

Big Data: Experiences and Challenges

Gabriel Quirós

Deputy Director, STA, IMF

*Harnessing Big Data & Machine Learning Technologies for
Central Banks*

Banca d'Italia

Video Conference, 26 March 2018



Outline

I. Background

II. High Potential

III. Considerable Challenges

IV. Big Data for Statistics

V. Big Data and Artificial Intelligence

N.B. There is no common definition of Big Data. Traditional Administrative Data can be considered a broad category within Big Data; see IMF SDN paper (Hammer, Kostroch and Quirós, 2017). In this PPP, I refer to Big Data originated in the private sector.



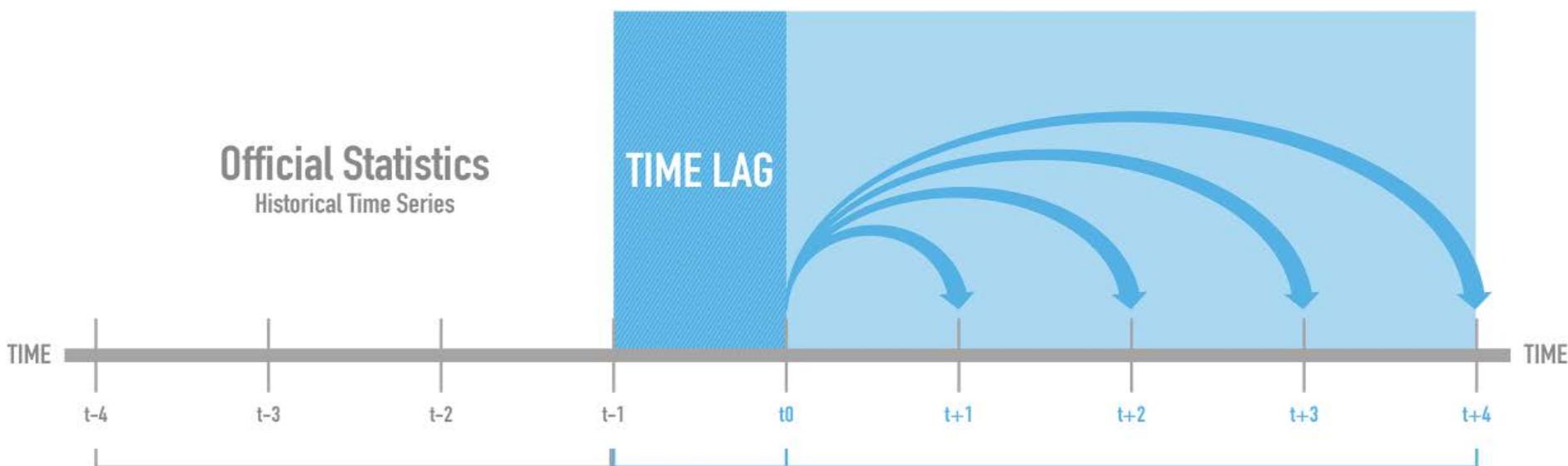
I. Background



- big data are not static, but a far-ranging *evolving concept* that requires a long-term vision
- a strategic organizational plan to deliver measurable and high-scale results



II. High Potential



3. Big data as an innovative data source in the production of official statistics

2. Big data to bridge time-lags of official statistics and support the forecasting of existing indicators

1. Big data to answer “new questions” and produce new indicators

III. Considerable Challenges

Data Quality

- quality assessments of indicators will be crucial to minimize governance, political, and reputational risks
- statistical techniques and methodologies best practices are needed to specifically address *veracity* and *volatility*
- metadata are key to assess and interpret new data sources



III. Considerable Challenges

■ Data Access

- Proprietary data held by the private sector
- Data that companies own may evolve from a byproduct to becoming a major asset
- Regular licensing costs come in addition to substantial investments into processing and storage solutions



IV. Big Data for Statistics

- In connection to the respective statistical domains, a number of Dos and Don'ts from big data can be identified, which are unevenly distributed across statistical domains
- The Dos and Don'ts are largely driven by the essence of big data: by-products of private technological and business models that capture behavior of consumers, corporates, banks, individuals or government agencies
- Big data are particularly promising to enhance directly or indirectly statistics on transactions, less so on stocks

IV. Big data for statistics

Dos

Big data, particularly promising at helping measure:

- ✓ “soft” information: sentiment, alerts, reactions...
- ✓ consumer behavior and patterns (e.g. Amazon, Google searches and ‘clicks’, social networks,...)
- ✓ Tourism (e.g. roaming information, Google searches, credit cards, click-stream data, ...)
- ✓ Financial flows (e.g. SWIFT, mobile phones, ...)
- ✓ Prices (scanner data,...)
- ✓ Job vacancies and labor skills (e.g. LinkedIn,...)
- ✓ big data provides granular, microdata
- ✓

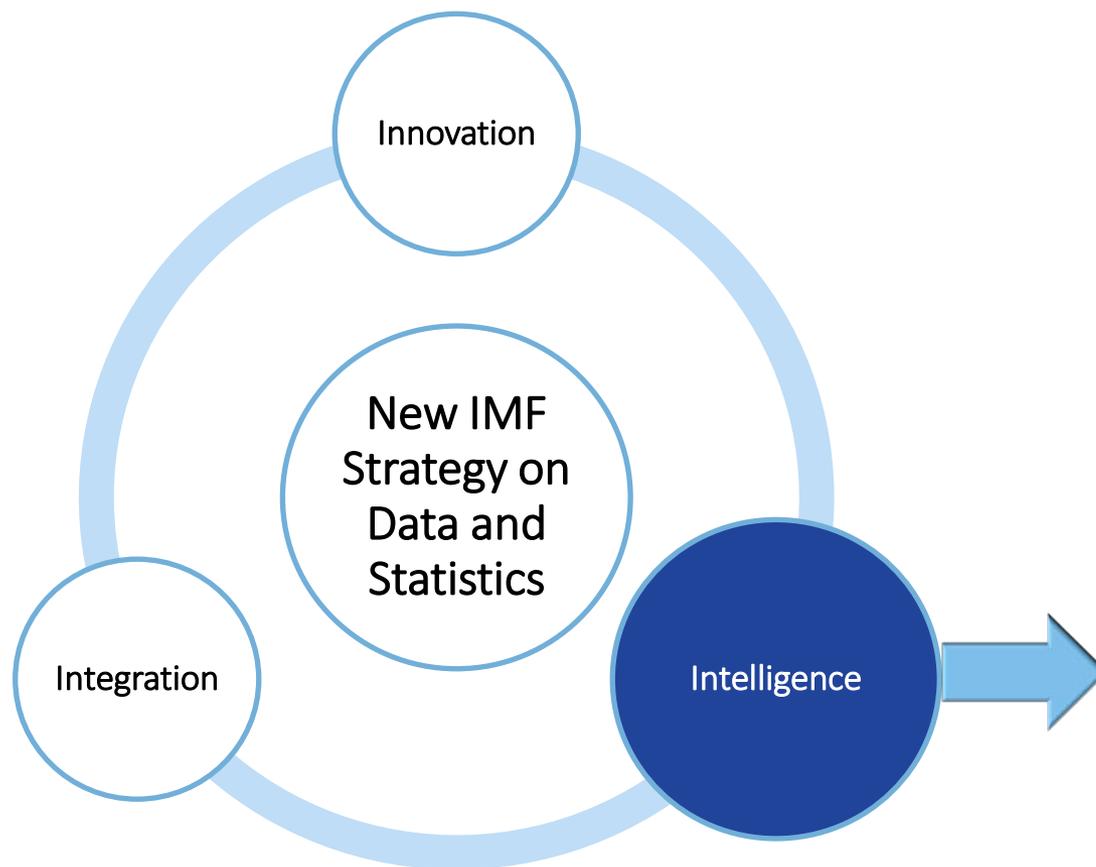
IV. Big Data for statistics

Don'ts

- Sample representativeness: bias towards more modern and dynamic economic activities and social behavior
- Big data less suited for stocks, i.e. total financial assets and liabilities of firms, households, government, non residents, both at micro and macro levels
- Revaluation and other volume changes, particularly important in monetary and financial statistics
- As by-product, long time-series based on big data are inexistent and will be fragile because instability from business and technological changes, discontinuity in data provision
- Privacy and confidentiality of personal, firm-level data



V. Big Data and Artificial Intelligence



Big data and artificial intelligence (AI) can help the Fund better tailor its policy advice to member needs

V. Big Data and Artificial Intelligence

- Potential of AI in the Fund
 - Machine learning and nowcasting
 - Text mining of Fund reports
 - Natural language generation to generate Fund reports
 - Automated data validation
 - Monitoring timeliness in data standards

- “AI and ML Innovation Challenge” just launched internally to submit proposals for applications in the Fund