

An Evolving Federal Statistical System

Robert M. Groves
Georgetown University
September 12, 2012



5 Observations

1. The difficulties of measuring the busy, diverse, and independent modern society and economy are increasing every year (that is, it costs more money to do the same things the statistical agencies have done for years).

2. The demands by business, state, local, and community leaders for timely statistics on their populations are continually increasing.

3. New technologies are being invented almost daily that can be used to make it more convenient for the public to participate in these efforts to inform us about the status of the country.

4. New digital data resources are being created both from national-state-local government programs, private sector transactions, and internet-related activities.

5. Near-term central government budgets are likely to be flat or declining.

Profound Conclusion

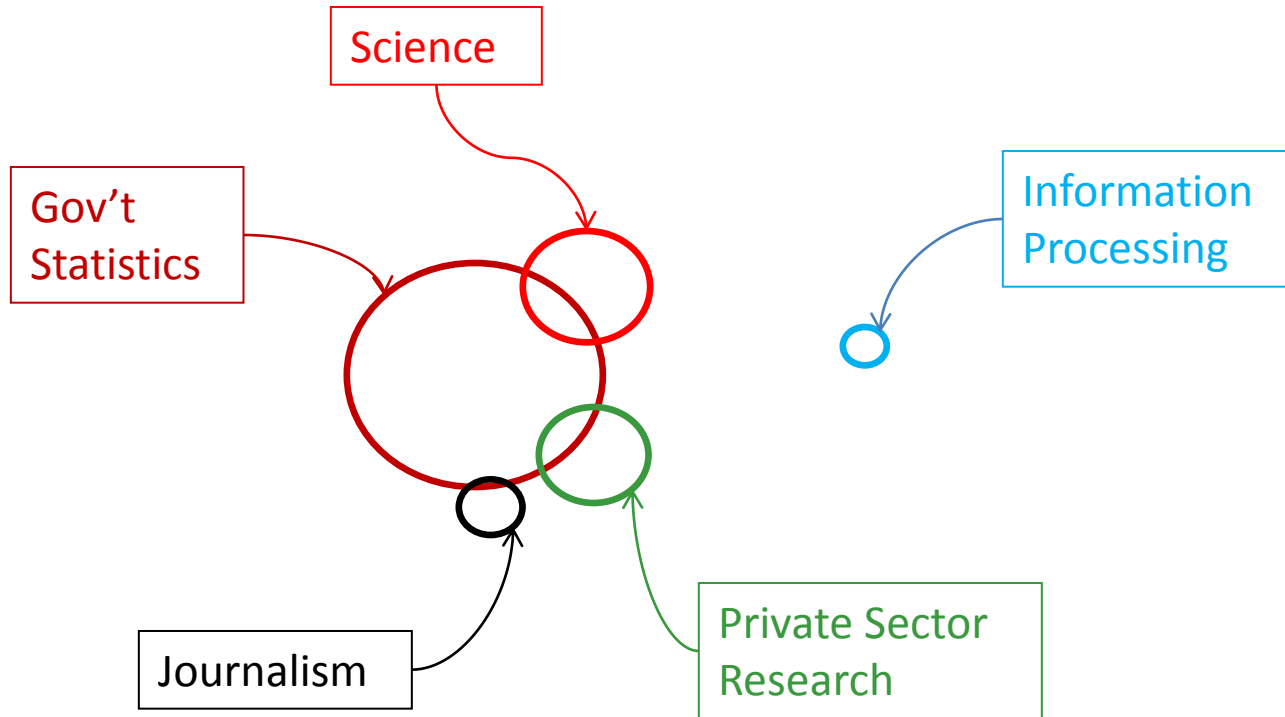
1. Higher costs
2. More demand for timely statistics
3. New technologies
4. New data resources
5. No new money

Conclusion: current practices are unsustainable

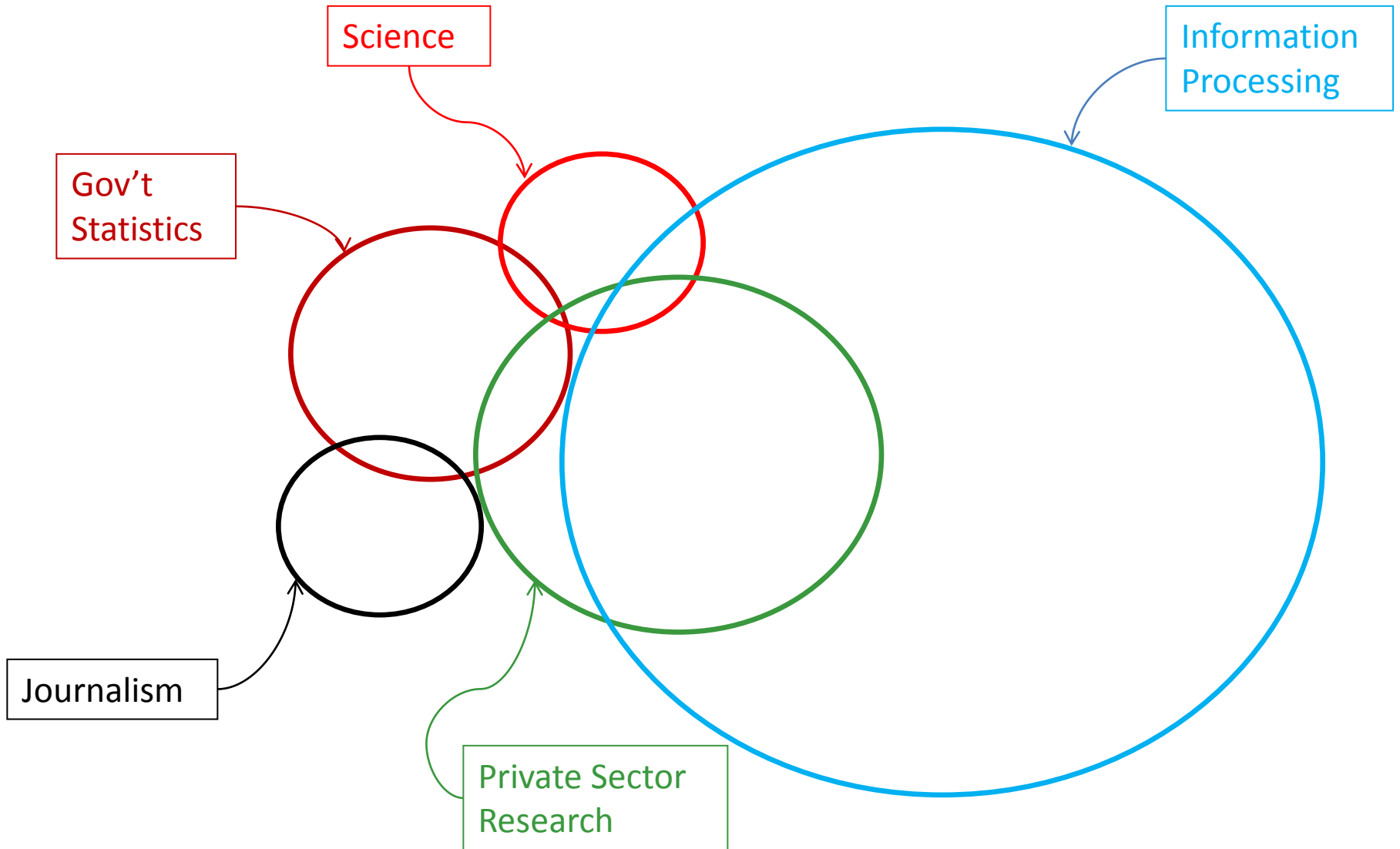
Outline

- The rise of “organic” data
- Organic data and new statistical information
- A possible future

Relative Sizes of Digital Data Production, c.1960



Relative Sizes of Digital Data Production, 2010



A Self-Monitoring Social and Economic Eco-System

- Organic data
 - Those produced auxiliary to processes, to record the process
- Designed Data
 - Those produced to discover the unmeasured

Examples of Organic Data

- Google searches (Google Flu)
- “Scraped” data from websites
- Tweets
- CCTV, traffic camera data
- Retail scanner data
- Credit card transaction data
- Data.gov

Common Features of Organic Data

- New data sources provide looks at interesting new phenomena
- They tend to be behaviors, not direct measures of internalized states
- The data are near real-time relative to the behaviors measured
- The data tend to be lean in variables
- They are grossly incomplete on coverage of usual target populations

[Google.org home](#)

Flu Trends

[Home](#)

United States

National

[Download data](#)

[How does this work?](#)

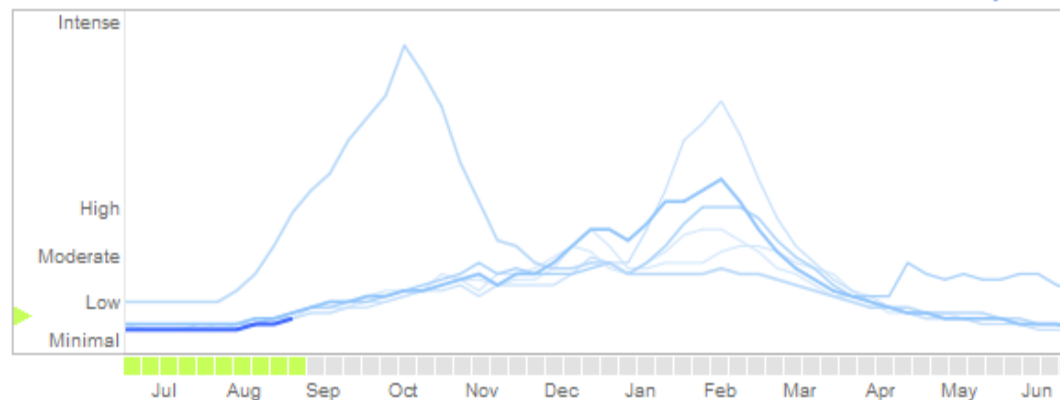
[FAQ](#)

Explore flu trends - United States

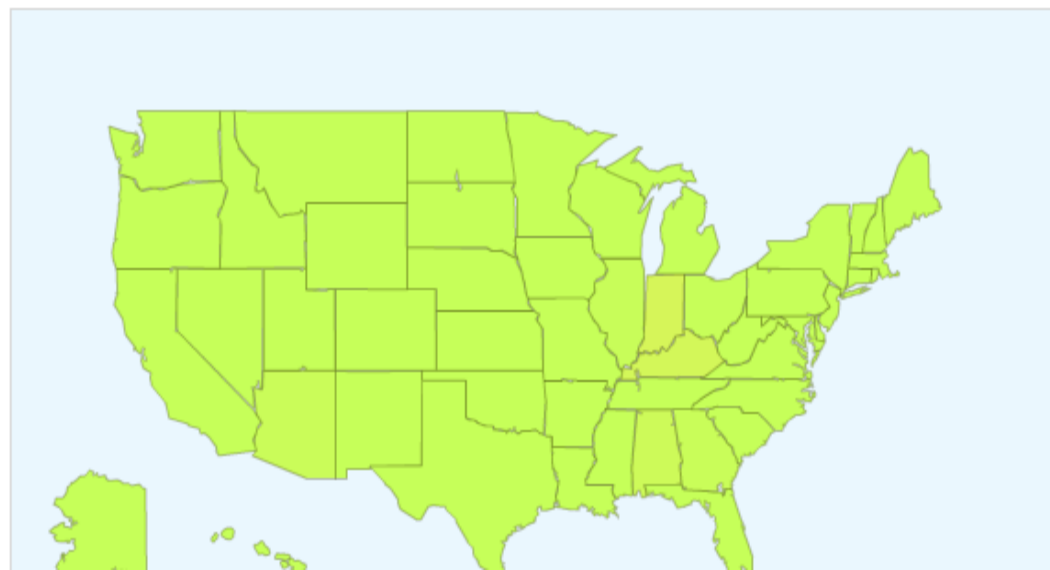
We've found that certain search terms are good indicators of flu activity. Google Flu Trends uses aggregated Google search data to estimate flu activity. [Learn more »](#)

National

● 2011-2012 ● [Past years](#) ▼



States | [Cities](#) (Experimental)



Fight influenza

CDC urges you to take these steps to protect yourself and others from the flu:

1. Get vaccinated against flu – it's your best defense.
2. Cover your cough, wash hands often.
3. Take antiviral drugs if your doctor recommends them.



Animated Flu Trends in Google Earth

[Download and explore](#) Flu Trends data in Google Earth. Need Google Earth? [Download it here.](#)

The Billion Prices Project @ MIT

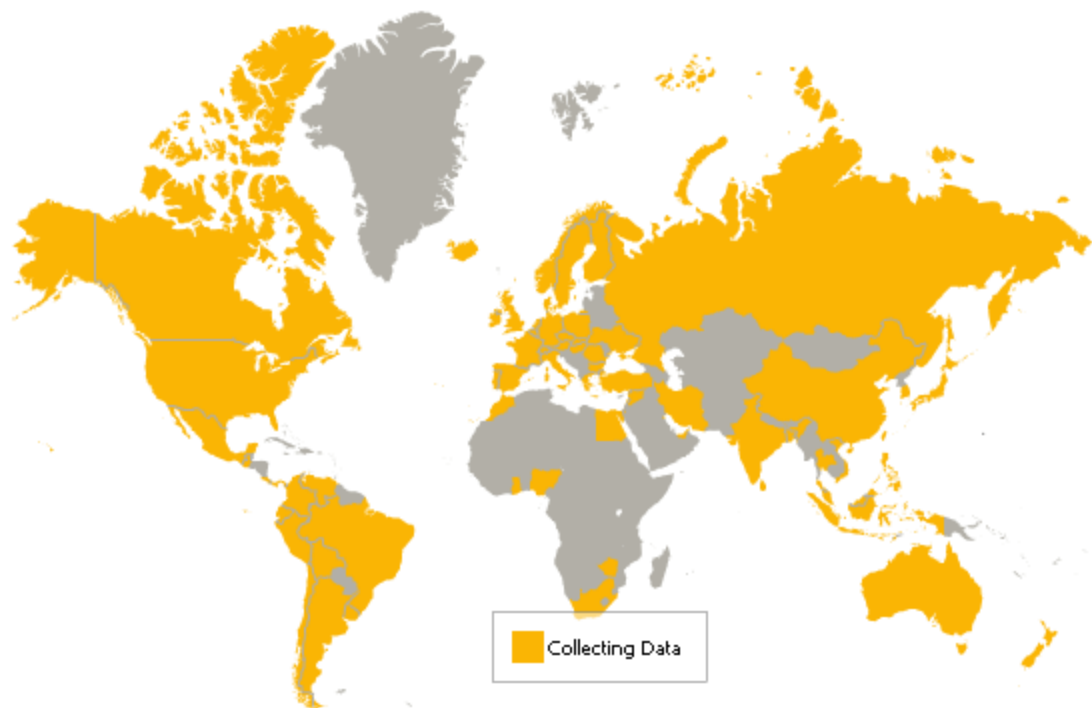
[Home](#) | [Log in](#)

[Home](#) [US Daily Index](#) [Blog](#) [About Us](#)

The Billion Prices Project is an academic initiative that uses prices collected from hundreds of online retailers around the world on a daily basis to conduct economic research.

This page shows our most recent research leveraging high-frequency price data, as well as the US daily inflation index (updated monthly on this page).

BPP Geographic Coverage



Email:

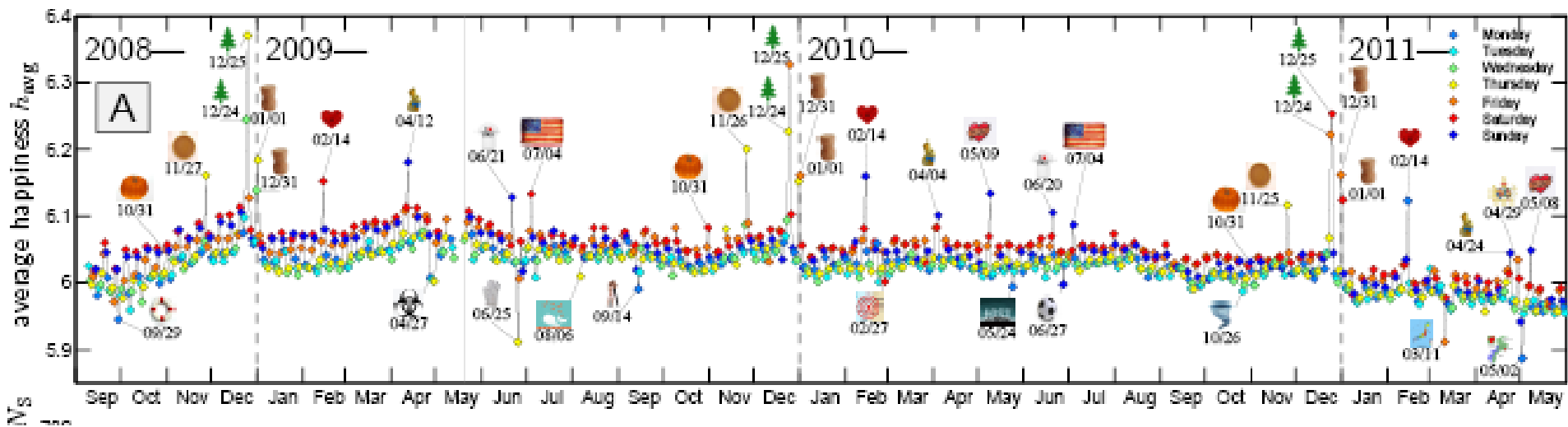


HOW TO USE OUR GRAPHS

- Click on each graph's legend icons to show/hide a variable
- Place the mouse over the graph lines to see daily values
- Click and drag the mouse to zoom in on a range of dates

FINANCIAL AND RESEARCH SUPPORT



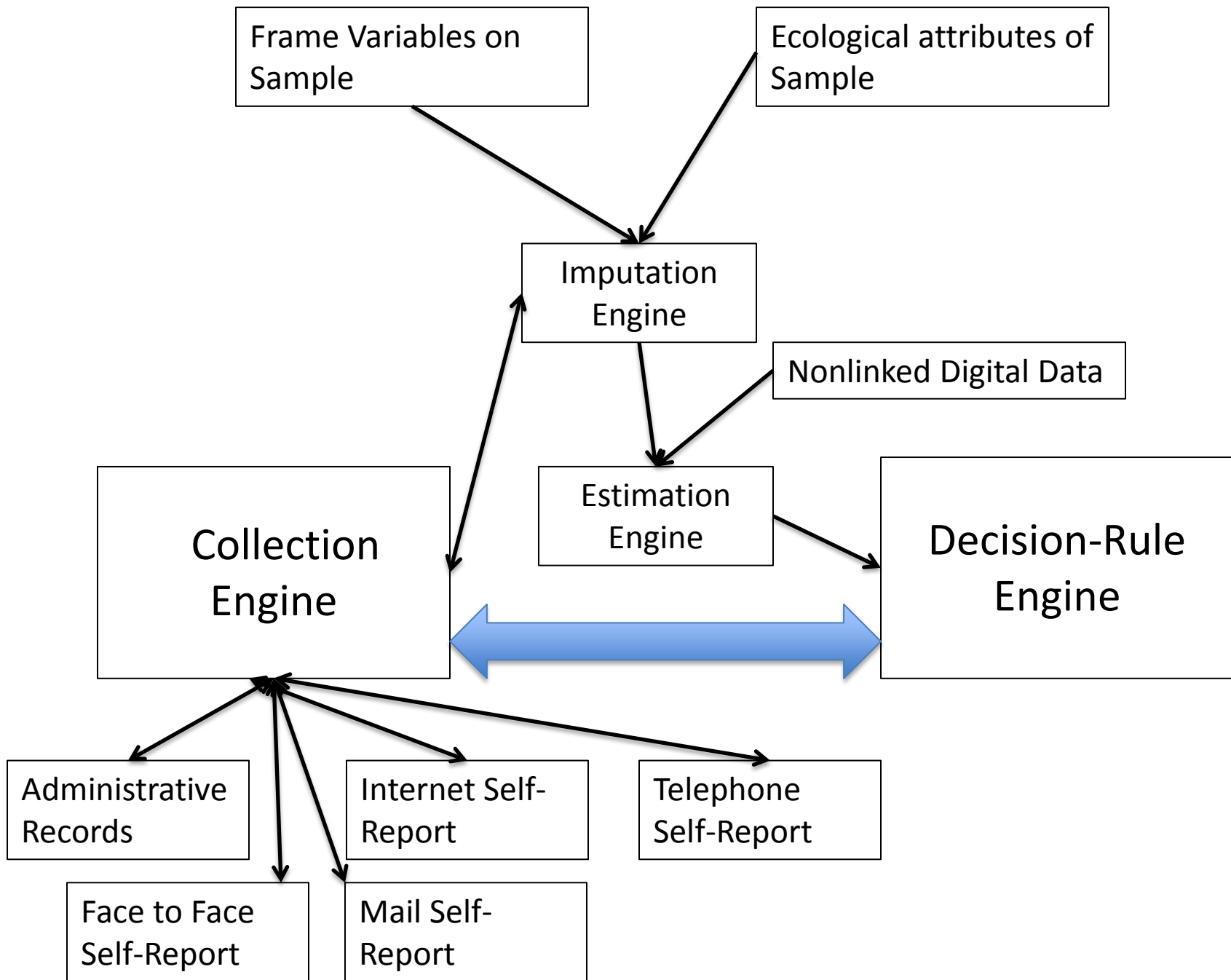


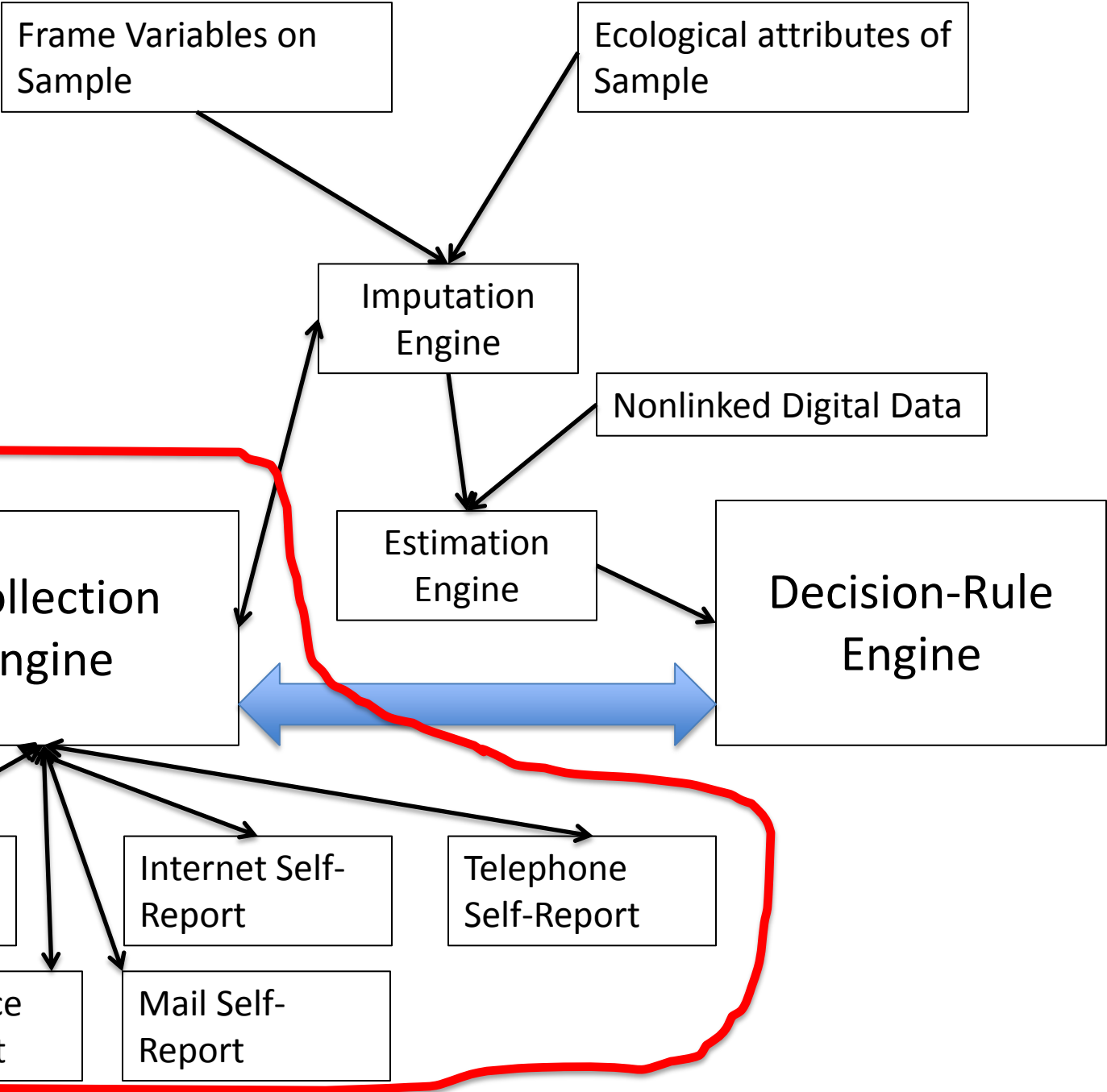
Temporal patterns of happiness and information in a global social network:
Hedonometrics and Twitter

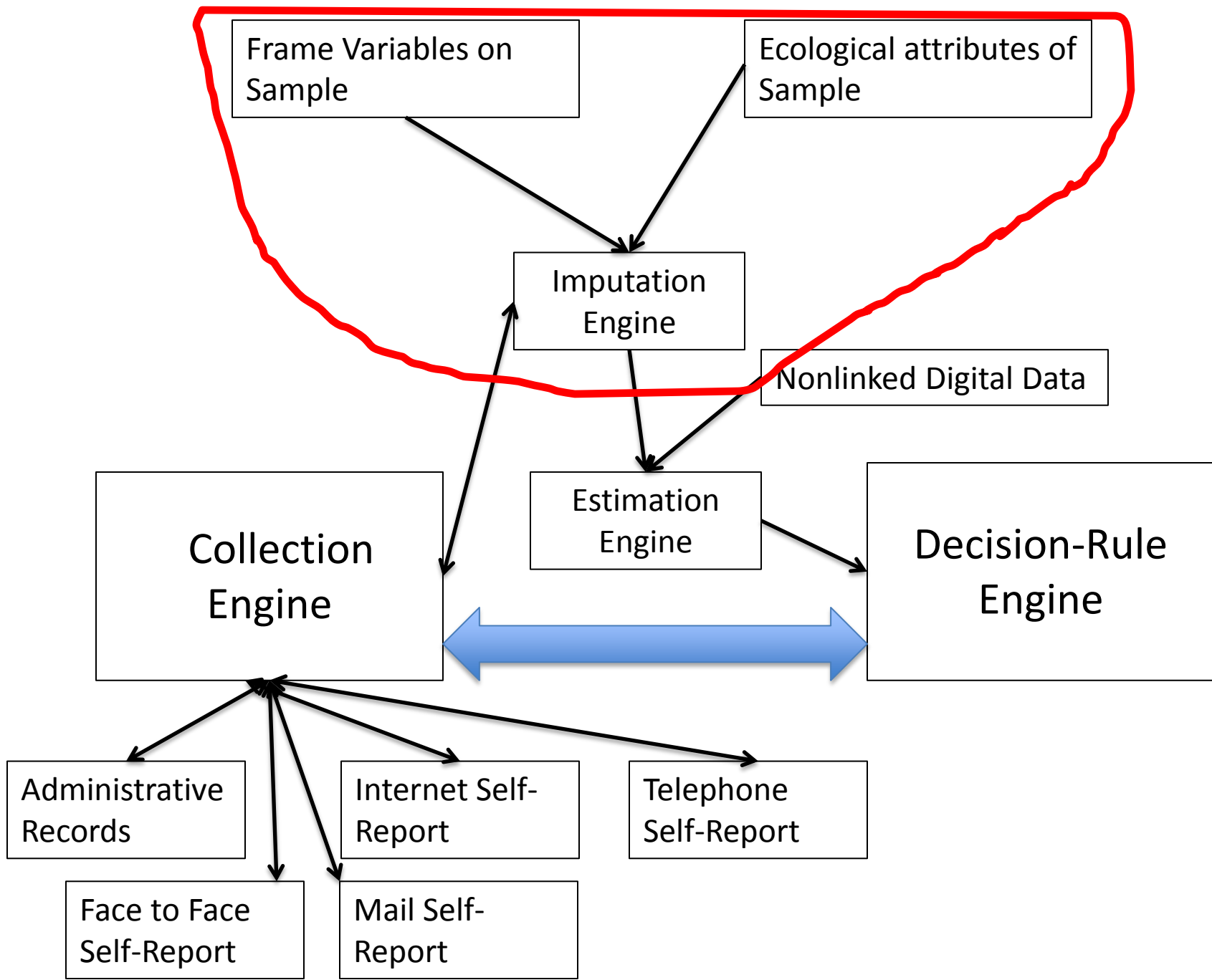
Peter Sheridan Dodds,^{1,*} Kameron Decker Harris,¹ Isabel M. Kloumann,¹ Catherine A. Bliss,¹ and Christopher M. Danforth^{1,†}

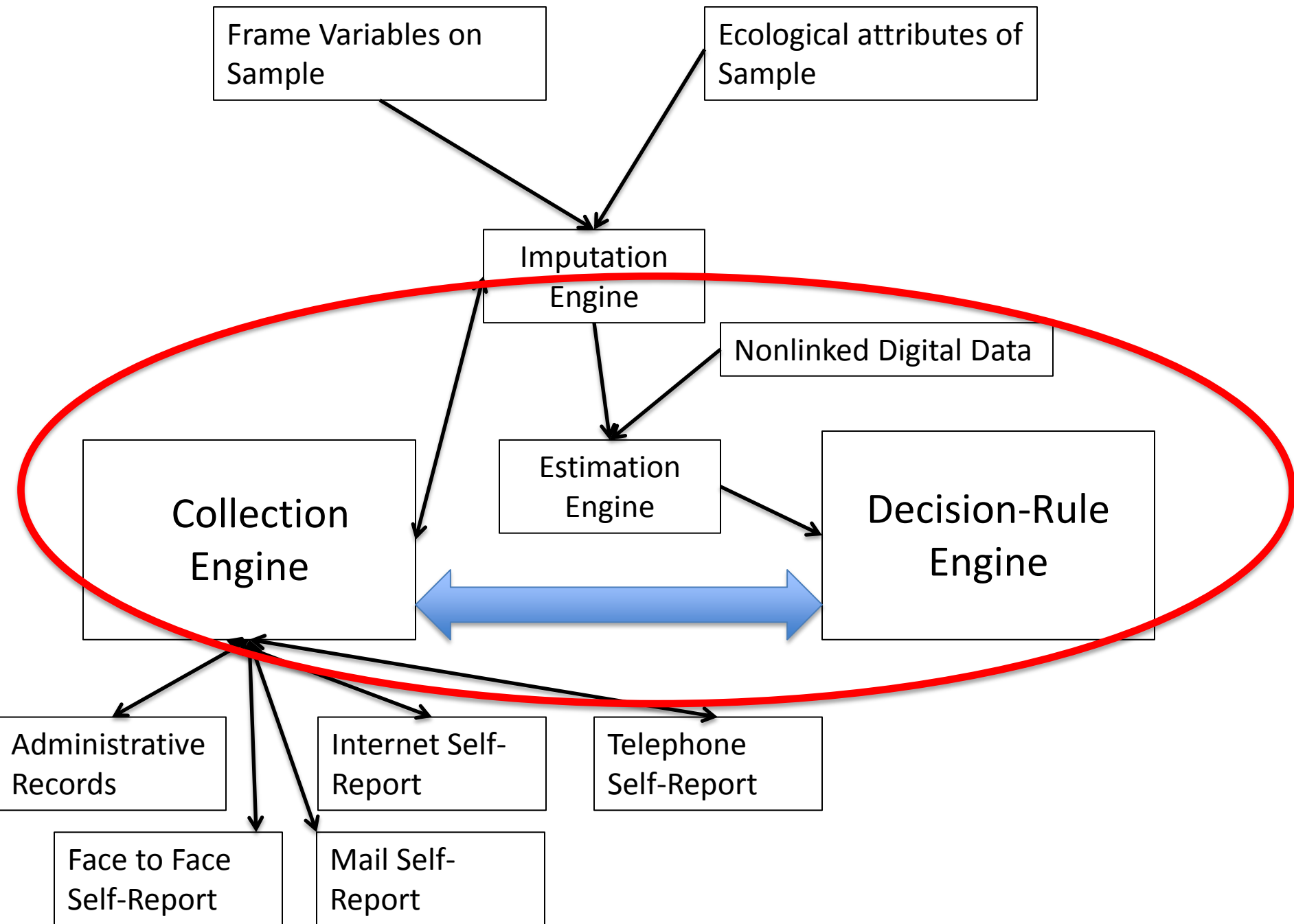
A Vision of the Future

- Multiple modes of data collection/acquisition
 - Internet behaviors
 - Administrative records
 - Internet self-report
 - Telephone, face-to-face, paper
- Real-time mode switch to “fill-in” missing data
- Real-time estimation









Attributes of the Vision

- Real time mode-switch, imputation, estimation
- Empirical stopping rules for continued self-report efforts
- Statistical modeling to combine survey data with external, relevant other digital data
- Reduced cost, increased timeliness