

EDCT GE2550: DATA SCIENCE IN EDUCATION

Big Data, Learning Analytics & The Information Age

2/5/16 2:57 PM

Today

Data Systems 10min 6:45 - 6:55

Quiz 10min 6:55 - 7:05

Data Acquisition 10min 7:05 - 7:15

Consent Policy 10min 7:15 - 7:25

Setup?/Zotero 10min 7:25 - 7:35

Git & Github 5min 7:35 - 7:40

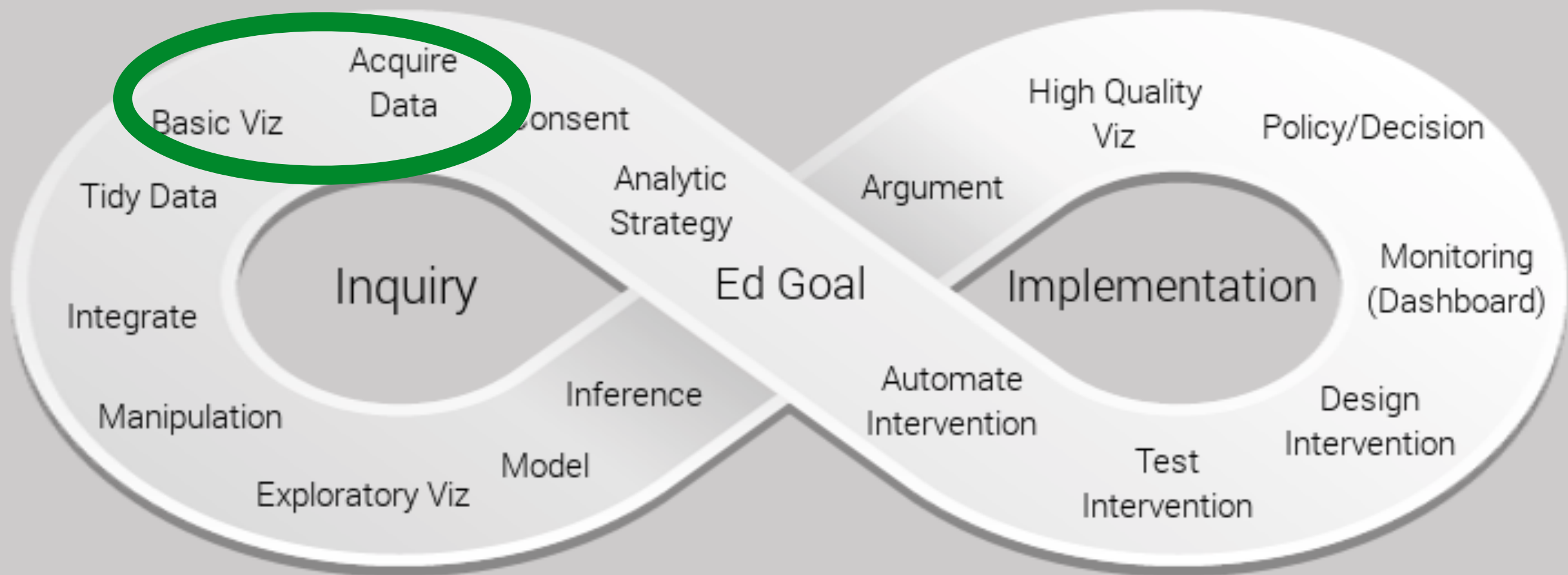
Open Syllabus 5min 7:40 - 7:45

RStudio Basics/Markdown 5min 7:45 - 7:50

Cloning & Plotting Twitter 30min 7:50 - 8:20

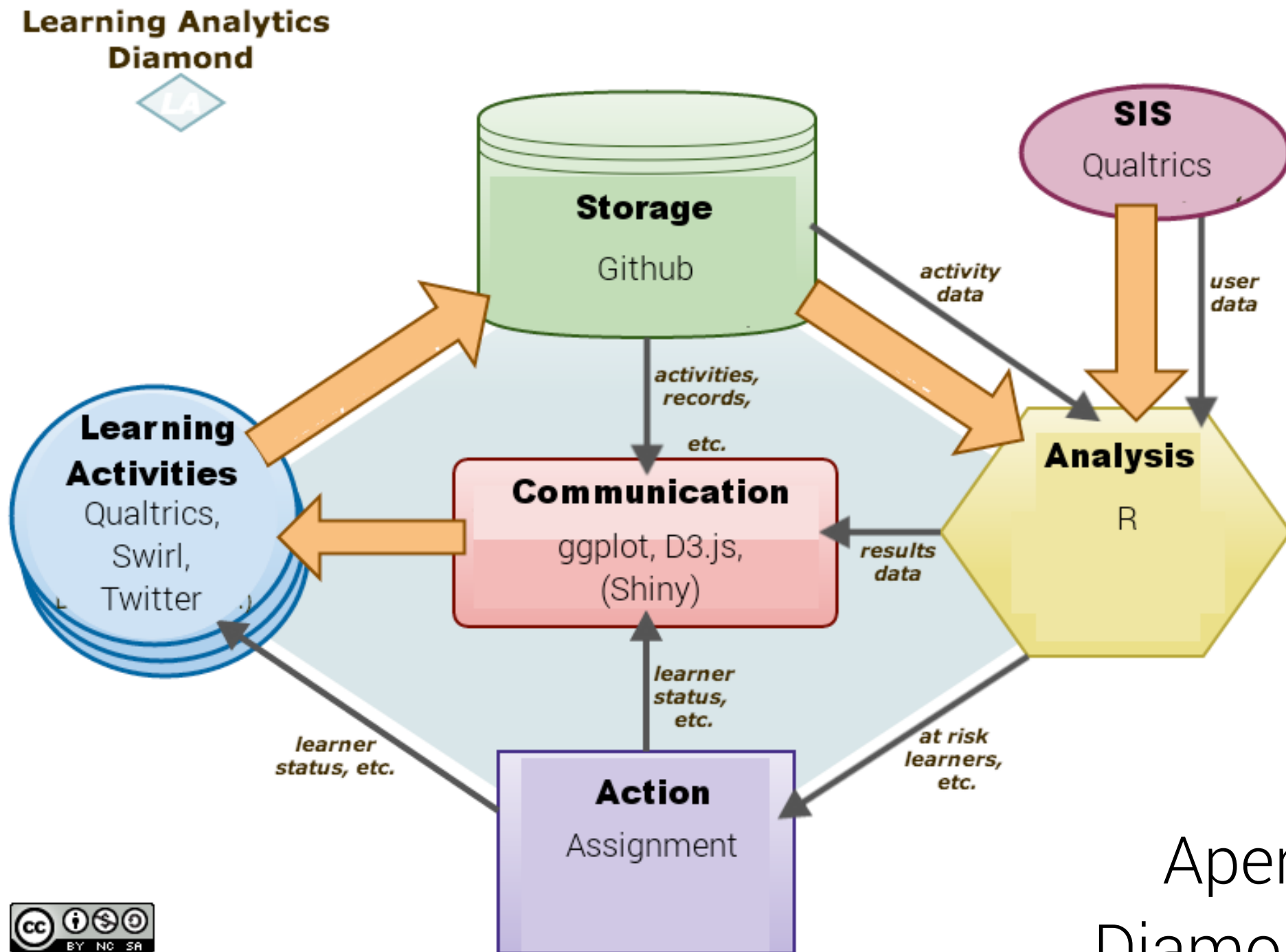
Document in Zotero 5min 8:20 - 8:25

Ed Data Science Cycle

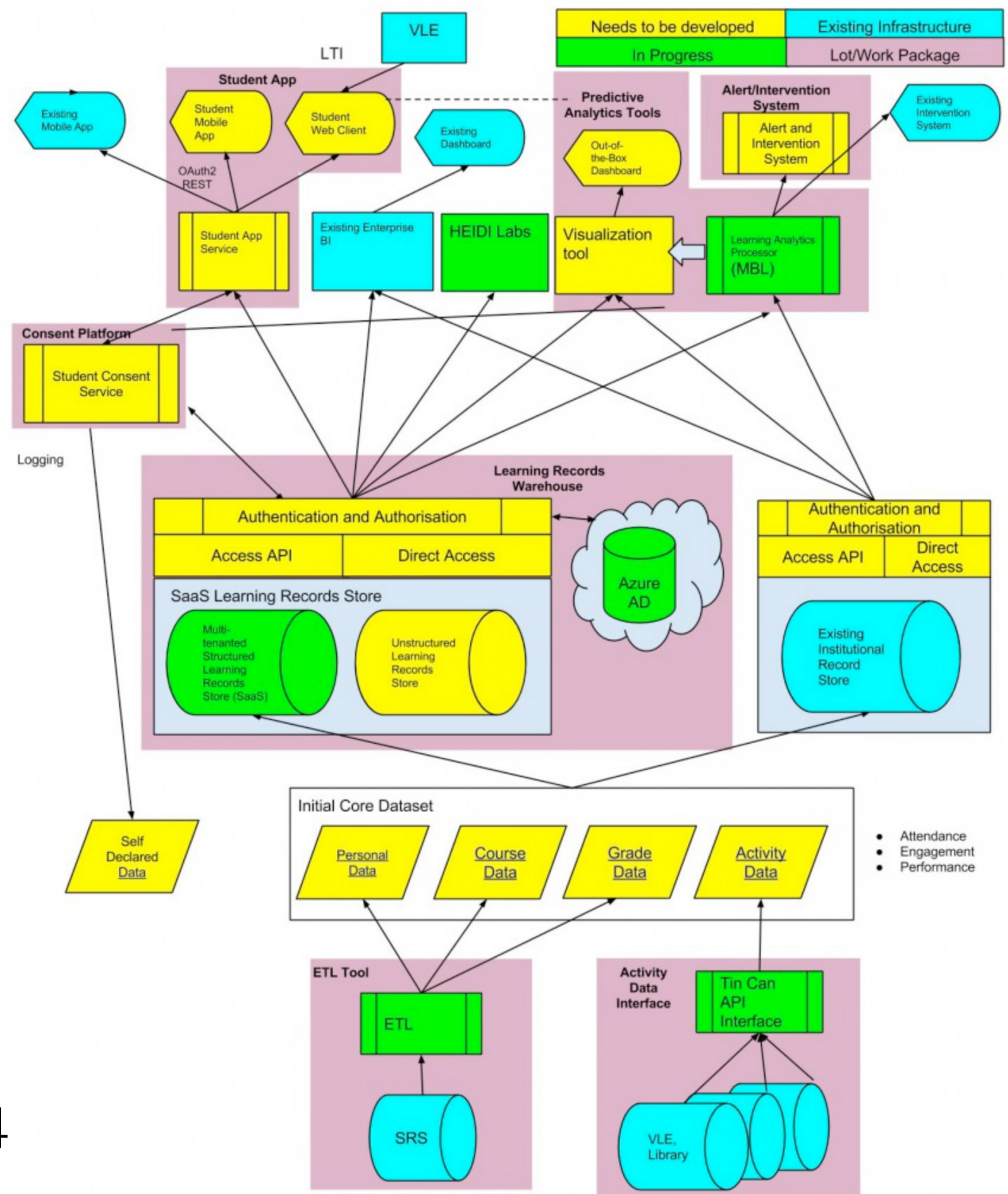


WARNING

Where does data come from?



JISC Learning Architecture, 2014



Standardizing Data Communication

- Limited vocabulary for describing learning
- So that LMS can communicate with tools with SIS, etc
- Scorm = limit information
- Tin Can = limit syntax
- *All this comes from the US Defense Dept?*



I call it **Tin Can** | **xAPI**.

In Class Survey

<http://bit.ly/1SvQ7r7>

What is BIG Data?

- It is relative
- Process vs Study
- Depends on the domain of study: ed (MB-GB) vs ed tech (TB) vs astrophysics (PB) vs business (EB)

Common File Formats

DBF:

- Database format
- Microsoft Access, some freeware
- Table

Common File Formats

XML:

- Extensible Markup Language
- Export web page data
- Hierarchy like HTML with tags to delimit

<row>

<Year>2016</Year>

<Course>EDCTGE2550</Course>

<Price>Priceless</Price>

</row>

Common File Formats

JSON:

- JavaScript Object Notation
- Similar to XML, most common server-browser format

```
[  
  [10, 12, 15, 100],  
  [100, 200, 150, 500],  
  [9, 8, 8, 7],  
];
```

Common File Formats

Fixed Width:

- Create a grid with text using spaces

```
Year.....Course.....Price.....  
2016.....EDCTGE2550....Priceless.....
```

Common File Formats

CSV (TSV):

- Comma Separated Value (Tab Separated Value)
- Most common data format
- Lightweight, easy to interpret - but you can run into trouble with text

```
Year,Course,Price  
2016,EDCTGE2550,Priceless
```


Yeah, but where do WE get data?

Open Data Sets

- *Government*: NYC (<https://nycopendata.socrata.com/>), Whitehouse (<https://open.whitehouse.gov/>), UK (<https://data.gov.uk/>)
- *Research Labs*: ASSISTments (<https://sites.google.com/site/assistmentsdata/>), PSLC DataShop (<https://pslcdatashop.web.cmu.edu/>), Learnsphere (learnsphere.org)
- *Private release*: Harvard/MIT MOOC Data (<https://dataverse.harvard.edu/dataverse/mxhx>)

Yeah, but where do WE get data?

Cut a Deal

- Happens often but will lose autonomy/\$\$\$/control of results

Automated acquisition

- APIs, web scraping, beacons

Generate

- Make your own!

So what variables do we have?

- *Anonymous*: Current degree
- *Anonymous*: Educational background in computer science & statistics
- *Anonymous*: Concerns about the course

Is the “All for One” policy the
right policy?

The GitHub Octocat logo is a stylized blue cat-like creature with a large head, small body, and a long, curved tail. It is centered in the background of the slide. The text "Git & Github" is overlaid on the logo in a light gray, sans-serif font.

Git & Github

2/5/16 2:57 PM

THIS IS GIT. IT TRACKS COLLABORATIVE WORK
ON PROJECTS THROUGH A BEAUTIFUL
DISTRIBUTED GRAPH THEORY TREE MODEL.

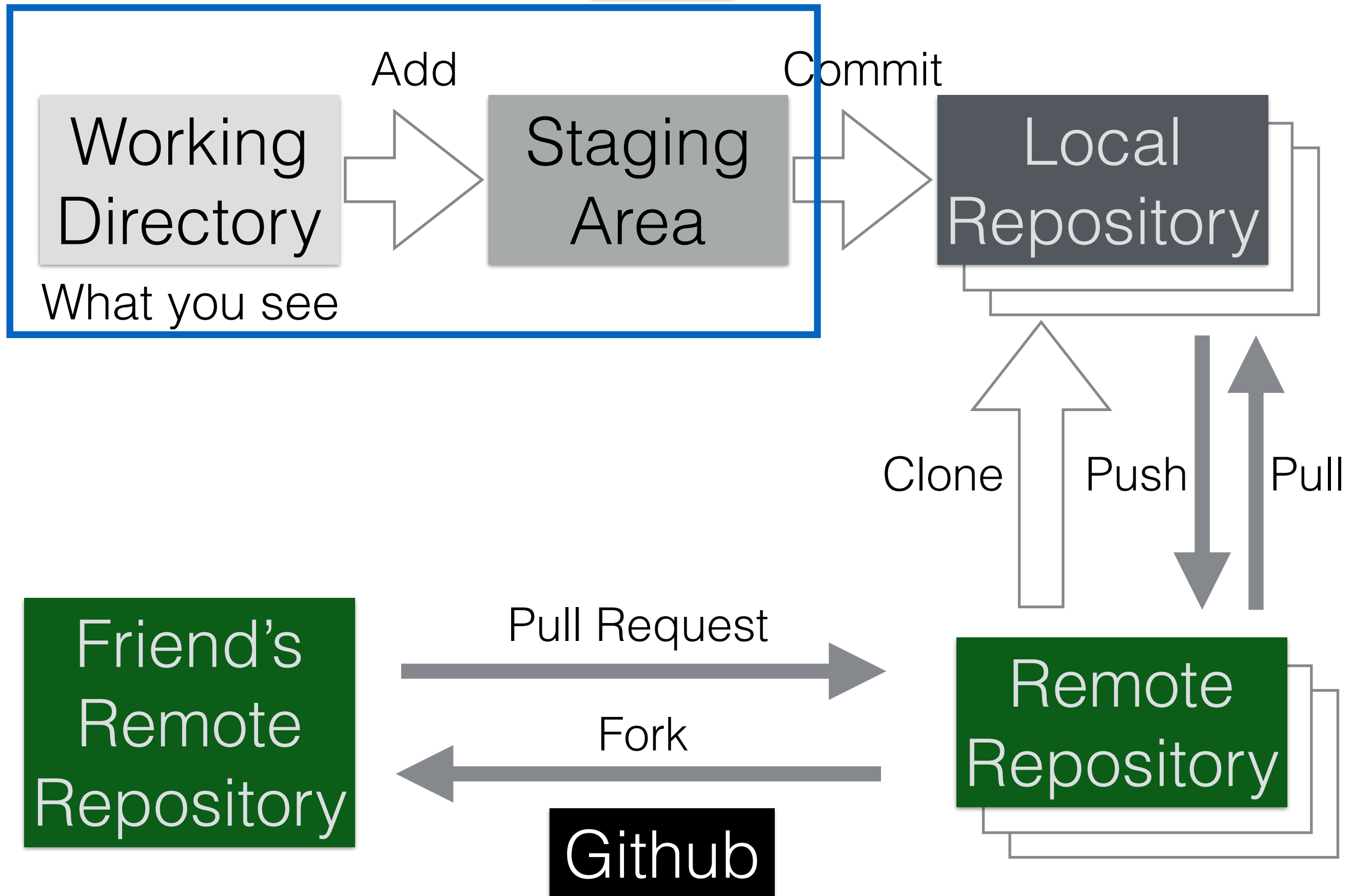
COOL. HOW DO WE USE IT?

NO IDEA. JUST MEMORIZE THESE SHELL
COMMANDS AND TYPE THEM TO SYNC UP.
IF YOU GET ERRORS, SAVE YOUR WORK
ELSEWHERE, DELETE THE PROJECT,
AND DOWNLOAD A FRESH COPY.



RStudio


Git



Functions

- Open syllabus
- RStudio integration

Github Forking

- Log into Github
- “Fork” data-science-in-ed/Syllabus page to your personal account
- Click on the “README.md” file in your clones version
- Find something you would like to change
- Click  and change it!
- Scroll down and click “Commit Change”
- Click on the “Pull Requests” tab and then “New pull request”
- Describe the change you made in the message and click “Create Pull Request”