

# PHPM 672 Data Science for HSR

## PHPM 677 Data Science in Public Health

Hye-Chung Kum

Population Informatics Research Group

<http://research.tamhsc.edu/pinformatics/>

<http://pinformatics.web.unc.edu/>

**License:**

Data Science in the Health Domain by Hye-Chung Kum is licensed under a  
[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/)

**Course URL:**

<http://pinformatics.tamhsc.edu/phpm672>

# Who am I ?

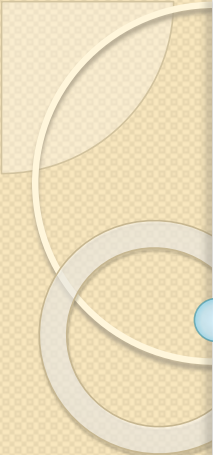
- PhD in Computer Science
  - Datamining
  - KDD (Knowledge discovery and datamining)
- MSW: Policy & Management Track
  - Certificate of nonprofit management
- 9 years
  - School of Social Work, UNC-CH
  - Department of Computer Science, UNC-CH
- Teaching
  - I love teaching. I put a lot into it & I expect a lot from students
  - Slides are my personal notes so I wont forget (don't use as example of good presentations)
- Questions?

# Who are you ?

- Program
  - PhD in HPM ?
  - MPH?
  - Anyone else?
- Majors before
- Experience in statistics
  - Class in statistics
  - Have used STATA
  - Have used SAS
  - Have any programming experience
- What would you like to get from this class

# What is this class about

- Website
  - <http://pinformatics.tamhsc.edu/phpm672>
  - Very Important to check regularly!
- Blackboard
  - Submit homework
  - Post grades
- Syllabus
- Schedule
- Resource on group website



POPULATION  
INFORMATICS  
RESEARCH GROUP



# WARNING:

## Be Prepared to Work Hard

- This was the only part of the class that will go at this nice pace
- There is a LOT of materials for me to cover
- You will not have another dedicated time to learn this, but programming will hit you as you work on your dissertation, when you are looking for a job, and on the job.
- So I want to teach you as much as I can this semester.
- If you feel you are lost, please come talk to me. I will try to help. If majority of you come talk to me, I will slow down. If you don't give me input, I will assume the pace is fine.
- You will get AS MUCH AS you put into this class



# Last thoughts

- Programming
  - Bottom line, you have to DO this.
  - READ (lecture), WATCH (lab), DO (assignment).
  - Not easy, but really worth it to take the time to learn. Like your multiplication tables.
- Data Science
  - Very new. I didn't read and textbooks, no one taught me
  - So mostly my opinion on an evolving topic
  - Share your thoughts. Younger generation born into the digital world have something I don't.
- Data & Programming CAN be FUN !!
  - It's my favorite hobby

# Agenda

- What is Data Science/Population Informatics?
  - How does it relate to HPM? HSR?
  - How does it relate to Public Health?
- Examples of population informatics
  - NC-DHHS (Dept. Health & Human Svc) Management Assistance Project
  - County self evaluation
  - Research



# Agenda

- What is Data Science/Population Informatics?
  - How does it relate to HPM? HSR?
  - How does it relate to Public Health?
- Examples of population informatics
  - NC-DHHS (Dept. Health & Human Svc) Management Assistance Project
  - County self evaluation
  - Research



POPULATION  
INFORMATICS  
RESEARCH GROUP



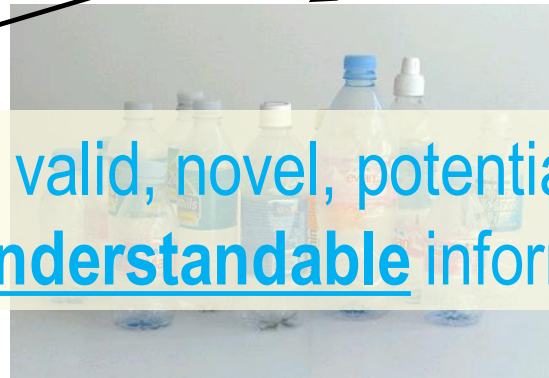
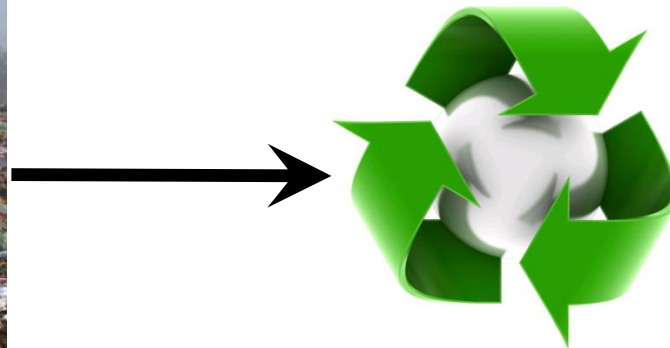
# Data Science

## Knowledge Discovery & Data mining (KDD)

**Big Data**

**KDD**

**Clean, Merge, Reprocess**



Human consumable, valid, novel, potentially useful,  
and ultimately understandable information

# Properties of BIG DATA : 4V

- Volume : constantly generating
- Velocity : constantly changing
- Variety : expressed in many ways
- Veracity : lots of errors

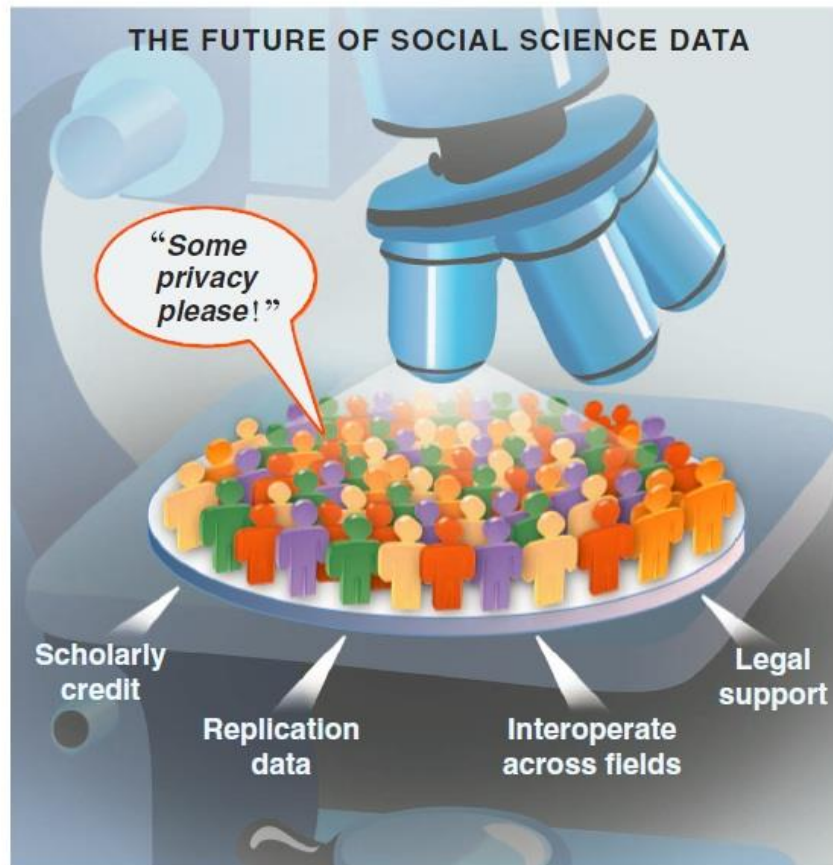
**What do you do to find information/knowledge on the Internet?**



POPULATION  
INFORMATICS  
RESEARCH GROUP



# Population Informatics (Massive secondary data analysis)

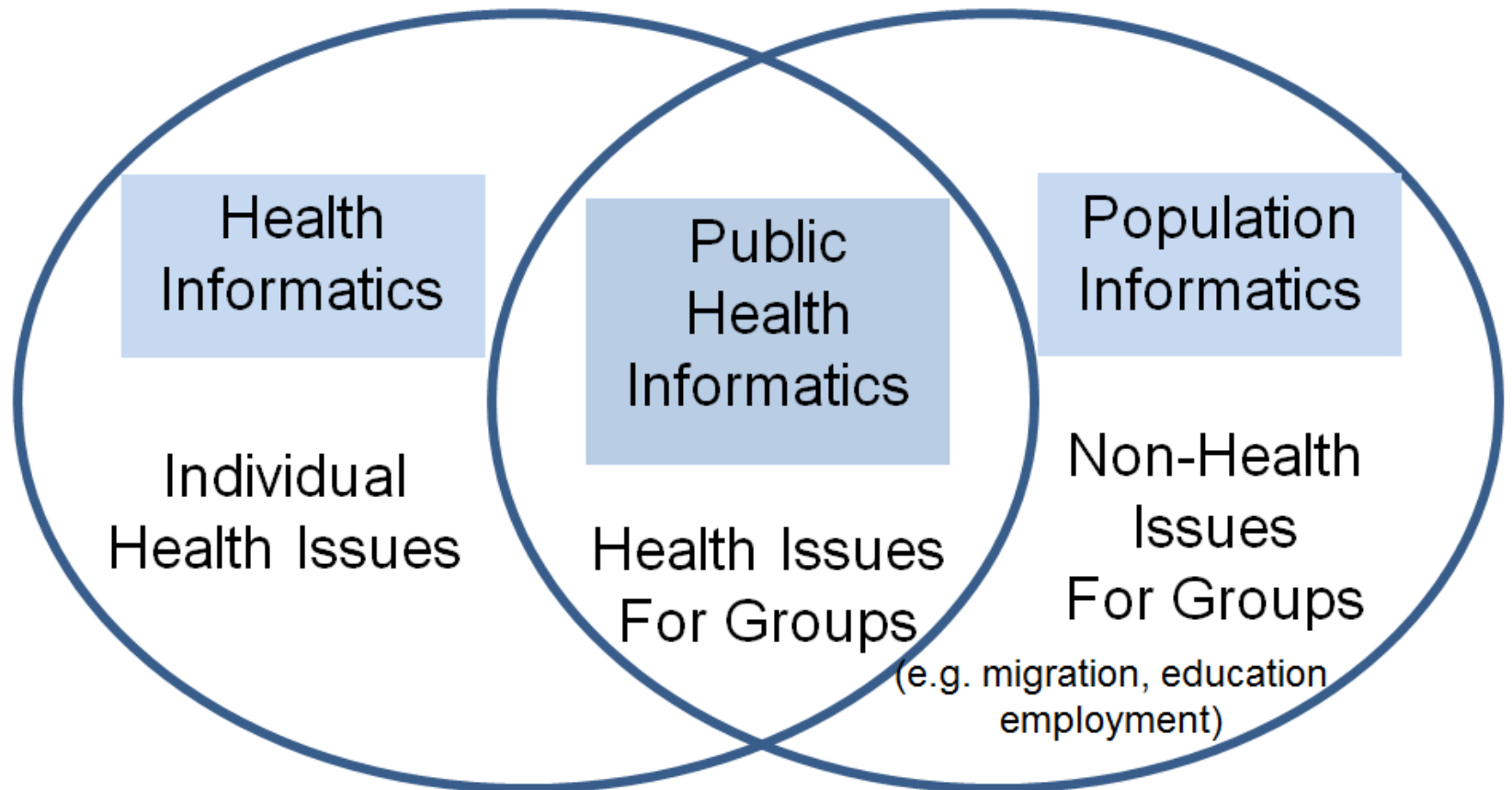


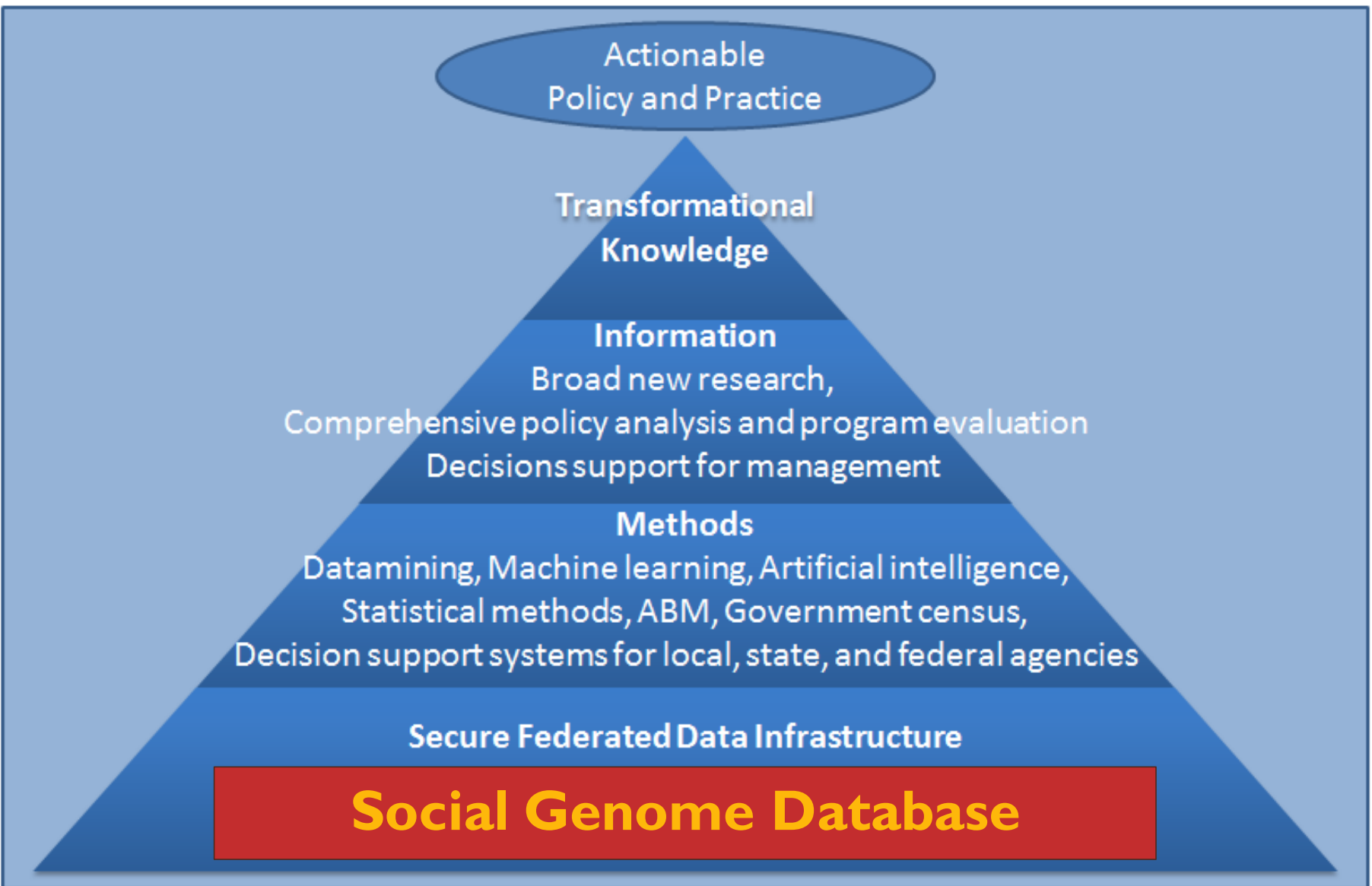
- CS+Statistics+Social Science
- Big data analysis about people
- **Health Population Informatics: Analyzing Big Data about People for Better Healthcare**
- **E-government: Analyzing Big Administrative Data about People to better manage government resources**
- Gary King. Ensuring the Data-Rich Future of the Social Sciences, *Science*, vol 331, 2011, pp 719-721.

**Fig. 1.** New types of research data about human behavior and society pose many opportunities if crucial infrastructural challenges are tackled.



# Different areas in Population Informatics

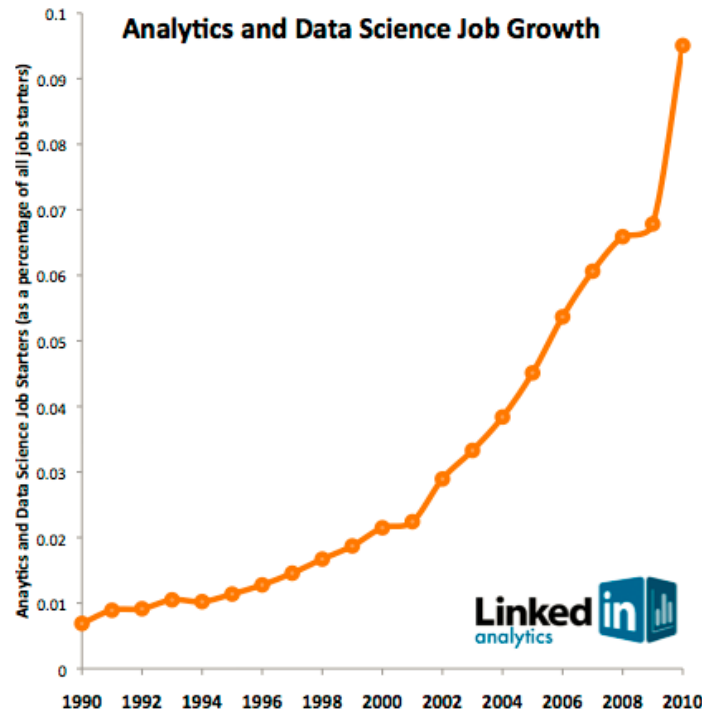






# Job market of data scientists

- statisticians will be the next sexy job
  - Google Chief Economist Hal Varian
- shortage of 190,000 data scientists by the year 2019
  - McKinsey Global Institute



Courtesy LinkedIn Corp.



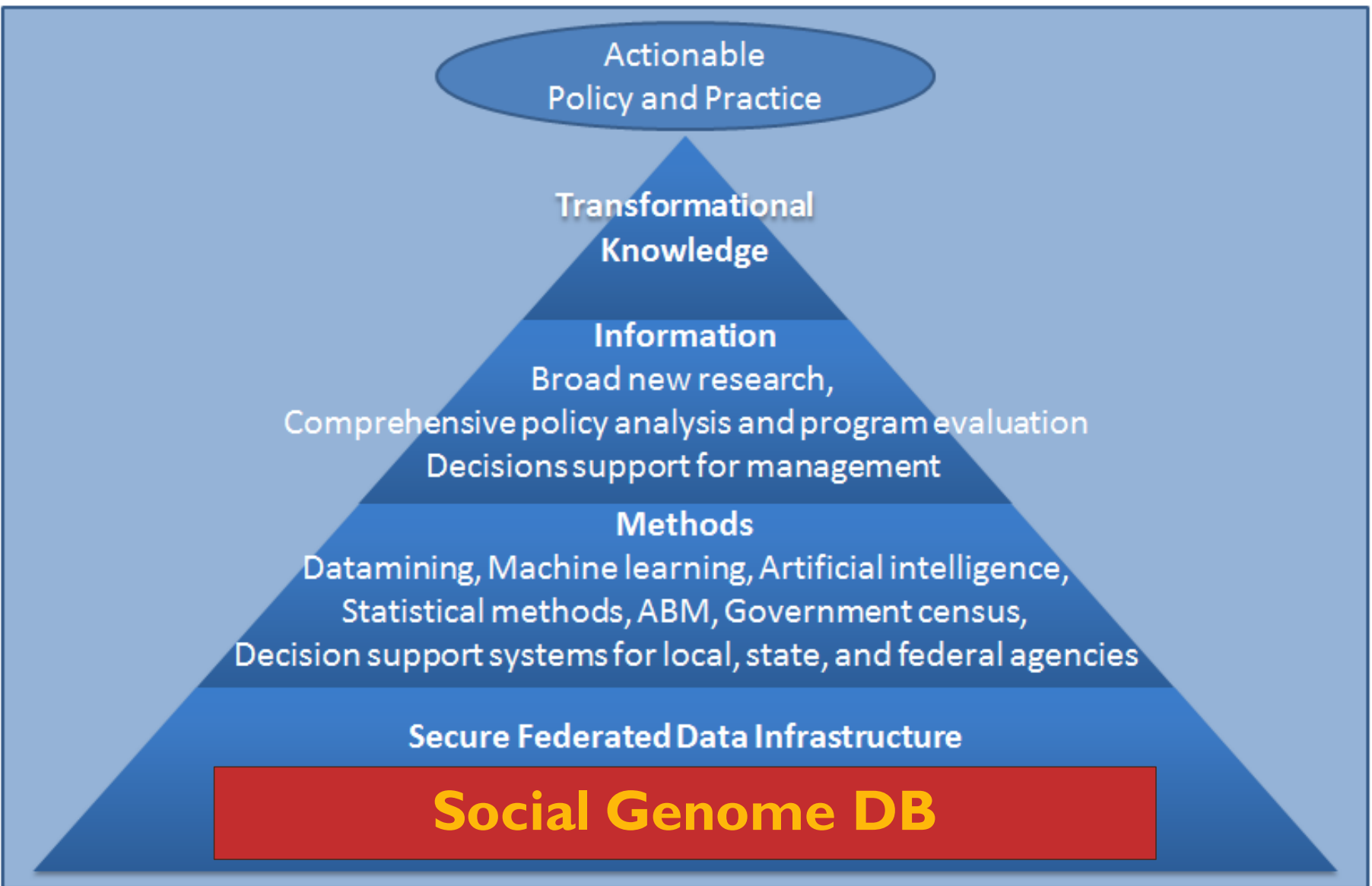
POPULATION  
INFORMATICS  
RESEARCH GROUP

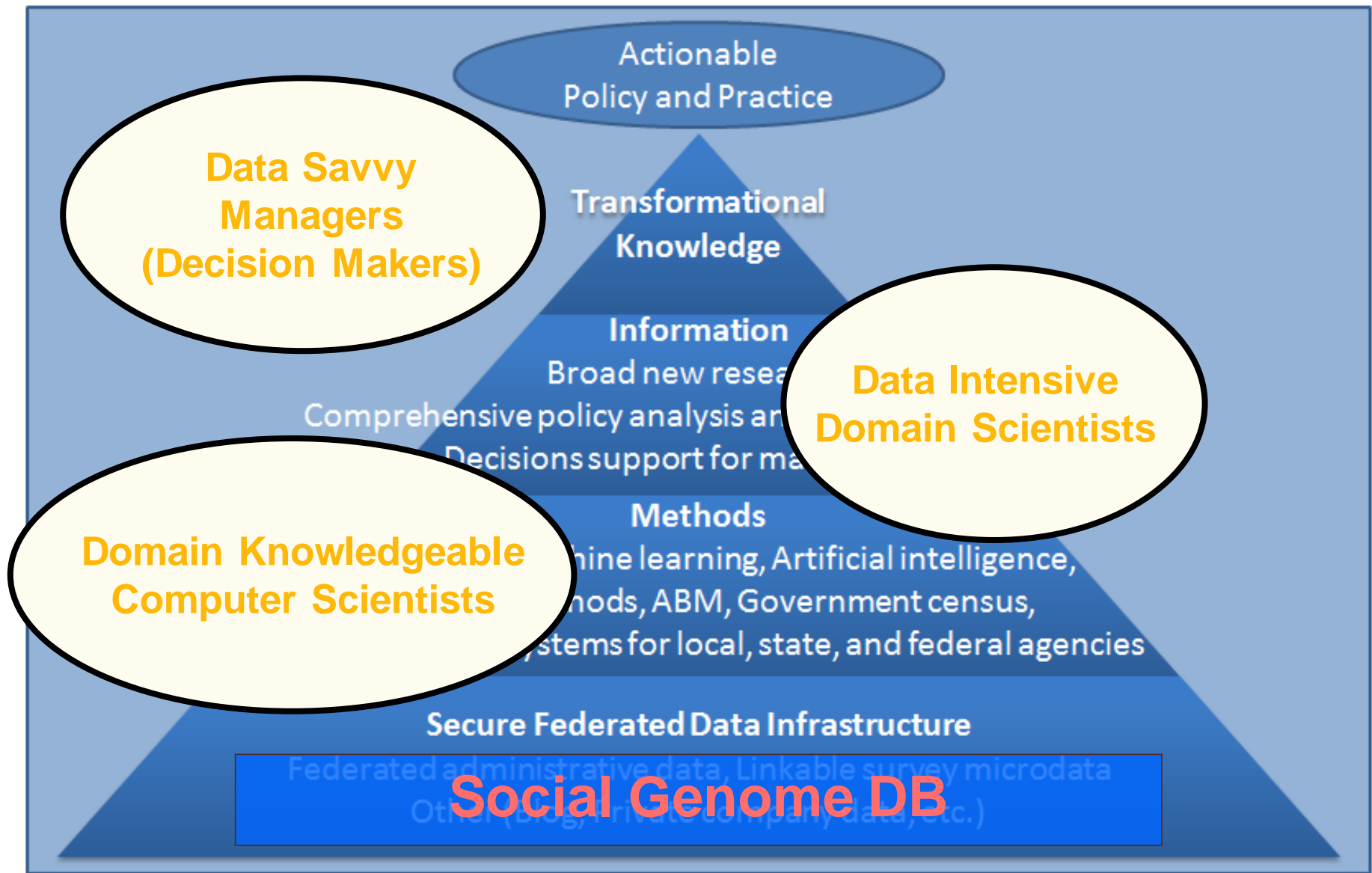


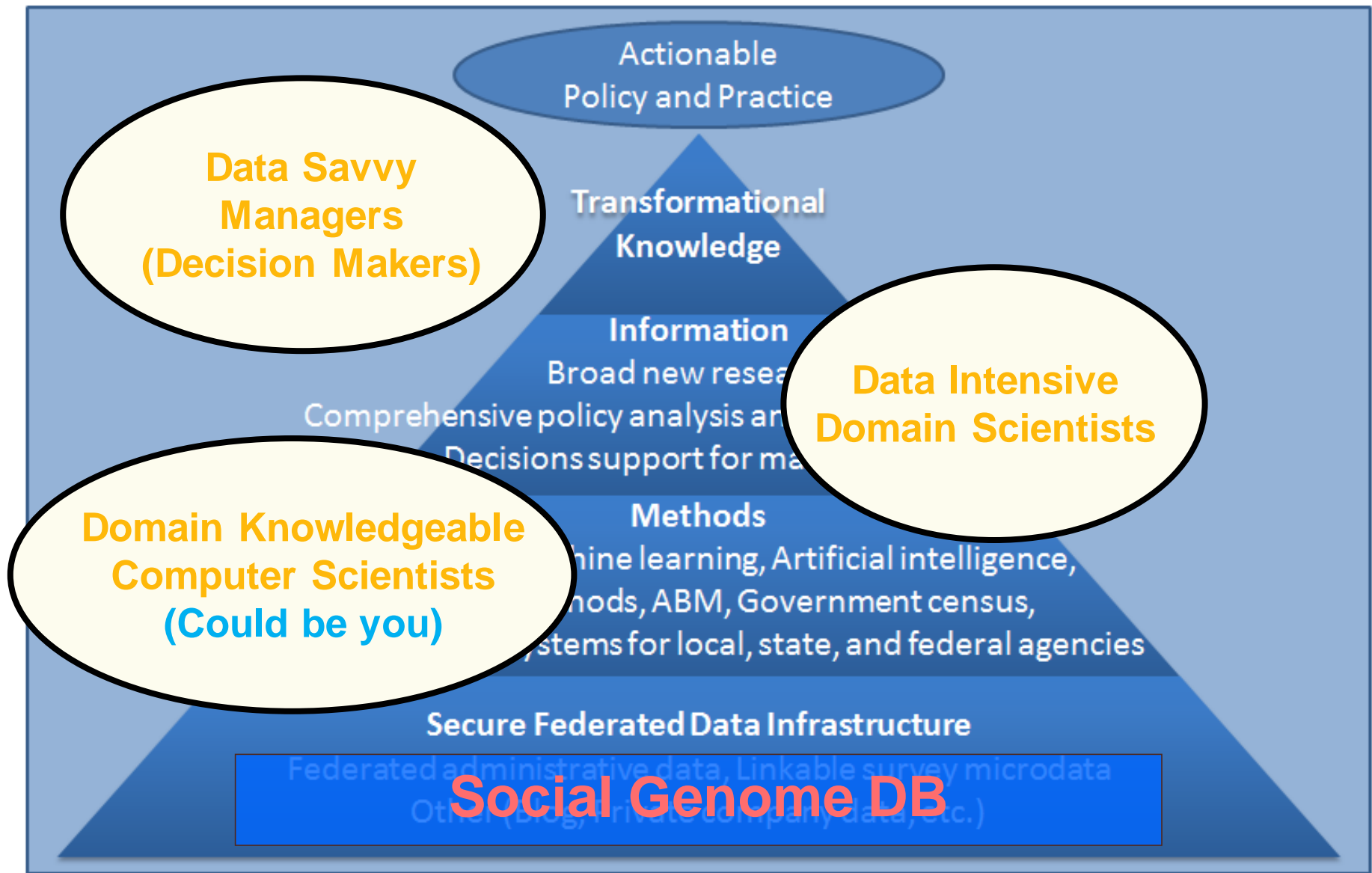
# Data experts in the next century

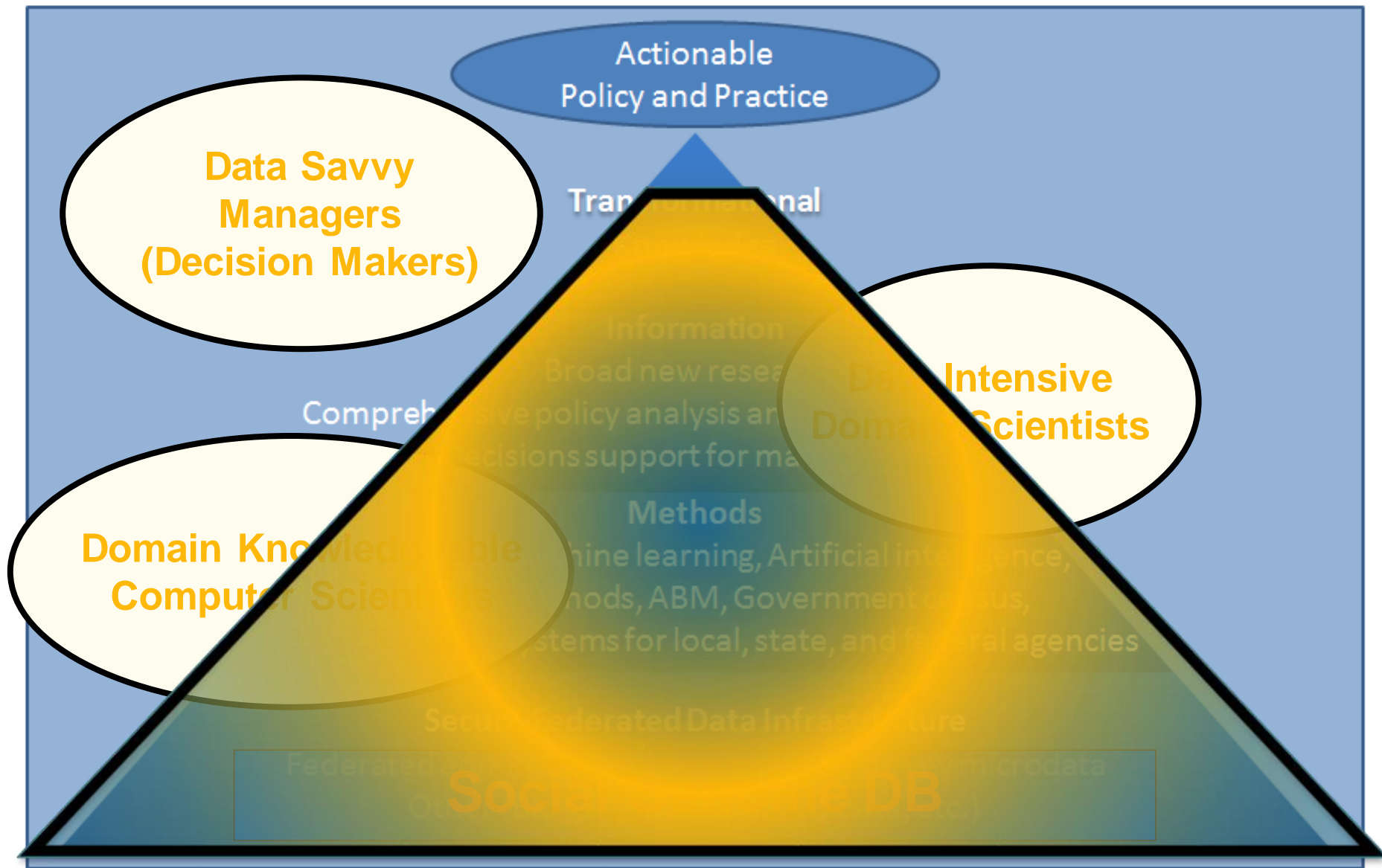
- **Data Savvy Mangers (decision makers)**
  - **MHA & MPH**
  - who can understand and use data for decisions and actions.
- **Data intensive domain scientists**
  - **PhD, MSPH, MPH**
  - experts in the domain science with intensive training on data science and analytics
- **Domain knowledgeable computer scientists**
  - Collaborators in CS (undergraduates, MS, PhD).
  - **HPM department role : teach them the domain science!**
  - Build tools, manage data, and run analytics











# Thomas Davenport

## *Competing on Analytics*

- Skill set for good data scientists
  - IT & Programming skills
  - Statistical skills
  - Business skills:
    - Understand pros/cons of decisions & actions
    - Communication skills
    - Excel / PowerPoint
  - Intense curiosity: the most important skill or trait.  
“a desire to go beyond the surface of a problem, find the question at its heart, and distill them into a very clear set of hypothesis that can be tested”

# New Era in Science : Big Data Science

- **Data** is the new raw material of business: an economic **input almost on par with capital and labor**.(Microsoft's Craig Mundie)
- **Those who can harness the power of data will lead the next century** and drive innovation in commerce, scientific discovery, healthcare, finance, energy, government, and countless other fields.
- Students who learn to be a data science will be in high demand.





# High International Interest Doing Good Research with Big Data

- Figuring how to do good research with Big Data
- EUDAT (Oct 2011): EU
  - €16-million initiative to develop an international data management infrastructure over 3 years
- White house (Mar 2012) : US
  - a national effort to fund Big Data research across the federal agencies including NSF, NIH, DOD (Dept of Defense), and DOE (Dept of Energy)



# International Population Health Informatics Research

- US : LEHD (Census Bureau) – 2010 Nobel Prize in economics
- Australia & New Zealand
  - National Centre for Epidemiology and Population Health (NCEPH), The Australian National University
  - Australian Institute of Health and Welfare
  - Centre for Health Record Linkage
  - Centre for the Study of Assessment and Prioritisation in Health, School of Medicine and Health Science (NZ)
- EU
  - Health Information Research Unit, School of Medicine, Swansea University, Wales, UK
  - Health Services Research Unit, University of Aberdeen, Scotland
- Canada
  - Canadian Institute for Health Information
  - Child and Youth Data Lab, Alberta Centre for Child, Family and Community Research

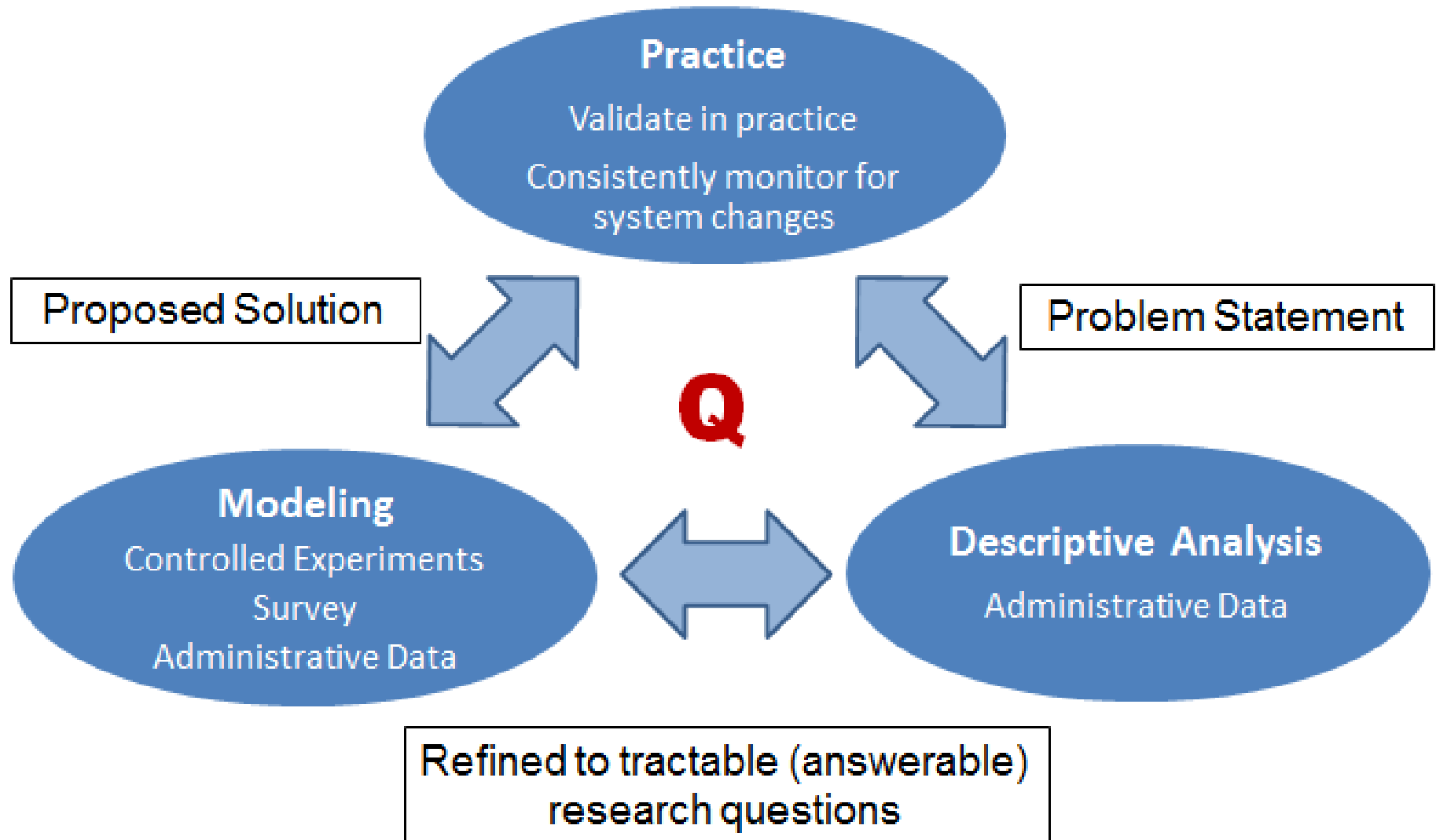


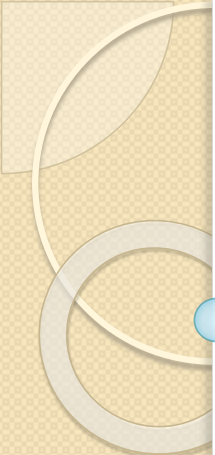
# LEHD : US Census Bureau

- Vertically integrated in one domain
  - Wage : UI (Unemployment Insurance) Data
- Decision support : LEHD website
- By building an integrated data that “permits the real world of the US economy to be interrogated by the models of unemployment dynamics” Peter Diamond, Dale Mortense, and Christopher Pissarides shared the Nobel Prize in economics last year (David Warsh, [economicprinciple.com](http://economicprinciple.com))

# Iterate until balance is reached & maintained

Q: Where is the sweet spot for  
**balancing access, cost, & quality** of health care





POPULATION  
INFORMATICS  
RESEARCH GROUP



# Reminder

- Read article on social genome
  - Post of blackboard
    - 3 interesting things you learned
    - 1 thought/opinion
- Look at class website

# How to post reading log

- Click on the appropriate forum (P1 for this week)
- "Create Thread" and type in your feedback
- Don't forget to save by clicking on "Submit"



POPULATION  
INFORMATICS  
RESEARCH GROUP

