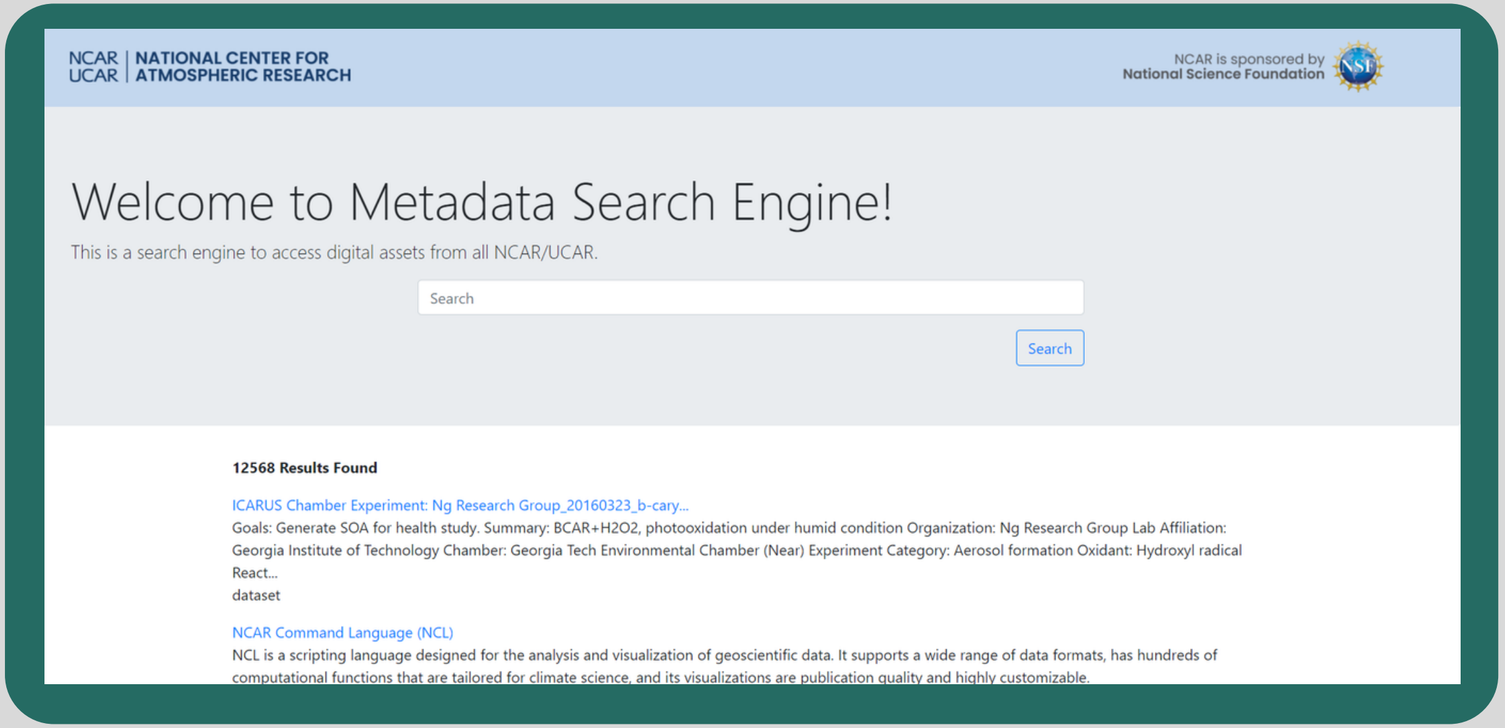
Harvester Automation for Metadata Search Web Application

Part One: Methodologies by Terry Yuan, SIParCS Intern

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

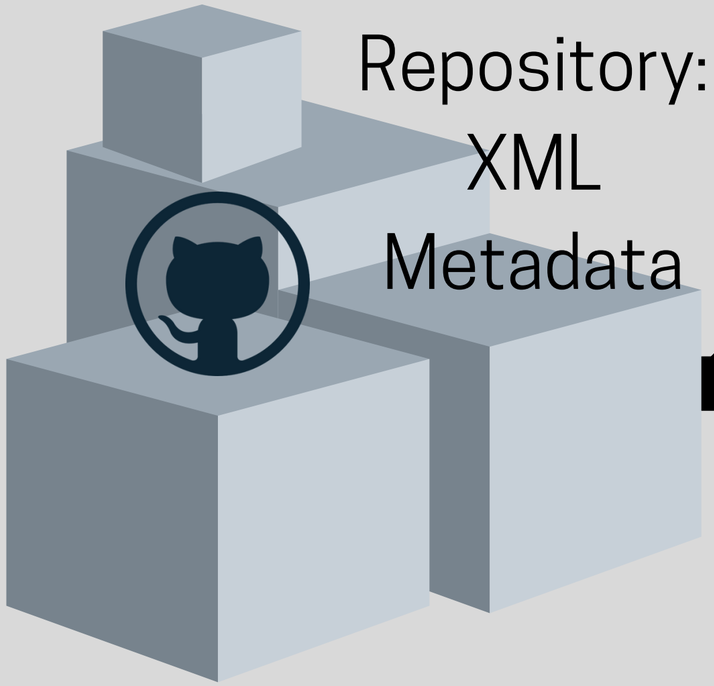
Background

A Search Tool
Built Under NCAR
Made For NCAR

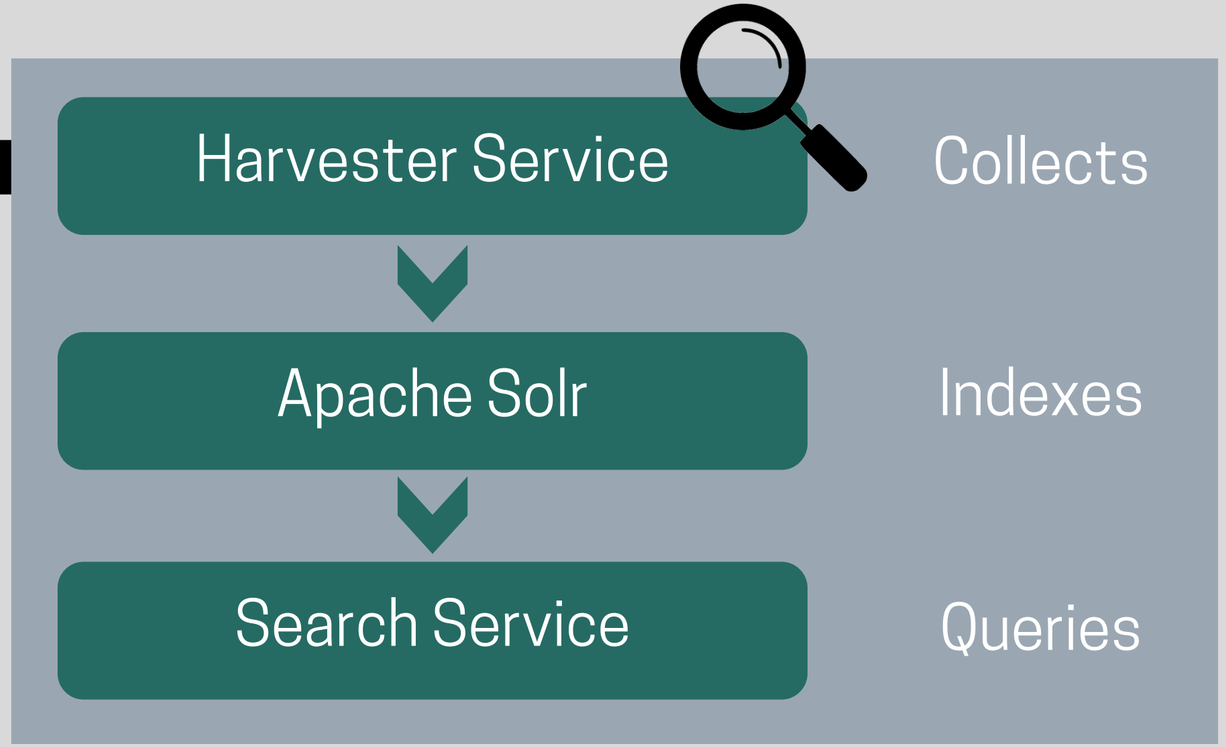


NCAR UCAR
Sama Manalai / **Terry Yuan**
SIParCS 2021: Project 7

Background

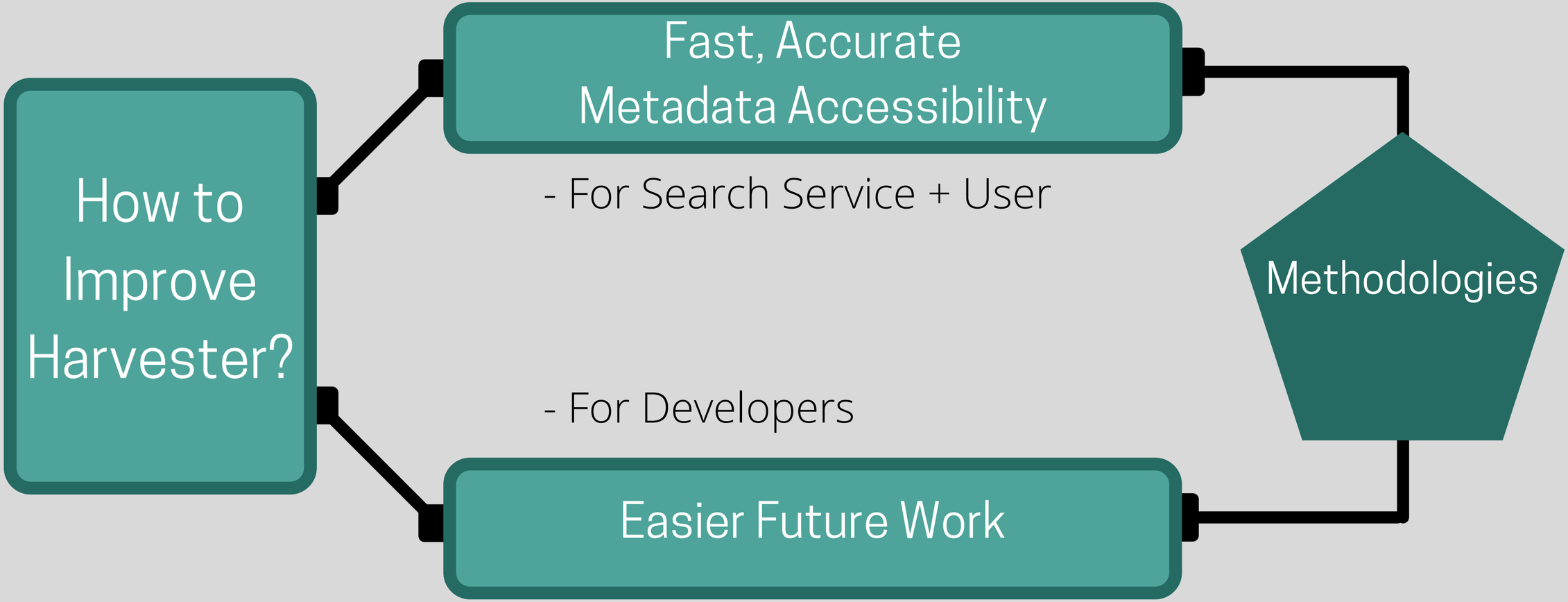


GitHub Repository: XML Metadata

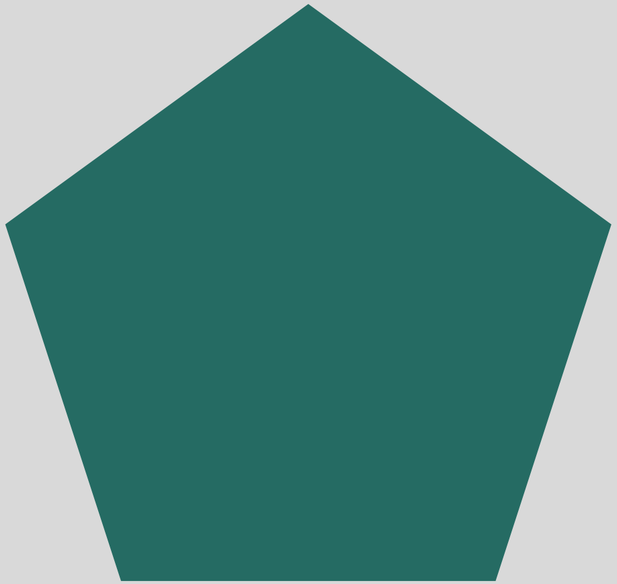


Metadata Search Web Application

Motivation


How to Improve Harvester?
Fast, Accurate Metadata Accessibility
- For Search Service + User
- For Developers
Easier Future Work
Methodologies

Methodologies



Structure:

- Method
- Implementation

Methodologies: Agile Scrum

The Plan:

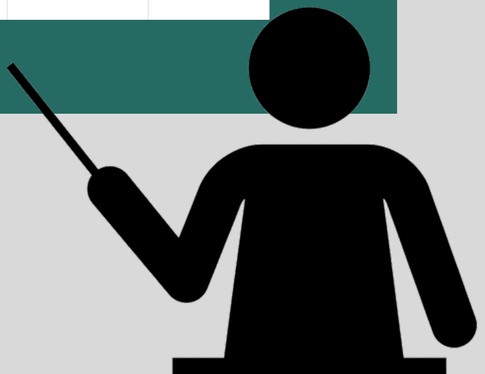
Weekly Sprints: Main Objective: Harvester

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

Methodologies: Agile Scrum

Plan Sprints

Sprint	Actual	Key	Summary	Points	Information
1	19	DSETSP-101	Ignore Solr Core files during Commit	3	Update .gitignore in solr
1	19	DSETSP-103	Clean Up Extra XML Files in Harvester Project	N/A	Remove ISO XML files that were originally placed in harvester for testing
1	19	DSETSP-107	Update Readme File	0.5	Reread and amend README files for harvester, search, solr
1	19	DSETSP-108	Setup External Instance Properties Files for Development	3	External application.properties file
2	20	DSETSP-113	Read YAML File Via Program Argument and Print Out Contents Via System.out.println()	5	Build a YAML file that contains the information on repositories
2	20	DSETSP-114	Git Clone Repositories from YAML File	3	Use JGit to clone repositories
2	20	DSETSP-115	Git Pull Updates from Github on Startup	3	Use JGit to perform pull updates
3	21	DSETSP-102	Harvester re-runs constantly in Sagedockerdev VM	3	Restart the harvest process via a webpage so that harvester can run again without full app reset
3	21	DSETSP-116	Unable to Access Harvester Website on sagedockerdev.ucar.edu:8081	5	Docker deployment for harvester and give it a webpage on port 8081
3	21	DSETSP-118	Refactor the Git Clone/Pull and Harvest Code into a Service	3	HarvesterService and ReharvestEvent
4	22	DSETSP-117	Create Web Page/Controller That Can Trigger a Re-Harvest	5	Build confirmation and interactable webpages + controller for the reharvest process
4	22	DSETSP-123	Automating Reharvest via Polling	8	Cronjob for reharvest
5	23	DSETSP-78	ISO XML Metadata Can Be Viewed	5	Debug mode on search query enables github file link, modifications made to all 3 services
5	23	DSETSP-109	Keywords in Solr Index Contain XML Encoded Greater Than Symbols (>)	3	Fix the XML-encoding issue by toying with Jsoup + parsing
5	23	DSETSP-124	Refactor GitRepository Class to Handle cloneOrPull Behavior and Fire Events	5	Event firing after clone/pull
6	24	DSETSP-119	Unable to Easily See All Search Results	3	See all indexed results in search when you don't type anything - new query strategy
6	24	DSETSP-126	Refactor HavesterService to No Longer Directly Call the GitRepository Classes	5	Refactor HarvesterService, now it only manages indexing XML files.
6	24	DSETSP-127	Determine How Github Webhooks Work and Build Testing Infrastructure	3	Webhook research, able to replicate webhook interaction on localhost
7	25	DSETSP-56	Automagically Pull New Updates from Github	5	Github Webhook full implementation, only using accessible test repositories
7	25	DSETSP-128	Clean and Full ReHarvest all Records	5	New button that clears Solr indexed documents and reharvests
8	26	DSETSP-130	Code Cleanup	3	Refactoring, renaming, reformatting, removing comments, housekeeping
8	26	DSETSP-131	Add Unit Tests	3	Unit tests added for three mediators



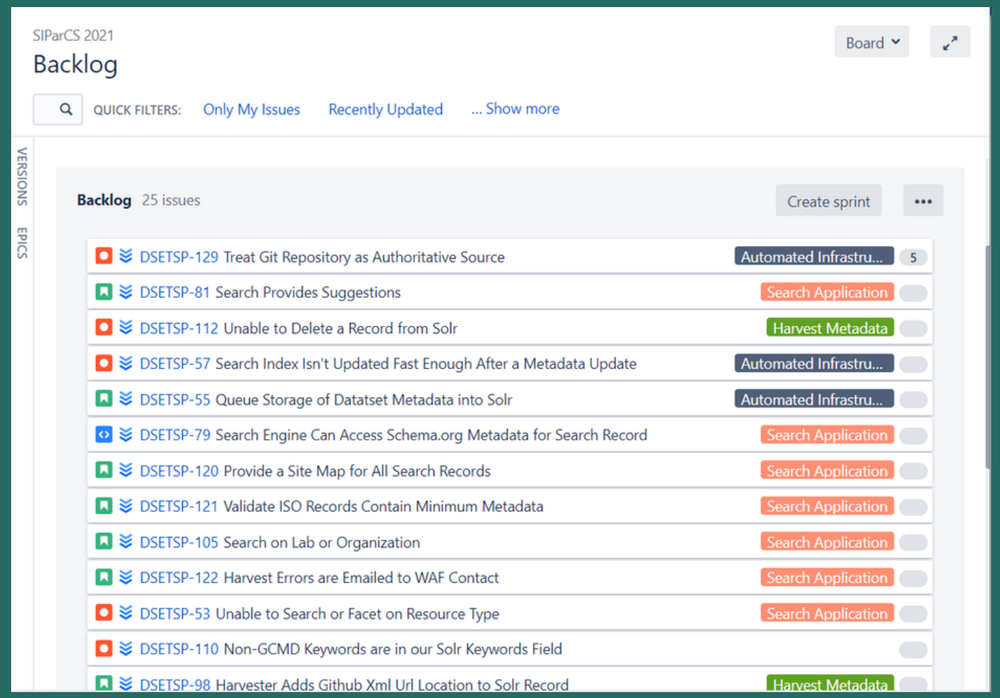
NCAR UCAR

Sama Manalai / **Terry Yuan**

SIParCS 2021: Project 7

Methodologies: Agile Scrum

Refine Backlog



Reflect Retrospective

Happy Face

Frowny Face

Ideas

Flowers

NCAR UCAR

Sama Manalai / **Terry Yuan**

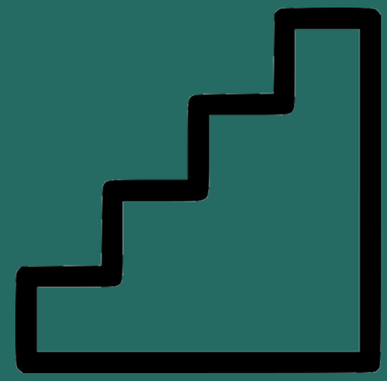
SIParCS 2021: Project 7

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

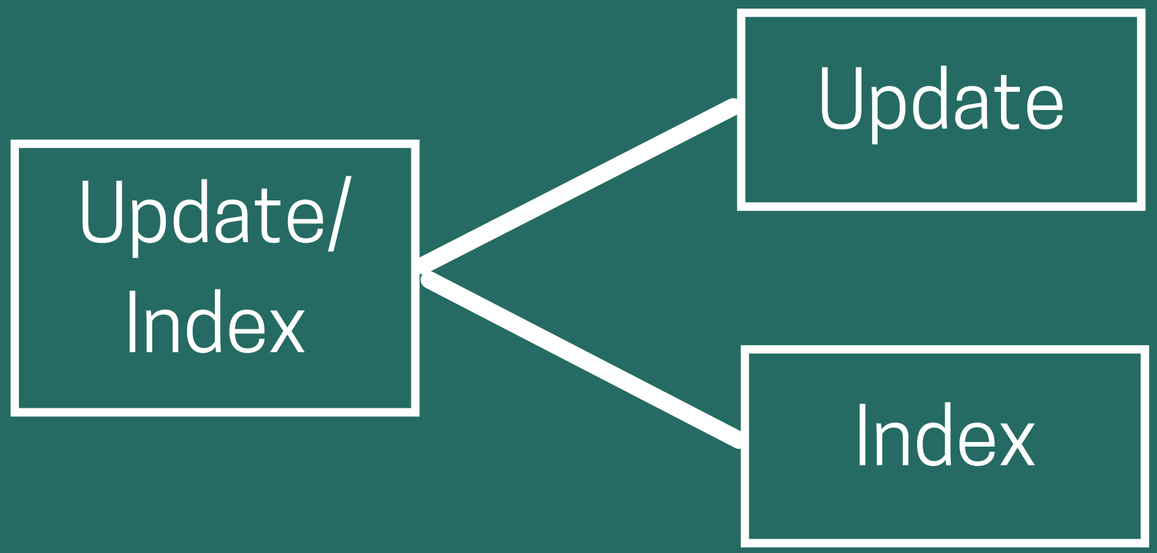
- One object has one purpose
- Reduce knowledge = Be modular

Not every principle was followed, and that's okay

Methodologies: SOLID Principles

One object has one purpose

Example: Evolution of a Class's Reponsibilities

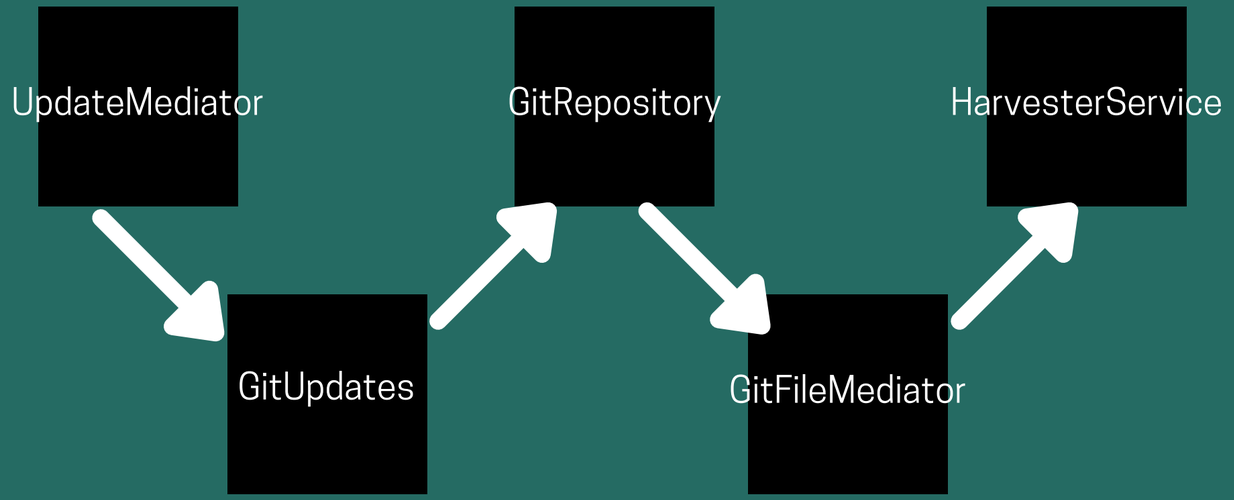


Because if you use "AND", you know it's too much

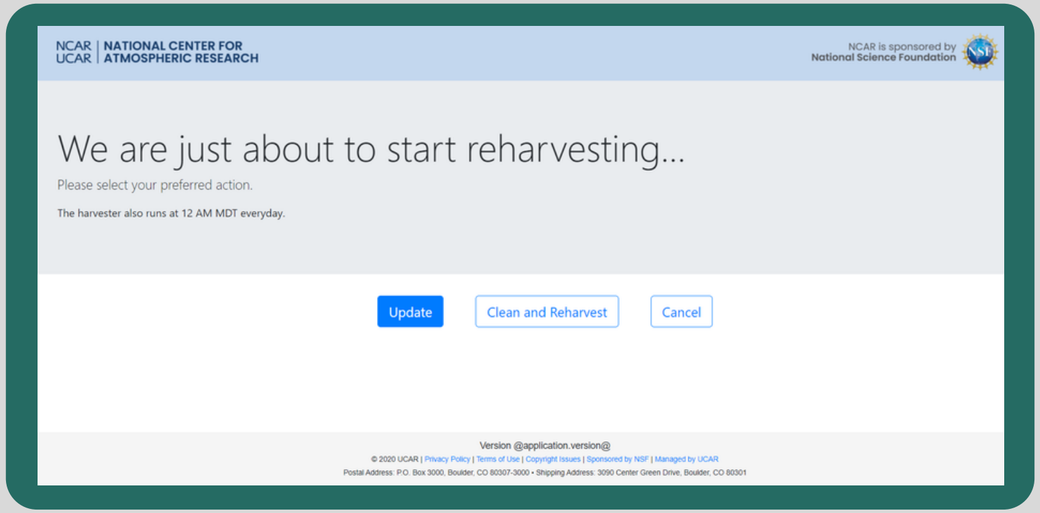
Methodologies: SOLID Principles

Reduce knowledge = Be modular

Example: Manual Update from GitHub



What does it take to do one thing?



NCAR UCAR

Sama Manalai / **Terry Yuan**

SIParCS 2021: Project 7

Methodologies: Layered Architecture

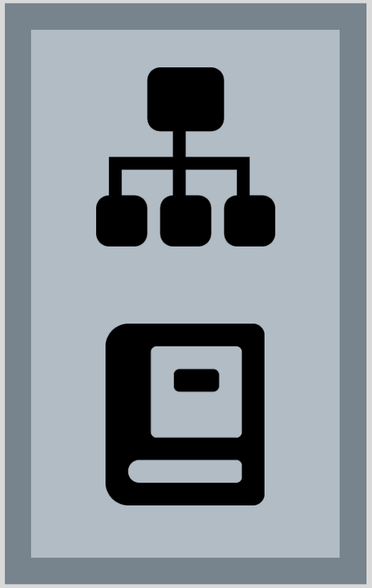
Example: Automatic Update from GitHub (+ Restaurant Parallel)

Presentation (Customer) **Application/Domain (Waiter)** **Persistence (Kitchen)**

Search Results (END)

Update Request (START)

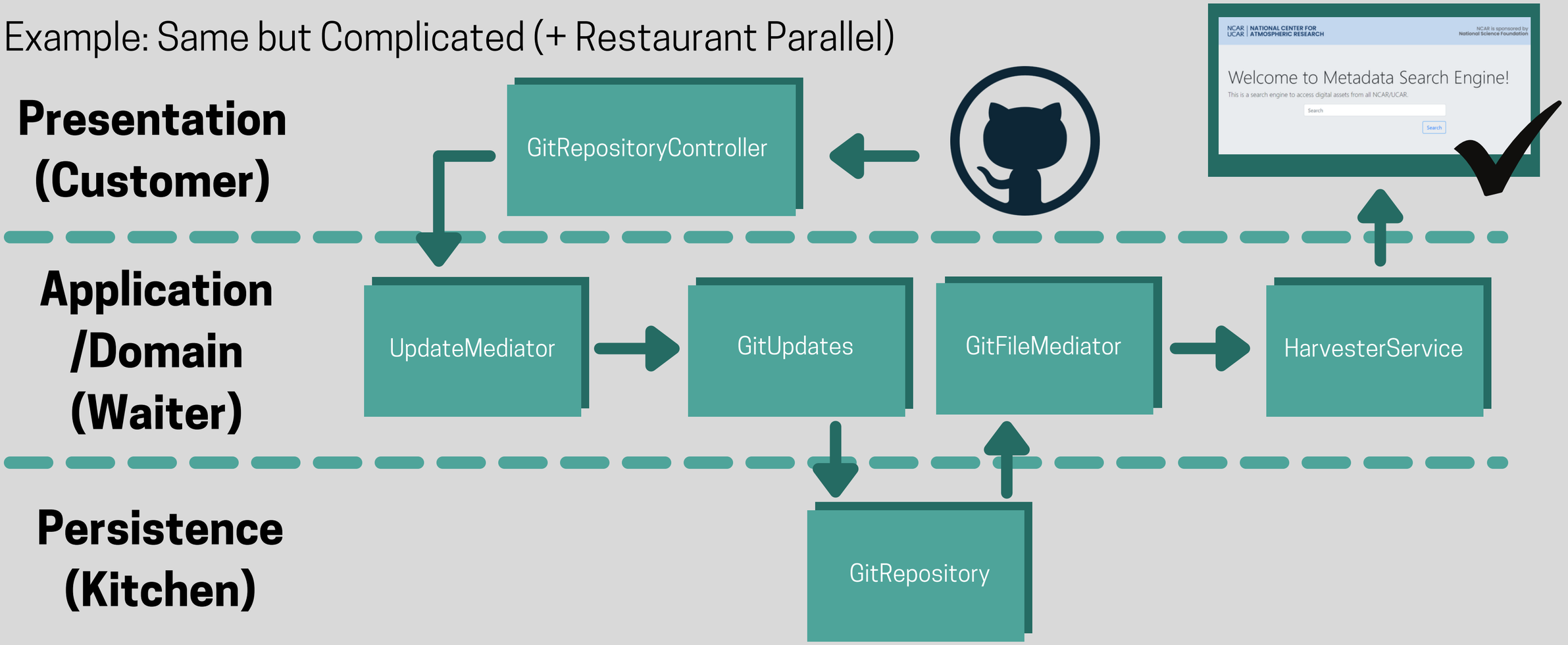
Repository Update Check



Repository Database

Methodologies: Layered Architecture

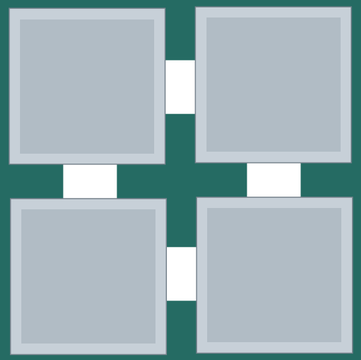
Example: Same but Complicated (+ Restaurant Parallel)



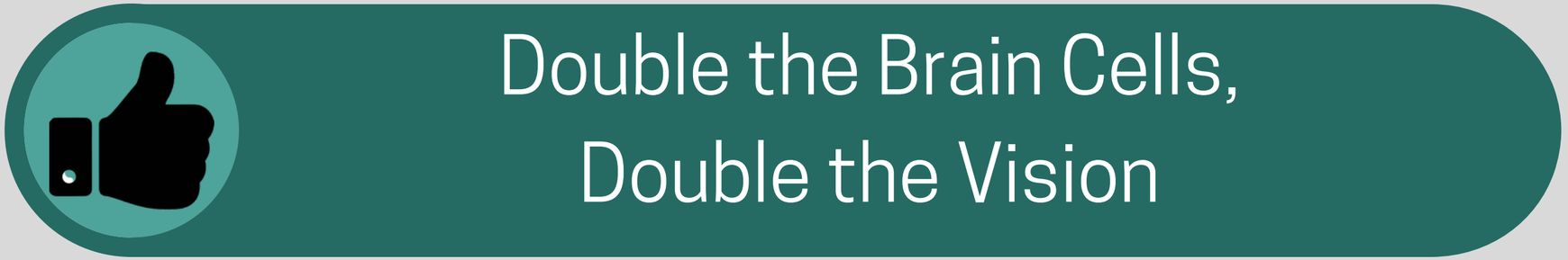
Methodologies: Layered Architecture

Objective: Implementing SOLID "theory"
(Complicated? Yes!)

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



Methodologies: Pair Programming



Literally just teamwork...

Methodologies: Pair Programming

Interactive Conversations

Double Engagement

Teamwork

Awkward Silences

Network Issues When Virtual

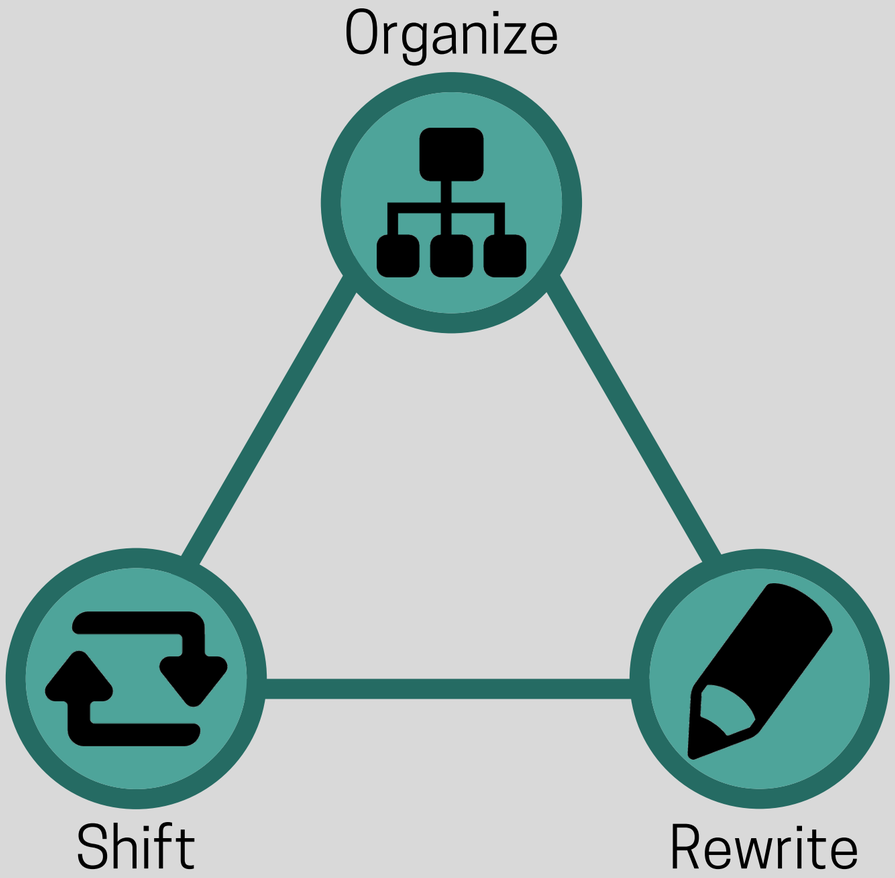
Not Rated E For Everyone

NCAR UCAR

Sama Manalai / **Terry Yuan**

SIParCS 2021: Project 7

Methodologies: Refactoring Code

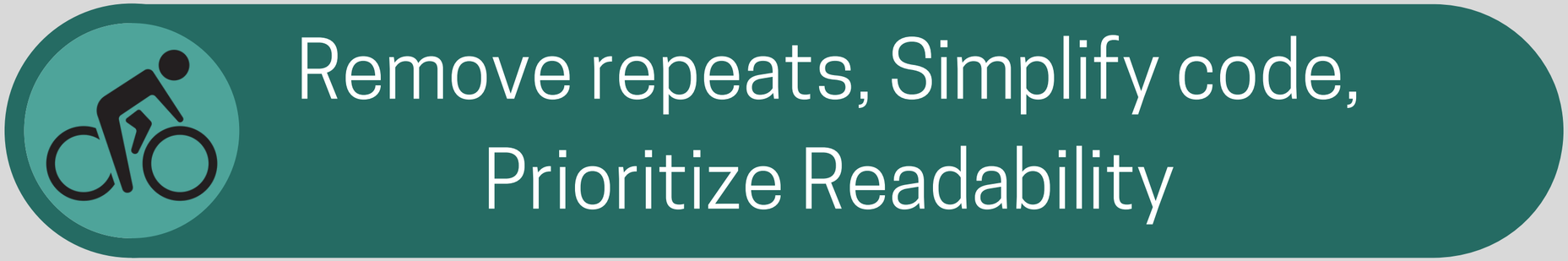


Balance Between:

Readability Optimization

Methodologies: Refactoring Code

Unit Tests



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,
Christy Grant,
Eric Nienhouse

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://lvivity.com/what-is-code-refactoring

Digital Media

- https://www.hiclipart.com/free-transparent-background-png-clipart-dphnr - GitHub logo
- https://www.nicepng.com/ourpic/u2q8i1q8a9w7w7o0_teacher-teacher-icon-png/ - Teacher icon
- 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