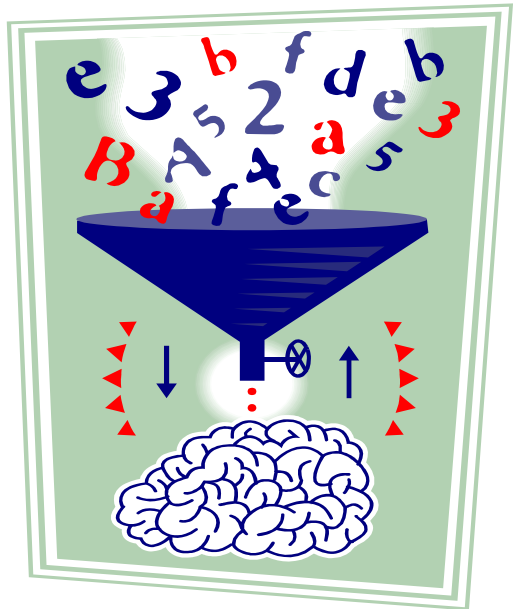
Open Science, Data and Publications

Sylvie Brouder
Professor And Wickersham Chair
Purdue University

Steven Daley-Laursen
Professor And Senior Research Executive
NAREEE Advisory Board, Open Data/Science Chair
University Of Idaho

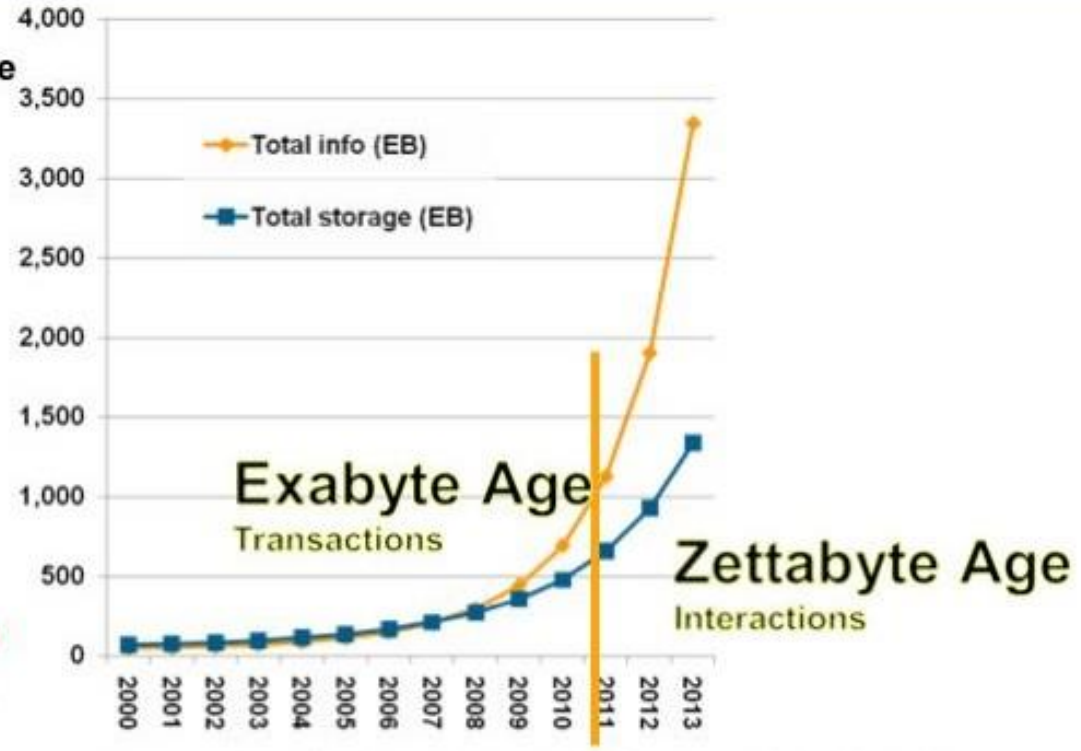
Many Types of Data to be Managed by Universities

- **Research and Outreach Data & Computing**
- On-Line Teaching
- University Business Analytics
- Student Data



1. Data Explosion; Volume, Variety, Velocity

- More data has been created in the last three years than in all past 40,000 years.
- Almost all of this data has a location
- Business and government decision-makers must have a strategy for dealing with location based data



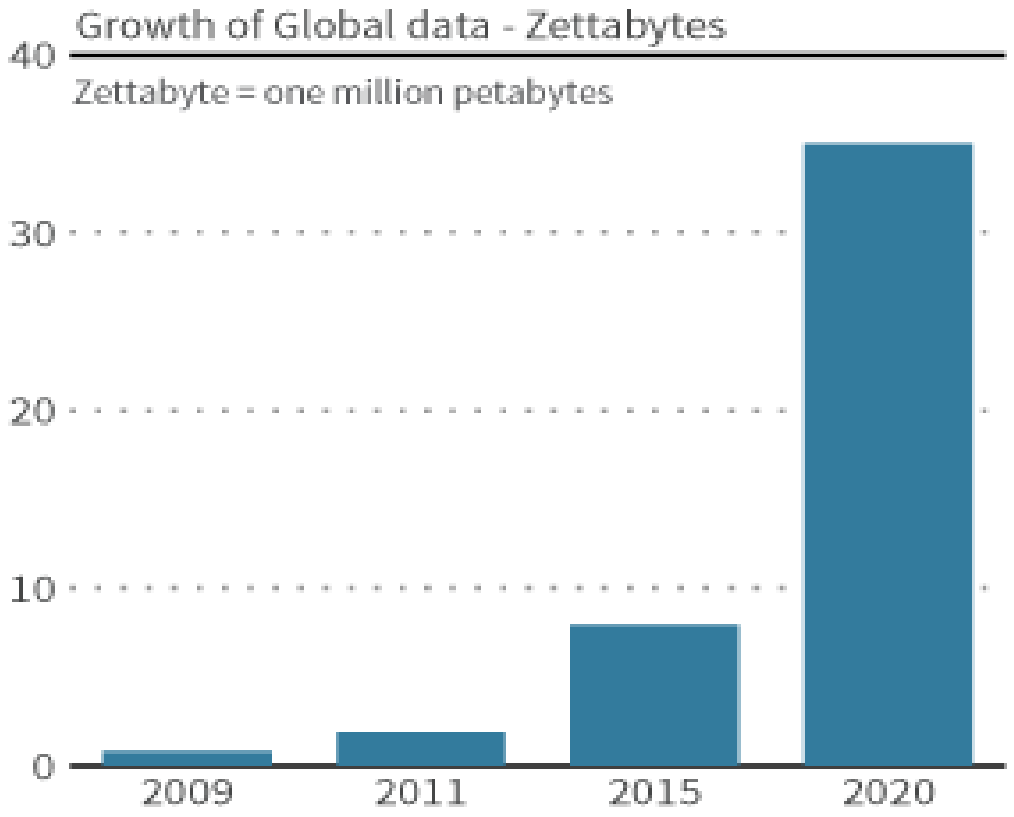
One Zettabyte (ZB) = 1,000,000,000,000,000,000,000 bytes = 10^{21} bytes.
Based on IDC and UC Berkeley data growth estimates.

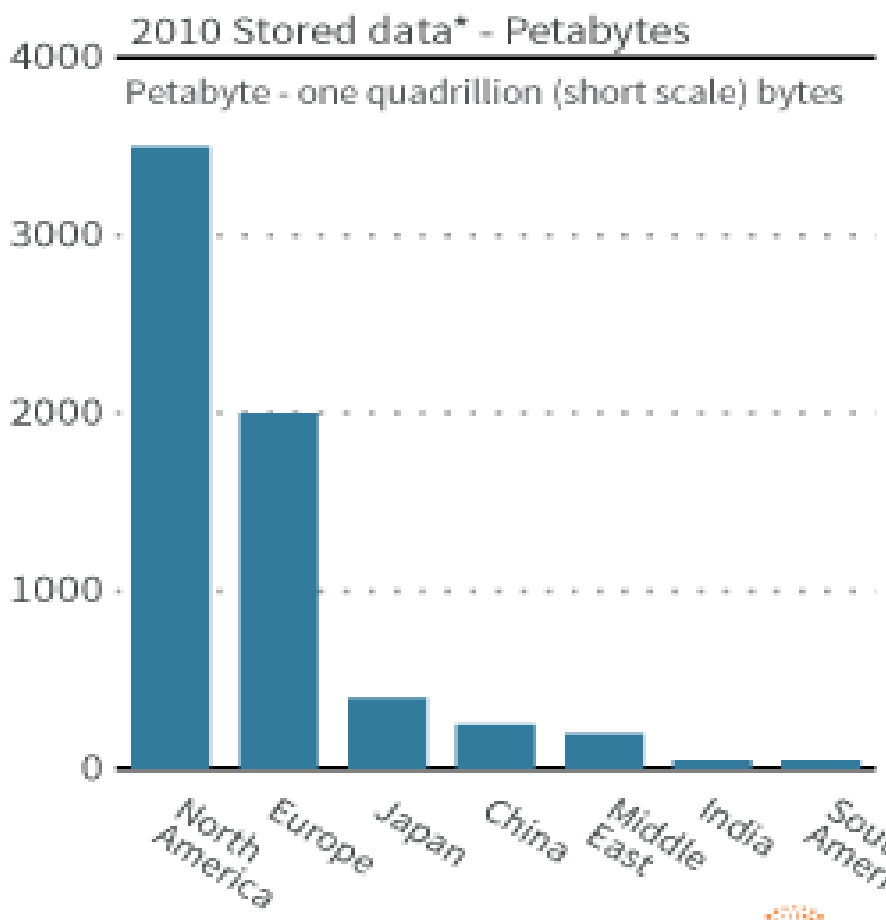
Technology Trend: (1) Sensor data and mobility apps are creating more data tagged with location. (2) Increasing number of apps are location-aware, so queries involve spatial dimension. High confidence that analytic apps will include who-what-when-**where** dimensions.

TERADATA
Raising Intelligence

Big data growth

Big data market is estimated to grow 45% annually to reach $25 billion by 2015





*greater than

Sources: Nasscom -CRISIL GR&A analysis

REUTERS

Reuters graphic/Catherine Trautman 05/10/12

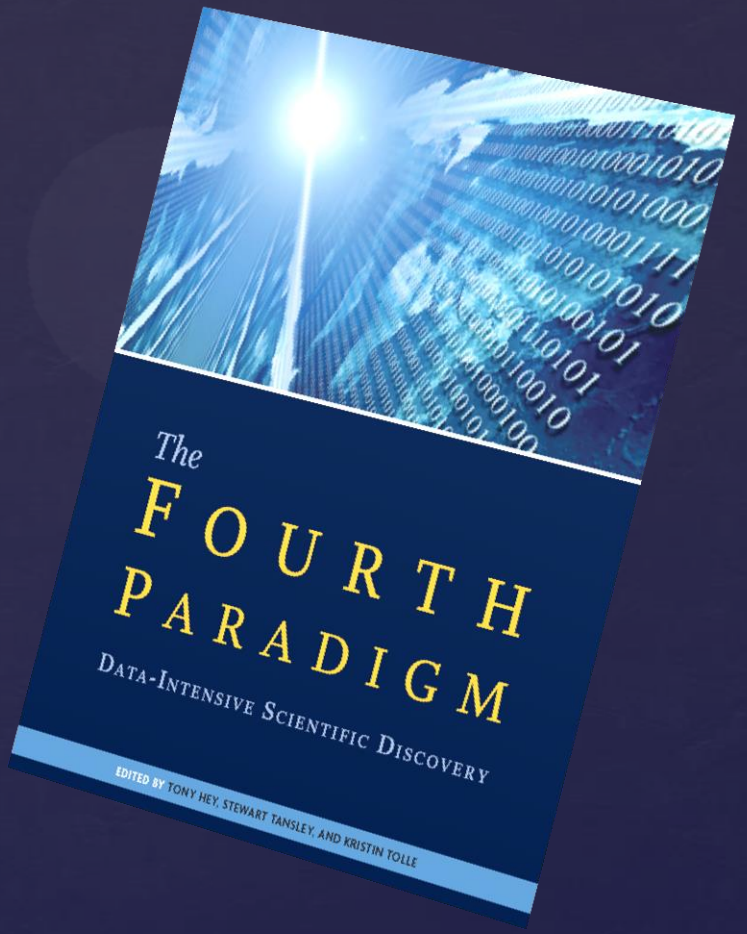
2. Science More Integrated, Computational, Data Intensive

"…data and software are redefining what it means to do science."

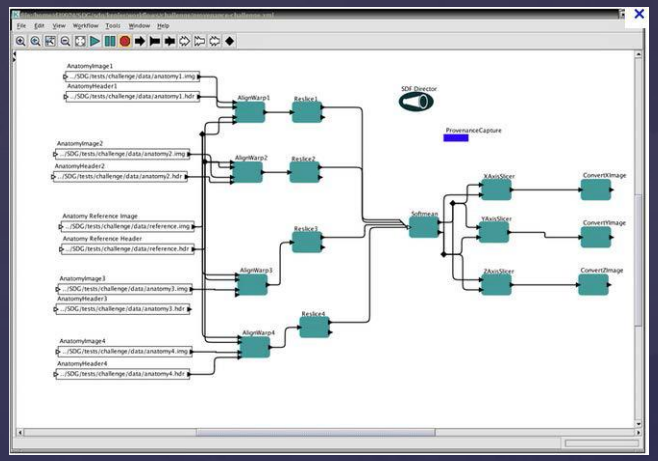
— **Bill Gates,** Chairman, Microsoft Corporation

"…greatest challenge for 21st-century science is responding to the new era of data-intensive science … a new paradigm beyond experimental and theoretical research and simulations of nature, requiring new tools, techniques, and ways of working."

— **Douglas Kell**, University of Manchester



Hey, Tansley, Hawley, Fourth Paradigm, http://research.microsoft.com/fourthparadigm/



"…everything about science is changing because of the impact of information technology. Experimental, theoretical, computational science are all being affected by the *data deluge*, and a fourth, ***data intensive science*** paradigm is emerging.

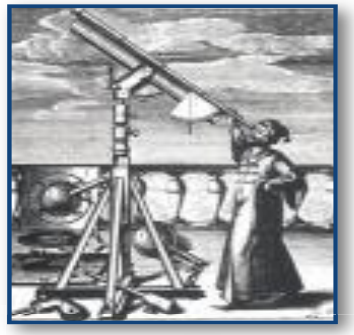
The goal is to have a world in which all of the science literature is online, all of the science data is online, and they interoperate with each other.

Lots of new tools are needed to make this happen."

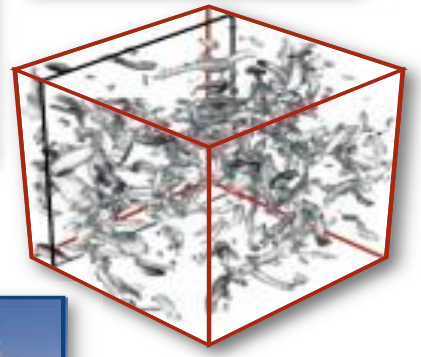
- Jim Gray, Microsoft Research

Science Paradigms

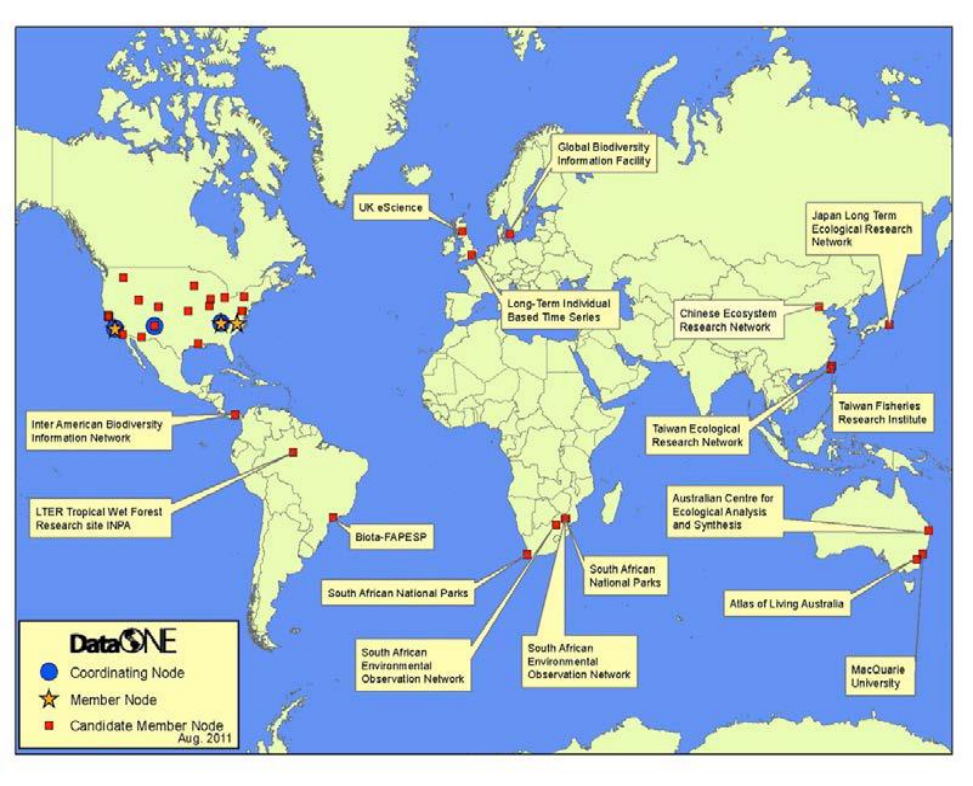
- Thousand years ago:
  science was **empirical**
  describing natural phenomena
- Last few hundred years:
  theoretical branch
  using models, generalizations
- Last few decades:
  a **computational** branch
  simulating complex phenomena
- Today: **data exploration** (eScience)
  unify theory, experiment, and simulation
  - Data captured by instruments
    or generated by simulator
  - Processed by software
  - Information/knowledge stored in computer
  - Scientist analyzes database/files
    using data management and statistics



$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G p}{3} - \mathrm{K}\frac{c^2}{a^2}$$



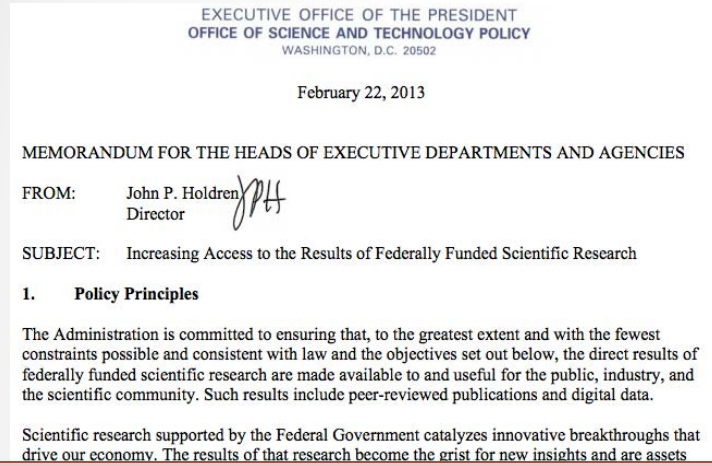
* Image from "*The Fourth Paradigm: Data Intensive Scientific Discovery*", Microsoft Research, 2009



3. Scientists and issues are geographically spread.

4. Open Data/Science Mandates

…governments and funding agencies are requiring data accessibility and encouraging data intensive use…



EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Holdren
Director

SUBJECT: Increasing Access to the Results of Federally Funded Scientific Research

1. Policy Principles

The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.

Scientific research supported by the Federal Government catalyzes innovative breakthroughs that drive our economy. The results of that research become the grist for new insights and are assets

2-2013

OSTP Policy: "Increasing Access to the Results of Federally Funded Scientific Research" Requires a plan to support increased public access to the results of research (scholarly publications and science data) funded by the Federal Government

5-2013

OMB: "Open Data Policy—Managing Information as an Asset"

·May 9: WH Executive Order: "Making Open and Machine Readable the New Default for Government Information"

Why are data not reused? Real costs…

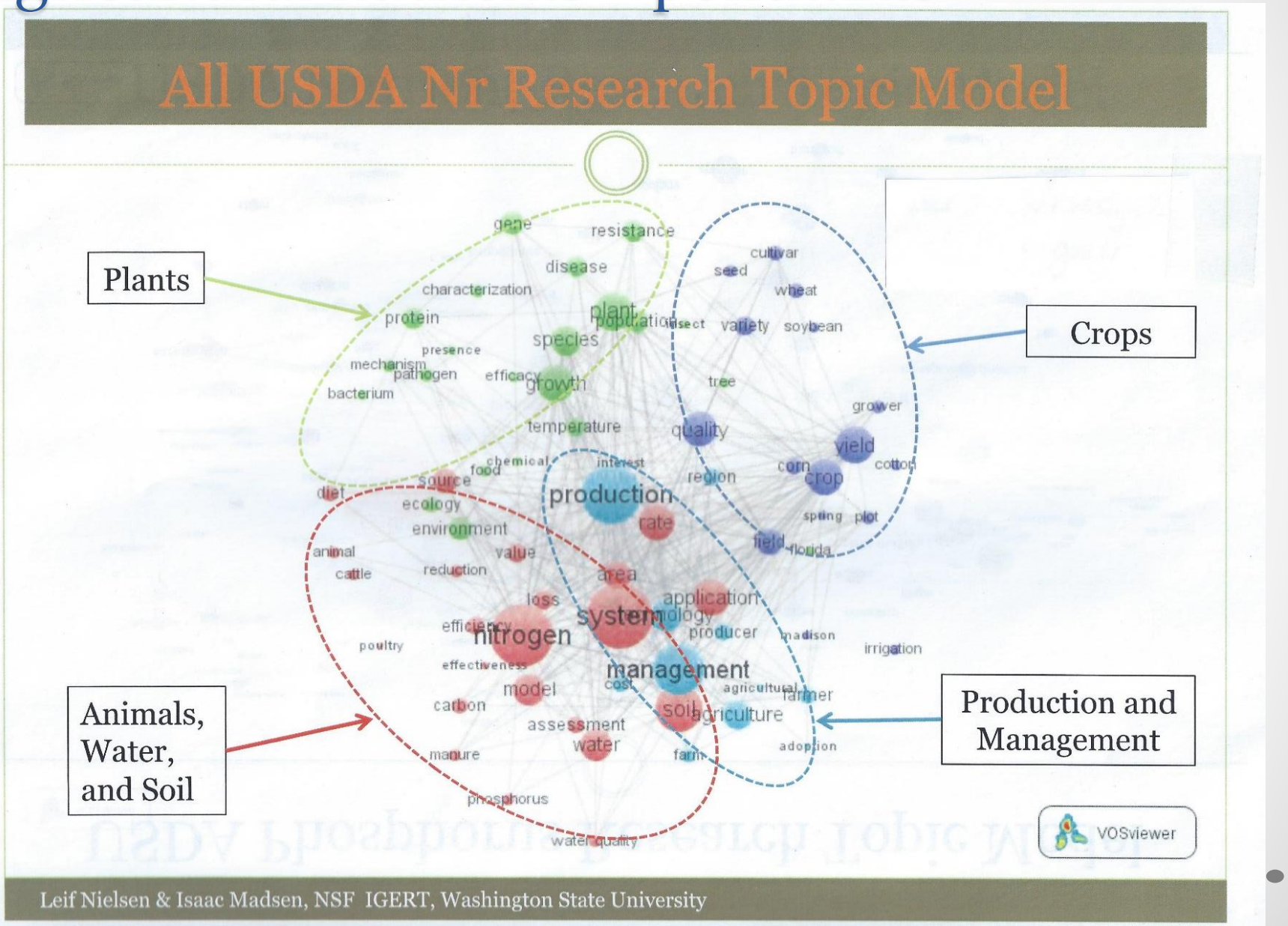
- **Too much work?** Lack of data workflow tools…
  - Diekmann interviews (J. Ag. & Food Info., 2012):

  "[Another group of scientists and I] were talking about, can we get our data and pull it together? They wanted that data, [but] ***it's the annotation that's really the hard part*** *[for] them [to be] able to make sense of it. I would be happy to give [out the data], but [then]* ***I have to explain whatever I did****."*
- **Too expensive?** > 80% of scientists surveyed in 2010 indicated that they did not have resources to make their data open access (Science. Feb. 2011)

Question of Money, Motivation, and Mechanics…

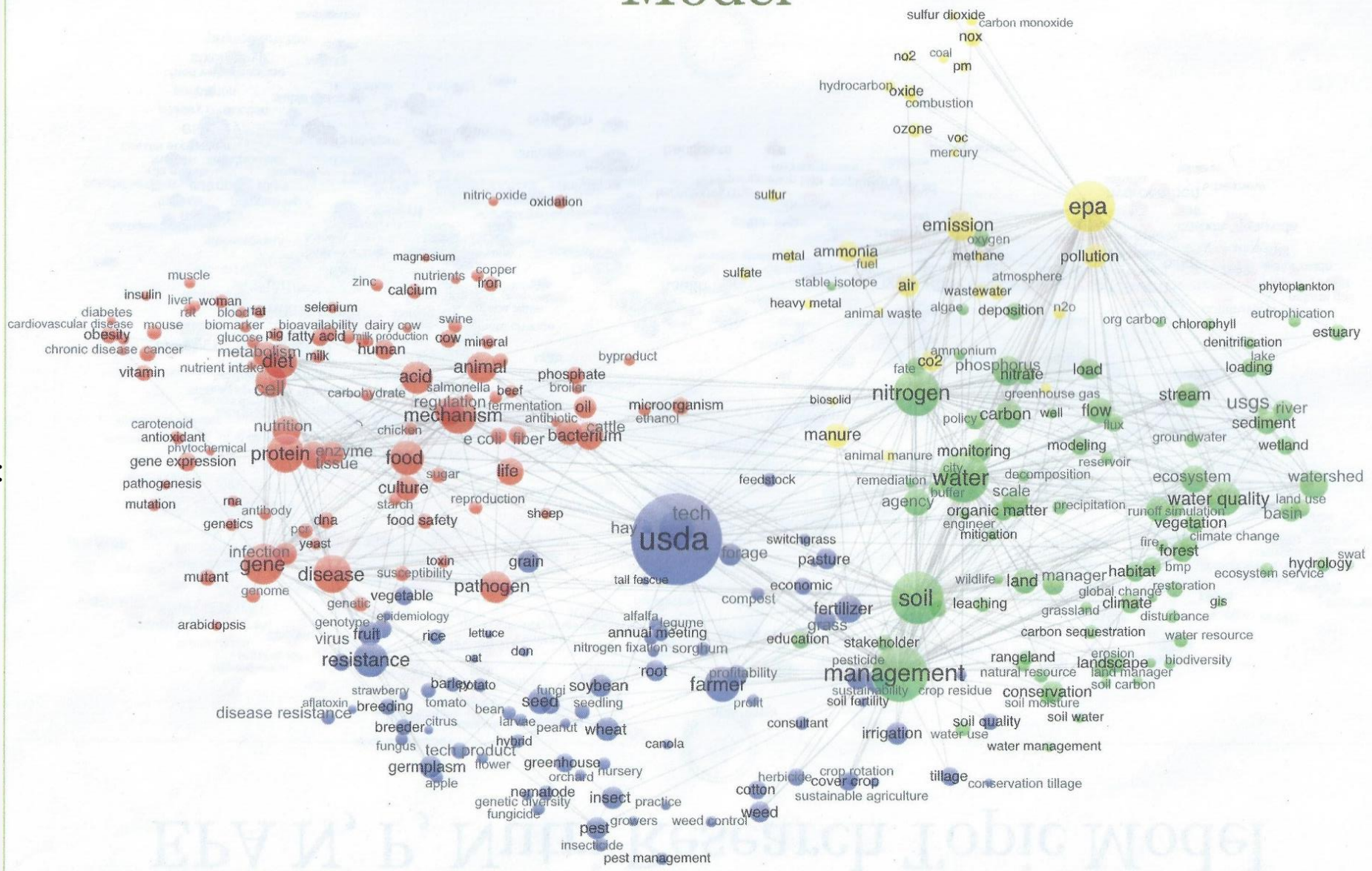
What do we know we know? Less than we could…
Agricultural nutrients = pollutants

Topic model of funded research shows USDA has invested a lot BUT what does it all mean?



EPA, USDA, USGS N, P, Nutri Research Topic Model

Investment across 3 Agencies significant BUT can it be translated into useful knowledge for: Management? Policy? Regulation?



Global Responsibility

Produce more food with fewer resources

- **Pilot commodity optimization program:** We collaborated with 15 large suppliers – representing 30% of our food and beverage sales in North America. By providing farmers with data and tools, they're able to develop plans to optimize fertilizer and tilling practices in corn and soy crop rotations. This saves money, reduces greenhouse gas (GHG) and, ultimately, delivers more sustainable products to our customers. The pilot commodity optimization program includes 2.5 million acres, with the potential to reduce GHG by 2.3 million metric tons (MMT).
- **Fertilizer optimization:** While our work on fertilizer optimization has been foundational, we are exploring opportunities to scale this and other pilots across food commodities. We are in the process of developing new relationships that could total 14 million acres, with the potential to reduce GHG by an estimated 7 MMT.

Hunger & Nutrition
Diversity & Inclusion

Leveraging our global scale

Related Links

WERF
Water Environment Research Foundation
Collaboration. Innovation. Results.

Learn about WERF | Search Research Publications & Tools

Home | About WERF | News | Kno

About WERF

The Water Environment Research F
research organization dedicated to v
have developed a portfolio of more t

We are a nonprofit organization that
Our subscribers include wastewater
equipment companies, engineers an
subscribers. WERF takes a progress
subscribers, environmental professic
experts.

NATIONAL CORN GROWERS ASSOCIATION

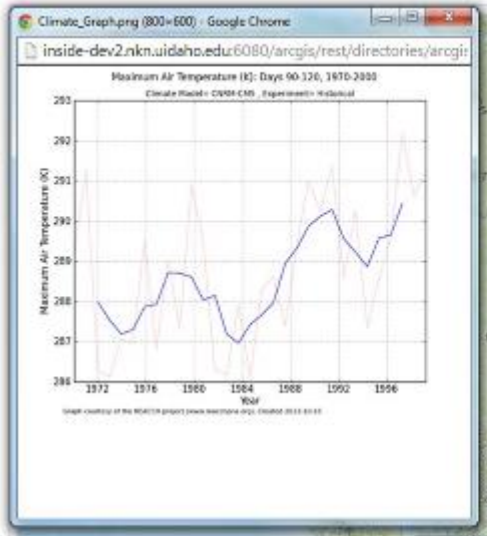
WERF
Water Environment Research Foundation
Collaboration. Innovation. Results.

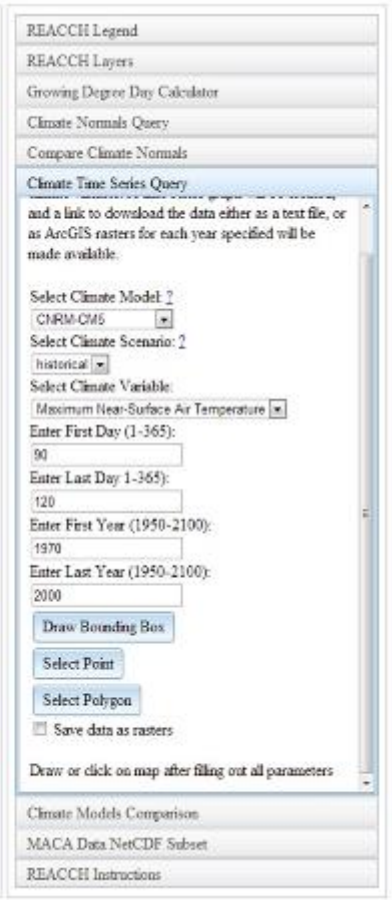
Development of a National Agricultural BMP Database

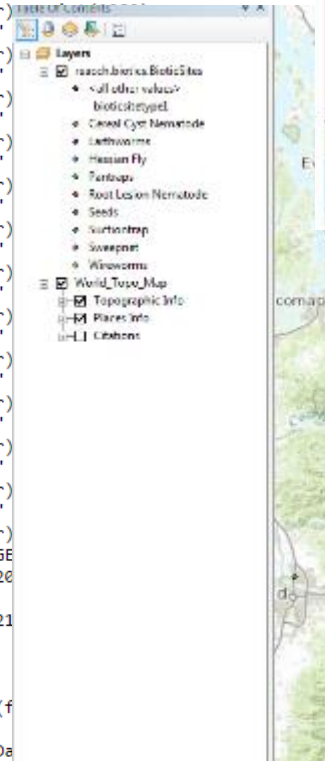
The Water Environment Research Foundation (WERF), the National Corn Growers Association (NCGA), and the Missouri Corn Growers Association (MCGA) have partnered to undertake the development of a national Agricultural Best Management Practices (BMP) Database. The purpose of the Agricultural BMP Database is to develop a centralized repository of agricultural BMP performance studies to provide scientifically-based information on practices that reduce pollutant loading from agricultural sites. The database will include performance data and meta data that document the many field-based and practice-based variables that affect BMP performance. The long-term goal of the project is to provide agricultural advisors, planners, consultants and producers with information that enables them to better select systems of BMPs for their operations and to support improvements in agricultural BMP design and implementation.

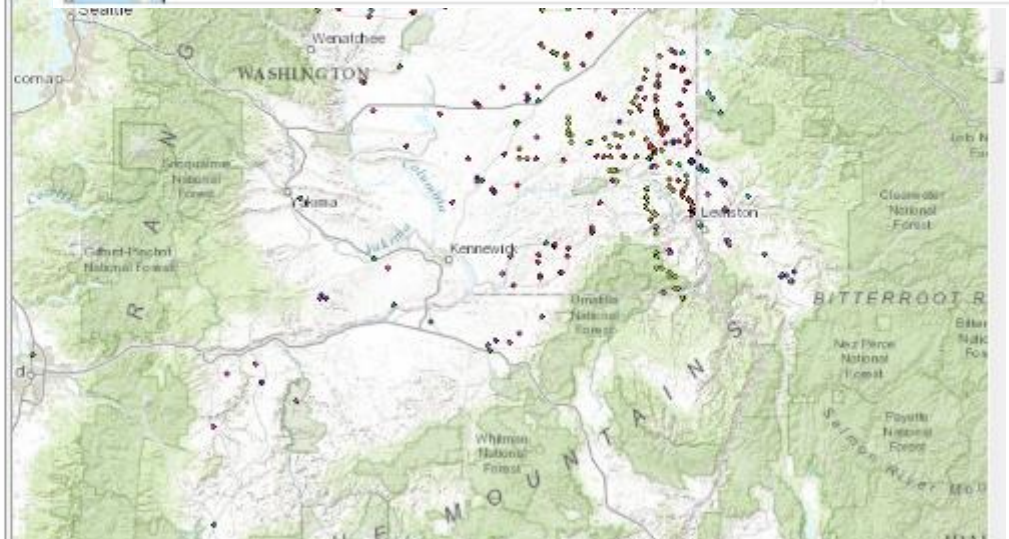
svn.nkn.uidaho.edu/svn/REACCH/trunk/reacch

```
Fimport arcpy
from pydap.client import open_url
import numpy as np
arcpy.env.overwriteOutput= True
from numpy import nonzero
import os
import sys
import shapefile
import pylab
from pylab import *
from netcdftime import utime
from matplotlib import pyplot as plt
import datetime as DT
from datetime import date
from datetime import timedelta
import numpy.ma as ma
import webbrowser


def AddmacaOneDay(polygon,GcmModel,GcmScenario,GcmVar,GcmYear
    if int(1950) <= int(GcmYear) <= int(1959):
        TimePeriod= "1950_1959"
    if int(1960) <= int(GcmYear) <= int(1969):
        TimePeriod= "1960_1969"
    if int(1970) <= int(GcmYear) <= int(1979):
        TimePeriod= "1970_1979"
    if int(1980) <= int(GcmYear)
        TimePeriod= "1980_1989"
    if int(1990) <= int(GcmYear)
        TimePeriod= "1990_1999"
    if int(2000) <= int(GcmYear)
        TimePeriod= "2000_2005"
    if int(2006) <= int(GcmYear)
        TimePeriod= "2006_2015"
    if int(2016) <= int(GcmYear)
        TimePeriod= "2016_2025"
    if int(2026) <= int(GcmYear)
        TimePeriod= "2026_2035"
    if int(2036) <= int(GcmYear)
        TimePeriod= "2036_2045"
    if int(2046) <= int(GcmYear)
        TimePeriod= "2046_2055"
    if int(2056) <= int(GcmYear)
        TimePeriod= "2056_2065"
    if int(2066) <= int(GcmYear)
        TimePeriod= "2066_2075"
    if int(2076) <= int(GcmYear)
        TimePeriod= "2076_2085"
    if int(2086) <= int(GcmYear)
        TimePeriod= "2086_2095"
    if int(2096) <= int(GcmYear)
        if GcmModel[0:6]=="HadGE
            TimePeriod= "2096_20
        else:
            TimePeriod= "2096_21

    fYear= TimePeriod[0:4]
    firstYear= int(fYear)
    numYears= int(GcmYear)-int(f
    extraDays= numYears*365
    Day= int(GcmDay)+int(extraDa
```









REACCH Data Access Tier 1 and 2 REACCH members

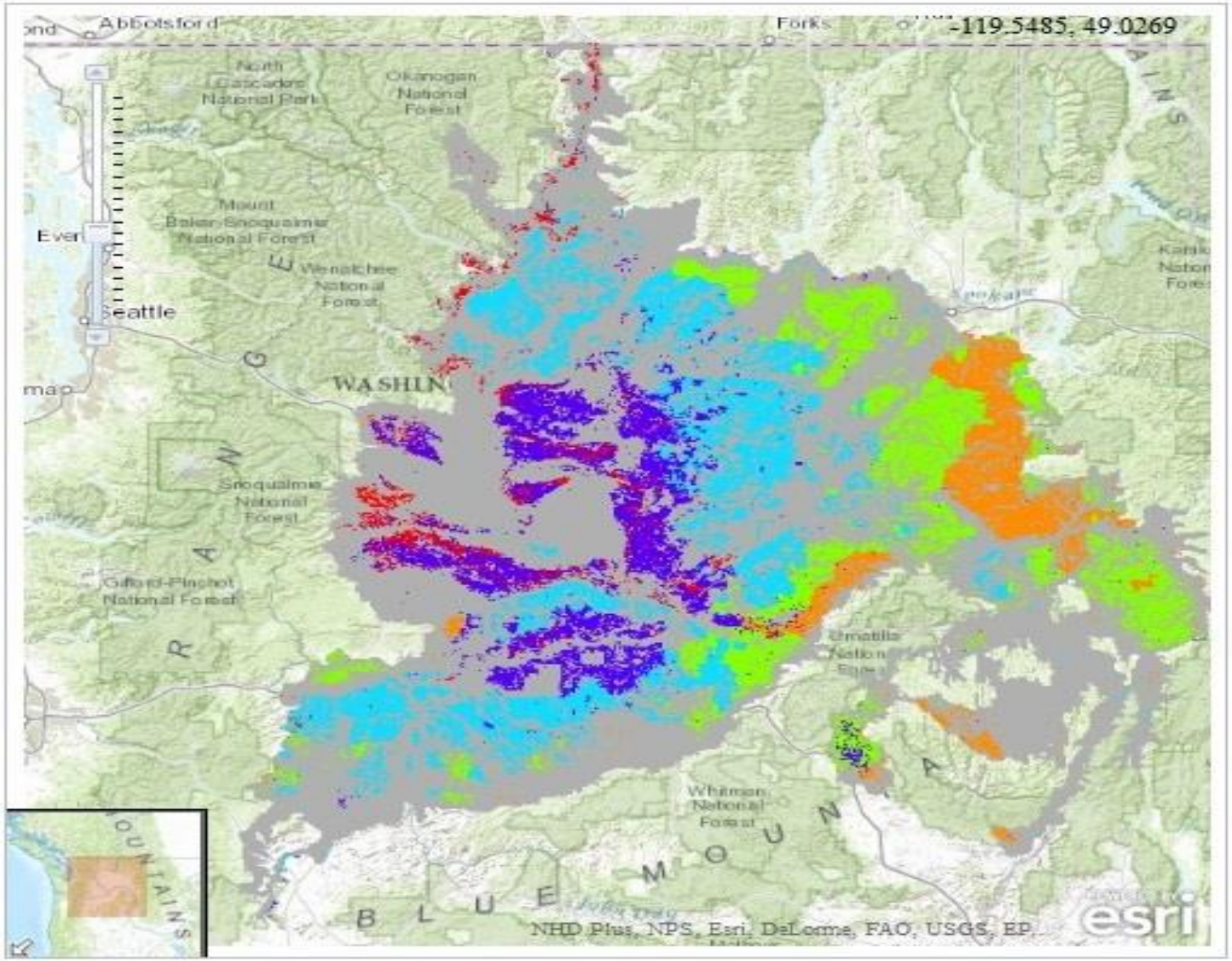
REACCH
Regional Approaches to Climate Change – PACIFIC NORTHWEST AGRICULTURE

REACCH Climate Data Viewer

Version 2.0: All users need to authenticate to the REACCH Extranet

Climate Data Viewer Functions: MACA Data NetCDF Subset
Climate Normals Query
Climate Time Series Query

Contact REACCH
How do I?
REACCH Analysis Library
REACCH Data Library
REACCH Biotics Data Viewer
REACCH Social Factors Data Viewer



REACCH Legend

reacch.models.AEZ_2011

- Background
- Annual Cropping Area
- Intermediate Cropping Area
- Grain/Fallow
- Irrigated
- Orchards

REACCH Layers

Growing Degree Day Calculator

Climate Normals Query

Compare Climate Normals

Climate Time Series Query

Climate Models Comparison

MACA Data NetCDF Subset

REACCH Instructions

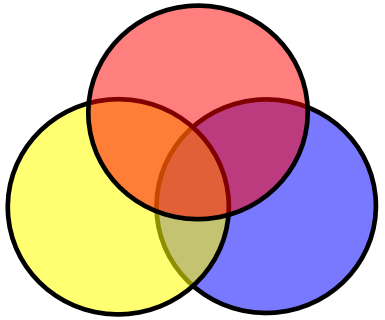
Climate Science

- Climate scientists from three Universities
- Multiple gridded downscaled climate scenarios for several hundred years for the entire US
- Code to perform dynamic, data-intensive analyses across multiple data sources
- Publish resulting dataset/metadata back to home base

Dickersin: Knowledge translation: From clinical research to practice decisions

The Medical Model

Evidence generation → Evidence Synthesis (systematic reviews) → Clinical policy (guidelines) → Evidence-based healthcare



Clinical trials, observational studies

Cochrane Collaboration, others

Professional Societies, others

<u>Application of policy:</u>
Evidence
Clinician expertise
Patient values

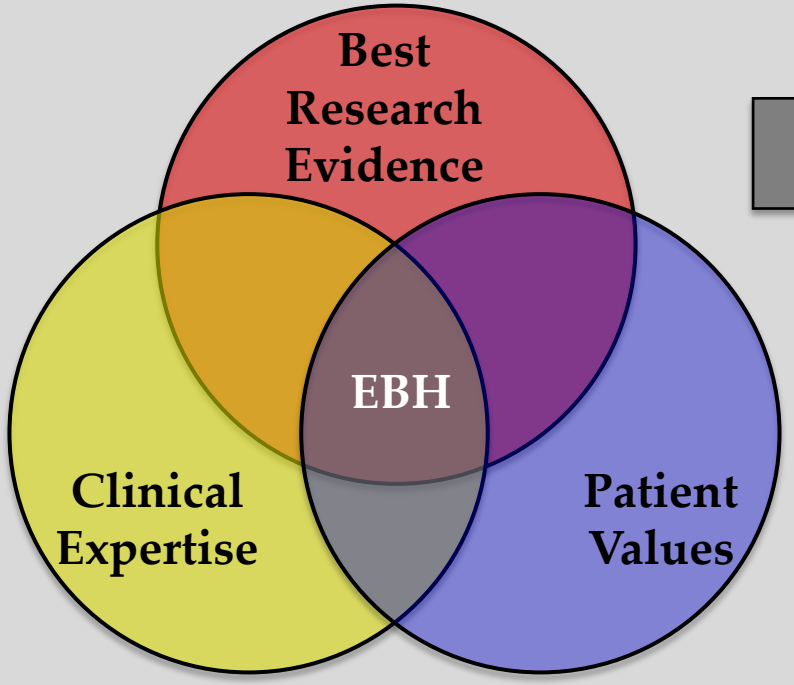
Knowledge translation

US government has 1.3 billion $$$ stockpile… Reduces symptoms by 17 hours (7 to 6.3 d), no effect on mortality

http://www.cbsnews.com/videos/tamiflu-may-have-little-effect-in-pandemic-study-says/

Evidence-Based Healthcare

"The integration of best research evidence with clinical expertise and patient values"

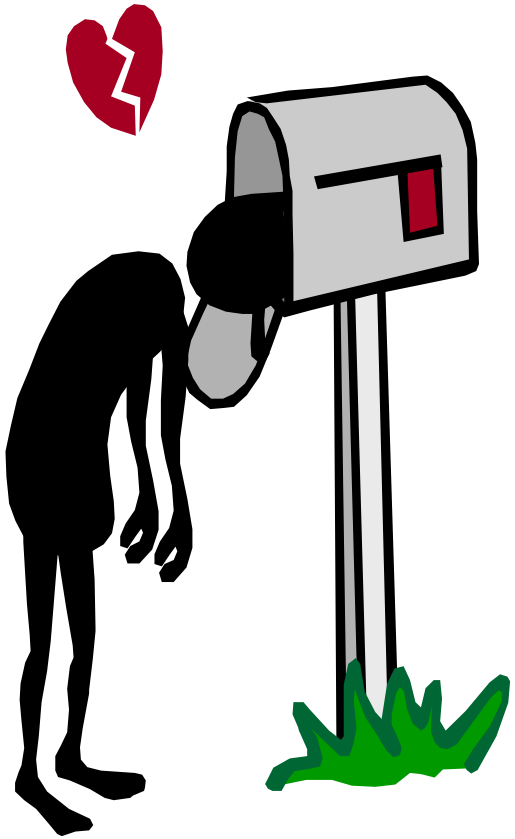


Sackett, 2000. Referenced in Dickersin, K. and M. Mayer. 2012. Understanding evidence-based healthcare: A foundation for action US Cochrane Center. Available online at http://us.cochrane.org/understanding-evidence-based-healthcare-foundation-action

Evidence-Based Agriculture

"The integration of best research evidence with management expertise and stakeholder priorities?"

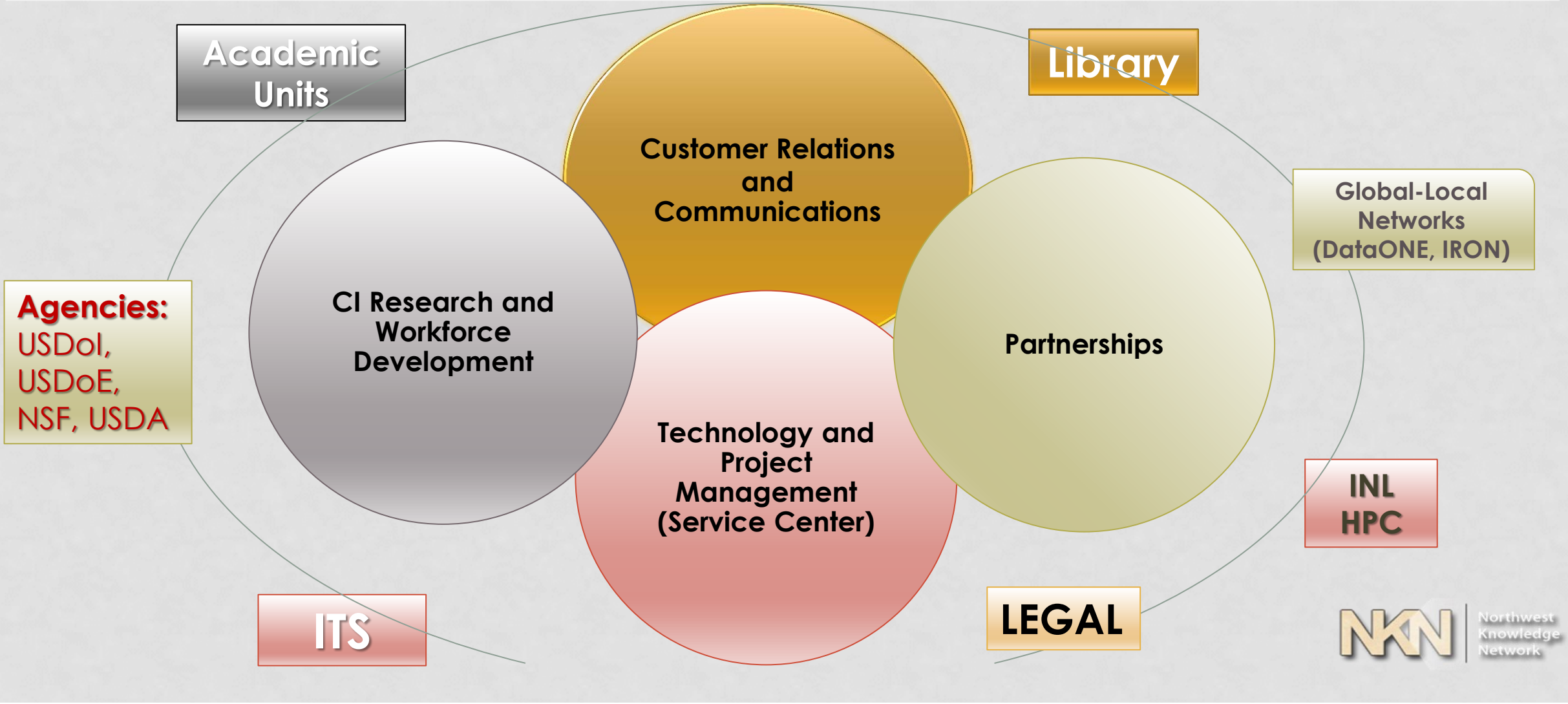
T. Scott Murrell, IPNI



I'm Lonely and Unsure

Who Else is Doing This That I Need to Connect with at My Campus???

BIG/OPEN DATA NETWORKS AND TEAMS



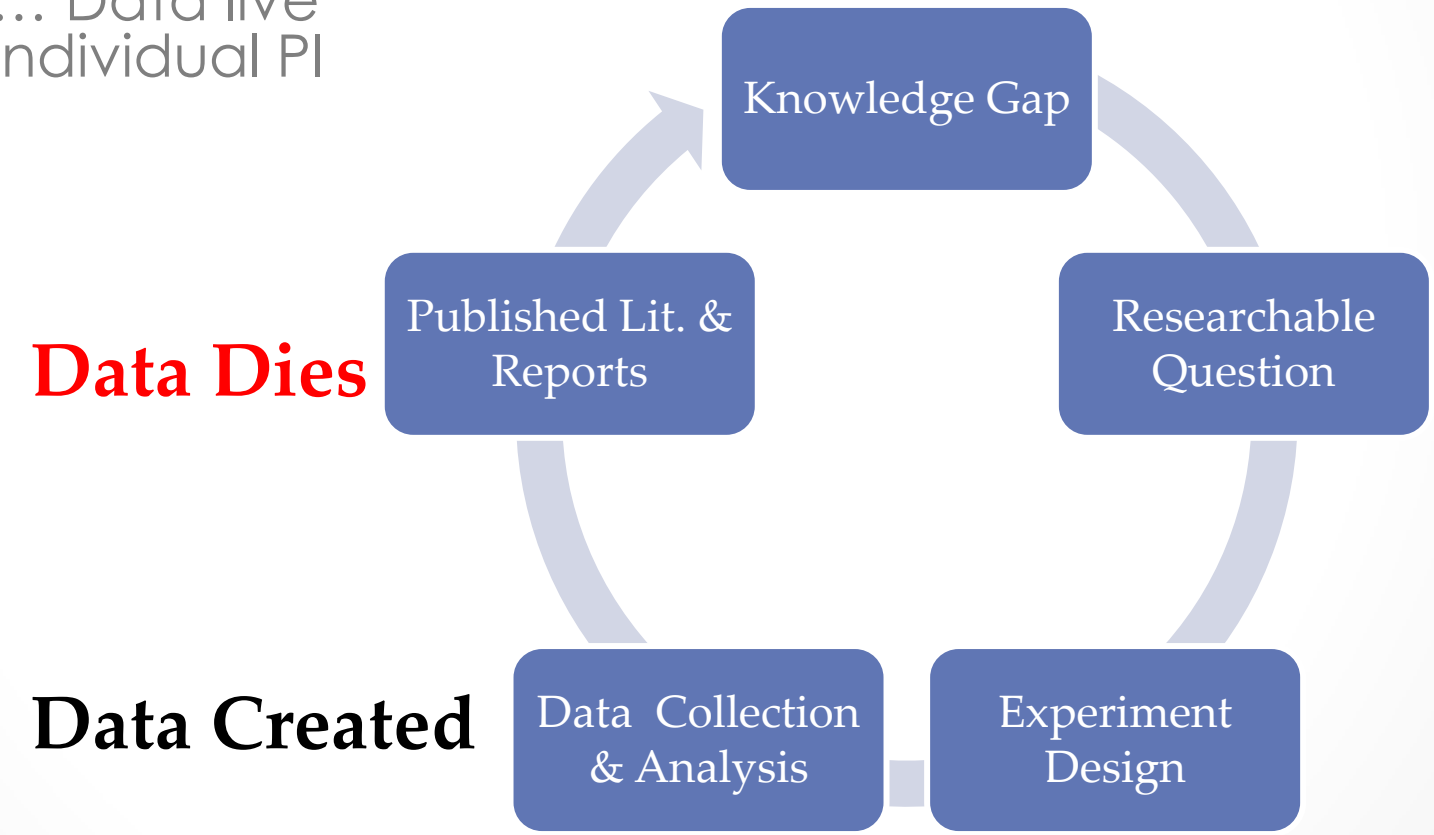
Enabling Data-Intensive Activity Dictates the Cooperators

- Quality data and metadata throughout lifecycle
- Data management policies
- Data/Pub cataloging, serving, application tool services
- Centralized IT, access to HPC, pipelines
- Research; Interoperability (TEK-BioP-Social) and Virtualization
- Workforce development; domain and software

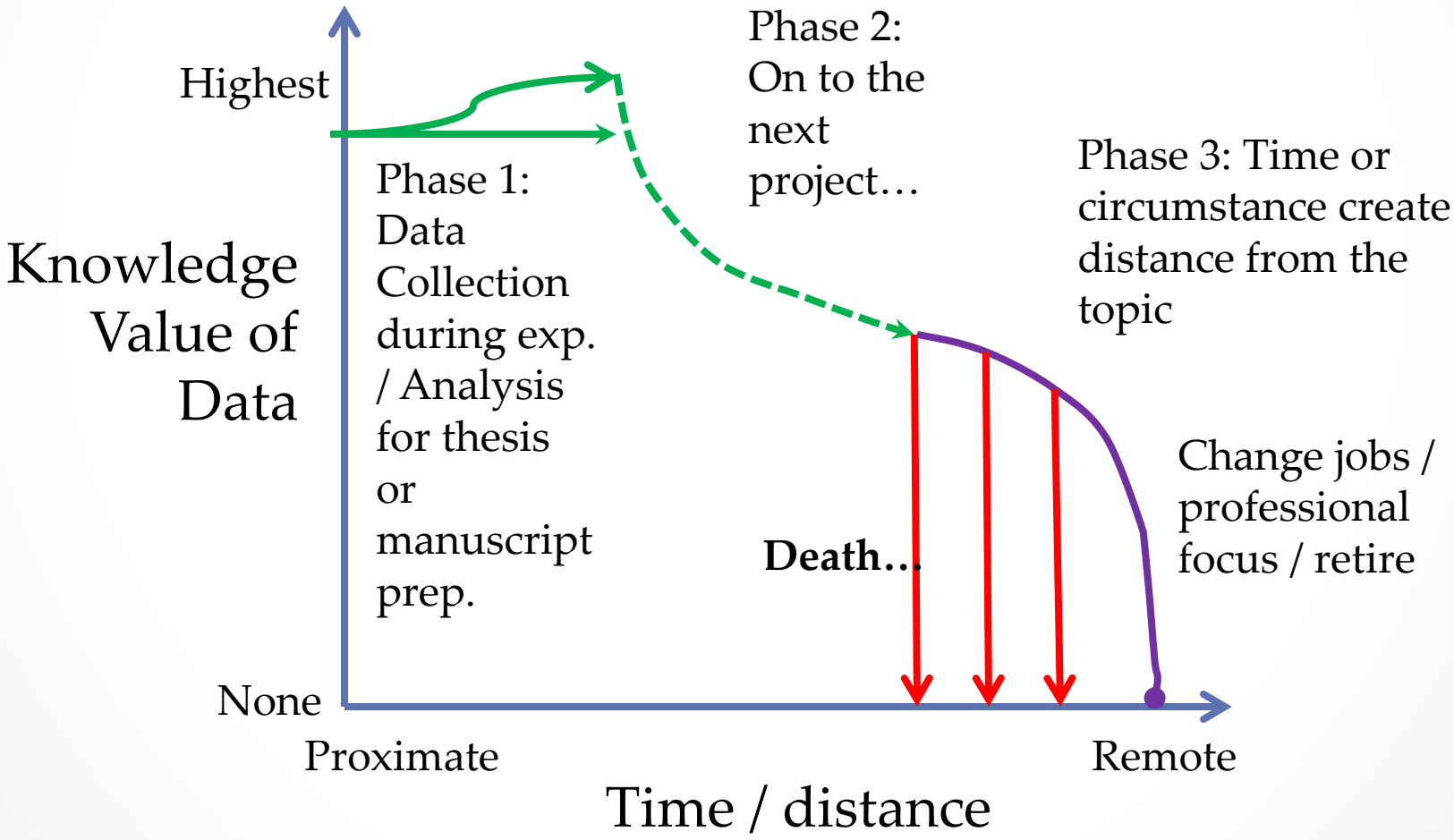
= RESEARCH – LIBRARY – ITS – ACADEMICS –GCOUNSEL
REGIONAL-GLOBAL NETWORKS

Culture of short data "lifecycles" in agronomic research…

Business as usual in Agronomy / Applied Research… Data live and die within an individual PI lab

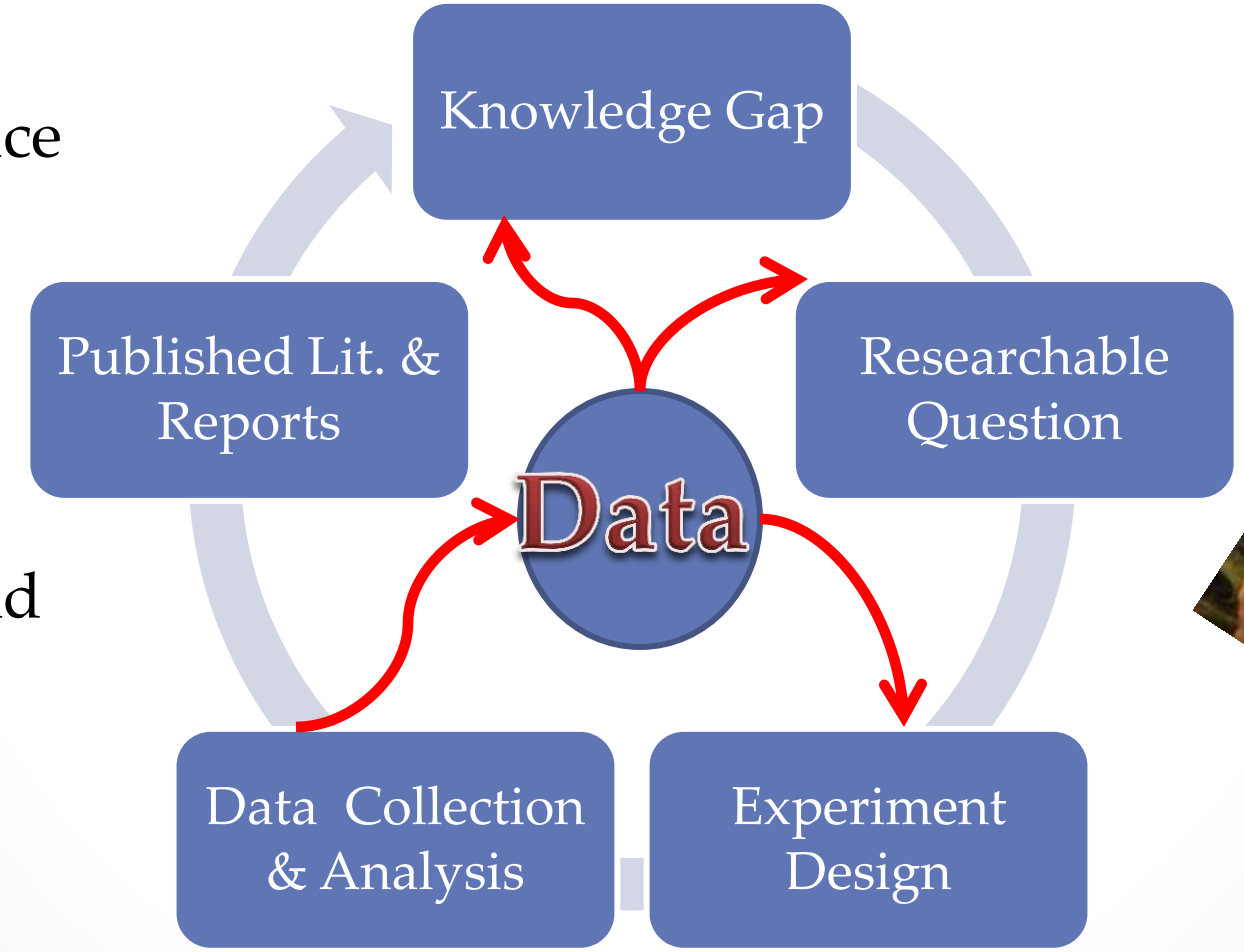


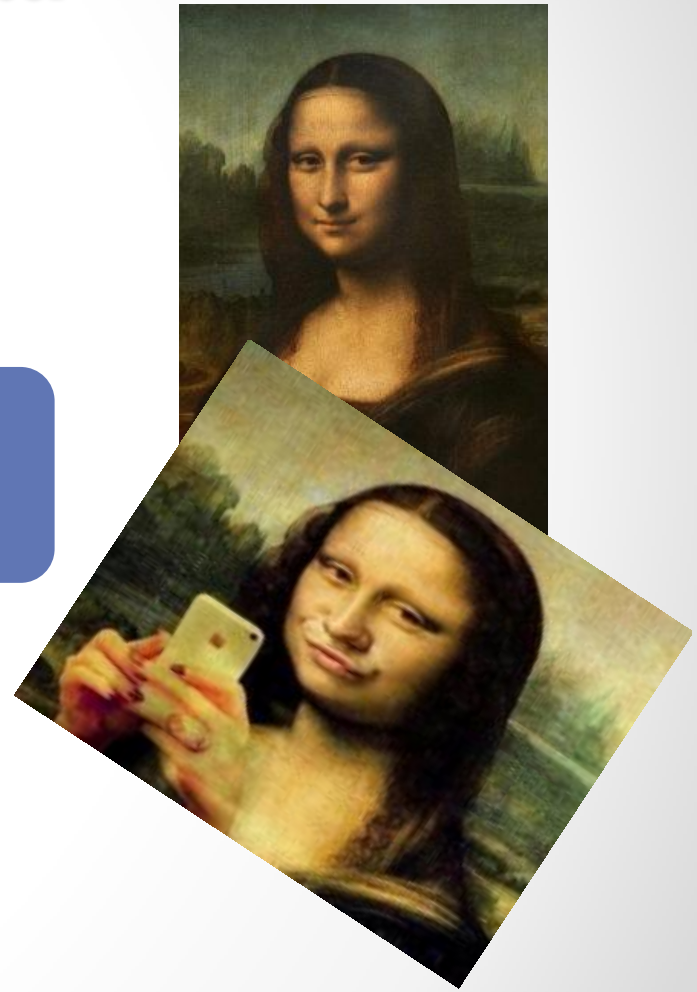
Precarious Nature of Typical Ag. Data Lifecycle: Scientifically proven that my ability to understand and find these data will erode extremely rapidly!



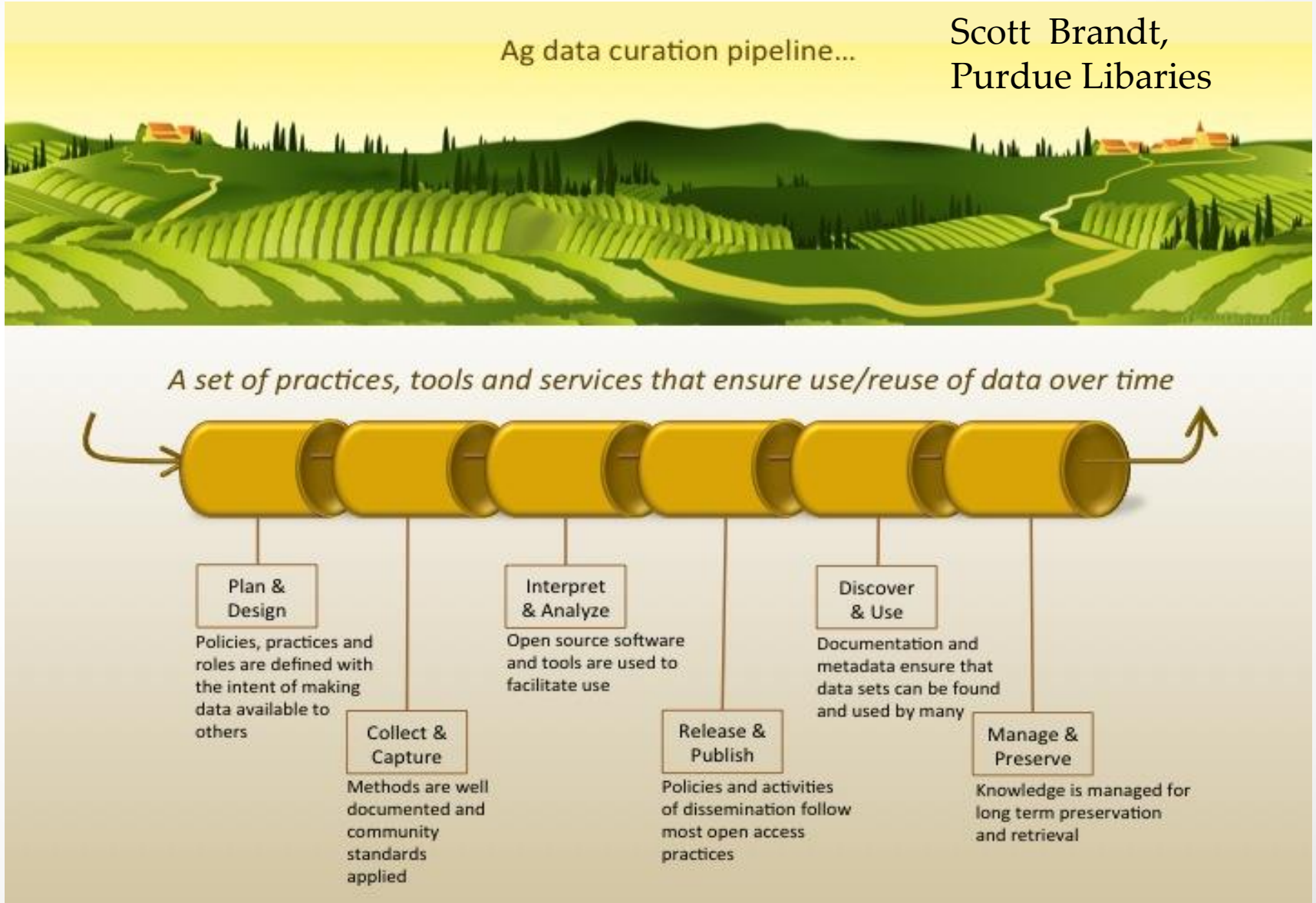
Applied research model with a longer data lifecycle … More "hands" on the data

Need someplace to put data w/ sufficient workflows & policies to ensure correct recognition and reuse





Why start w/ Libraries: Know how to organize & store so something can be discovered / accessed / used . They have the desired attributes for a data “destination”…



Purdue University Research Repository: What libraries are to books, PURR is to data (plus so much more!)

2010 to 2014

NORTHWEST KNOWLEDGE NETWORK

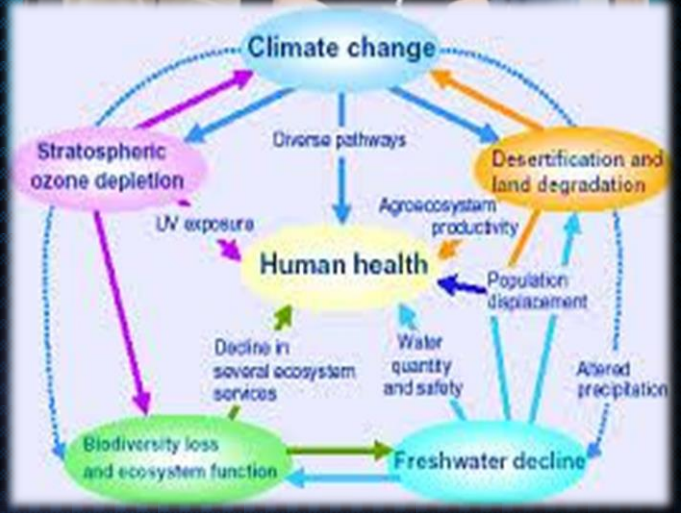
UNIVERSITY OF IDAHO AND COOPERATORS

www.northwestknowledge.net

NKN Mission

Enable research teams to address complex societal problems by facilitating quality metadata, and the storage, discovery and dynamic analysis of data as long term , dependable assets.

Advance research and education in support of data intensive science.



User Portal
Find
Upload
Collaborate

An Integrated Cyberinfrastructure Framework

Beneficiaries

- New CI Research — Visualization and Data Interoperability
- Education — Graduate Education and Research, REUs
- Stakeholders Decision Makers Public
- Federal, State, Private Industry Partners

Science

- Data Collection — Sensors, remote sensing, SES data
- Web Portal — Community Interface
- CAVE — Interactive Visualization and Decision Making
- Workstations — (at any Campus or INL) Visualization and Analytics
- Mobile — on-site scenarios and augmented reality
- Integrated Interactive Visualization Platform
- Modeling — Integrated Modeling and Simulation

Core Infrastructure

- High Performance Computing (HPC) — (local institutional clusters, INL, XSEDE)
- Long Term Data Life Cycle Management — Northwest Knowledge Network, DataONE, NEON, Others — Secure Persistent Storage, Data/Metadata Curation, Discovery, Access, Sharing, Data Mining
- Short Term Data Storage (Scratch Space) — Dropbox Teams, NKN iFolders, Others
- Idaho Regional Optical Network (IRON) — High Speed Optical Network, GigaPOP Access and Connectivity, Enabling Large-Scale Research

NKN Northwest Knowledge Network

Northwest Knowledge Network

- Lifecycle management for heterogeneous research data
  - Tiered, distributed data storage
  - Metadata Tools, Standards
  - Data discovery and retrieval
  - Data-centric researcher collaboration tools
  - Interoperability across scale, time, data discipline **(incl TEK, Social)**

- NKN Big Data Functions
  - ✓ Capture
  - ✓ Storage
  - ✓ Curation
  - ✓ Search
  - ✓ Sharing
  - Analysis
  - Visualization

- Collaborative regional data partnerships (NIFA USDA, NW CSC USDoI, INL USDoE, EPSCoR NSF, NW Climate Hub USDA, NW Forest Fire Science Center and Sustainable NW Dairies Center.

- Network of resources, services, and expertise
  - Policies, protocols, standards in support of effective data/metadata;
  - Systems admin, software development for data-intensive science;
  - Stable and enduring storage and access to data and metadata;
  - Hosting of virtual machines, applications, websites databases; and
  - Consulting/technical services for data and metadata management.
  - NSF DataONE, access to HPC and national high-speed data networks

The Next Phase: Online Data Observatory

- Enable investigators to visualize and intercompare heterogeneous datasets without struggling with file formats, unit conversion, subsetting, scales
- New research with existing data
- Important Components
  - Data representation/interoperability
  - New tools
  - Web service APIs

Case Study in Regional Data Management

- Startup venture; partner/institutional funding.
- Learned critical-minimal level of staffing hardware and software to sustain core services.
- Demand for services exceeding capacity. Venture ending.
- Established Service Center.
- Need dependable revenue flow for data services, and more sophisticated partnership between universities and federal government, on behalf of the PI's (the triangle of value propositions).

Seeking a Sustainable Fiscal Model

Northwest Knowledge Network FY2013 through FY2020 Budget Plan

	FY13	FY14	FY15	FY16	FY17	FY18	FY19	FY20
Revenue								
VP Funds	102,413	133,609	242,460	-	-	-	-	-
PI F&A Return	1,728	23,069	26,500	15,000	5,000	15,000	15,000	15,000
UI Central FA	-	-	-	-	-	-	-	-
Service Center	-	-	100,000	156,000	179,400	206,310	237,257	272,846
EPSCoR	39,720	78,727	172,631	164,154	84,284	-	-	-
USGS Grants	215,868	301,521	212,861	42,558	-	-	-	-
Misc Grants/Dept Funds	61,628	63,915	63,742	7,191	7,188	7,332	7,478	7,628
New Grant Funds	-	-	-	-	200,000	200,000	200,000	200,000
New Equipment Funding	-	-	-	-	-	100,000	100,000	100,000
Revenue Total	421,357	600,841	818,195	384,903	475,872	528,642	559,735	595,473
	-	-	-	-	-	-	-	-
Expense								
Payroll	402,116	563,468	798,734	785,220	788,737	804,328	820,229	836,450
Operating	19,058	18,861	46,017	38,517	38,517	38,517	38,120	38,517
Computer Equipment	-	89,203	45,000	40,000	40,000	140,000	180,000	180,000
Office Furniture/Equip	183	120	20,000	-	-	-	-	-
Expense Total	**421,357**	**671,651**	**909,751**	**863,737**	**867,254**	**982,845**	**1,038,349**	**1,054,967**
	-	-	-	-	-	-	-	-
Net FY Balance	-	(70,811)	(91,556)	**(478,834)**	**(391,383)**	**(454,204)**	**(478,614)**	**(459,494)**

Seeking a Sustainable Business Model via University-Agency Cooperation

Need activity-interaction on all three sides of a the value triangle; Federal agencies, PIs and universities must relate to each other.

- Agency require PI's to do DM planning; specific actions, costs, reporting;
- Agencies/Universities require PI's to dedicate direct costs for DM;
- Universities provide PI's with essential DM services or referrals;
- Universities/agencies convene national workshop on joint sustainable data management; cooperate on priorities, policies, protocols, costs.

USDA NAREEE Big/Open Data and Science

- USDA provide NAREEE with copy of USDA (OSTP) Open Data Plan
- USDA/NAREEE expand stakeholder involvement process, beyond scoping of individual REEE agencies
- ERS provide guidance on implementation of Open Data process
- USDA expand interagency collaboration on key topics like climate
- USDA place NAREEE representative on the OSTP Open Data Council

USDA NAREEE Big/Open Data and Science

- USDA provide glossary of terms, more definition(s) about what is required, preferred
- USDA provide basics on the value, best practices, benefits of managing Open/Big Data
- USDA gather input from universities re: their capacity for providing Open/Big data services
- USDA engage Capacity programs as a special case; get input from leaders
- USDA incentivize researcher for data preparation (offering scrubbing and other services)
- USDA provide guidance to universities on how Open and Big Data mandates will be enforced
- ARS conduct joint planning exercises with land grant universities leading data management
- USDA RFPs explicitly require data management activity and hold accountable
- USDA RFPs instruct PIs to include data management expenses in direct costs
- USDA work with smaller/medium sized universities to minimize negative economies of scale

The Case of Capacity Programs

Should Open Data mandate apply to Hatch, Smith-Lever, McIntire-Stennis, Evans-Allen, Animal Health, Renewable Resources (RREA), 1890, and Tribal?

A. $.5 billion in applied, regional and demonstration research programs and their data may be important;

B. Could be cumbersome, questionably effective and time consuming for data to be organized and called for from this community.

Capacity leaders need to provide input on whether to be included, and if so, how would they help design an approach that will work.

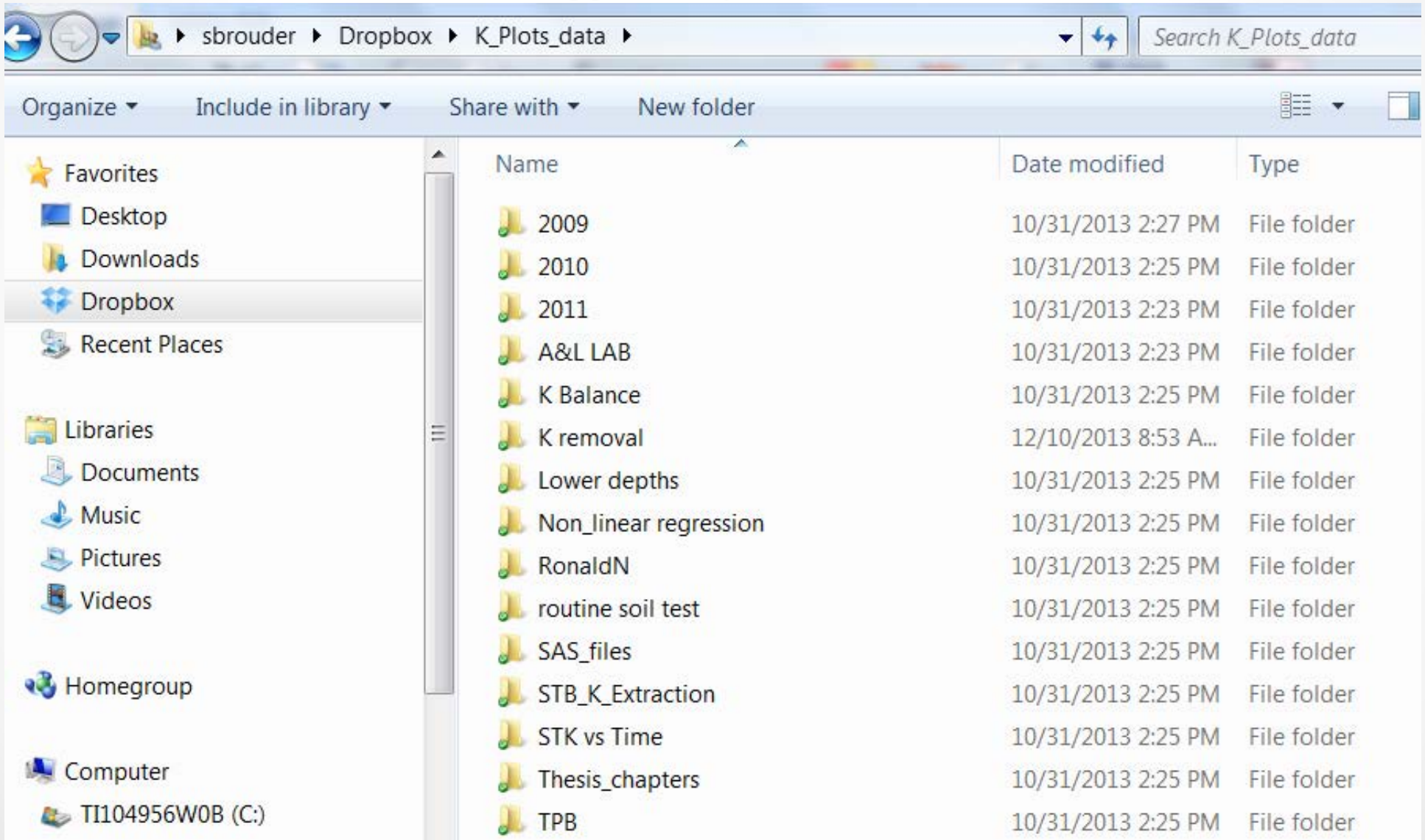
What is "big data" (vs "conventional")?

Ward & Barker (arXiv:1309.5821v1 [cs.DB] 20 Sep 2013)

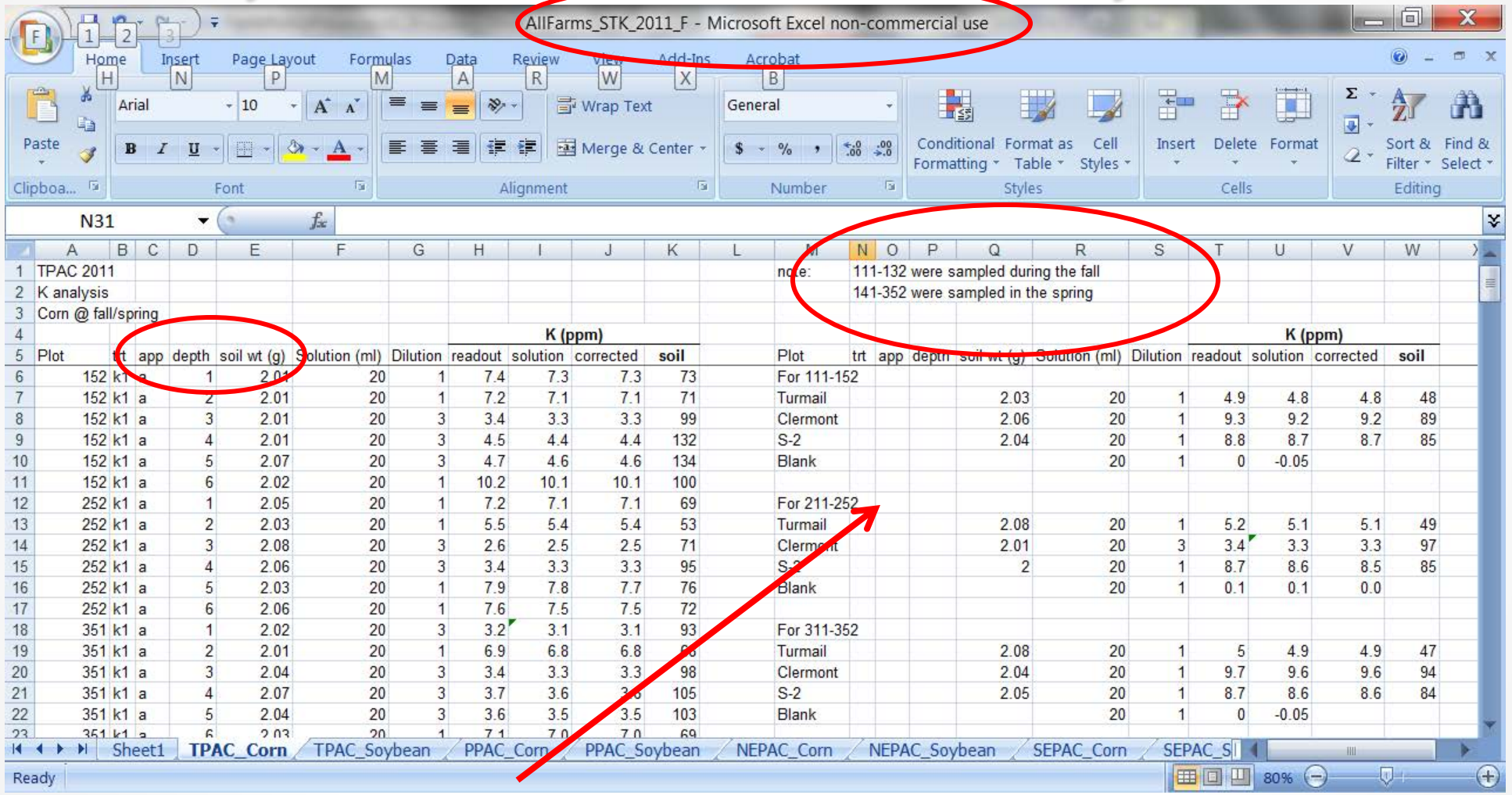
- Anecdotally: associated w/ data storage & analysis
- Gartner (2001): 3Vs ~ **Volume**, **velocity, variety**; (2012) **Veracity**
- Others: **Oracle** (structured w/ unstructured (e.g. social media)); **Intel** (generation of 300+ terabytes weekly); **Microsoft** (machine learning & artificial intelligence)
- Authors' conclusions: Size, complexity, technologies to process sizeable/complex
- **SB conclusions: 3Cs ~ Stuff that is cumbersome, costly (time, storage, whatever) & confusing to deal with.**

Yesterday's "big" is today's "conventional" ~ once we figure it out, it isn't big anymore… (Sonka, 2014 agrees w/ me on big data for ag.)

Status Quo: Taking a peek at data caretaking in AGRY… K Team Fellow (PhD student supported by Mosaic and PCS)



Today, I can tell you what this spreadsheet means but you can't understand all of it on your own…



What is this???

Tomorrow, we may both be in the dark…

12 Core Data Competencies for Data Information Literacy (Carlson et al., 2011, Libraries and the Academy, 11(2), pp. 629-657)

- Introduction to databases & data formats
- Discovery & acquisition
- Data management & organization
- Data conversion & interoperability
- Quality Assurance
- Metadata
- Data curation & reuse
- Cultures of practice
- Data preservation
- Data analysis
- Data Visualization
- Ethics including citation of data

Blending different ag data streams at different ed. levels requires new skills & DIL curricula **("Library Sciences should be solicited to educate all…")**

Future farmer or ag. industry employee (BS level)

- Everyone needs environmental info. mgmt that teaches how data are produced/used ("data in my life")
- Array of educational trajectories are needed from most basic level to specific endpts.
- Future farm managers need data skills in context of business mgmt & systems analyses
- Be able to understand data from outside their degree & be able to ascertain data quality

Future consultant, CCA, policy maker, Agent, Ext. Specialist (MS, PhD level)

- Understand exp. design, statistics & probability (risk)
- Understand geospatial data
- Curricula should use open-source software & "workforce-available" statistical tools
- Be able to translate science into lay language w/ context
- CCA: Certificate in Ext. Prgm should cover 12 data competencies
- Capstone data experience
- Ext. Spec. competent in Systematic Reviews; data mgmt plans / repositories part of degree

Extension Delivery and Application

- Help producers, managers and policy makers with the application of data to scenario building, modeling, visualization….
- Pursue cooperative arrangements between industry, producers and universities on the collection, storage, access, use of data.
- Pursue RFPs with integrated Extension and Research in the data-intensive context.

Why are data not reused (FHF (Faculty Hrmph Factor))?

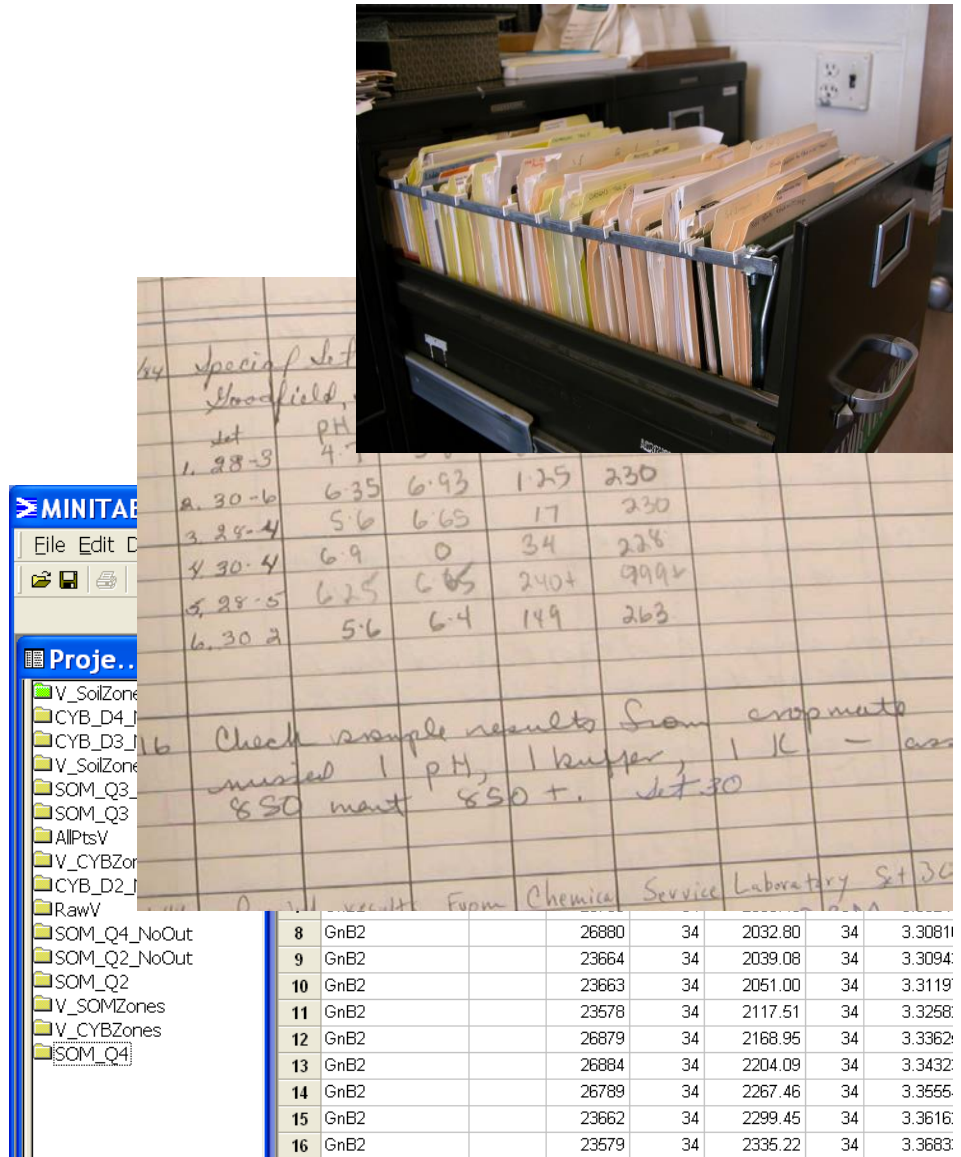
- **Not useful?** Question has changed… Hmmm: Yes & No
- **Not accessible?** Poor data hygiene…
  - Diekmann interviews (J. Ag. & Food Info., 2012):

*"The researcher wanted to reanalyze data from another figure and **I couldn't find it**. And I couldn't; **I lost it**. It was done on an old computer system and the **technician who did [it, had] moved on** and I wasn't able to find it."*

*"We have had a lot of problems in the past of losing data, or **just misplacing it**. And then we have to backtrack it and that's taken literally days or weeks to find where this data was stored. So it has been a real problem for us."*

Pressing technological challenges to informatics for all agronomic efforts concern data workflow…

- **Data dispersion**
  - Take advantage of small datasets collected by many researchers (not everything is “BIG”)
- **Data heterogeneity**
  - Varied protocols reflecting local culture & variation in 1º purpose
- **Data provenance**
  - Need to track data through multi-step process of aggregation, modeling, analysis



ODE (Opportunities for Data Exchange) 2012 D6.1 Summary of the Studies, Thematic Publications & Recommendations ~

Manifestation of data can take 5 different forms…

Past 10 Yr to today

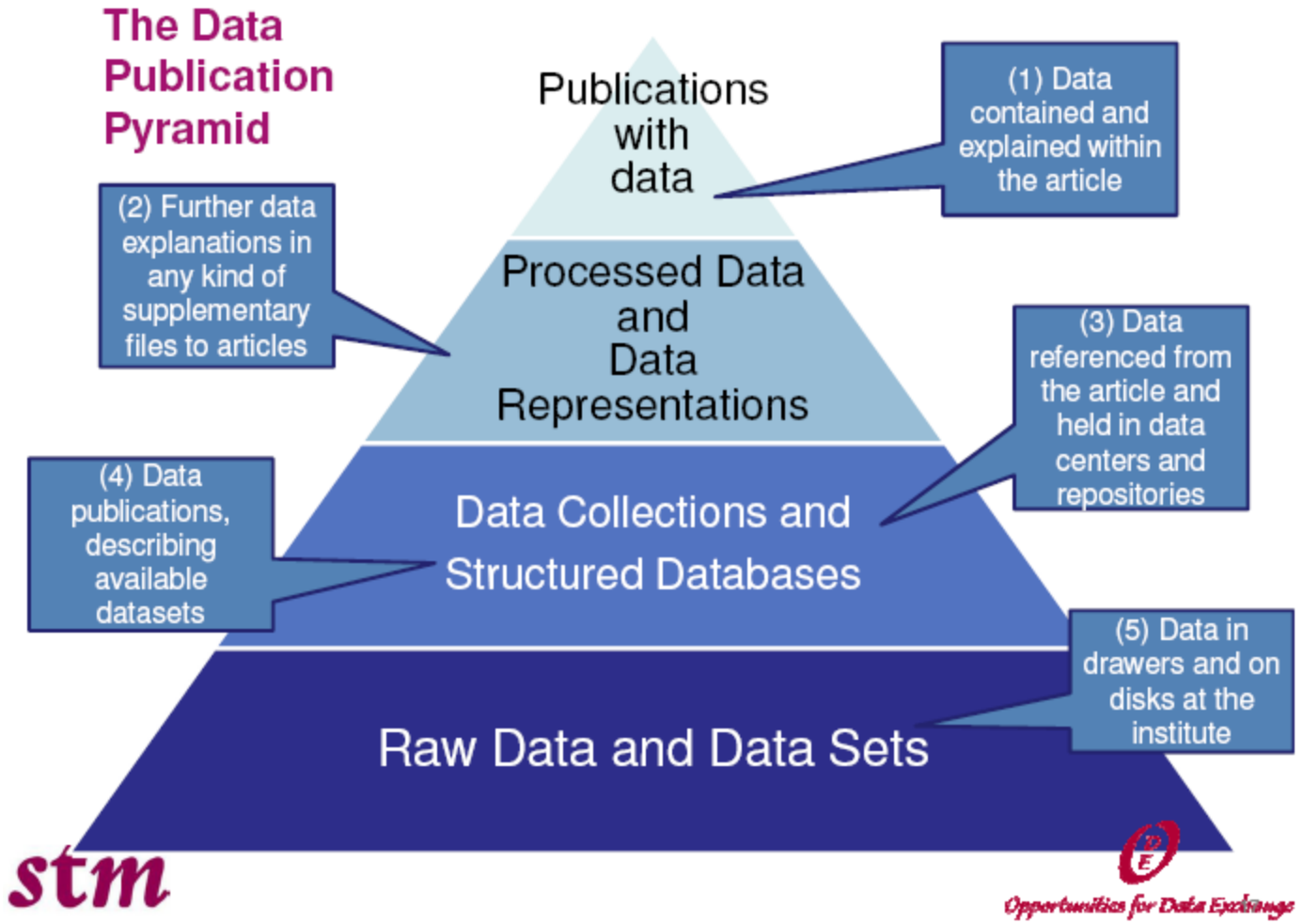


Illustration 1: Data Publication Pyramid

Stm=science, technical, medical publishing

http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/11/ODE-WP6-DEL-0001-1_0.pdf

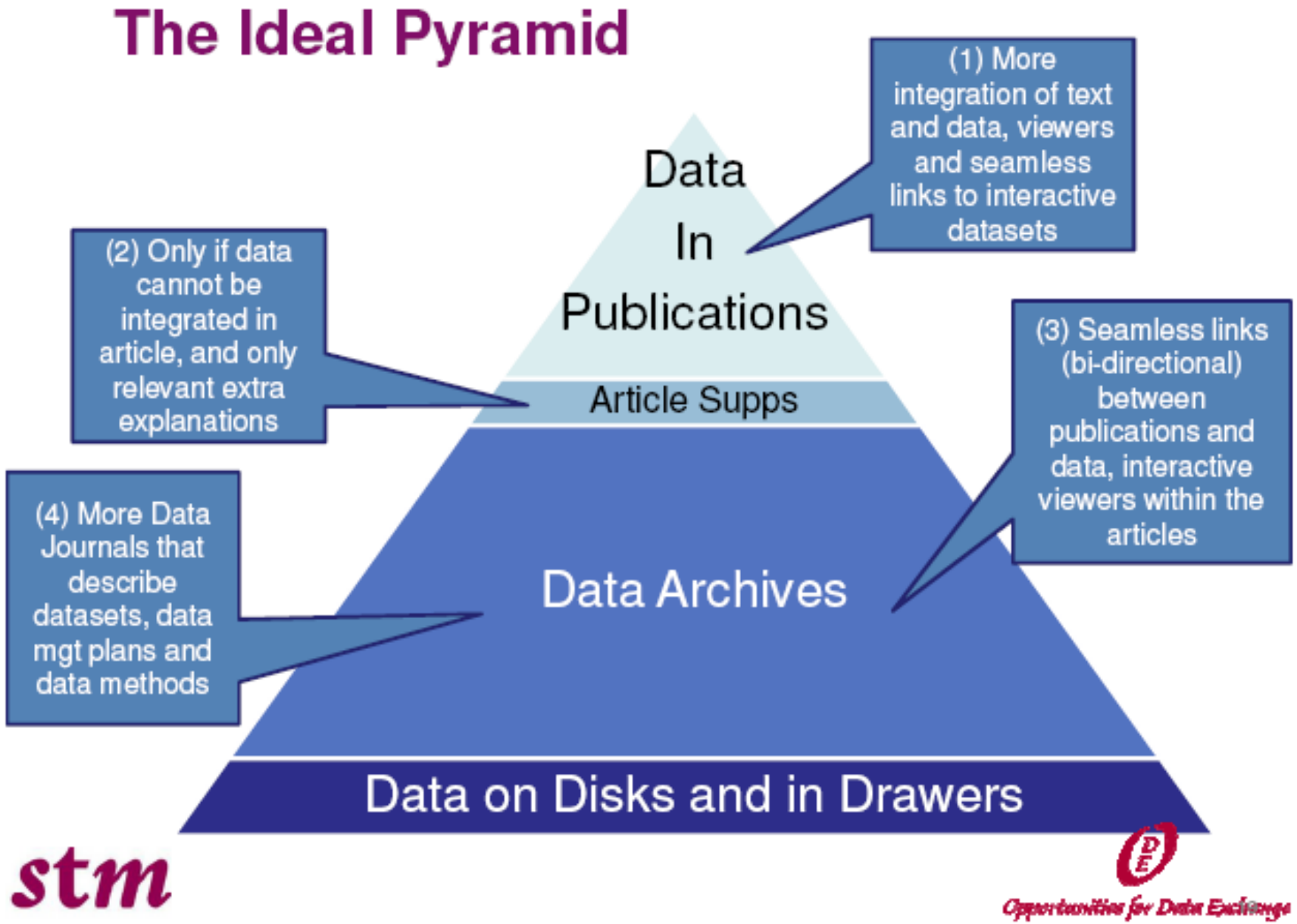


Illustration 3: The ideal Data Publication Pyramid

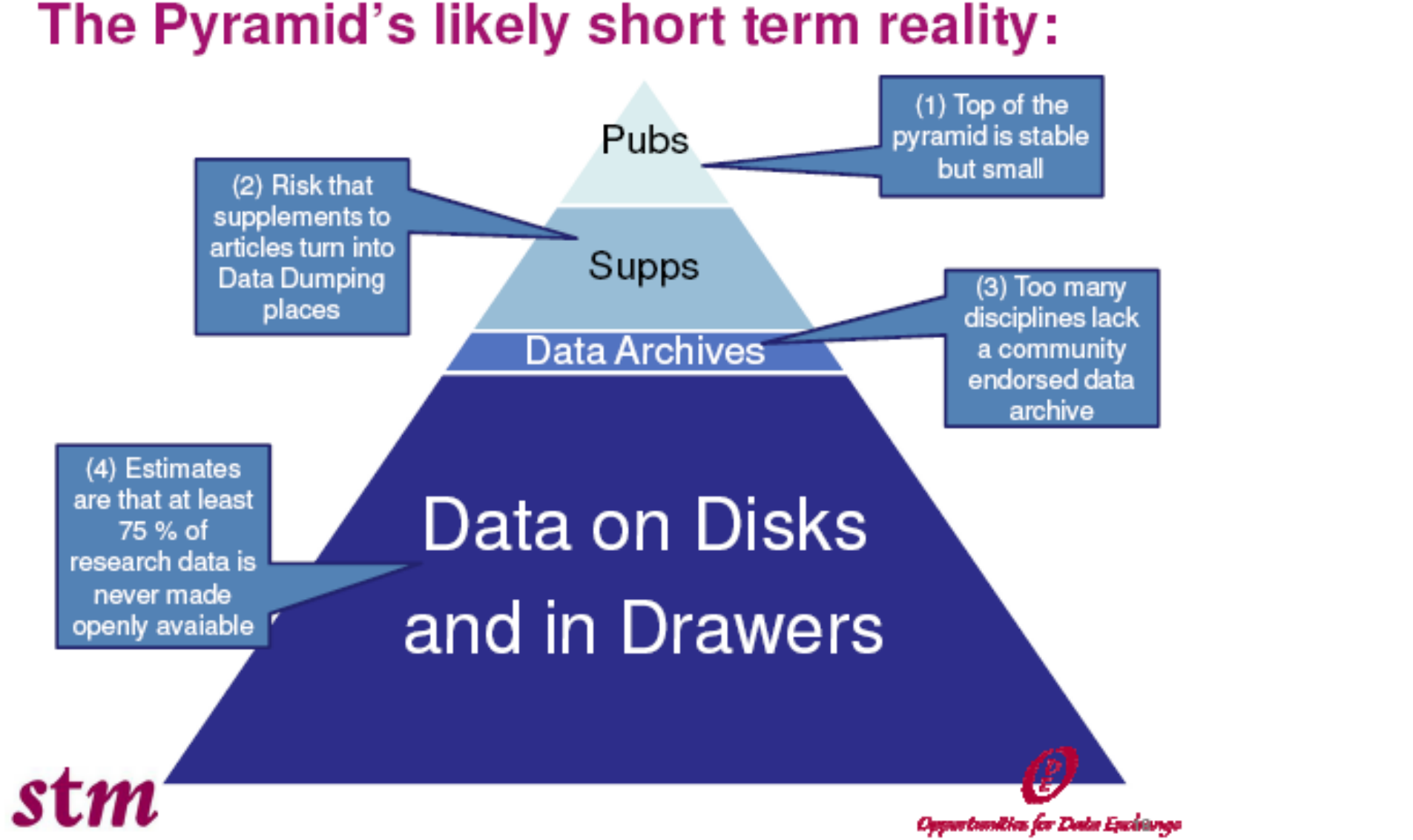


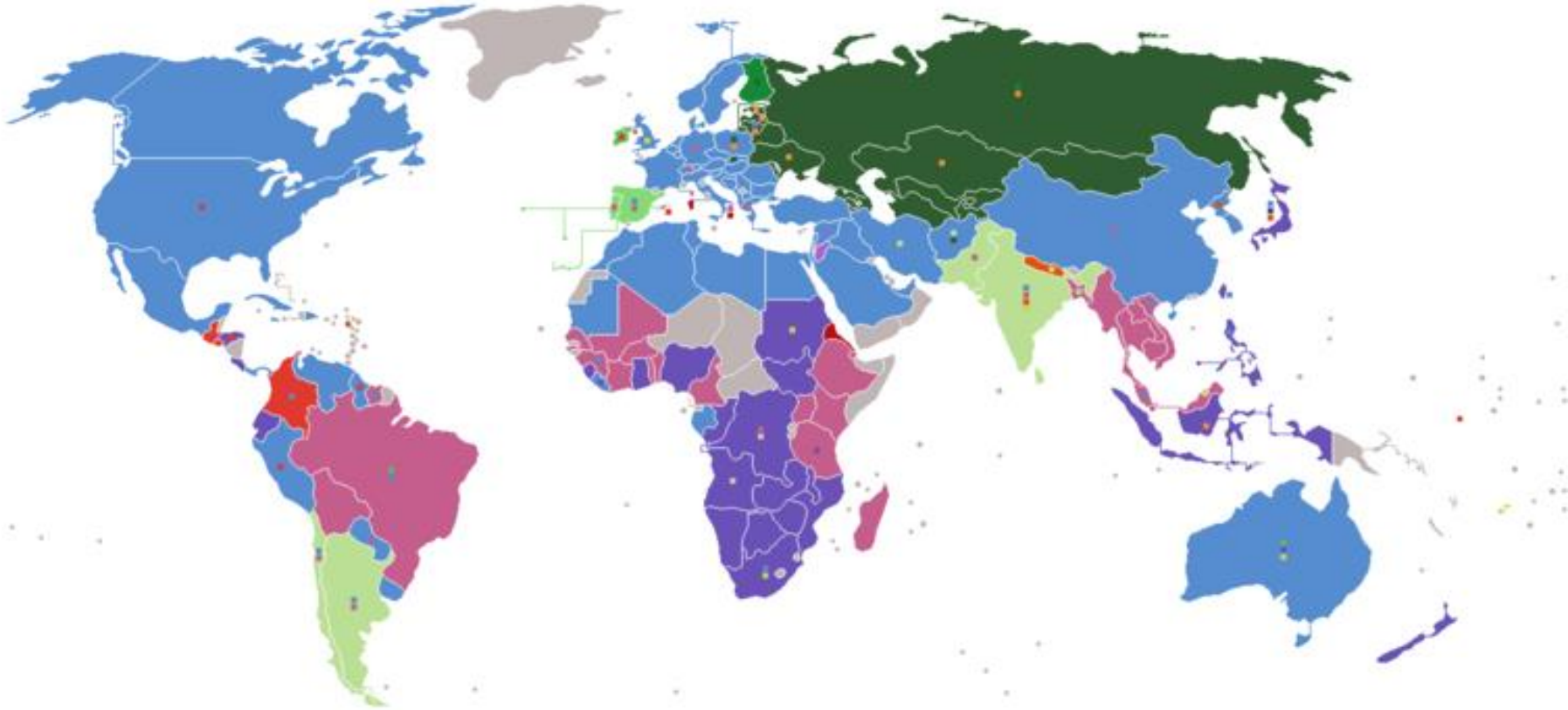
Illustration 2: The likely short term reality for the Data Publication Pyramid

Why Standards: What is “yield”…?

Is the width of a chariot at Pompeii the best determinant of gauge for railways?

Without standards you could not get “there” from “here”

Maps of Standards: World Rail Gauges



mm	1676	1668	1600	1524	1520	1435	1372	1067	1050	1000	950	914	762	750	610	600
ft in	5'6"	5'5.67"	5'3"	5'	4'11.8"	4'8.5"	4'6"	3'6"	3'5.3"	3'3.4"	3'1.4"	3'	2'6"	2'5.5"	2'	1'11.6"

http://oegeo.wordpress.com/2012/01/13/maps-of-standards/