

The Product & Asset Specificity of Measuring Impact Some Indicators of Data Use in a Research Data Archive

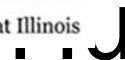
Nic Weber Andrea Thomer Matt
Mayernik

Steven Worley Robert Dattore Zaihua

CIRSS

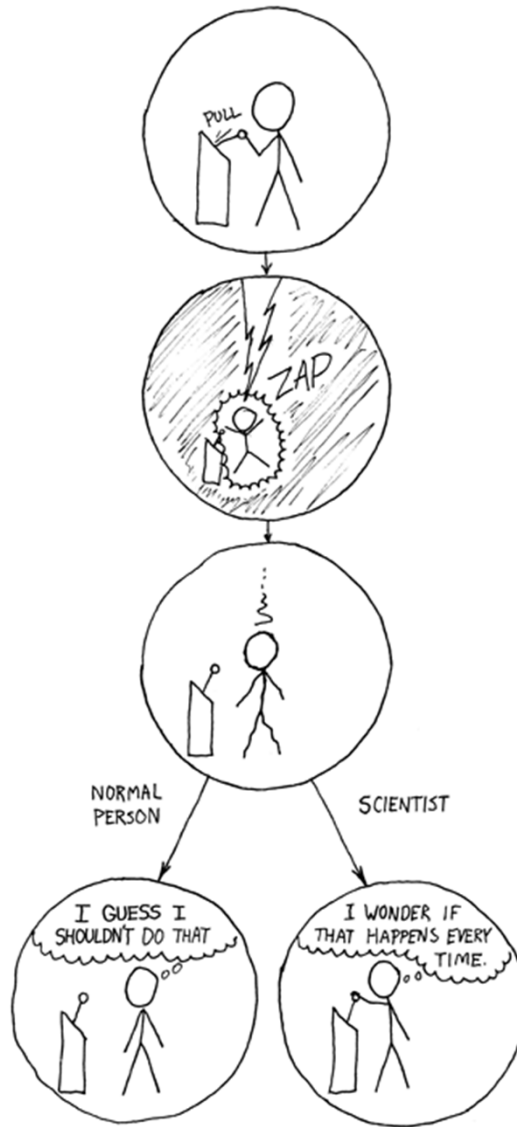
GRADUATE SCHOOL OF LIBRARY AND
INFORMATION SCIENCE

The iSchool at Illinois



NCAR

NATIONAL CENTER FOR ATMOSPHERIC RESEARCH



RDA @ NCAR

Curating data before data was curated

Holdings > 1.3 PB , including static and dynamic datasets

2011 served ~1 PB to ~1500 unique users from 127 different countries.

The Challenge

This --> “*2011 served ~1 PB to ~1500 unique users from 127 different countries*”

Impressive, but not very meaningful.

At best it's an incomplete picture of the RDA's curation work, and its impact on earth science.

Impact

What counts?

What to count?

RDA:

What to count?

Events and traces

Datasets served, Files served, Datasets accessed, Datasets browsed, Datasets searched, Number of users served, Number of unique users registered, Homepage hits, Page redirects, Dwell time, Format requested, Access type, Subsets, Subset requests

Data Usage Index

(Ingwersen and Chavan, 2011)

With abundance of traces and events the question becomes not what to count, but what are the most meaningful arrangements

14 usage based **indicators** that attempt to leverage user logs in creating comparable within and between metrics of repository impact

*Indicators

Data Download:Use

::

Article Download:Read

Indexes

Complications with extending data-based indexes turns on the issue of what we mean by use

Not all indicators are replicable in all repository / archival environments

Specificities

(Williamson 1981)

Externalities that lock an organization into certain ways of doing work.

Product Specificities

Qualities and properties of datasets (their file structure, format, and size) that affect the way a user can interact with the archive in consuming and discovering data.

System Specificities

The architecture and arrangement of datasets that prescribe, and limit the way a user can interact with the content of an archive.

DUI @ RDA

Product Specificities: Hourly and daily, model output (necessary to download component files), TB scale.

System Specificities: faceted browsing, subsetting functionality, programmatic access.

What counts?

(closing)

Efficiency makes work visible (not amount)

Density of downloads (not raw counts)

'Maturity' of dataset (not size)

Thanks

(please see the paper)

nmweber@illinois.edu

@nniicc

Response to questions

Complimentary to efforts of citations,
publishing and minting identifiers to track
data use