

Comments on:
**Economic Directorate's Big Data and Passive
Data Collection Overview**

Roberto Rigobon
MIT, NBER, CSAC, CNStat

General Overview

- Very impressive set of projects
 - System to System collection
 - Eliminate repeated requests
 - Reduces firm's burden
 - Tax Web scraping
 - Collect local tax revenues
 - Public
 - Identify relevant document
 - Organize data collection
 - Retail Payment Company
 - Large sample on US payments at POS
 - Private
 - Data curation still unclear
 - Engagement

New questions after my visit

- How to make the engagement between statistical offices and private sector sustainable?
 - Data or Sufficient Statistic?
 - Public versus Private?
 - In the case of private: Financial engagement?

Assumptions

1. In general, the statistical office does not need the full granularity the data exhibits.
 - We are interested in certain aggregates and geographical statistics
 - We are interested in improving and validating data we already collect
 - We are interested in reducing burden of surveys (both firms and statistical office)
2. In general, privacy and integrity of the data and data provider is crucial
3. In general, the access to data sources has a long run horizon.
 - Reliability in the collection of the data is an important dimension
4. In general, not everything can be a mandate

Forms of Engagement

- Public Data
 - Example
 - The tax revenue project
 - Web-scraping and other techniques are viable
 - Analysis can be done over the cloud: instead of saving and downloading all the data,
 - Only a sufficient statistic needs to be recorded
 - Privacy of the time series needs to be addressed
 - One data point is public, but is the time series also public?
 - Burden on companies or data providers
 - Technical burden due to web-scraping needs to be evaluated
 - Fear over how the data might be used
 - The data is public, but is it public due to a mandate or voluntary
 - No financial concerns
 - Payment is not expected

Forms of Engagement

- Private Data
 - Example
 - The Retail Payment Company
 - Questions
 - Access: System-to-system?
 - Where the computation takes place?
 - The data transferred to the statistical office and then the economic and social indicator is computed
 - The data is not transferred only a sufficient statistic
 - » Economic indicator is developed in collaboration with the firm
 - How to compensate the firm for the data?
 - Financial:
 - » Pay for data: Not incentive compatible
 - » Mandate or collaboration?
 - Privacy
 - » How to guarantee?

Two ways to organize our thinking

- Private Data

	Micro Data
Collection	<ul style="list-style-type: none">• Collect micro data from firms
Computation	<ul style="list-style-type: none">• Computation at the SO
Knowledge	<ul style="list-style-type: none">• Resides at the SO
Technology	<ul style="list-style-type: none">• Burden on the SO
Privacy	<ul style="list-style-type: none">• Guaranteeing Privacy is a Concern
Financial	<ul style="list-style-type: none">• Payment for micro data is sustainable

Two ways to organize our thinking

- Private Data

	Micro Data
Collection	<ul style="list-style-type: none">• Collect micro data from firms
Computation	<ul style="list-style-type: none">• Computation at the SO
Knowledge	<ul style="list-style-type: none">• Resides at the SO
Technology	<ul style="list-style-type: none">• Burden on the SO
Privacy	<ul style="list-style-type: none">• Guaranteeing Privacy is a Concern
Financial	<ul style="list-style-type: none">• Payment for micro data is sustainable

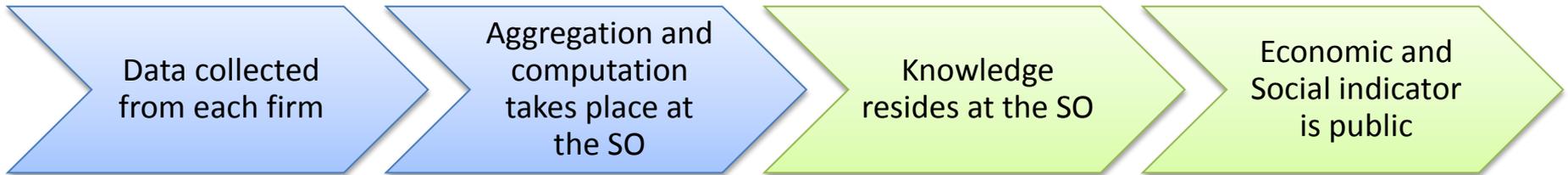
Two ways to organize our thinking

- Private Data

	Micro Data	Sufficient Statistic
Collection	<ul style="list-style-type: none">• Collect micro data from firms	<ul style="list-style-type: none">• Collect a sufficient statistic
Computation	<ul style="list-style-type: none">• Computation at the SO	<ul style="list-style-type: none">• Computation at the firm level
Knowledge	<ul style="list-style-type: none">• Resides at the SO	<ul style="list-style-type: none">• Join ownership: Collaboration
Technology	<ul style="list-style-type: none">• Burden on the SO	<ul style="list-style-type: none">• Small burden for companies
Privacy	<ul style="list-style-type: none">• Guaranteeing Privacy is a Concern	<ul style="list-style-type: none">• Privacy is better protected
Financial	<ul style="list-style-type: none">• Payment for micro data is sustainable	<ul style="list-style-type: none">• Not incentive compatible

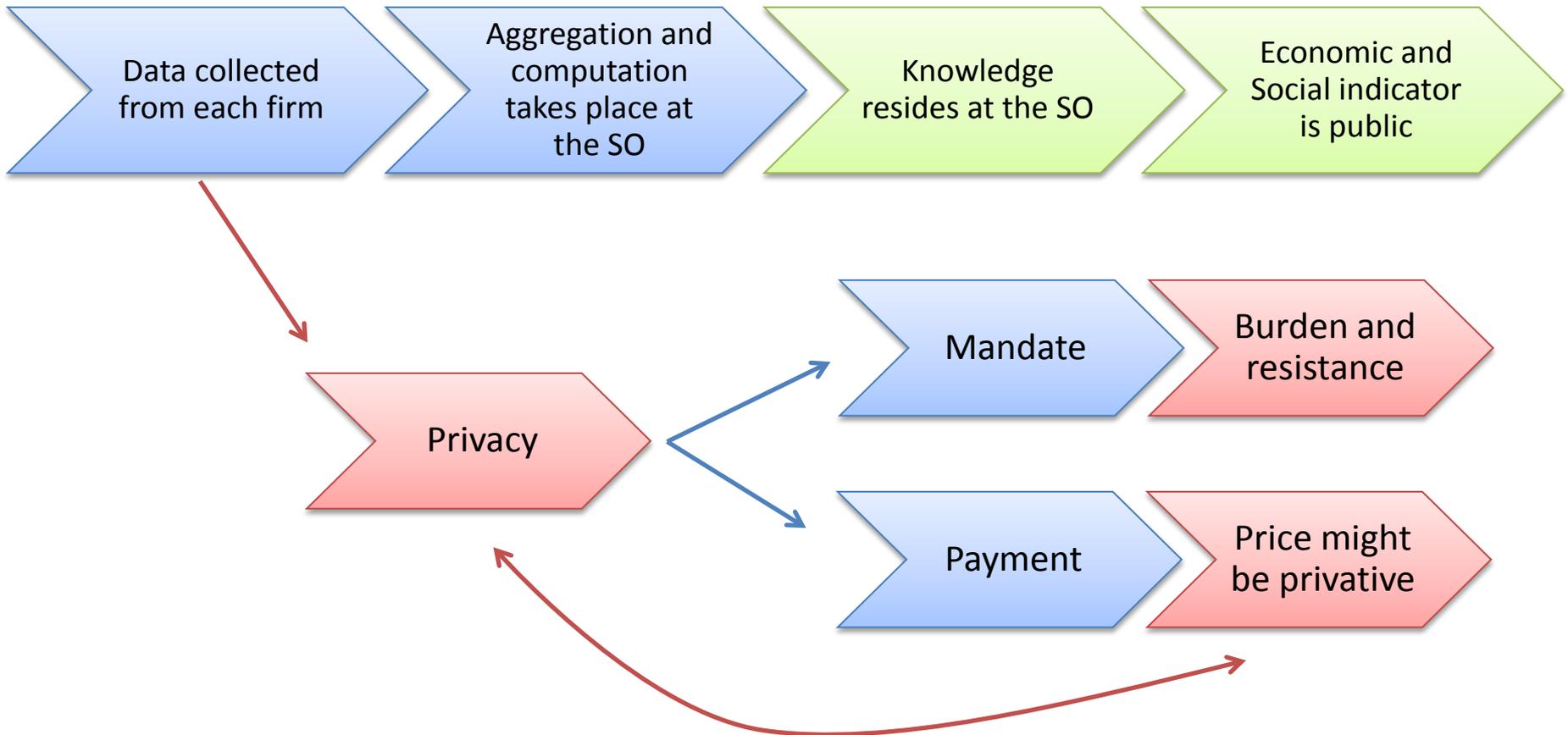
Public good

- Micro Data



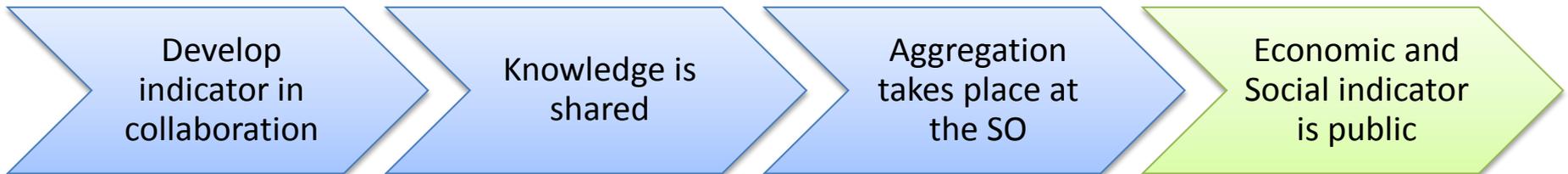
Public good

- Micro Data



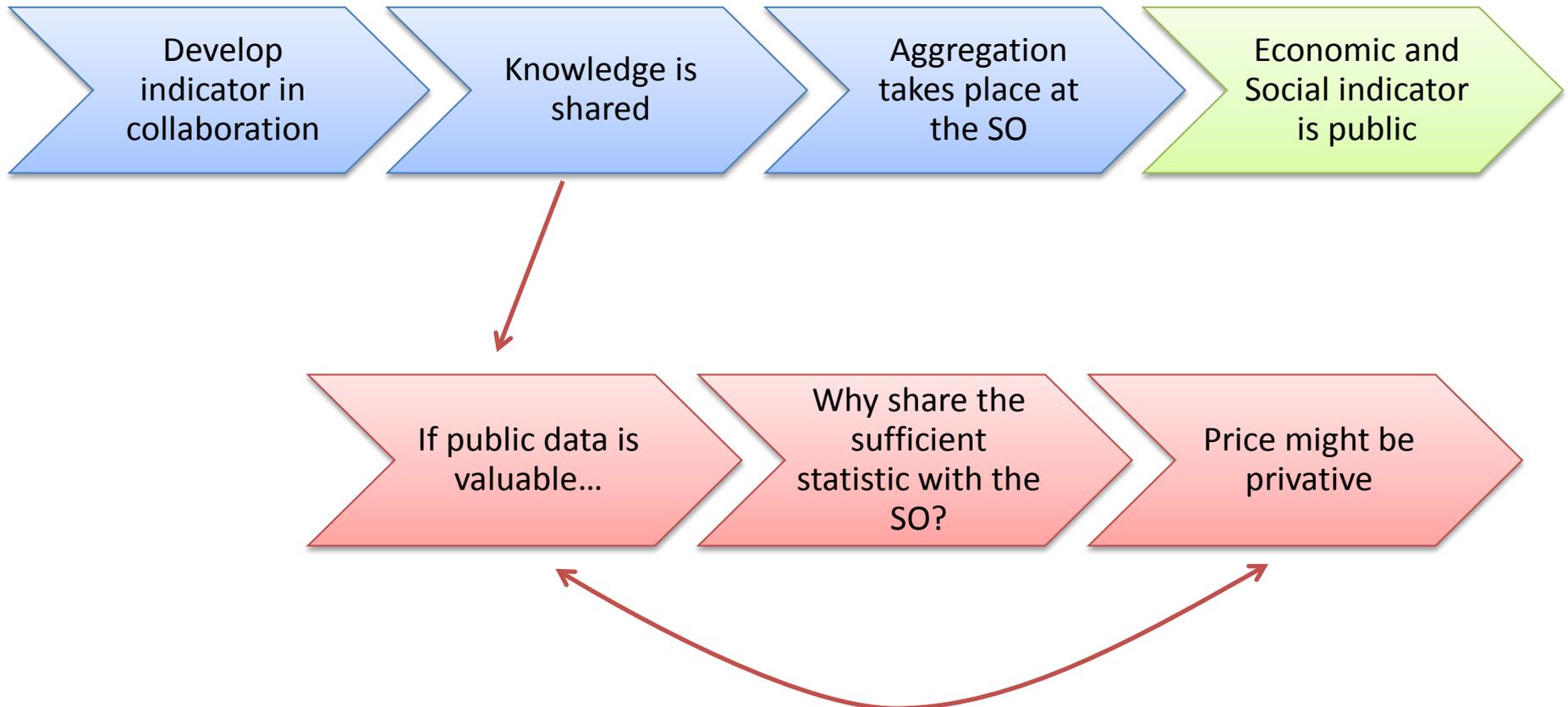
Public good

- Sufficient Statistic



Public good

- Sufficient Statistic



Final comments

- These questions will be addressed in the next face of the Retail Payment System in the very short run
 - I do not have the answer
 - What the Census decides to do might become guidelines for other statistical offices
- Excellent set of projects
 - Proof of concept
 - They highlight what can be done
 - They will improve economic indicators
 - Engagement is the next step