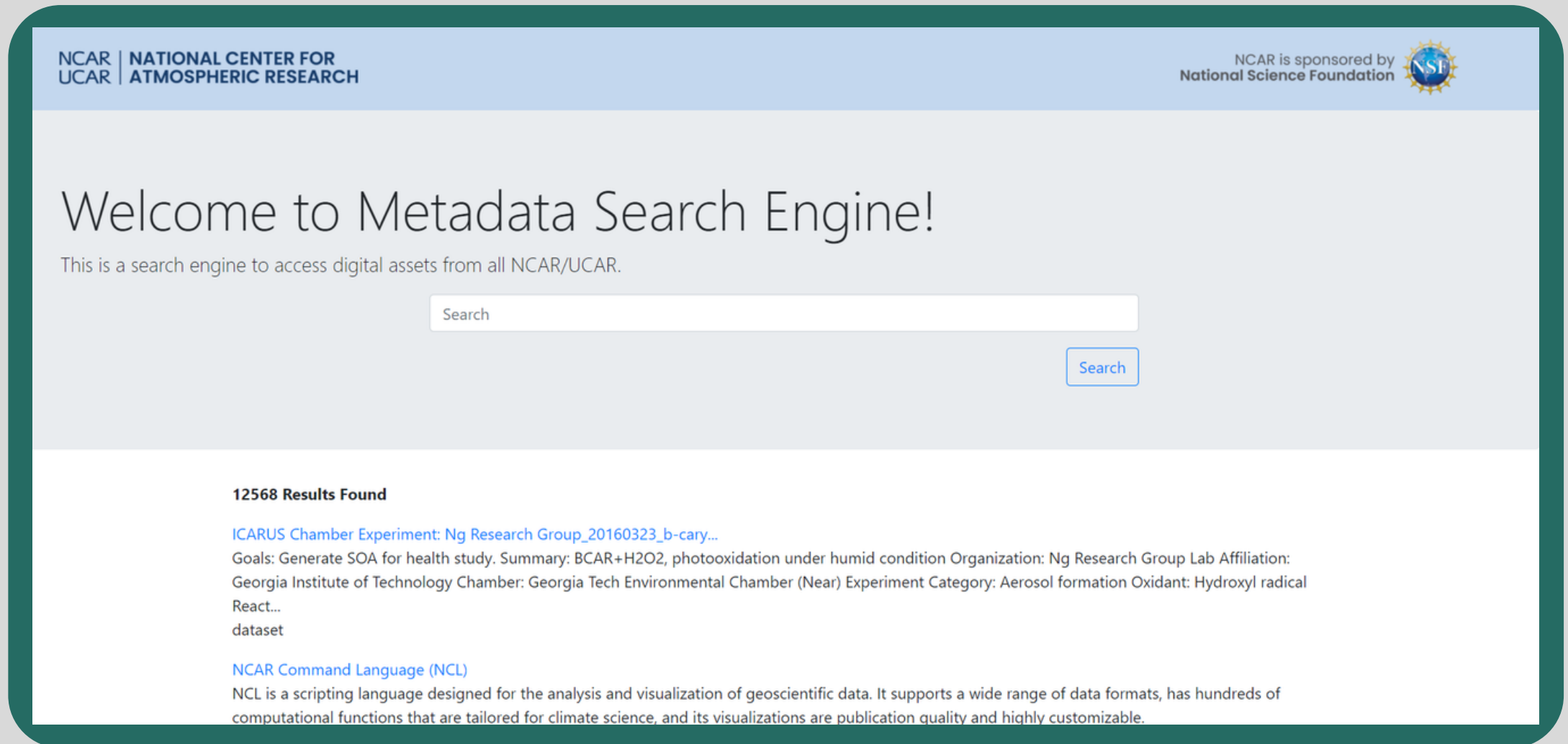
NCAR
UCAR

Sama Manalai / **Terry Yuan**
SIParCS 2021: Project 7



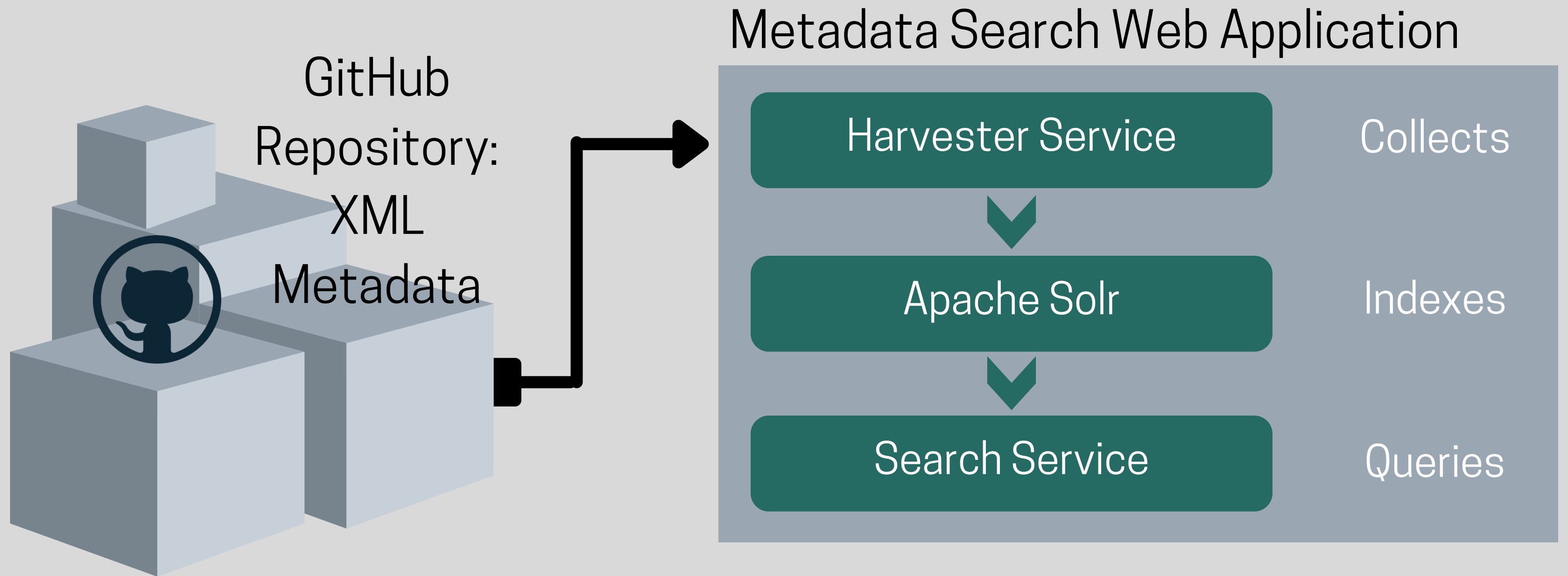
Background

A Search Tool
Built Under NCAR
Made For NCAR

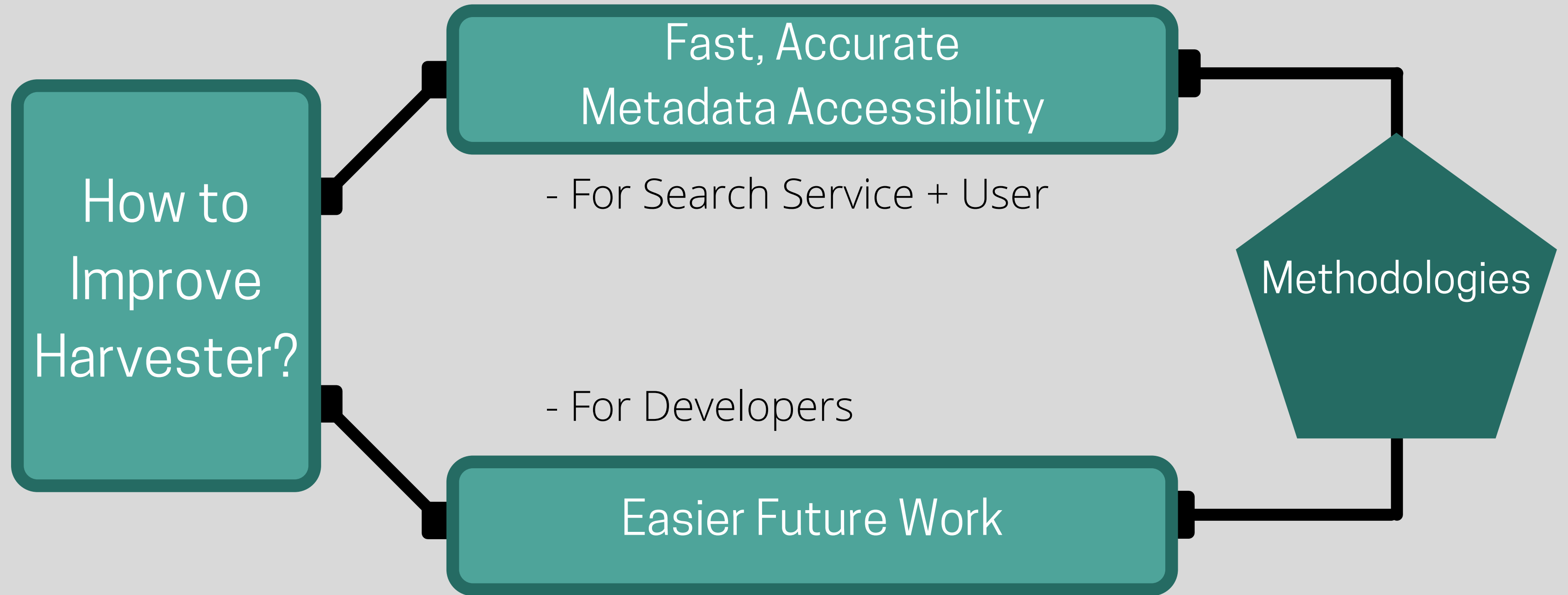


The screenshot shows the NCAR Metadata Search Engine interface. At the top left, it says "NCAR | NATIONAL CENTER FOR UCAR | ATMOSPHERIC RESEARCH". At the top right, it says "NCAR is sponsored by National Science Foundation" with the NSF logo. The main heading is "Welcome to Metadata Search Engine!". Below this, it says "This is a search engine to access digital assets from all NCAR/UCAR." There is a search input field with the placeholder text "Search" and a "Search" button. Below the search bar, it says "12568 Results Found". The first result is "ICARUS Chamber Experiment: Ng Research Group_20160323_b-cary...". Below the title, it says "Goals: Generate SOA for health study. Summary: BCAR+H2O2, photooxidation under humid condition Organization: Ng Research Group Lab Affiliation: Georgia Institute of Technology Chamber: Georgia Tech Environmental Chamber (Near) Experiment Category: Aerosol formation Oxidant: Hydroxyl radical React... dataset". The second result is "NCAR Command Language (NCL)". Below the title, it says "NCL is a scripting language designed for the analysis and visualization of geoscientific data. It supports a wide range of data formats, has hundreds of computational functions that are tailored for climate science, and its visualizations are publication quality and highly customizable."

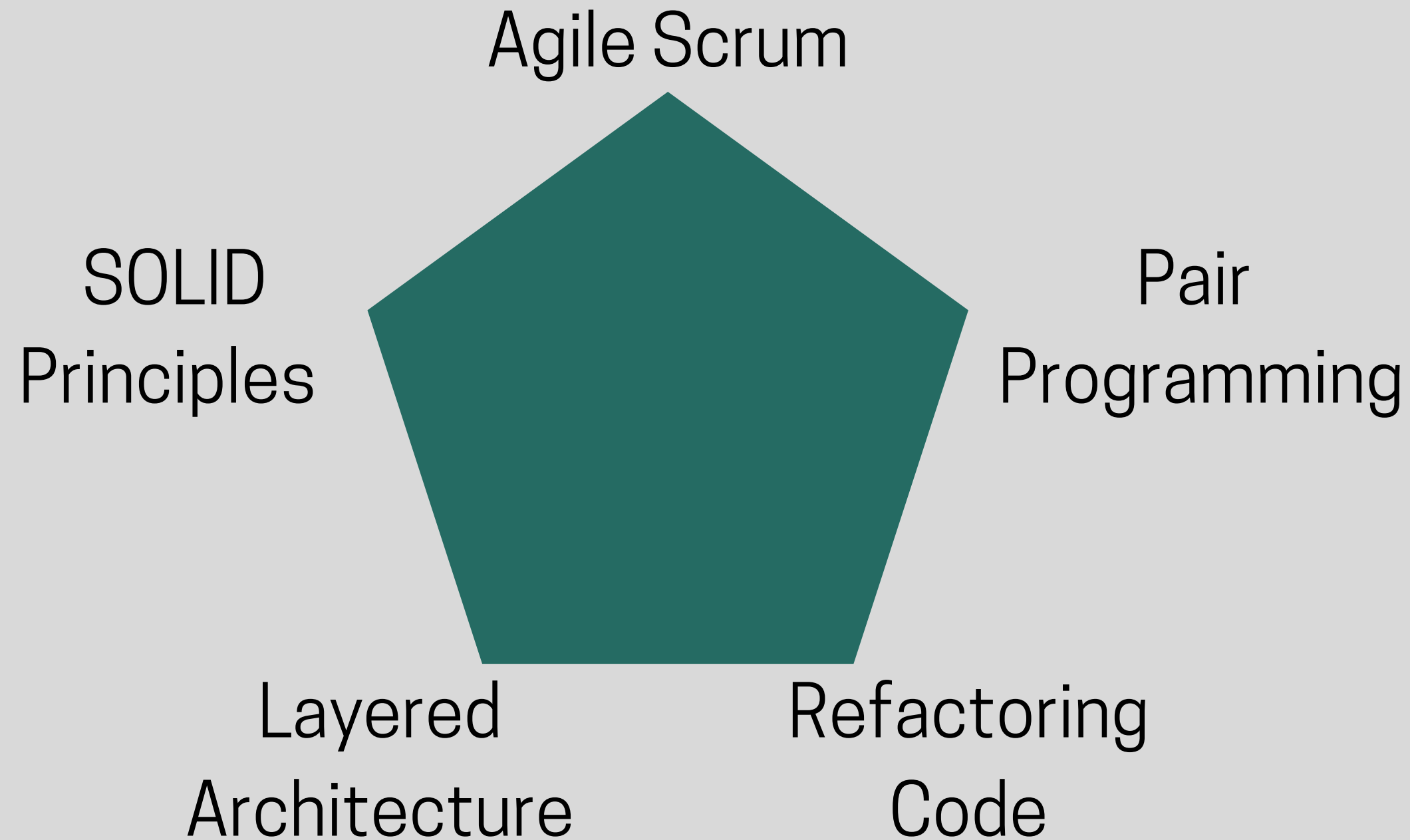
Background



Motivation



Methodologies



Structure:

- Method
- Implementation

Methodologies: Agile Scrum

The Plan:

Weekly Sprints: Main Objective: Harvester

- Plan, Refine, Reflect
- Immediate Feedback from the Team

Methodologies: Agile Scrum

Sprint Planning

Summary	Points	Information
Ignore Solr Core files during Commit	3	Update .gitignore in solr
Clean Up Extra XML Files in Harvester Project	N/A	Remove placeholder ISO XML files
Update Readme File	0.5	Reread and amend README files
Setup External Instance Properties Files for Development	3	External application.properties file

Example: Week 1 Sprint

Methodologies: Agile Scrum

The screenshot shows a Jira Backlog with 25 issues. The issues are listed in a table-like format with columns for issue ID, description, labels, and priority. The labels include 'Automated Infrastru...', 'Search Application', and 'Harvest Metadata'. The priorities are represented by colored icons: red (High), green (Medium), and blue (Low).

Issue ID	Description	Labels	Priority
DSETSP-129	Treat Git Repository as Authoritative Source	Automated Infrastru...	High (5)
DSETSP-81	Search Provides Suggestions	Search Application	Medium
DSETSP-112	Unable to Delete a Record from Solr	Harvest Metadata	Medium
DSETSP-57	Search Index Isn't Updated Fast Enough After a Metadata Update	Automated Infrastru...	High
DSETSP-55	Queue Storage of Dataset Metadata into Solr	Automated Infrastru...	High
DSETSP-79	Search Engine Can Access Schema.org Metadata for Search Record	Search Application	Medium
DSETSP-120	Provide a Site Map for All Search Records	Search Application	Medium
DSETSP-121	Validate ISO Records Contain Minimum Metadata	Search Application	Medium
DSETSP-105	Search on Lab or Organization	Search Application	Medium
DSETSP-122	Harvest Errors are Emailed to WAF Contact	Search Application	Medium
DSETSP-53	Unable to Search or Facet on Resource Type	Search Application	High
DSETSP-110	Non-GCMD Keywords are in our Solr Keywords Field		High
DSETSP-98	Harvester Adds Github Xml Url Location to Solr Record	Harvest Metadata	Medium

Refine Backlog

Know the Problems

- Past
- Present
- Future

Methodologies: Agile Scrum

Reflect Retrospective

- **** Start thinking about presentations?
- **** More hands-on learning (e.g. Docker tutorials)?
- ** Smooth demo. ++1 +1 +1
- * Use docker on local machines instead of vagrant?
- * Do S & T know about U Bob? [Watch naming video.](#)



Happy Face

Stories completed well before the end of sprint +1 +1
Moving harvester forward/good direction +1
Setting up asynchronous and/or cron jobs via spring
Annotation is surprisingly quick and easy+1 +1
** Running ahead w/ backlog review +1 +1



Frowny Face

Lots to learn when using frameworks. +1
Really misspotted scheduling story.



Ideas

Thinking about 2 separate presentations
** More asynchronous stories +1
*** Working on stories for search application +1
*** Jamboard meeting to kickoff presentation/poster ideas. +1



Flowers

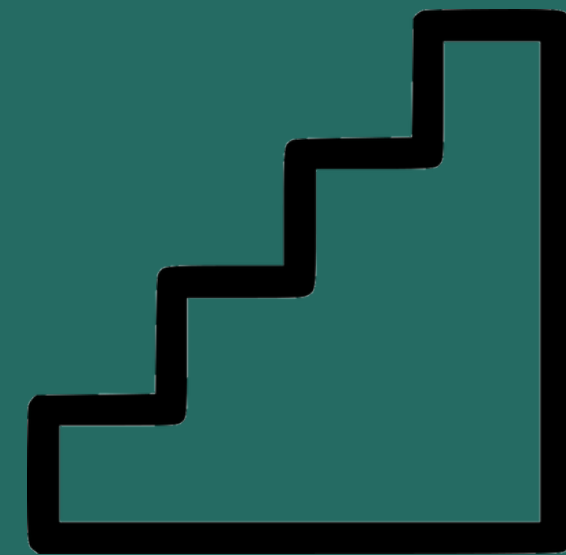
Sama and Terry working together. +1 +1 +1
Sama and Terry looking at stories in the backlog.

Methodologies: Agile Scrum

Focus on the Main Objective: Harvester Automation

Short Sprints = Progress!

- Fast Deadline
- Quick Results
- Small Steps



Methodologies: SOLID Principles

Objective: Make Software More Maintainable

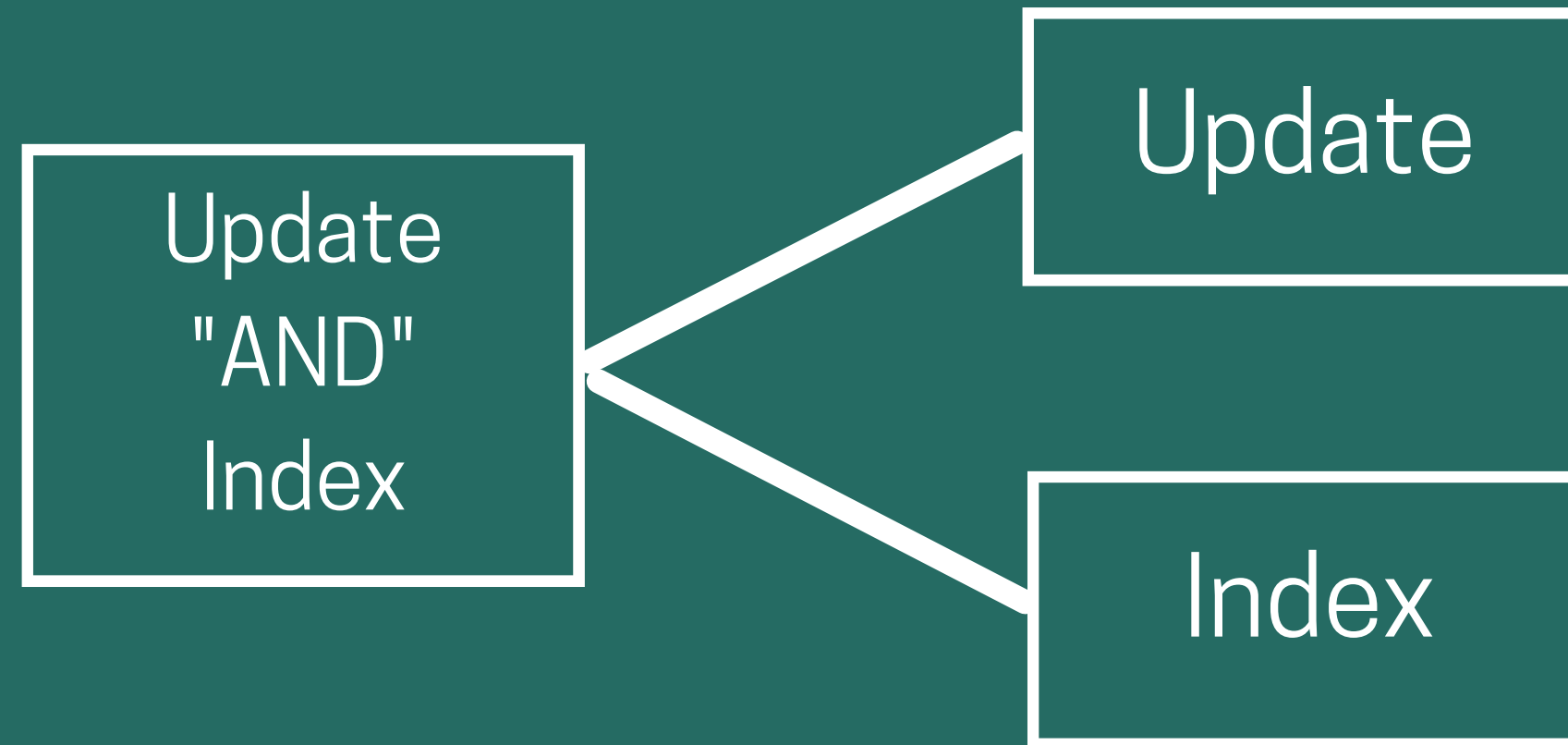
- One object has one purpose
 - Single Responsibility Principle
- Reduce knowledge = Be modular
 - Interface Segregation Principle

Not every principle was followed, and that's okay

Methodologies: SOLID Principles

One object has one purpose

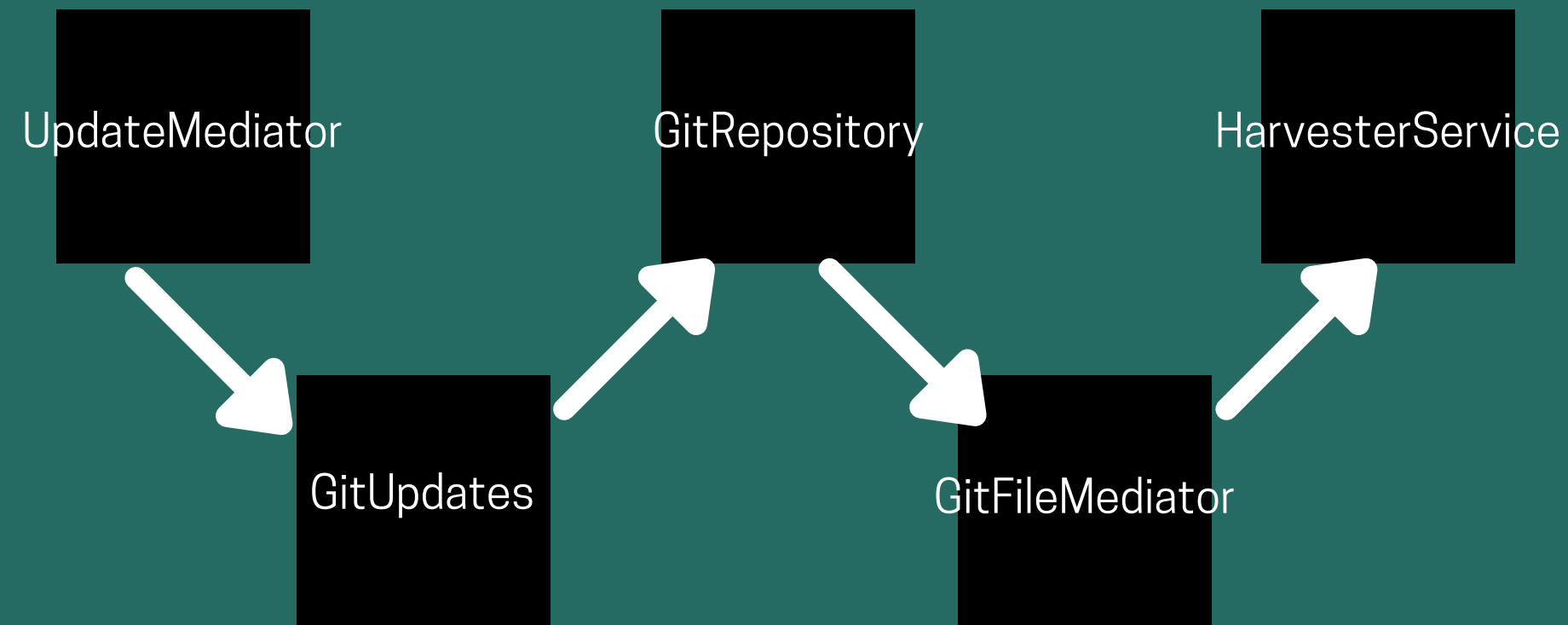
Example: Evolution of a Class's Responsibilities



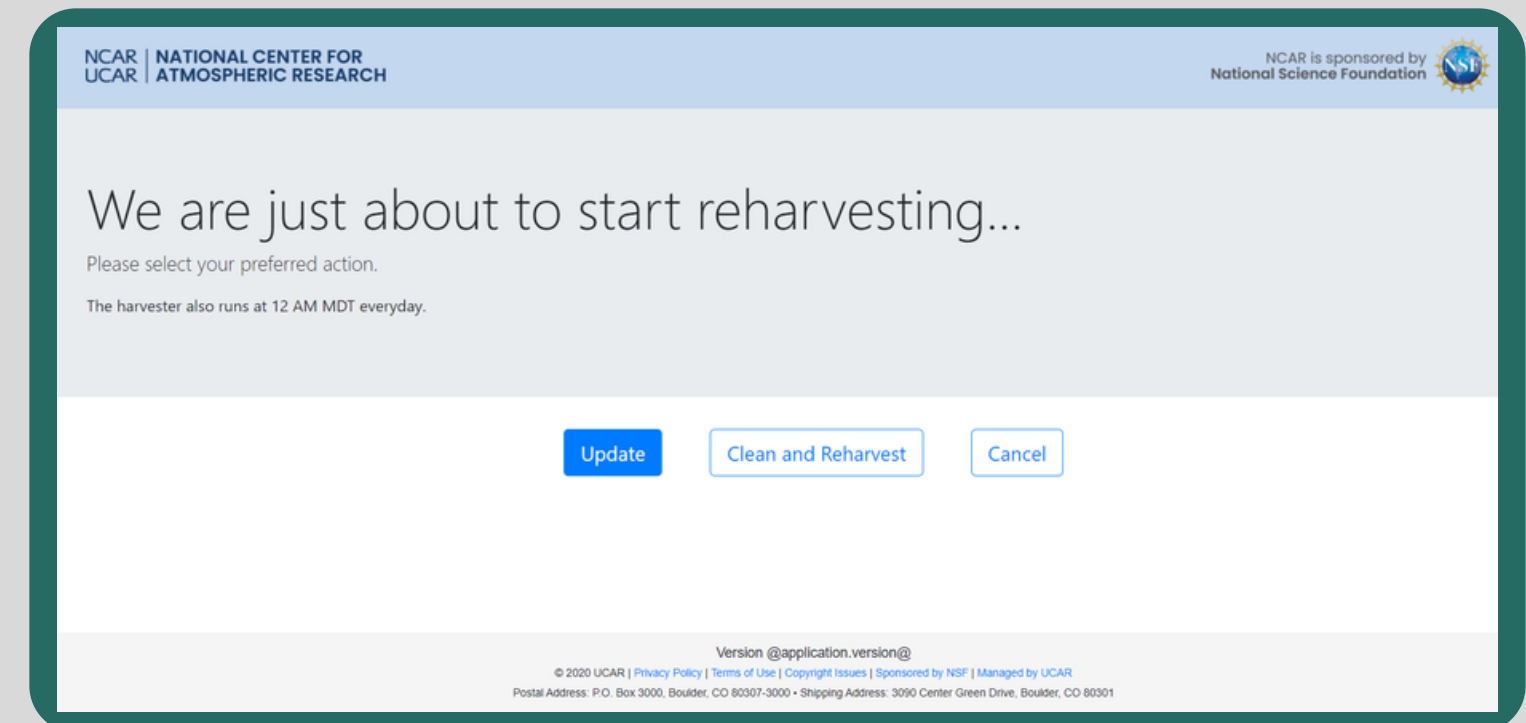
Because if you use "AND",
you know it's too much

Methodologies: SOLID Principles

Reduce knowledge = Be modular
Example: Manual Update from GitHub



What does it take to do one thing?



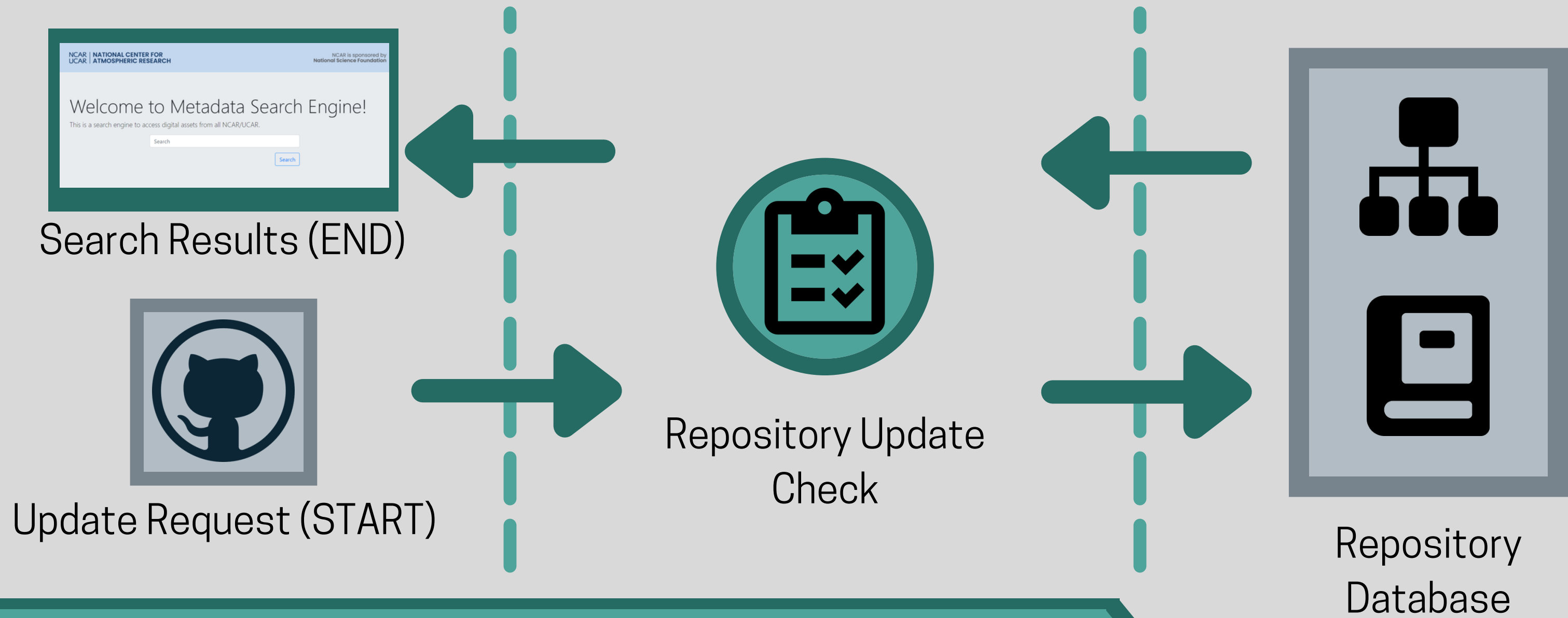
Methodologies: Layered Architecture

Example: Automatic Update from GitHub (+ Restaurant Parallel)

Presentation (Customer)

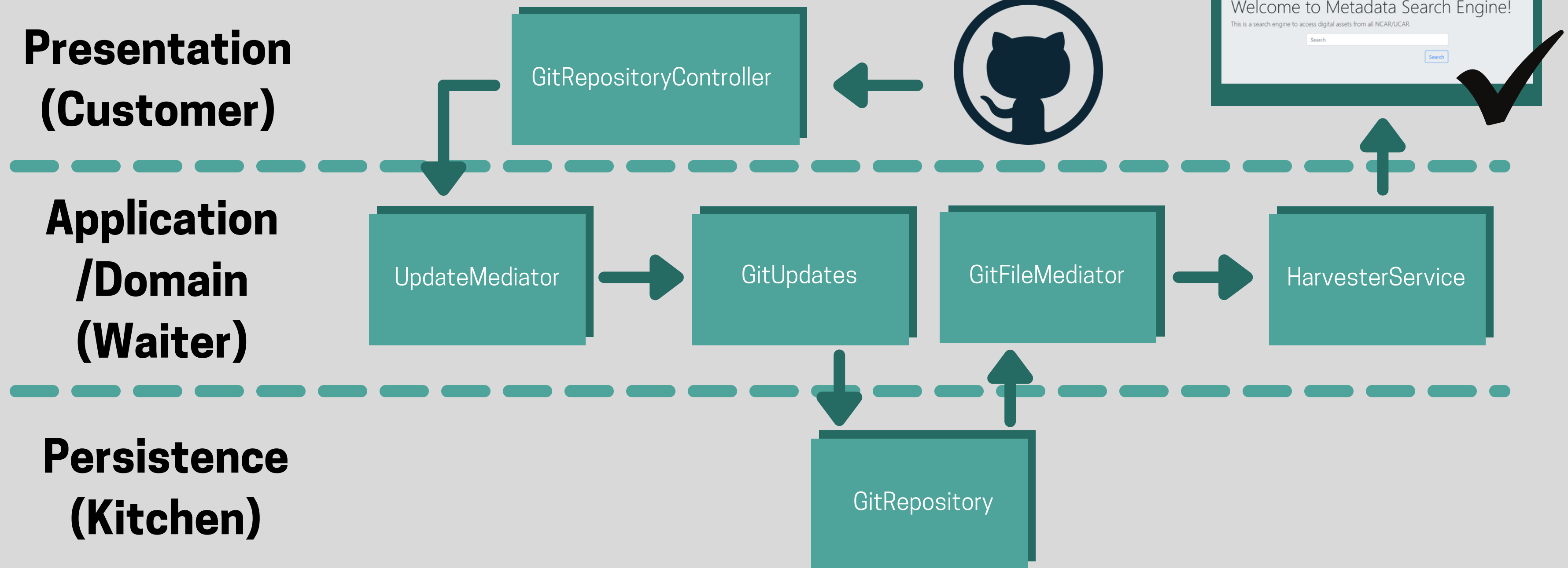
Application/Domain (Waiter)

Persistence (Kitchen)



Methodologies: Layered Architecture

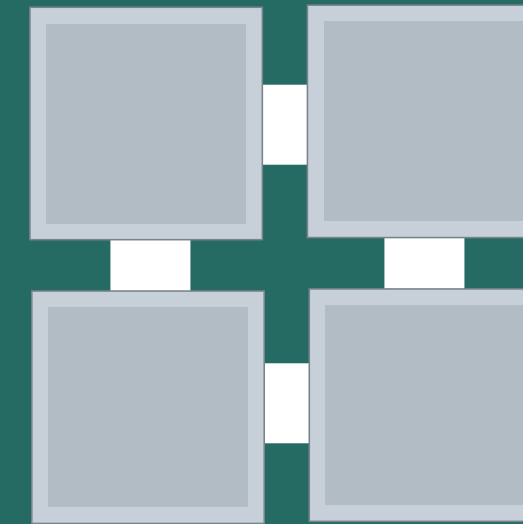
Example: Same but Complicated (+ Restaurant Parallel)



Methodologies: Layered Architecture

Objective: Implementing SOLID "theory"

- One Thing, One Job
- Communicate One Layer At A Time
- I Know My Code



Methodologies: Pair Programming



Double the Brain Cells,
Double the Vision

Literally just teamwork...



Variable Communication

Methodologies: Pair Programming



Interactive
Conversations



Double
Engagement



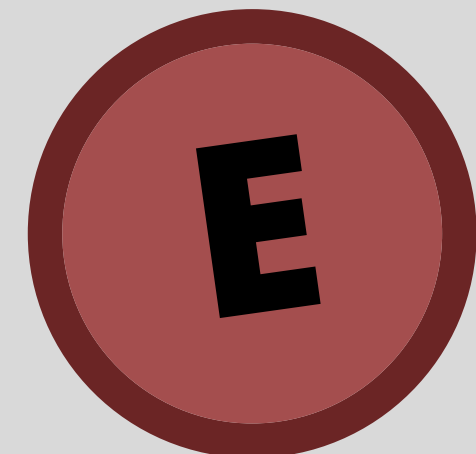
Teamwork



Awkward Silences

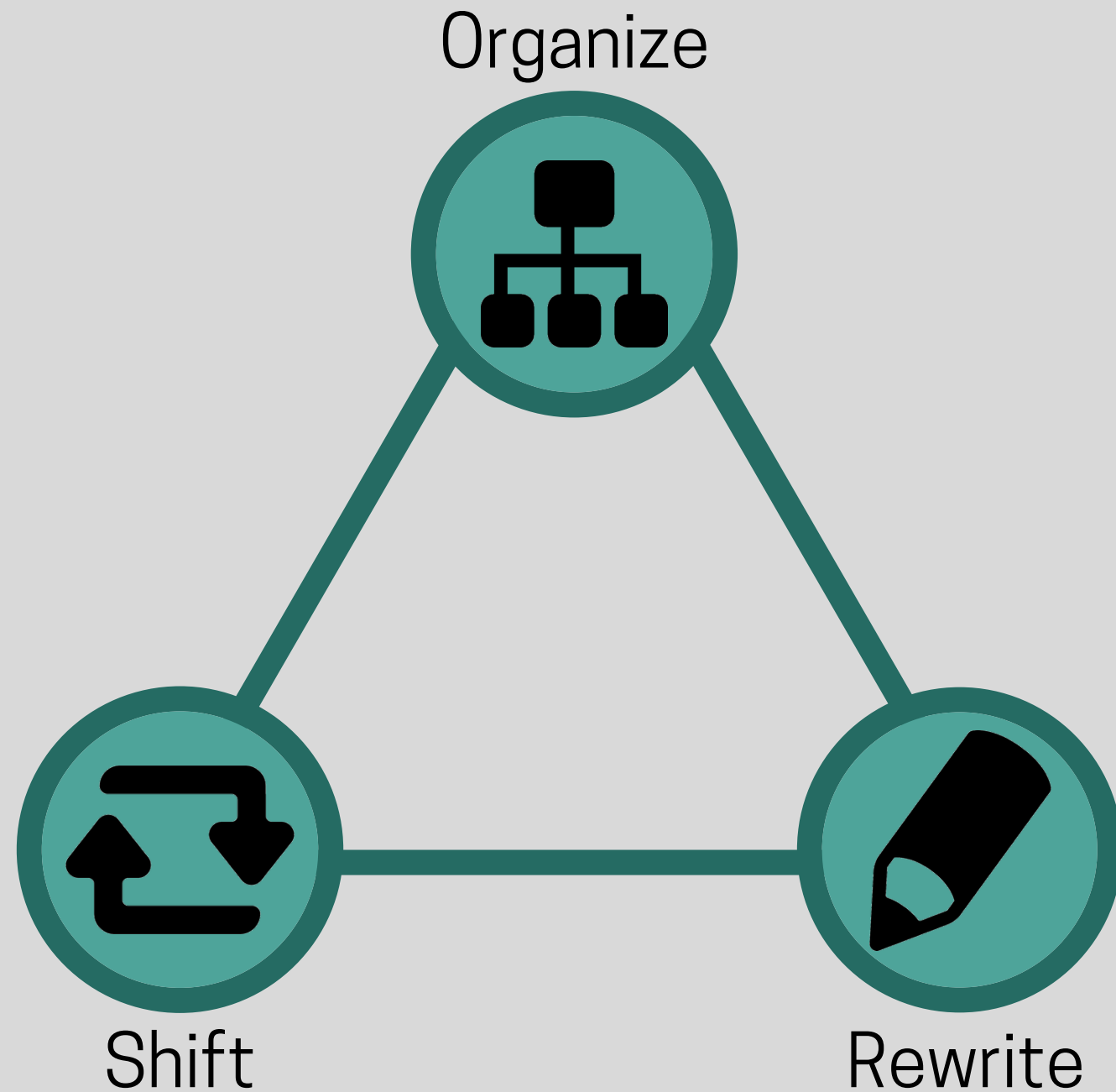


Network Issues
When Virtual



Not Rated E
For Everyone

Methodologies: Refactoring Code



Balance Between:



Readability



Optimization

Methodologies: Refactoring Code



Boost Confidence, Settle
the Bare Necessities

Unit Tests



Remove repeats, Simplify code,
Prioritize Readability

Iterative Cycles

Future Usage

What is your project?

Recommendations:



Modular?



Maintainable?



Team Size > 1?

Thank You For Listening

Thank You To...

SIParCS Mentors:

Nathan Hook,
Saquib Aziz Khan,
Eric Nienhouse,
Christy Grant

SIParCS Program Leads

AJ Lauer,
Virginia Do,
Jerry Cyccone,
Max Cordes Galbraith

...and everyone else for making this program happen

Shift to Technologies

Harvester Automation
for Metadata Search
Web Application

NEXT UP!

Part Two: Technologies
by Sama Manalai, SIParCS Intern

References

- <https://www.atlassian.com/agile/scrum>
- <https://medium.com/mindorks/solid-principles-explained-with-examples-79d1ce114ace>
- <https://hackernoon.com/solid-principles-simple-and-easy-explanation-f57d86c47a7f>
- <https://blog.ndepend.com/layered-architecture-solid-approach/>
- <https://dzone.com/articles/layered-architecture-is-good>
- <http://fewagainstmany.com/blog/introduction-to-layered-architecture-part-one>
- <https://www.agilealliance.org/glossary/pairing/>
- <https://lvivivity.com/what-is-code-refactoring>

Digital Media

- <https://www.hiclipart.com/free-transparent-background-png-clipart-dphnr> - GitHub logo
- <https://webiconspng.com/icon/5129> - Binoculars icon
- <https://thenounproject.com/term/disconnected/26774/> - Disconnected, by Ugur Akdemir
- https://www.flaticon.com/free-icon/thumbs-up-hand-symbol_25423 - Thumbs up icon
- <https://pluspng.com/png-82868.html> - Test/Checklist icon
- <https://www.searchpng.com/2019/02/21/cycle-bicycle-travel-ride-bike-icon-png-image-free-download/> - Bicycle icon
- <https://www.onlinewebfonts.com/> - For all other icons