#GITS2015 #MetaDataCAPTL

@all_metadata

Metadata Capital: Conceptual Understanding, Predictive Value

HITS—Hollywood IT Summit

May 14, 2015

Jane Greenberg, Professor
Director, Metadata Research Center
College of Computing & Informatics
Drexel University

DREXEL UNIVERSITY
Metadata Research Center
College of Computing & Informatics

What is capital to you?

Write something down

DREXEL UNIVERSITY
Metadata
Research Center
College of Computing & Informatics

How do you describe value?

Write something down

DREXEL UNIVERSITY
Metadata
Research Center
College of Computing & Informatics

Capital?

Value ?

Your data is only as good as your metadata

***Metadata** is a first class object*

DREXEL UNIVERSITY
Metadata Research Center
College of Computing & Informatics

The topic…

- Motivation – Making metadata work harder (DRYAD)
- ROI – return on investment (CAPITAL)

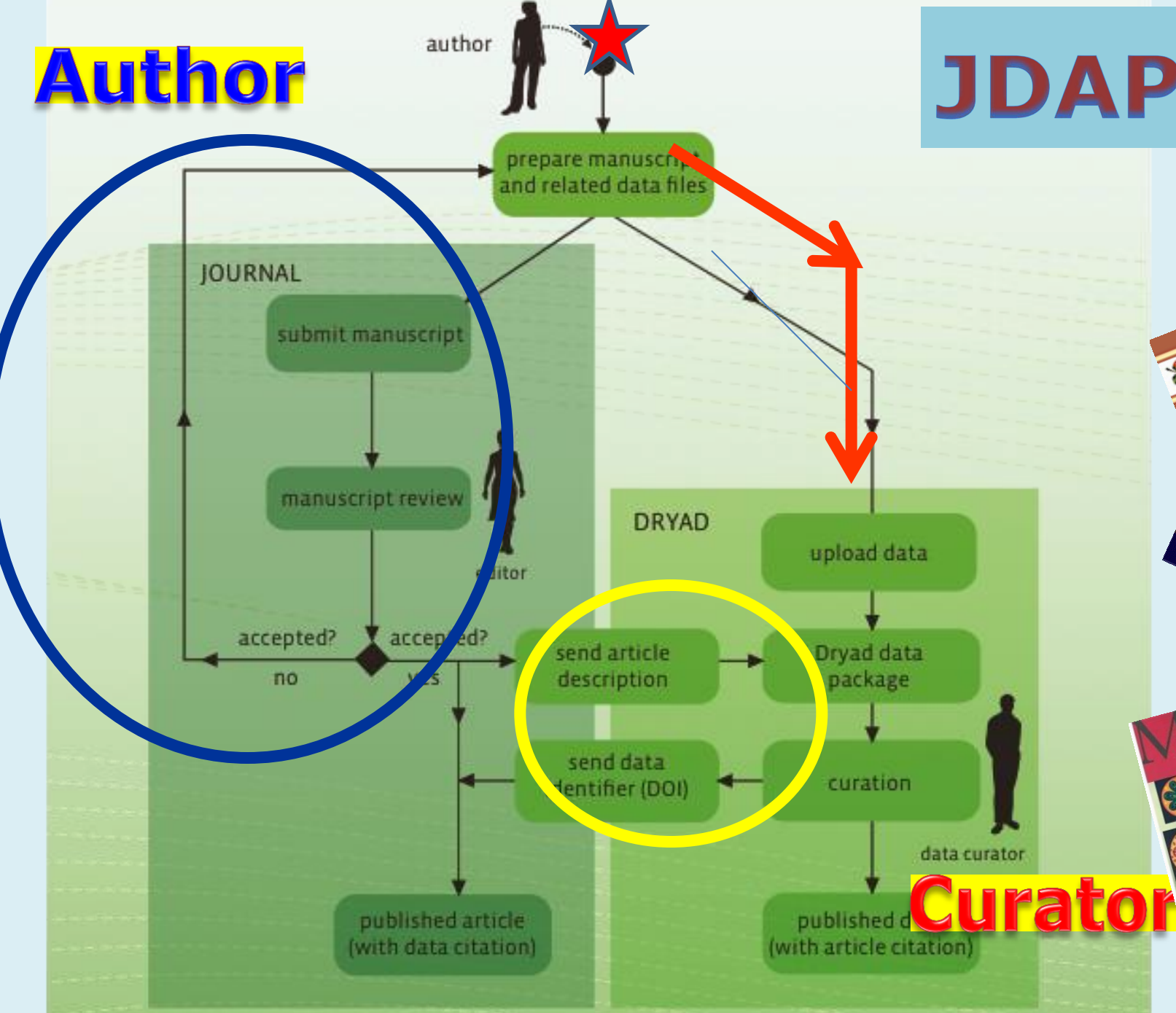
DREXEL UNIVERSITY
Metadata Research Center
College of Computing & Informatics

DRYAD

Not this →

Author

JDAP

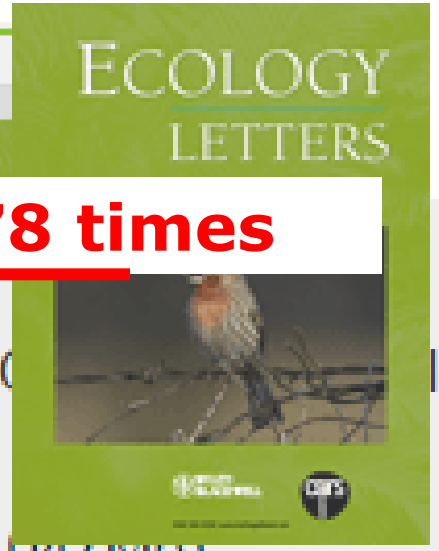


Curator

Data from: Towards a worldwide wood economics

Downloaded	12378 times
Description	Please direct all correspondence to C Gonzalez@leeds.ac.uk>
Download	GlobalWoodDensityDatabase.xls (2.047Mb)
Details	View File Details

Download 12378 times



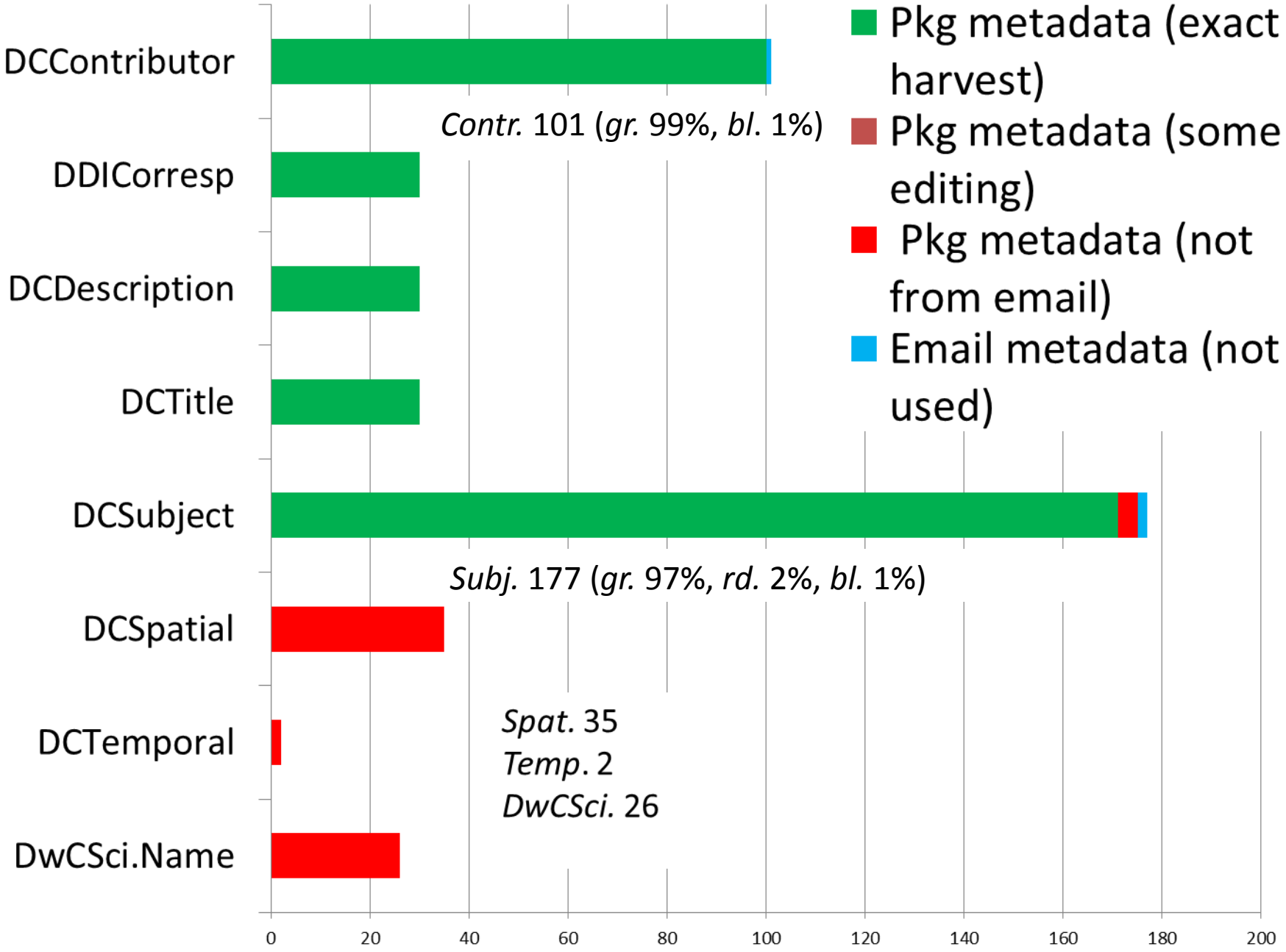
Observations, motivating study of metadata capital

1. Metadata generation costs money
2. Metadata reuse is **a BIG part** of Dryad's workflow
3. Metadata reuse via OAI
4. Metadata reuse via data sharing, reuse, and repurposing

Statistics

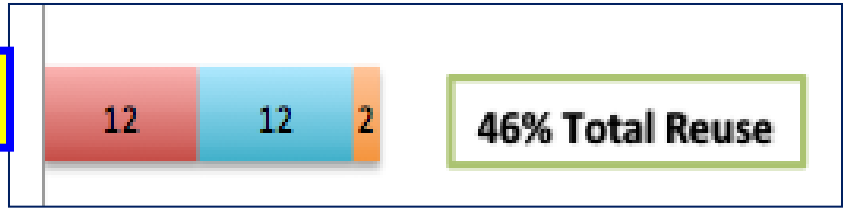
Type	Total	30 days
Data packages	7978	326
Data files	24903	1195
Journals	390	105
Authors	28714	4279
Downloads	772229	18043

Package metadata harvested from email

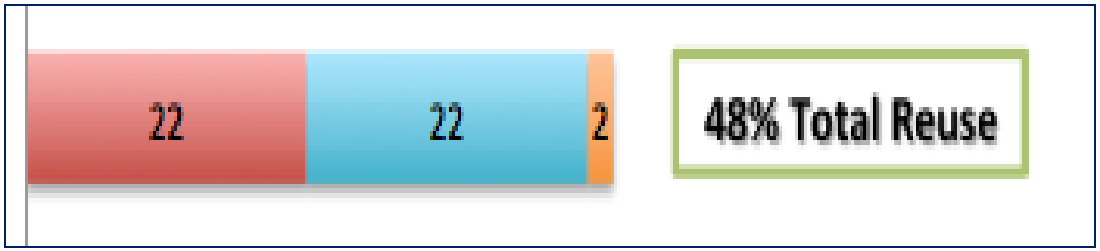


Author

Dcterms.spatial→



Subject



← DwC.ScientificName

MetaDataCAPT'L

ncds THE NATIONAL CONSORTIUM for DATA SCIENCE

$$R + \sum_{i=1}^{n} a_i = R + a_1 + a_2 + a_3 + \ldots a_n$$

R = value of the metadata record

i= number of usages

a = incremental increase in value

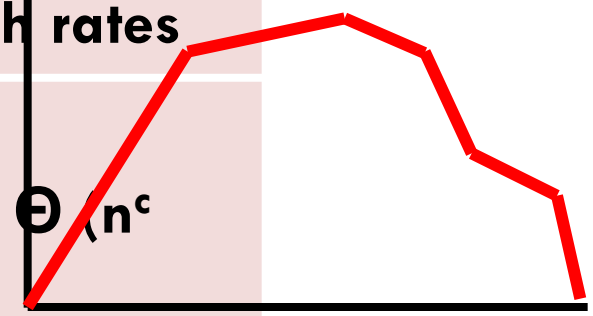
n = maximum number of reuse

Metadata duplication is inefficient, tedious

- An economic concept (Weber, 1905; Smith, 1776)

Successive growth rates

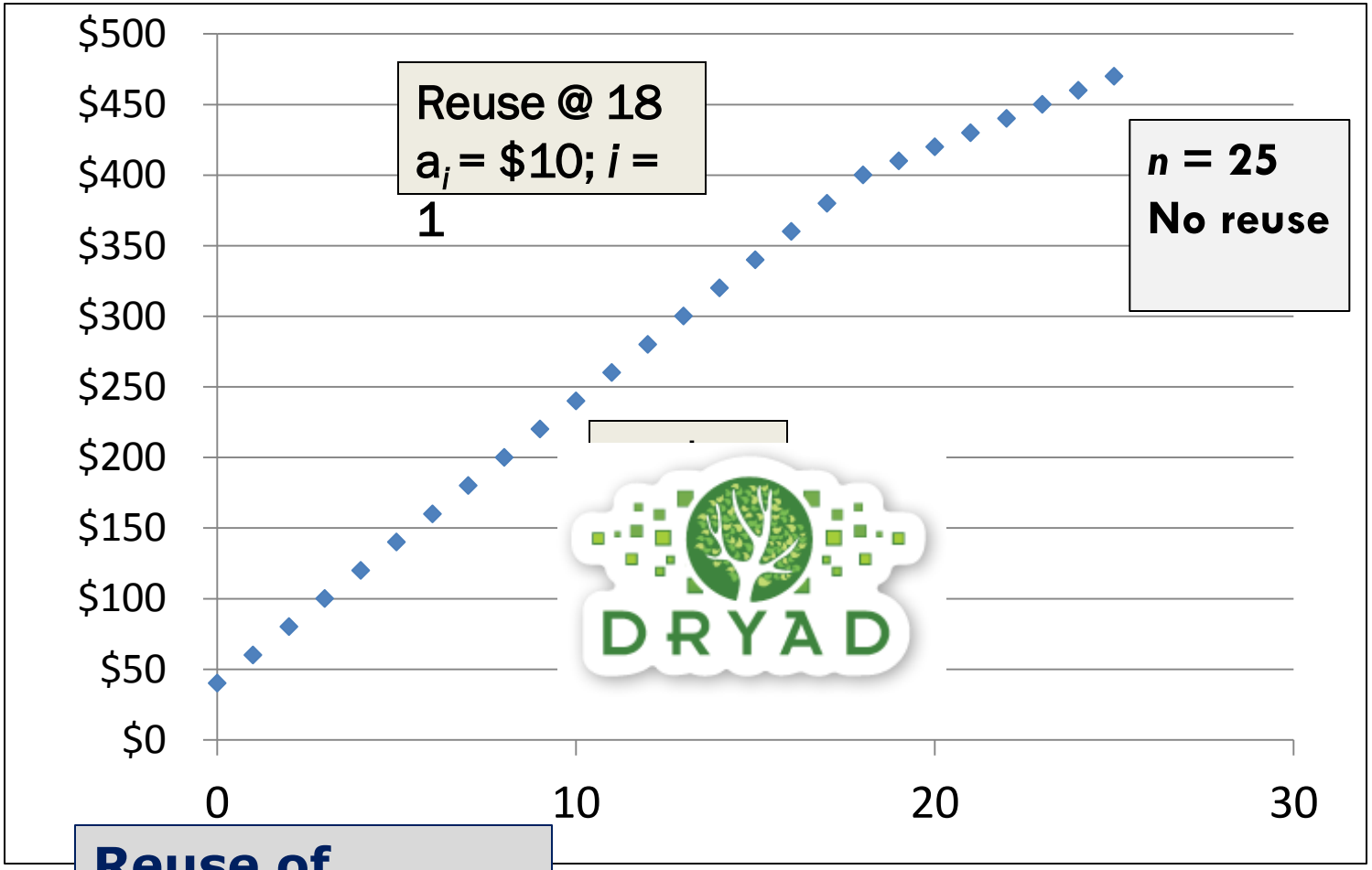
$$\sum_{i=1}^{N} i^c = \Theta(n^c + 1)$$



Cycles

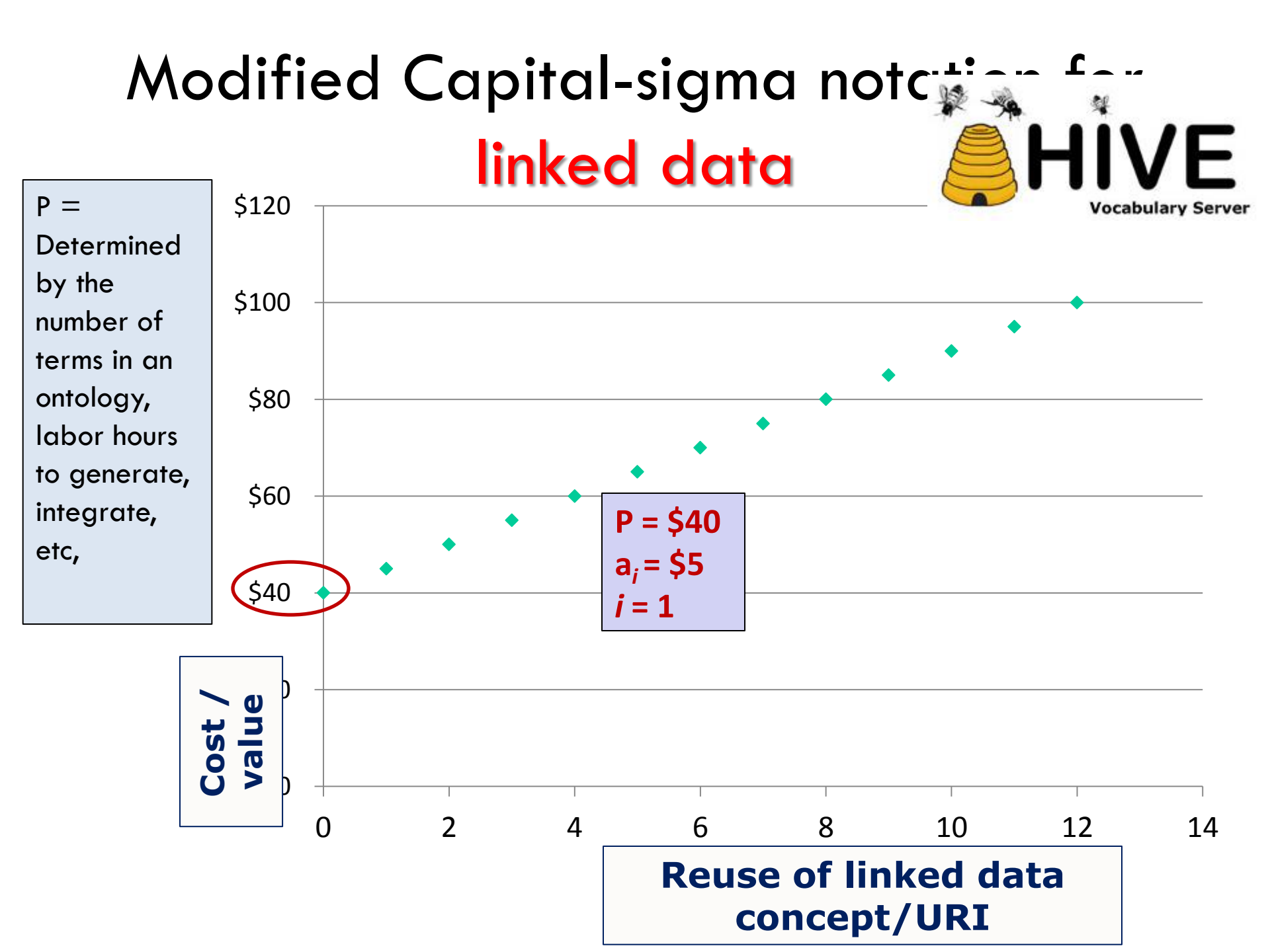
THE NATIONAL CONSORTIUM for DATA SCIENCE

MetaDataCAPT'L @ $440 for a data object



Reuse of metadata

Modified Capital-sigma notation for linked data



RTI INTERNATIONAL

SGHIx CONSORTIUM

	A	B	C	D	E	F	G	H
	Validic API Cate	Parameter in Us	BodyMedia	FatSecret	DailyMile	Fitbit	Fitbug	Fleetly
	Fitness	utc_offset	P	X	X	X	X	P
	Routine	timestamp	X	NA	NA	X	X	NA
	Routine	steps	X	NA	NA	X	X	NA
	Routine	distance	X	NA	NA	X	X	NA
	Routine	floors	X	NA	NA	NA	NA	NA
	Routine	elevation	NA	NA	NA	NA	NA	NA
	Routine	calories_burned	NA	NA	NA	X	X	NA
	Routine	utc_offset	X	NA	NA	X	X	NA
	Nutrition	timestamp	X	X	NA	X	X	NA
	Nutrition	calories	X	X	NA	X	X	NA
	Nutrition	carbohydrates	NA	X	NA	X	P	NA
	Nutrition	fat	NA	X	NA	X	P	NA
	Nutrition	fiber	NA	NA	NA	X	P	NA
	Nutrition	protein	NA	X	NA	X		NA
	Nutrition	sodium	NA	X	NA	X		NA
	Nutrition	water	NA	NA	NA	NA		NA
	Nutrition	meal	X	X	NA	X		NA
	Nutrition	utc_offset	X	X	NA	X		NA

Total Fields Referenced (FitBit), toward SGHIx

X Available: 39

P (Pending): 3

NA (not available): 42

(Caruso & Ogletree)

metadata

Most popular
timestamp
type
start_time
distance
duration
calories
utc_offset

Conclusion…other Valuation Approaches

Market cap of Facebook per user: $40 – $300

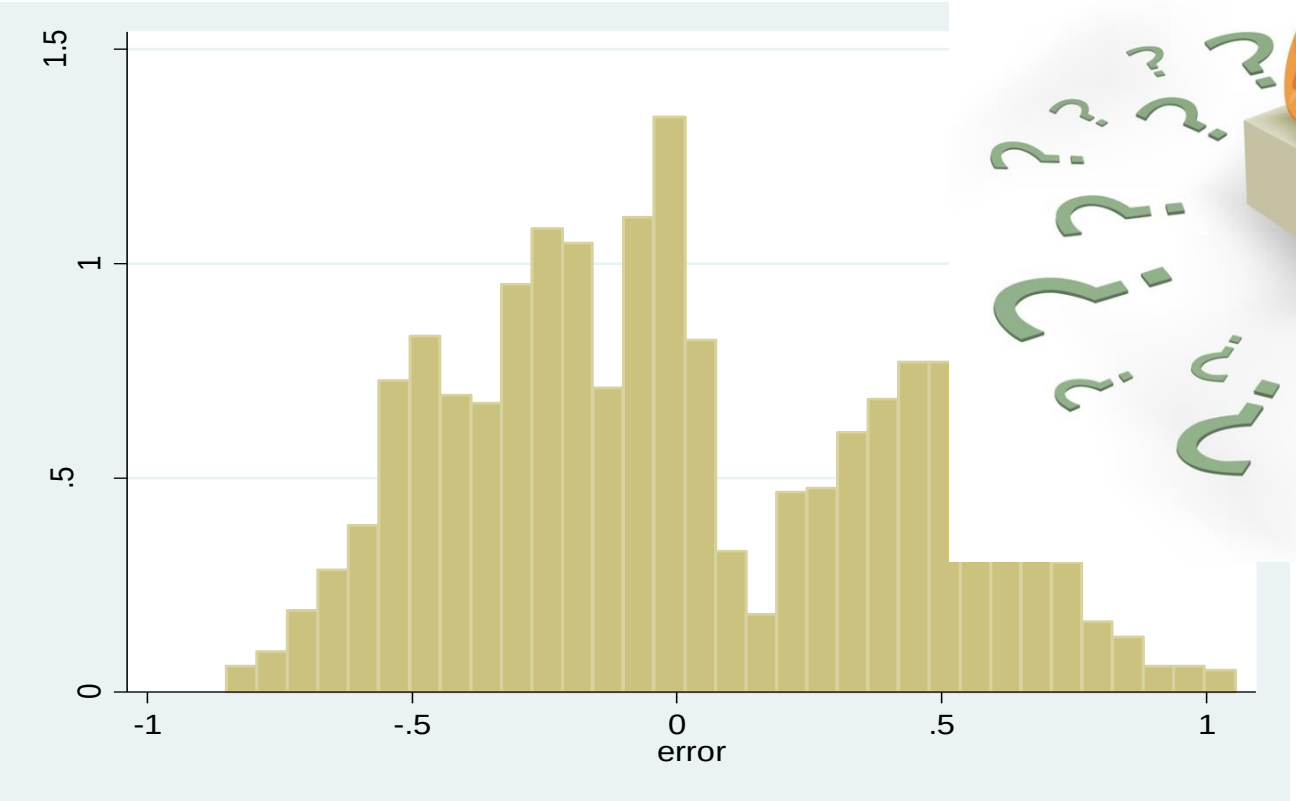
Revenues per record per user: $4 – $7 per year

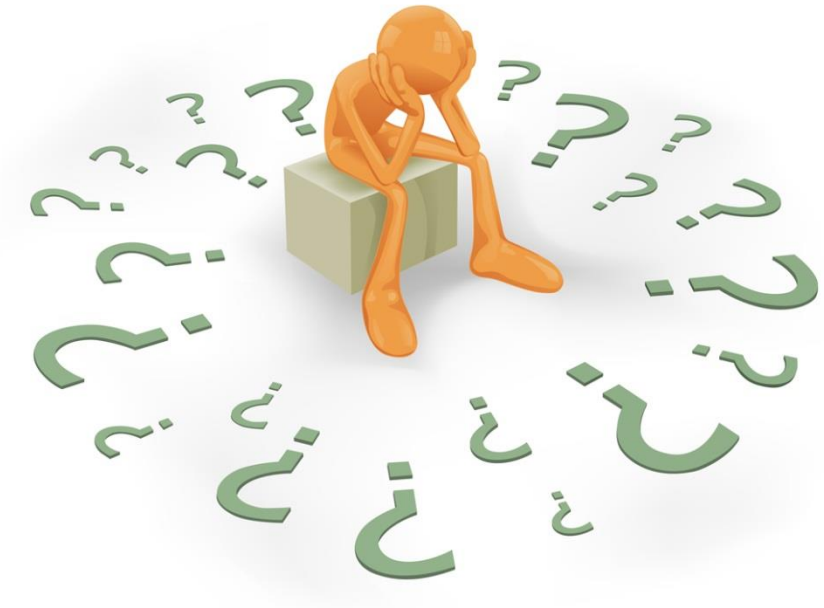
- Facebook
- Experian

Market prices of personal data:

- $0.50 for street address
- $2.00 for date of birth
- $8 for social security number
- $3 for driver's license number
- $35 for military record

SOURCE: OECD. *Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value*. OECD Digital Economy Papers. Office for Economic Cooperation and Development Publishing, 2013.





Fitbit data scenario:

patient's exercise data and environmental quality data can be combined with asthma condition data, we will get a better prediction of the way in which asthma evolves.

Working on metadata connection…

National Institute of Environmental Health Science

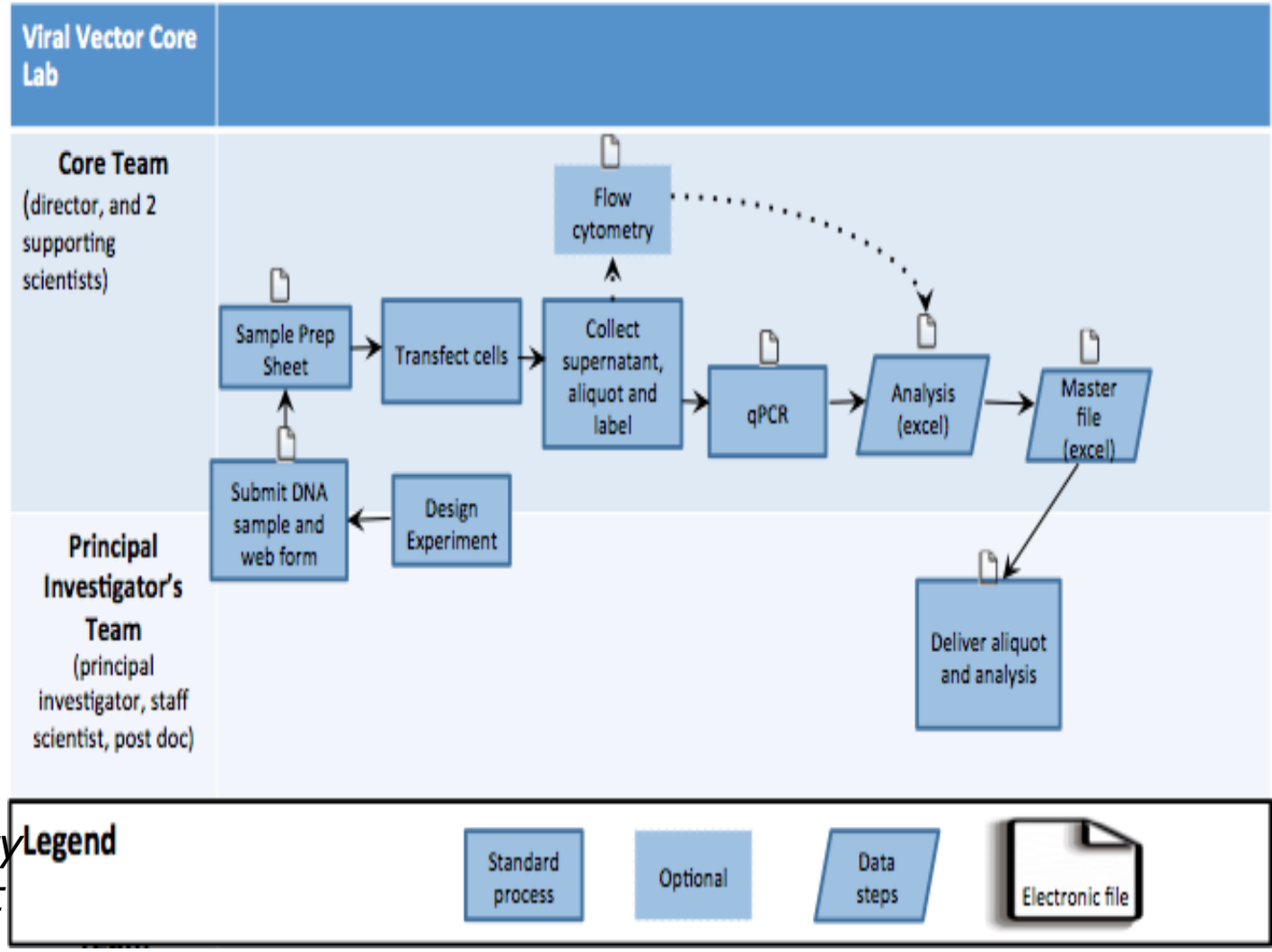
Goals of exploratory work

- Understand the Viral Vector Core Laboratory (VVCL) workflow.
- Map the VVCL metadata lifecycle.
- Explore machine-actionable rules that can support the VVCL metadata lifecycle.
- Create an iRODS prototype for the VVCL workflow, and explore the application of machine-actionable rules.

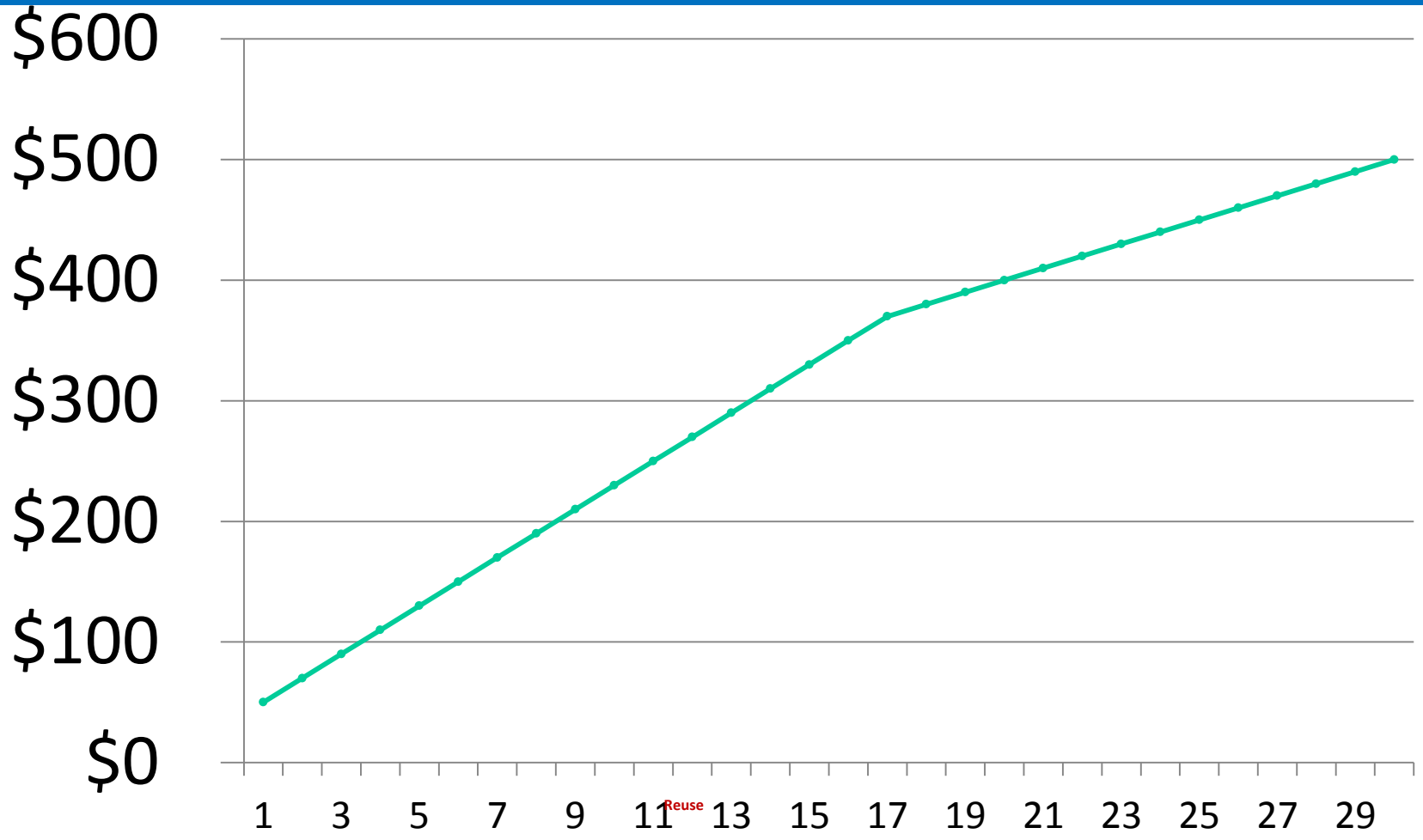
Methods

- Collaborative workflow modeling was used to capture the day-to-day workflow.
- A metadata analysis - identify basic metadata generated, automatically propagated during each workflow stage in the VVCL process.
  - Microservices articulation

- *Design experiment:*
- *Submit DNA sample and web form*
- *Transfect cells*
- *Collect the supernatant, aliquot, and label*
- *Present the qPCR*
- *Flow cytometry*
- *Deliver aliquot and analysis*



$50 USD per hour for an experiment



Some concluding observations

- Discover and advance the application of methods for quantifying the cost and value of metadata over time
- *Raise dialog*
- Advance nascent work on "metadata capital"

Information as an economic asset - Machlup's *The Production and Distribution of Knowledge in the United States*

Limitations

- Modified capital-sigma is only one dimensional; all metadata properties/concept are not equal
- Also, we know cost/value relationship is not 1:1.
- Metadata is only as good as your data

  not always true

- What about successive growth rate may be the way to go

HITS entertainment community

DREXEL UNIVERSITY
Metadata
Research Center
College of Computing & Informatics

Capital?

ANTIQUES
& OLD STUFF
Value?
www.bigstock.com · 164410

Friendship, social network

Capital, value?

Box Office Mojo

Search Site

Search

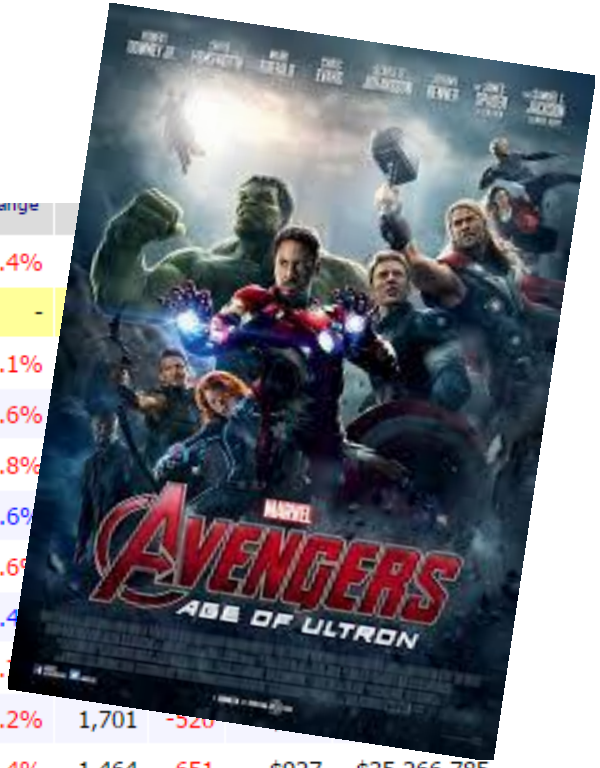
- Daily
- **Weekend**
- Weekly
- Monthly
- Quarterly
- Seasonal
- Yearly
- All Time
- Chart Watch
- International

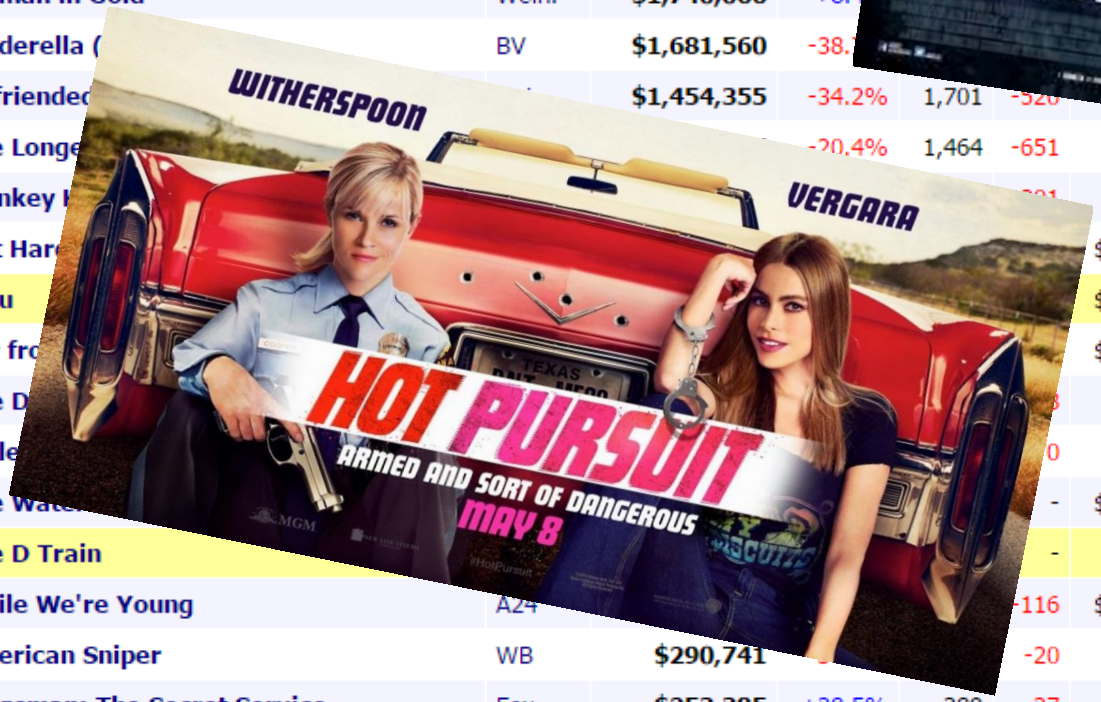
Indices

- Movies A-Z
- Studios
- People
- Genres
- Franchises
- Showdowns
- Oscar
- Theater Counts

TOP STORY

				BV	$77,746,929	-59.4%			
2	N	Hot Pursuit		WB	$13,942,258	-			
3	3	The Age of Adaline		LGF	$5,821,894	-6.1%			
4	2	Furious 7		Uni.	$5,407,200	-18.6%			
5	4	Paul Blart: Mall Cop 2		Sony	$5,306,041	-9.8%			
6	7	Ex Machina		A24	$3,510,224	+53.6%			
7	5	Home (2015)		Fox	$3,070,015	-11.6			
8	10	Woman in Gold		Wein.	$1,740,066	+8.4			
9	6	Cinderella (		BV	$1,681,560	-38.			
10	8	Unfriended			$1,454,355	-34.2%	1,701	-526	
11	9	The Longe				-20.4%	1,464	-651	$927	$35,266,785
12	11	Monkey							$875	$14,363,633
13	12	Get Har							$1,107	$87,831,115
14	N	Piku							$8,025	$938,938
15	23	Far fro							$7,842	$1,016,528
16	13	The D							$920	$127,623,243
17	14	Little							$892	$5,330,380
18	15	The Wat						-	$1,432	$3,226,006
19	N	The D Train						-	$444	$447,524
20	17	While We're Young		A24				-116	$1,161	$6,909,277
21	16	American Sniper		WB	$290,741			-20	$723	$349,324,606
22	21	Kingsman: The Secret Service		Fox	$253,395	+28.5%	289	-27	$877	$127,227,084





Highest Rated TV Series Released In 2015

1-50 of 2,713 titles. Next »

Sort by: MOVIEmeter | A-Z | **User Rating▼** | Num Votes | US Box Office | Runtime | Year | US Release Date

1. **Mutlu Ol Yeter** (2015 TV Series) Add to Watchlist
   9.9/10
   Dir: Kerem Çakiroglu With: Ali Atay, Asli Enver, Ertan Saban
   Comedy | Drama | Romance

2. **Bes Kardes** (2015 TV Series) Add to Watchlist
   9.5/10
   This is one of the five brothers, not like in the slightest way to another.
   Dir: Onur Ünlü With: Fatih Artman, Merve Ates, Tansu Biçer
   Comedy | Drama | Family 120 mins.

3. **Hunters of the South Seas** (2015 TV Series) Add to Watchlist
   8.6/10
   Writer Will Millard explores the extraordinary people of the Coral Triangle in the Western Pacific. Here whale hunters, spear fishermen and ocean traders all live in close connection to the sea.
   Dir: Will Lorimer With: Will Millard
   Documentary 59 mins.

4. **Çka ka Shpija** (2015 TV Series) Add to Watchlist
   9.2/10
   Comedy based on Characters,Situations and Style (Humor)
   With: Naser Rafuna, Gresa Pallaska, Leze Qena
   Comedy 50 mins.

Feedbac about thi

Discussion…

- Can we study cost?
- How do we convey value?
- Is there a connection between cost/value/quality?
- How does this all fit with media and entertainment?

DREXEL UNIVERSITY
Metadata Research Center
College of Computing & Informatics

Your data is only as good as your metadata

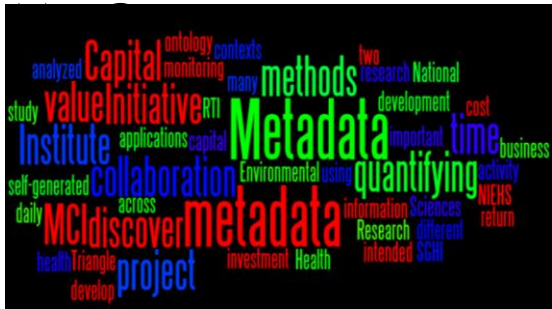
***Metadata** is a first class object*

DREXEL UNIVERSITY
Metadata Research Center
College of Computing & Informatics

The Team / acknowledgments

ncds | THE NATIONAL CONSORTIUM for DATA SCIENCE

- Tom Caruso, Health Information Liaison Research Associate, UNC-SILS/RTI
  - –Self-generated Health Information (SGHI)
- Rebecca Boyles, Data Scientist, NIEHS
  - –Common Core Vocabulary
- Jane Greenberg, SILS/UNC, MRC
- Herbie Huang, Ph.D. student, Economics Dep. UNC
- Austin Mathews, BSIS student, SILS/UNC, MRC
- Angela Murillo, Ph.D. student, ,SILS/UNC, MRC
- Adrian Ogletree, MSIS student, SILS/UNC,



DREXEL UNIVERSITY

Metadata Research Center

College of Computing & Informatics

http://cci.drexel.edu/mrc/

DREXEL UNIVERSITY
Metadata Research Center
College of Computing & Informatics