HUDK 4050:
CORE METHODS
IN EDM

9/14/17 9:36 AM

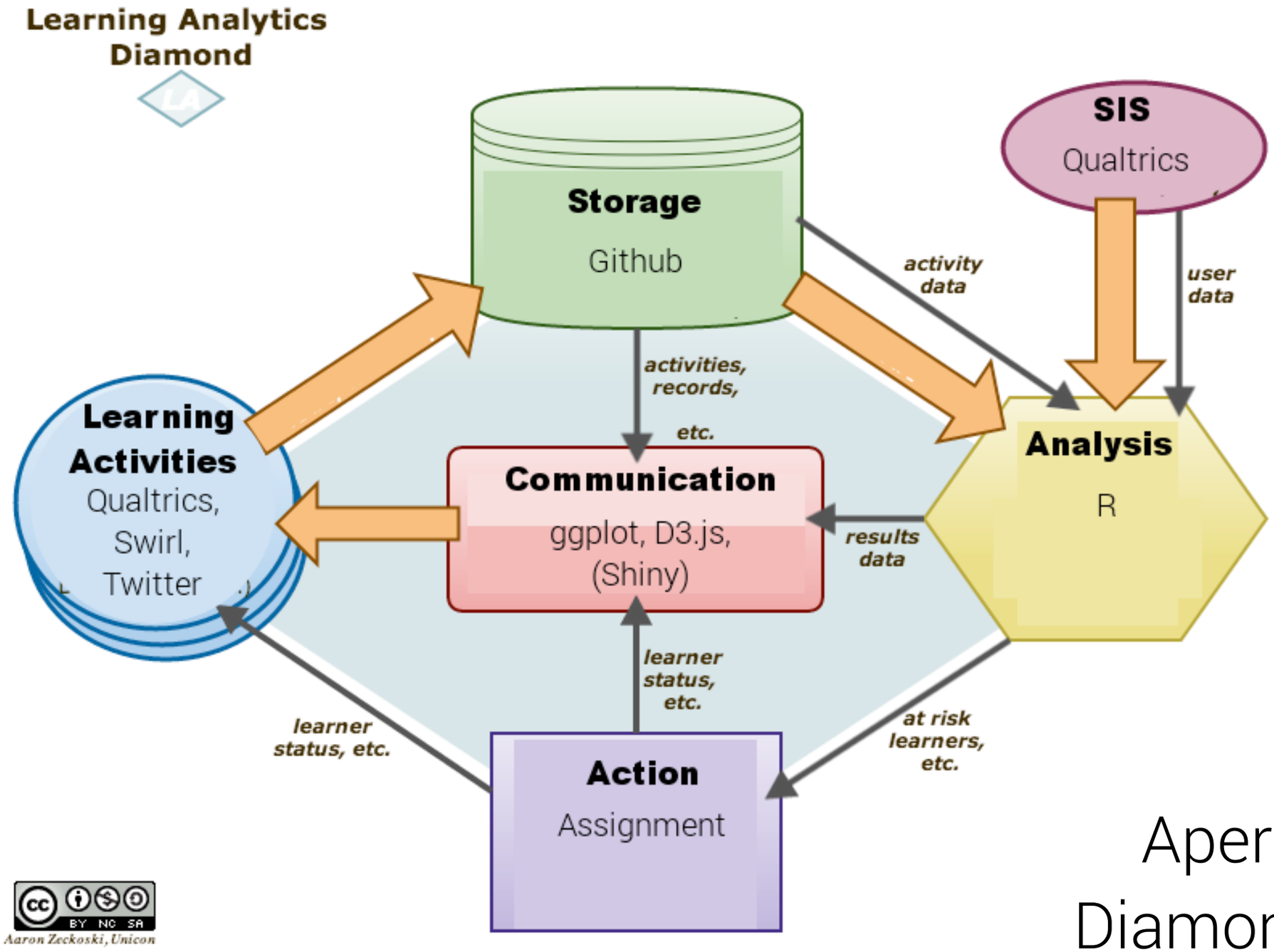
Today

- Data sources
- Zotero
- Assignment 1
- Complete survey

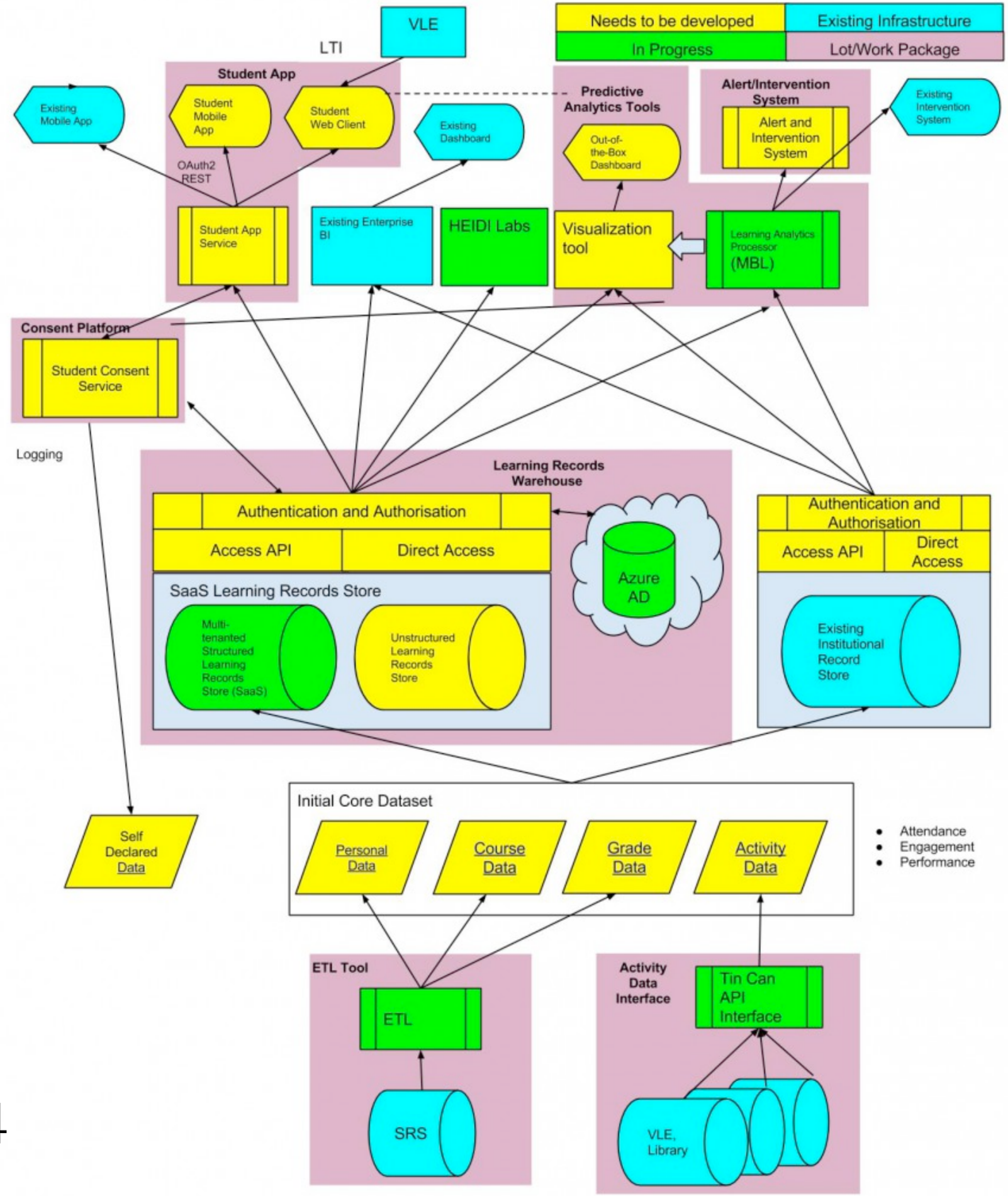
“We are what we measure.”

–Paolo Blikstein, 2013

Where does data come from?



Apereo LA Diamond, 2013



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?*

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)

9/14/17 9:36 AM

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,HUDK4050,Priceless
```

Zotero

- Install Firefox & Zotero (Assignment 1)
- Download the HUDK2017.rmd file from the syllabus GitHub repo
- Upload into Zotero folder
- Write a note within one of the articles you have read
- Right (cmd) click on the folder
- Choose “Export Collection…”
- Choose CSV from the drop down menu and choose a location you know the file path for
- Open R and type the code:

DF <- read.csv(“FILE PATH WHERE YOU PUT CSV FILE”, header = TRUE)

e.g.. DF <- read.csv(“~/Desktop/HUDK2017.csv”, header = TRUE)

- Behold! Your bibliography including your notes (under the “notes" column)

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)

- LearnSphere (https://learnsphere.org)

Yeah, but where do WE get data?

Cut a Deal

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

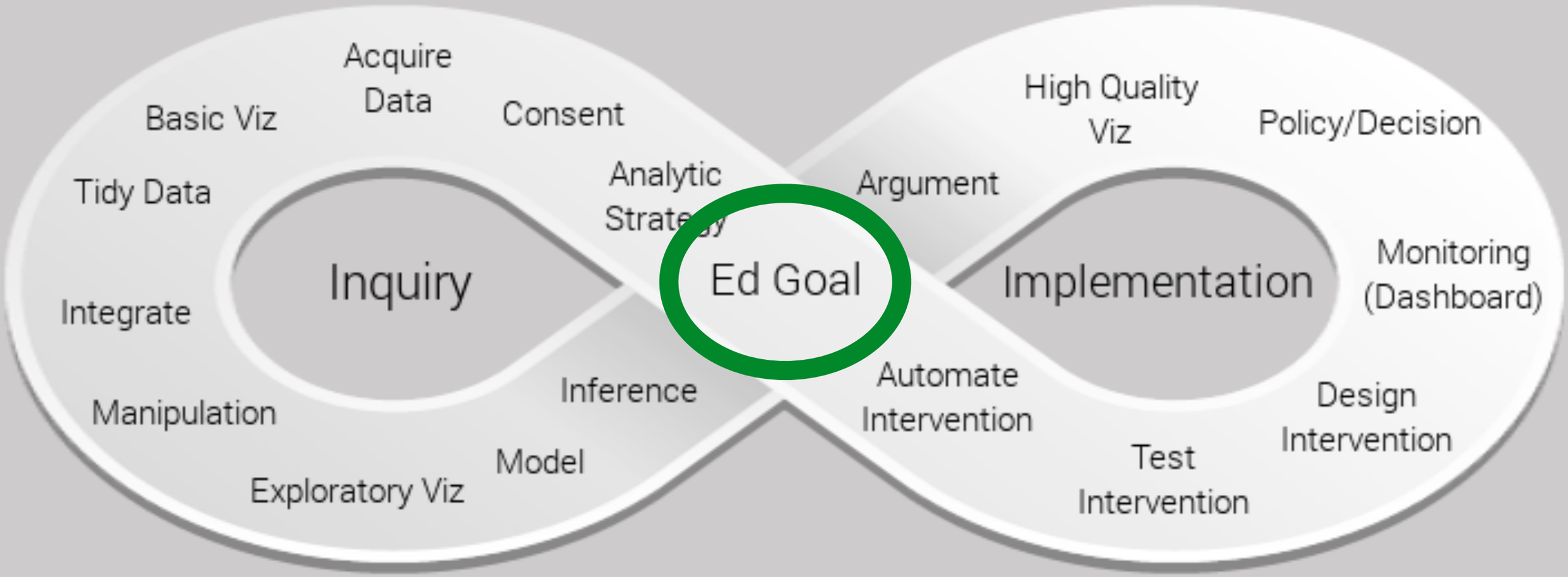
Automated acquisition

- APIs, web scraping, beacons

Generate

- Make your own!

Ed Data Science Cycle



Activity

- Consider your educational goal from last class
- What information would like to know today that might be useful to your analysis? Why?
- Convert that information to a data point/ variable

http://bit.ly/2flChcj