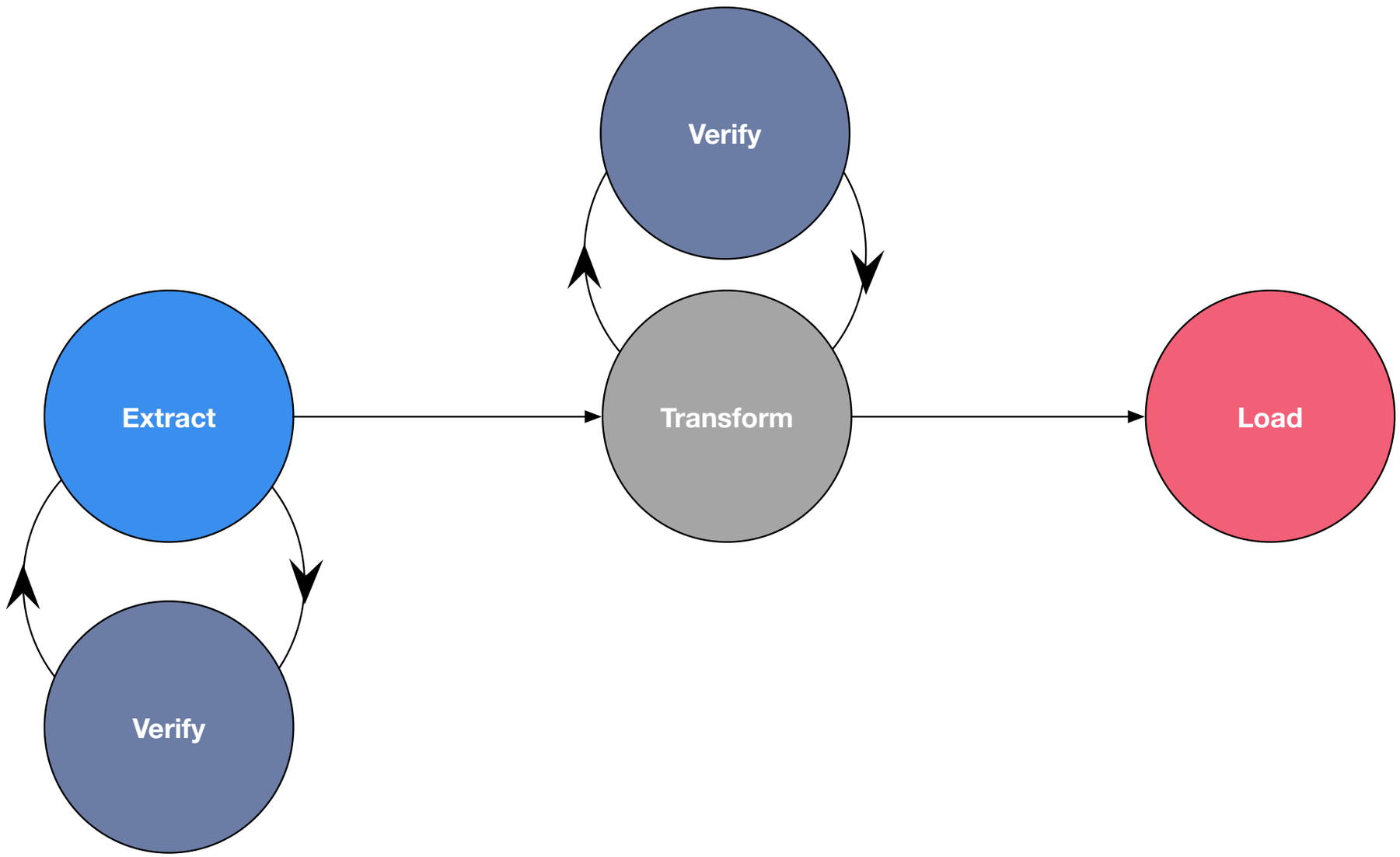
Repositories

INFX 551

- Agenda (this)
- Group work …
    - Github (briefly I promise)
    - Team repository and description
- Discussion / Lecture
- Activity
    - Data Curation Workbench…
    - Activity: Profile Repositories - CKAN + Dataverse
    - Activity: User stories
- Homework
    - User story and repo profile write up.
    - Data paper pitch

- Git set up.
    - Name your project.
    - Create a description of your project (the gauntlet has been thrown!!: https://github.com/RochelleLundy/INFX-551-Spring2017)
    - Add me as a collaborator.
    - 20 min… then we come back together and report out.

Data centric view…



What about a ‘systems’ centric view?

Repositories are sociotechnical infrastructures.

Data repositories are **sociotechnical** infrastructures.

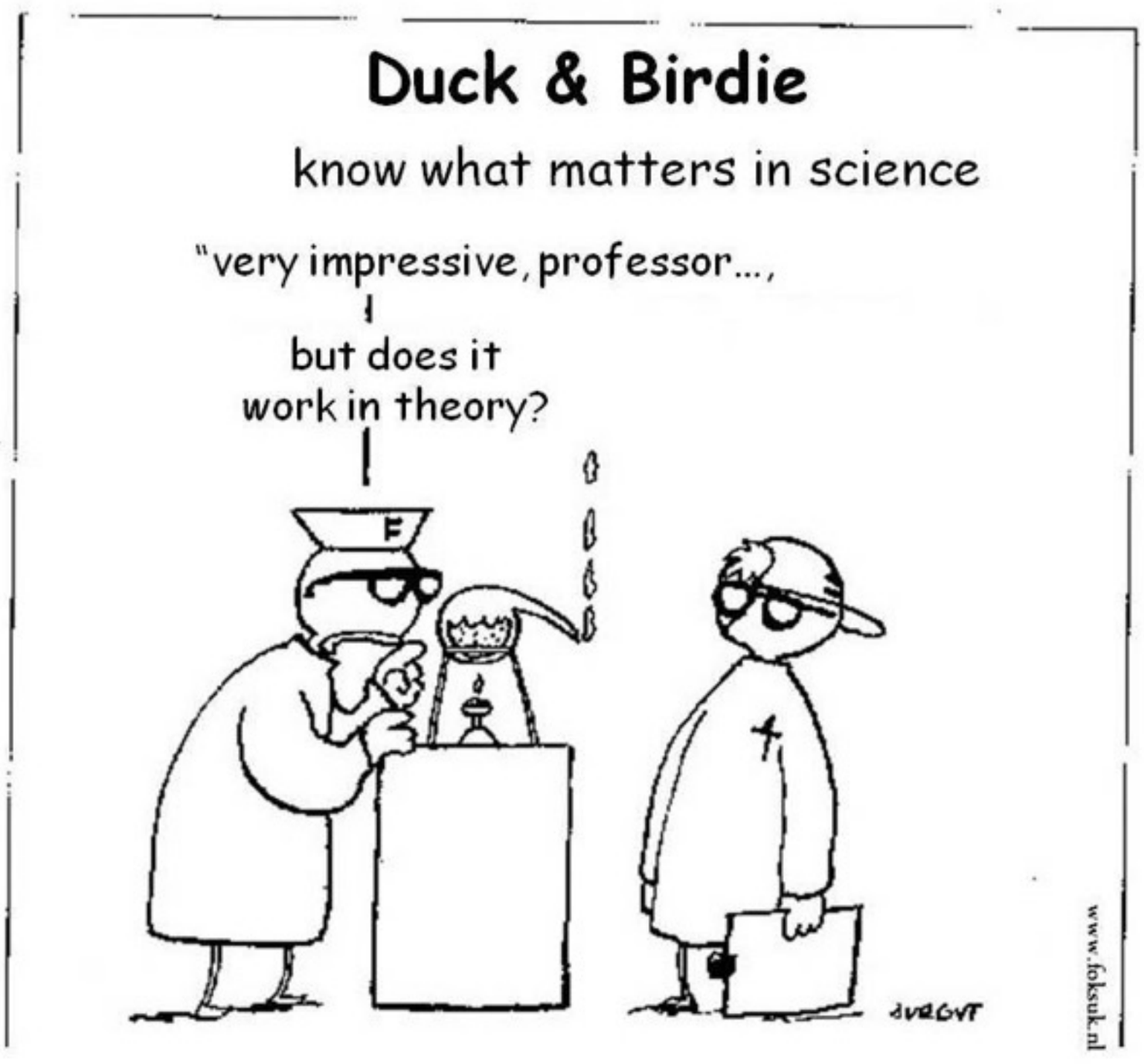
Mutual constitution of people and technologies, as they are embedded in political, economic, and ecological systems of collective action.

“ …the interdependent and inextricably linked relationships among the features of any technological object or system and the social norms, rules of use and participation (with that system) by a broad range of human stakeholders.”

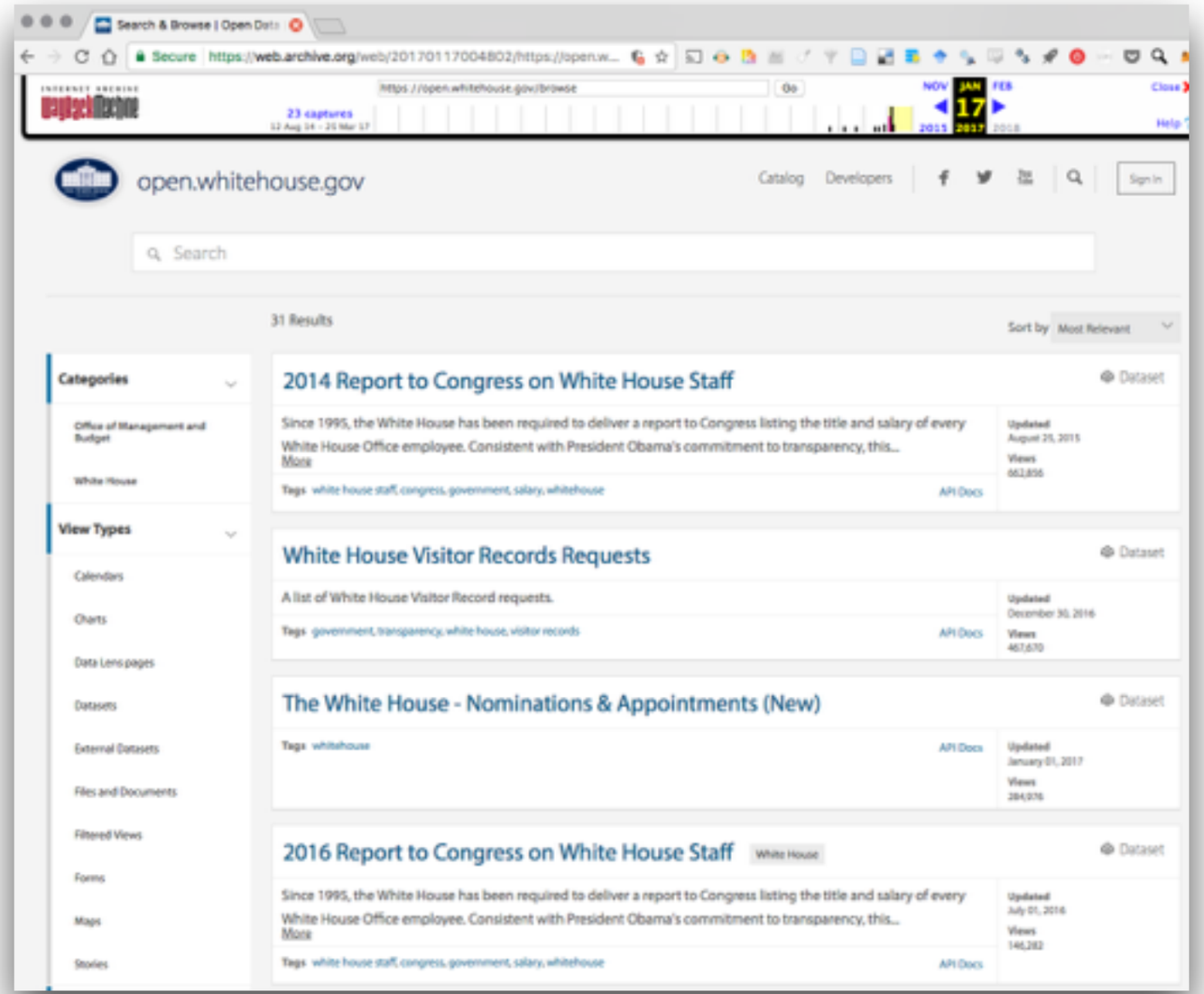
-Sawyer and Jarrahi (2014)

http://sawyer.syr.edu/publications/2013/sociotechnical%20chapter.pdf

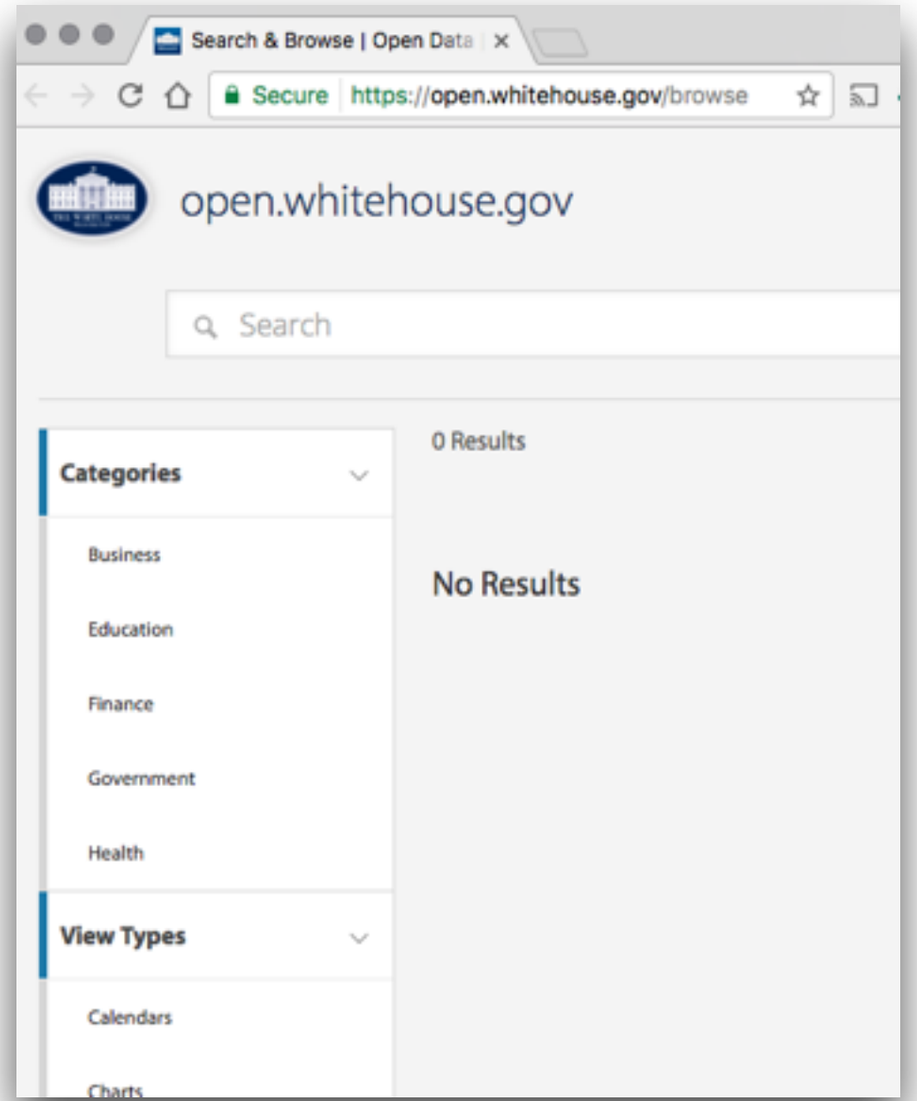
“Ok … what does this have to do with this class?”



January 17th



Current



Data repositories are sociotechnical **infrastructures**.

Infrastructures are

… relational
… embedded
… learned (as opposed to sensical)
… made up of standards
… built on other infrastructures
… often invisible (until break down)

Readings discussion

Lagoze

Data

“This perspective of data is reflexive; something (e.g.images, text, and Excel worksheet, etc.) is data because someone uses it as data in a specific context, and transcendent, it carries across the many disciplines, practices, and epistemologies of science.”

Lagoze

Big Data

Two perspectives:

3 V’s
3 “shifts”

Credible Science

Data Integrity

The Control Zone

What does this have to do with repositories?

Science

- Marcial, L. H., & Hemminger, B. M. (2010). Scientific data repositories on the Web: An initial survey. (Links to an external site.) Journal of the American Society for Information Science and Technology, 61(10), 2029-2048.

- 2015 Landscape on Research Data Repositories http://www.dlib.org/dlib/march17/kindling/03kindling.html (Links to an external site.)

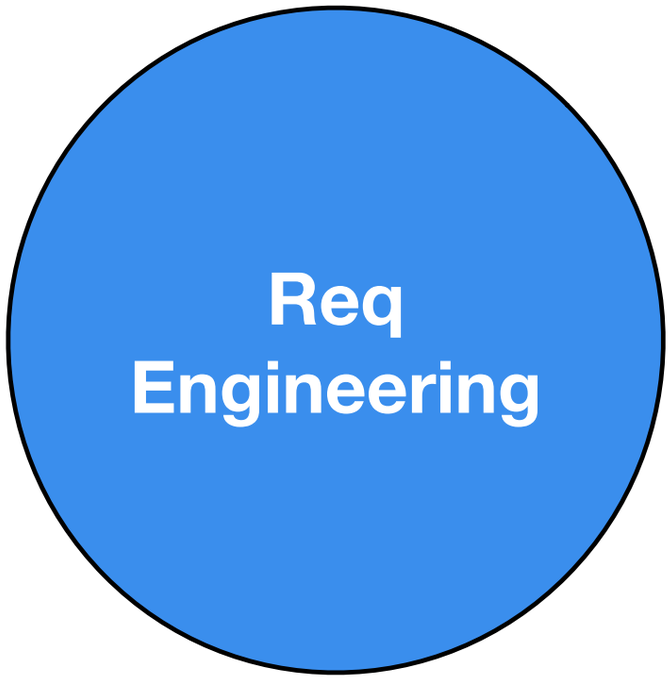
Humanities

- Svensson, P. (2011). From optical fiber to conceptual cyberinfrastructure. Digital Humanities Quarterly, 5(1). http://www.digitalhumanities.org/dhq/vol/5/1/000090/000090.html (Links to an external site.)

- Muñoz, T., & Renear, A. H. (2011). Issues in humanities data curation. https://www.ideals.illinois.edu/bitstream/handle/2142/30852/MunozRenear_2011_IssuesinHumanitiesDataCuration.pdf?sequence=2&isAllowed=y (Links to an external site.)

Public Sector Data

- Thorsby, J., Stowers, G. N., Wolslegel, K., & Tumbuan, E. (2016). Understanding the content and features of open data portals in American cities (Links to an external site.). Government Information Quarterly.


Req
Engineering

Functional aspects

- Deposit
- Ingest
- Preserve
- Create authenticity checks (object level management)
- Implement standards
- Documentation, Documentation, Documentation
- Serve out users…

Functional Requirements —> Technical Specifications

Systems Development for Data Curation

Typically we approach user functional requirements in two ways:

1. User Story
2. System / Architecture Profile

Both of these sets of documents help us build a use case (your second assignment).

Assignments:

- Write-up reflection on your activities
- Data Paper (pitch)