

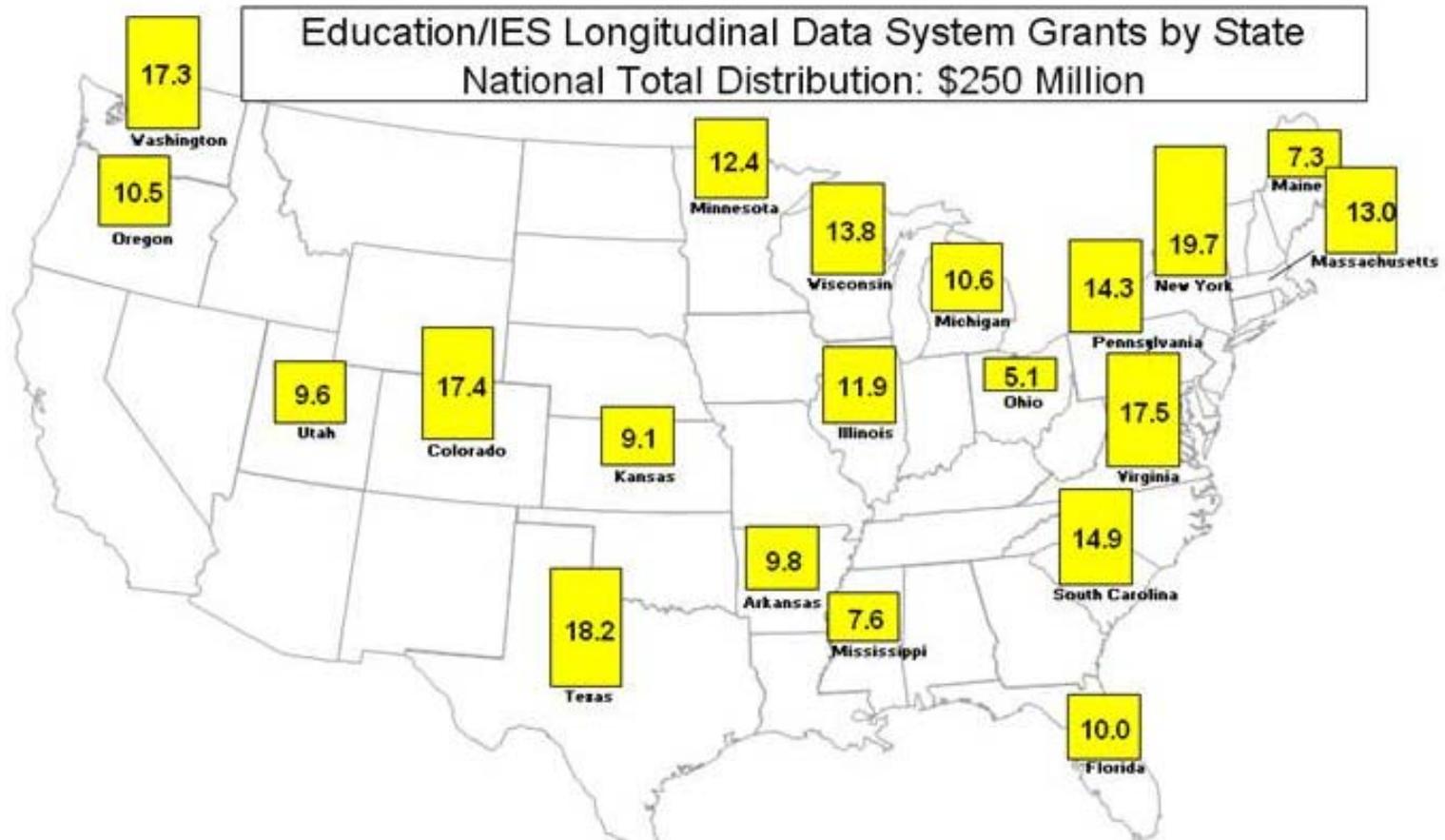
Longitudinal Data Systems: New Frontier in Statistics

Jeremy S. Wu, Ph. D.

**International Chinese Statistical Association
Indianapolis, IN
June 21, 2009**

May 21, 2010

Department of Education Grant Announcement

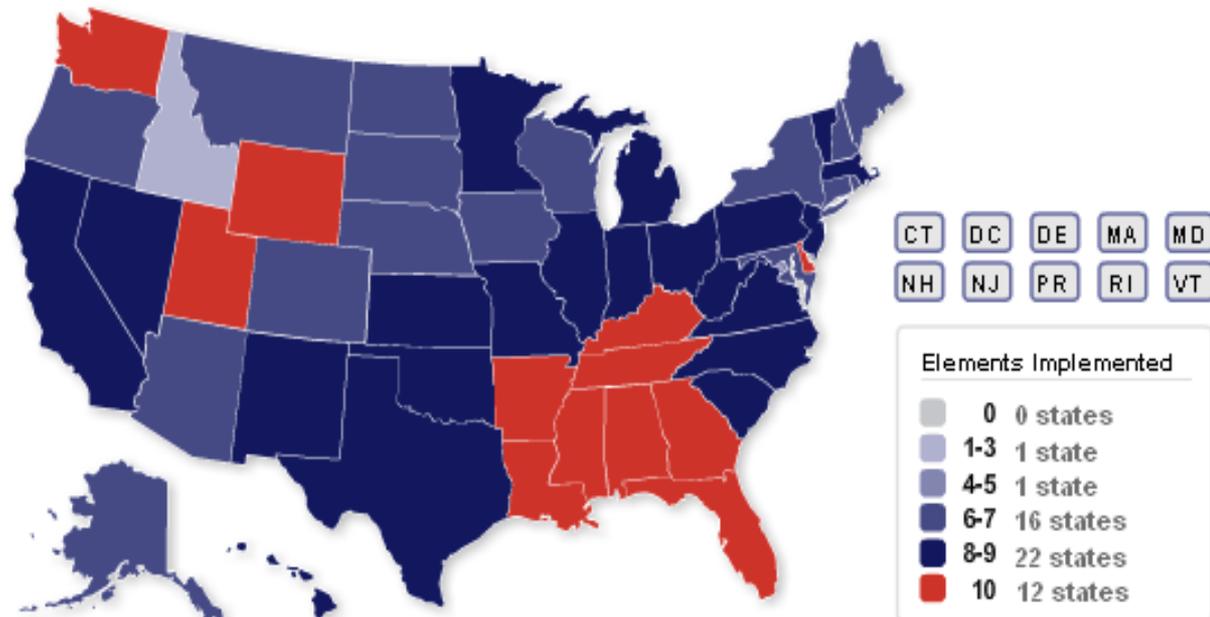


DATAQUALITY CAMPAIGN

Using Data to Improve Student Achievement

2009-10 Survey Results by State

States have made remarkable progress in developing longitudinal data systems that can follow student progress over time, from early childhood through 12th grade and into postsecondary education through implementation of the 10 Essential Elements. The 10 State Actions are the fundamental steps states must put in place to change the culture around how data are used to inform decisions to improve system and student performance. The maps below show responses to



Census – A long History

History of the Census From Ancient Romans to GPS Technology

© [Elizabeth Linehan](#)

May 27, 2009

Today's US Census is a far cry from its humble beginnings some 2500 years ago. And it's still evolving.



The first recorded census took place in the 5th century b.c. in Rome. Under the rule of Servius

http://americanaffairs.suite101.com/article.cfm/history_of_the_census

Chinese Census during Western Han Dynasty (206 BC – 220 AD)

Population ~ 56-57 million

http://www.guoxue.com/shibu/24shi/hansu/hsu_038.htm

京兆尹，故秦内史，高帝元年属塞国，二年更为渭南郡，九年罢，复为内史。武帝建元六年分为右内史，太初元年更为京兆君。元始二年，户十九万五千七百二，口六十八万二千四百六十八。县十二：长安，高帝五年置。惠帝元年初城，六年成。户八万八百，口二十四万六千二百。王莽曰常安。新丰，骊山在南，故骊戎国。秦曰骊山，高祖七年置。船司空，莽曰船利。蓝田，山出美玉，有虎侯山祠，秦孝公置也。华阴，故阴晋，秦惠文王五年更名宗秦，高帝八年更名华阴。太华山在南，有祠，豫州山。集天官，武帝起。莽曰李城也。郑，周宣王弟郑桓公邑。有铁官。湖，有周天子祠二所。故曰胡，武帝建元元年更名湖。下

BD6A，南陵，文帝七年置。沂水出蓝田谷，北至霸陵入霸水。北入渭，与泾水合。有周天子祠二所。名以章霸功。视子孙。奉明，宣帝置也。霸陵，故芷阳，文帝更名。莽曰水章也。杜陵。故杜伯国，宣帝更名。有周右将军杜主祠四所。莽曰饶安也。

Rise of Statistics

History of statistics

From Wikipedia, the free encyclopedia

http://en.wikipedia.org/wiki/History_of_statistics

Statistics arose, no later than the 18th century, from the need of states to collect data on their people and economies, in order to administer them. Its meaning broadened in the early 19th century to include the collection and analysis of data in general. Today statistics is widely employed in government, business, and the natural and social sciences.

Sampling – The Introduction

International Statistical Institute, 55th Session 2005

International Statistical Institute, 55th Session 2005: Colm O'Muircheartaigh

Balancing Statistical Theory, Sampling Concepts, and Practicality in the Teaching of Survey Sampling

Colm O'Muircheartaigh

*University of Chicago, NORC and Harris Graduate School of Public Policy,
1155 East 60th Street, Chicago IL 60637, USA*

colm@norc.uchicago.edu

抽样调查理论与实践 (1995), 中国统计出版社。
ISBN 7-5037-1670-3

抽样调查一百周年

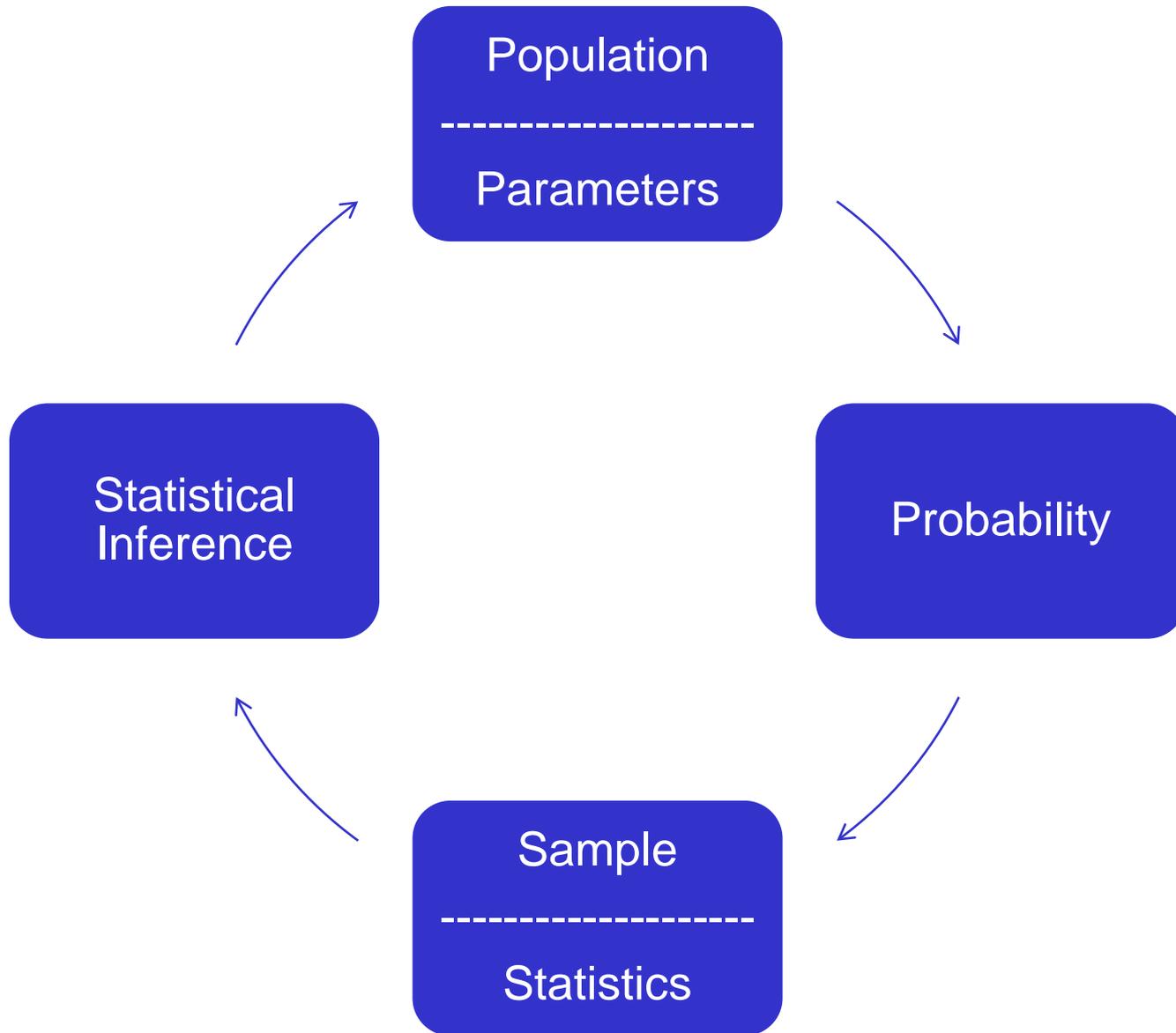
胡善庆 著
张尧庭 于祥 译

1995年将是统计史上一个重大转折的一百周年纪念。依照 You Poh Seng (1951)、以及卡尔顿 (Graham Kalton) (1988) 的论述, 这是因为 1895年, 在瑞士伯尔尼召开的国际统计学会的报告会上, 挪威统计局局长 A. N. Kiaer (凯尔) 在阐明他的“代表性调查”的思想时, 在历史上首次引入了抽样的概念, 并引起了国际统计学界的关注。

2. History of Survey Sampling

The origins of survey sampling are in official government statistics in the late 19th and early 20th centuries; the provenance is non-mathematical, and the motivation is practical and operational. [A more extensive version of the material in this section can be found in O'Muircheartaigh and Soon (1981).]

One of the finest discussions of the philosophy of sampling took place at the International Statistical Institute (ISI) meetings in Berne in 1895. This was the first recorded occasion on which the statistics establishment (then primarily the collectors of data for government purposes) was confronted with a coherent plea for the use of samples in data collection. A N **Kiaer**, the director of the Norwegian Bureau of Statistics, presented a report on his experience with sampling surveys and advocated further investigation of the field. Kiaer's definition of a representative investigation was that it was a partial inquiry in which observation was made on a large number of units distributed



Sampling – The Reception

Michael Cowles (1936, 2001). *Statistics in Psychology: an historical perspective*. Lawrence Erlbaum Associates, Inc. New Jersey.
http://books.google.com/books?id=8pRuRm6WAp8C&dq=a.+n.+kiaer&source=gbs_navlinks_s

pling, pressing for complete enumeration. It took almost 30 years for the utility and benefits of the methods to be appreciated. Seng (1951) and Kruskal and Mosteller (1979) give accounts of this most interesting period in statistical history. The latter authors give a translation and paraphrase of the remarks of Georg von Mayer, Professor at the University of Munich, on Kiaer's work on the *representative method*, which was presented at a meeting of the Institute in Berne in 1895:

I regard as most dangerous the point of view found in his work. I understand that representative samples can have some value, but it is a value restricted to terrain already illuminated by full coverage. One cannot replace by calculation the real observation of facts. A sample provides statistics for the units actually observed, but not true statistics for the entire terrain.

It is especially dangerous to propose representative sampling in the midst of an assembly of statisticians. Perhaps for legislative or administrative goals sampling may have uses – but one must never forget that it cannot replace a complete survey. It is necessary to add that there is among us these days a current in the minds of mathematicians that would, in many ways, have us calculate rather than observe. We must remain firm and say: no calculations when observations can be made. (von Mayer, quoted by Kruskal & Mosteller, 1979, pp. 174–175)

Sampling – Acceptance in the U.S.

Developing Sampling Techniques

The Census Bureau first used statistical sampling methods in the 1937 test survey of the United States during the Great Depression, but it also served as a "check" on a large-scale survey.

The Census Bureau implemented statistical sampling in a decennial census for the first time in 1950 due to the increasing cost or respondent burden. Enumerators asked a random sample of the population for demographic data for the entire United States.

Sampling became a fixture of the decennial censuses, with a certain percentage of the population being asked of a sample of the population.

The success of statistical sampling in the decennial census contributed to the expansion and renaming of the Current Population Survey in 1947. The Census Bureau's figures.

http://www.census.gov/history/www/innovations/data_collection/developing_sampling_techniques.html



[View larger image](#)

The Census Bureau used statistical sampling in the 1937 Survey of Unemployment.

Sampling: Evolution in U.S. Census

(Thanks to W. G. Cochran)

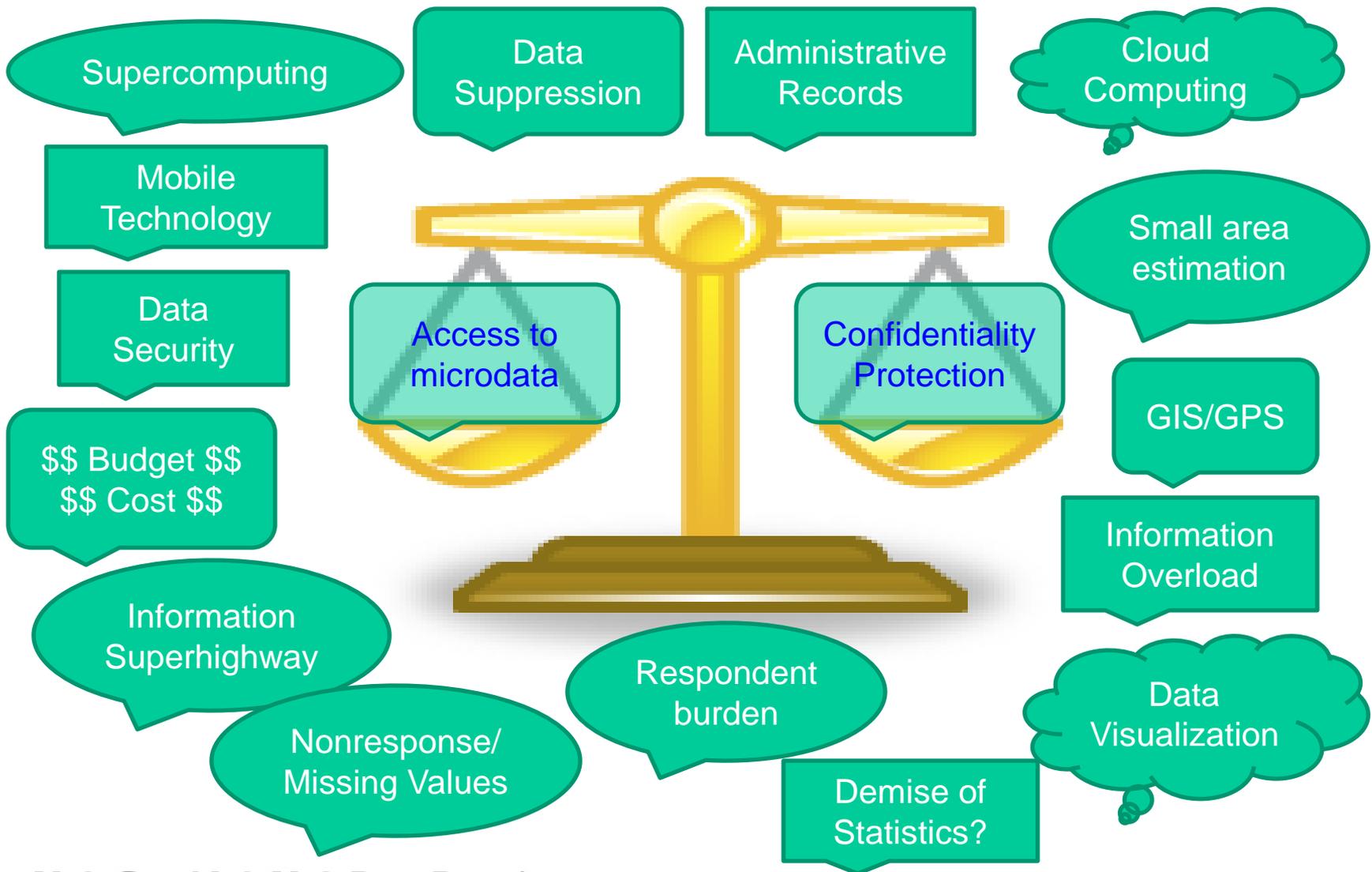
- 1940 5% sample on occupation, ...
- 1950 20% sample on income, schools, ...
1 in 6 sample of 20% on marriage, ...
- 1960 25% sample ...
- 2000 “Short” form – for all
“Long” form – 1 in 6 sample
- 2005 American Community Survey
~1 in 8 sample over 5 years

Two Lessons Learned

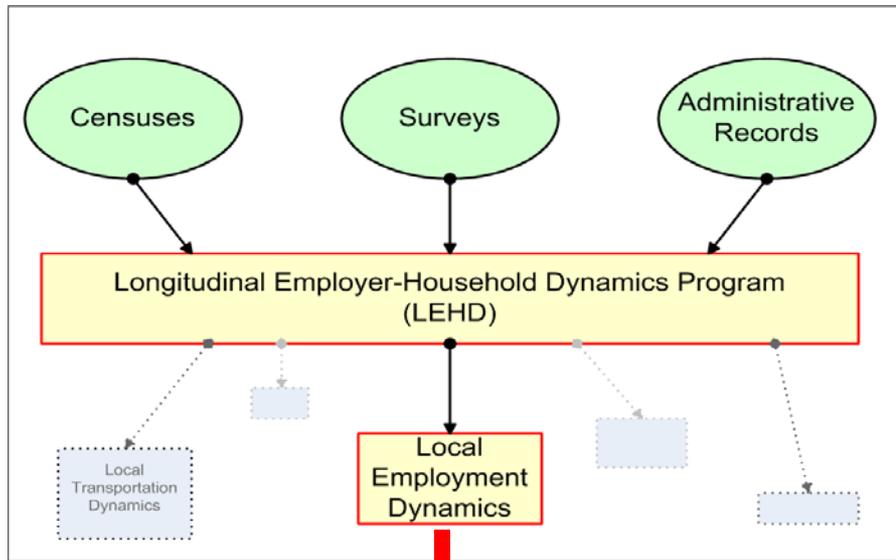
A 5% random sample
is “better” than
a 5% non-random sample
in measurable terms

Study of statistics was saved;
Mathematical Statistics was born

The Information Re/Evolution



The LED Approach

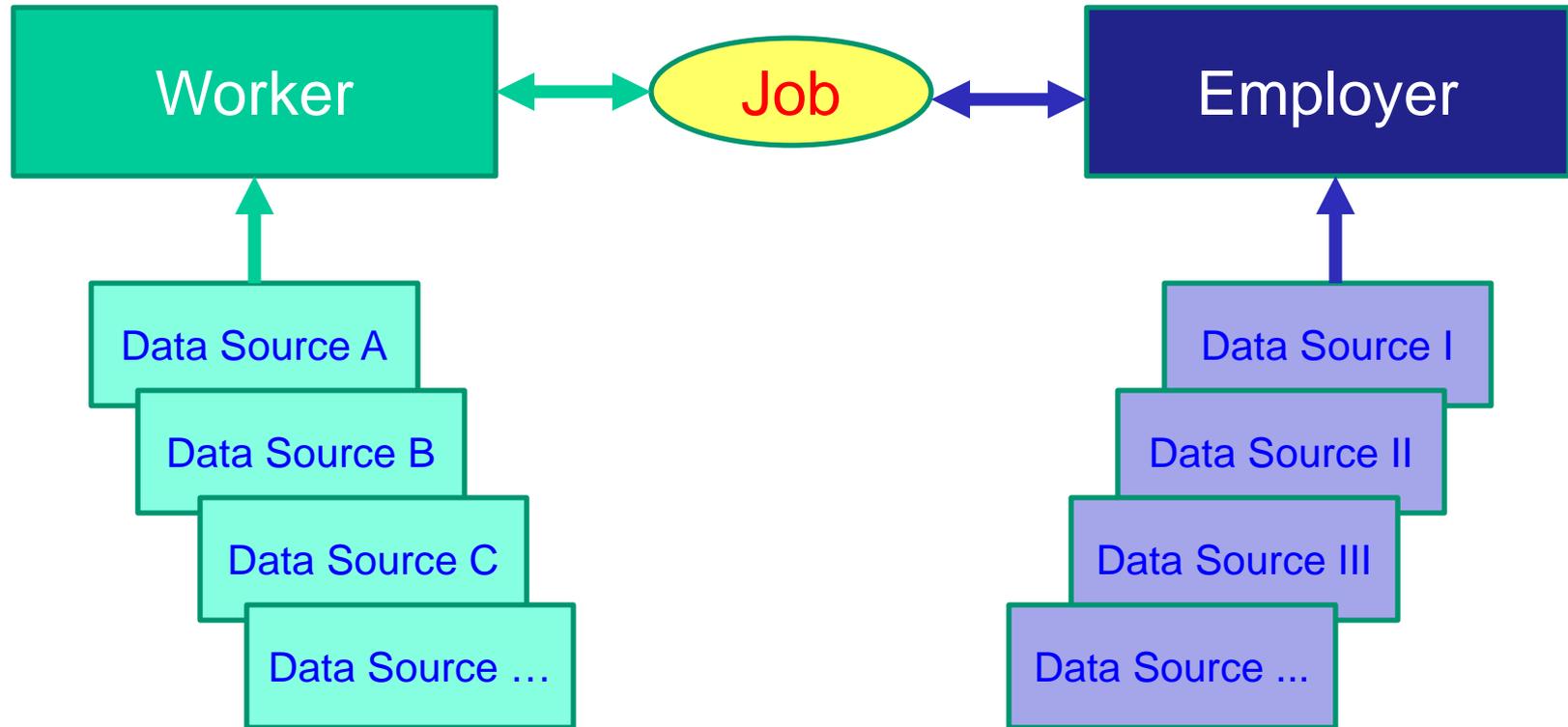


- Started in 1998
- Use existing data
- Create new data and products
- Make valid detailed data available and protect confidentiality

Longitudinal National
Frame of Jobs

New data and products

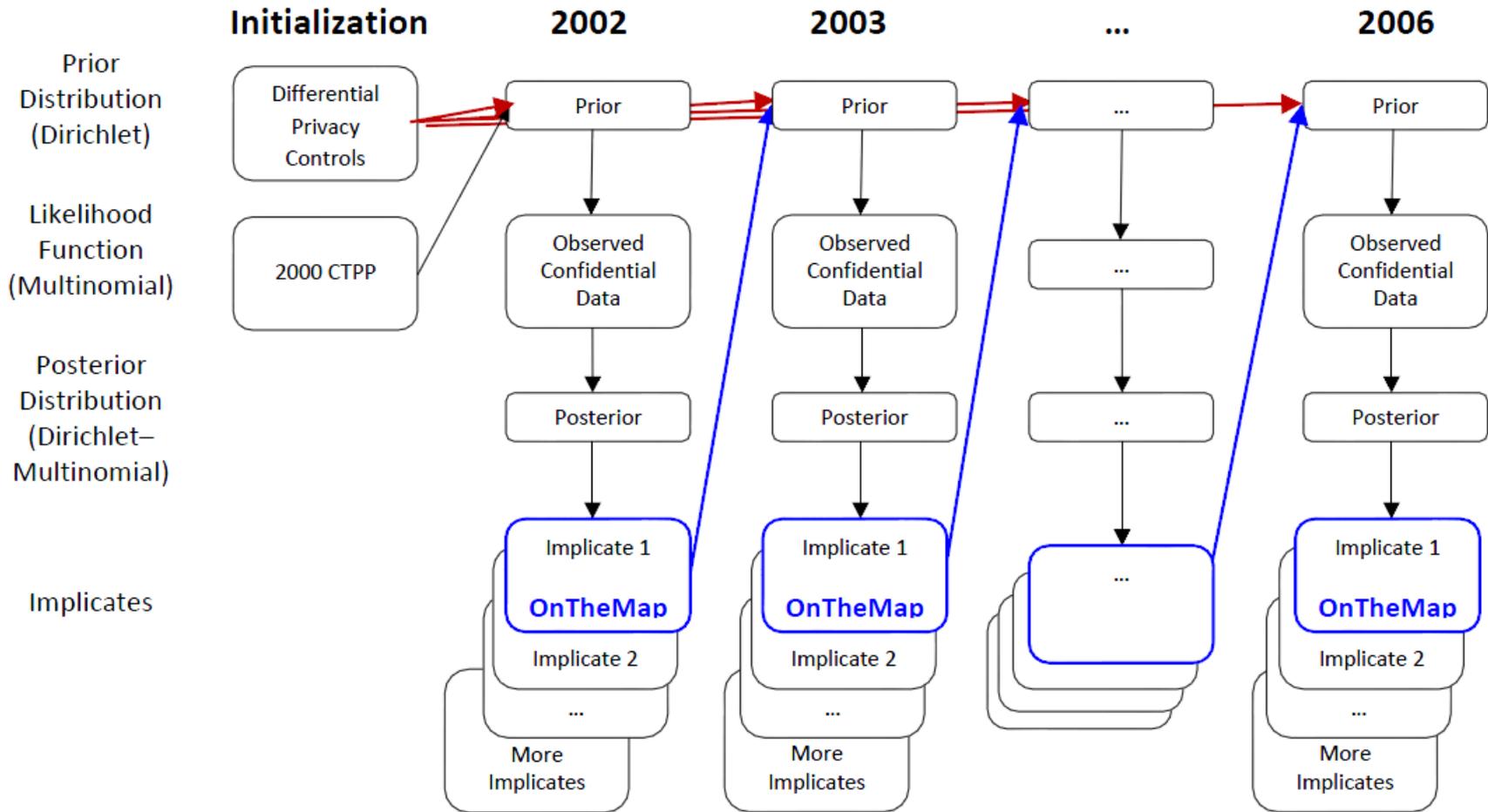
LED: Longitudinal Integrated Data



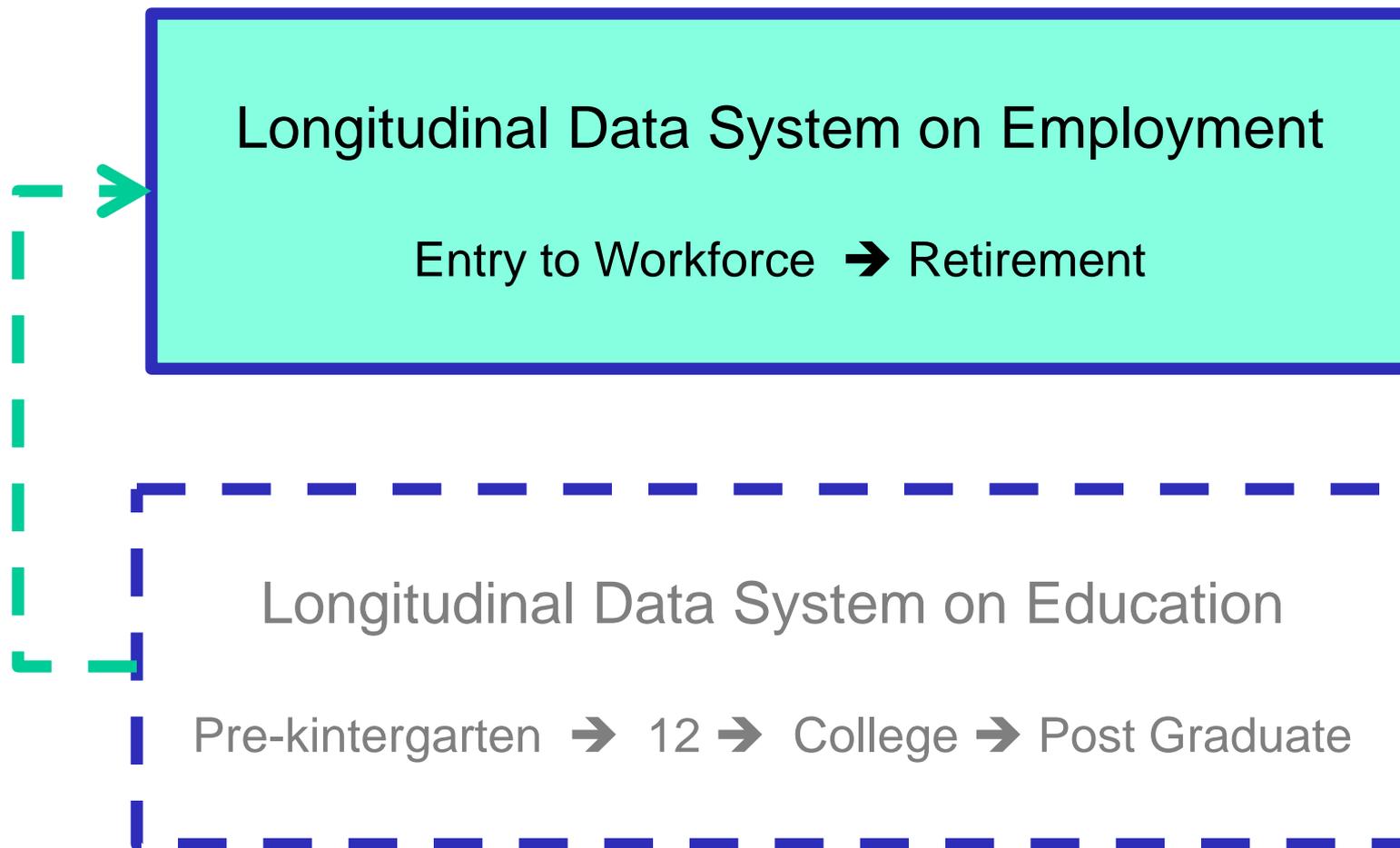
LED: Key Methodologies

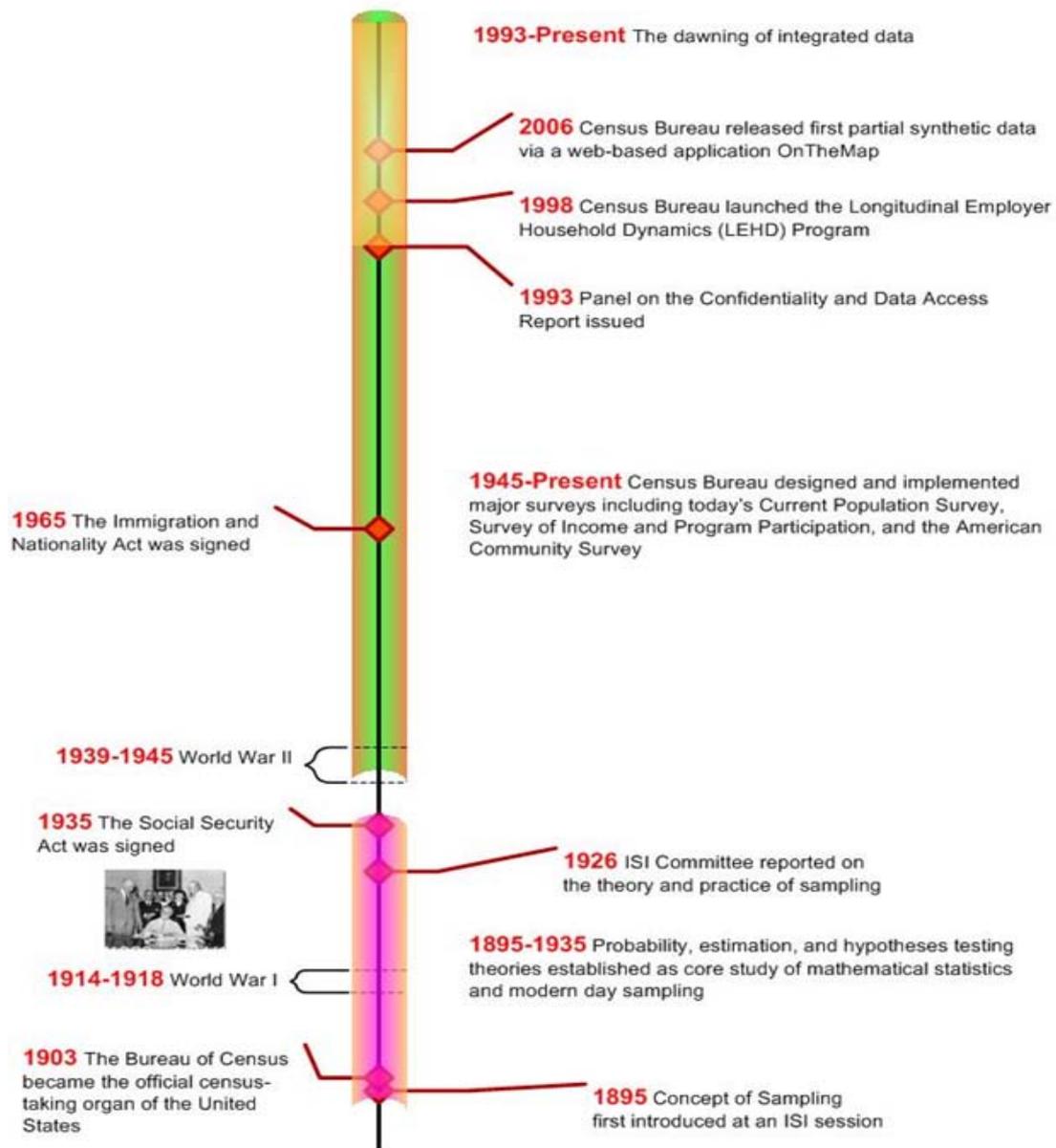
- System Design
- Record Linkage
- Noise Infusion
- Imputation
- Synthetic Data Modeling

OnTheMap: Partial Synthetic Data Product



Potentials of Longitudinal Data Systems on Flow Statistics





The Unknowns

- Is an 80% non-random sample “better” than a 5% random sample in measurable terms? 90%? 95%? 99%?
- Is the demise of statistics imminent? Or is this the dawning of another new field of statistical knowledge?

Contact Us

Comments/Suggestions

CES.local.employment.dynamics@census.gov

Jeremy.S.Wu@census.gov

Local Employment Dynamics Website

<http://lehd.did.census.gov>

Join the Listservs

Lehd-general@lists.census.gov



<http://www.census.gov>

People & Households [American Community Survey](#) · [Estimates](#) · [Projections](#) · [Housing](#) · [Income](#) | [State Median Income](#) · [Poverty](#) · [Health Insurance](#) · [International](#) · [Genealogy](#) · [Census 2000](#) · [More](#)

Business & Industry [Economic Census](#) · [Get Help with Your Form](#) · [Economic Indicators](#) · [NAICS](#) · [Survey of Business Owners](#) · [Government](#) · [E-Stats](#) · [Foreign Trade](#) | [Export Codes](#) · [Local Employment Dynamics](#) · [More](#)

Geography [Maps](#) · [TIGER](#) · [Gazetteer](#) · [More](#)



U.S. Census Bureau

Longitudinal Employer-Household Dynamics

Home | [Local Employment Dynamics](#) | [Data Tools](#) | [Research](#) | [Library](#) | [About Us](#)

[Site Map](#) | [Feedback](#) | [Ask Census](#)

Print

QUICK LINKS	WHAT'S NEW?
<ul style="list-style-type: none"> Tools Release Schedule OnTheMap Version 4 (Help) QWI Online (Help) Job Gains/Losses Industry Focus Older Worker Profiles 	<ul style="list-style-type: none"> 05/27/10: OnTheMap Version 4.1 Released 05/13/10: OnTheMap Version 4.1 Coming 04/21/10: New OnTheMap e-Learning Module Available 03/22/10: 2010 LED Partnership Workshop Successfully Conclud 02/26/10: OnTheMap Data Available for HTTP Download <p>2010 LED Partnership Workshop</p> <p>Click here to send an email to join the OnTheMap listserv.</p>

<http://lehd.did.census.gov>

U S C E N S U S B U R E A U