Big Data and Machine Learning at the IMF

WORKSHOP ON "BIG DATA & MACHINE LEARNING APPLICATIONS FOR CENTRAL BANKS"

BANK OF ITALY - OCTOBER 21-22, 2019

Marco Marini

IMF Statistics Department

Please note that the views expressed here are those of the IMF's staff, and not necessarily those of the IMF's management or Executive Board

Big Data and Machine Learning at the IMF

- **Current Research Projects**

- Behind the Scene
  - Information Sharing, Access, Infrastructure

Emerging Trends in using Big Data and Machine Learning at the IMF
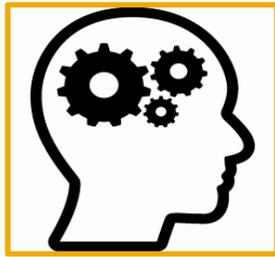
Artificial Intelligence/ Machine Learning for nowcasting and monitoring

Text Mining/Natural Language Processing for risk analysis and vulnerability exercise

Satellite images/ Geolocational data to improve economic measurement

Alternative data sources to improve timeliness and granularity

Innovative Methods   **Innovative Data**

Nighttime Lights to Measure Economic Growth

Hu and Yao, 2019 IMF Working Paper

Use satellite-recorded nighttime lights data to measure uncertainty in official GDP per capita

Nighttime lights processed by the National Oceanic and Atmospheric Administration (NOAA) from 1992 to 2013 and Visible Infrared Imaging Radiometer Suite (VIIRS) from 2014 to 2017 for 182 countries

Relationship between GDP and nighttime lights is nonlinear
- *Elasticity steadily decreases as real GDP per-capita increases, reflecting different developing mode at different stages of economic development*

Nighttime lights can play a big role in improving real GDP per capita measures for the majority of middle and low income countries

Nighttime Lights to Measure Economic Growth

Hu and Yao, 2019 IMF Working Paper

Figure 7. Real GDP per capita: Economic Disruption and Restoration
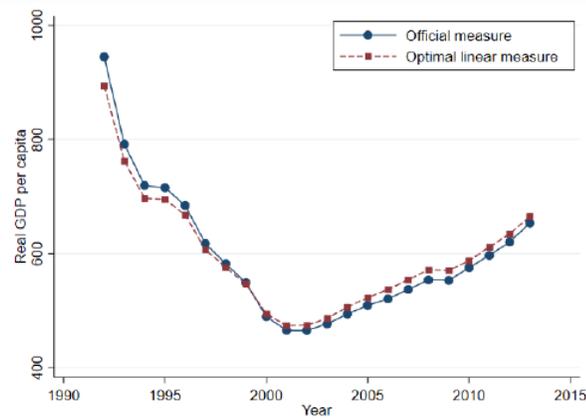
(a) Dem. Rep. of Congo
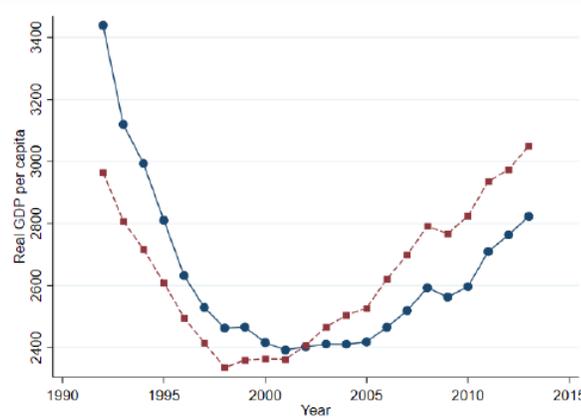
(b) Djibouti
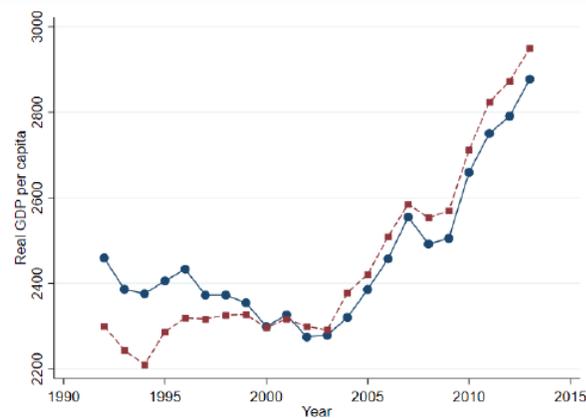
(c) Kenya
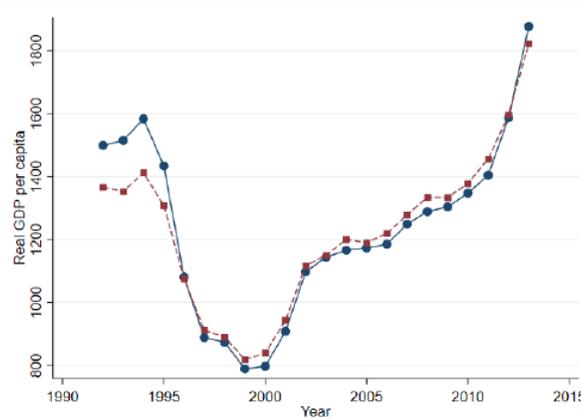
(d) Sierra Leone

GDP adjusted for nighttime lights shows lower levels during downturns and higher levels during recoveries

Further works on nighttime lights are progressing on measuring informality for low-income countries

Vessel Traffic Data to Measure Trade Volumes

Arslanalp, Marini, Tumbarello, 2019 IMF Working Paper, forthcoming

Nowcast trade statistics using port calls based on Automatic Identification System (AIS) data

Port calls for Malta international ports from 2015 to 2018. Source: MarineTraffic

Two-step approach to nowcast trade flows in real time.

- Filter to identify cargo ships involved in trade activity
- Derive a trade volume index based on information in the AIS data about the size of the vessel (i.e., deadweight tonnage) and changes in its cargo load (i.e., draught)

Vessel Traffic Data to Measure Trade Volumes

Arslanalp, Marini, Tumbarello, 2019 IMF Working Paper, forthcoming

Malta: Trade Volumes
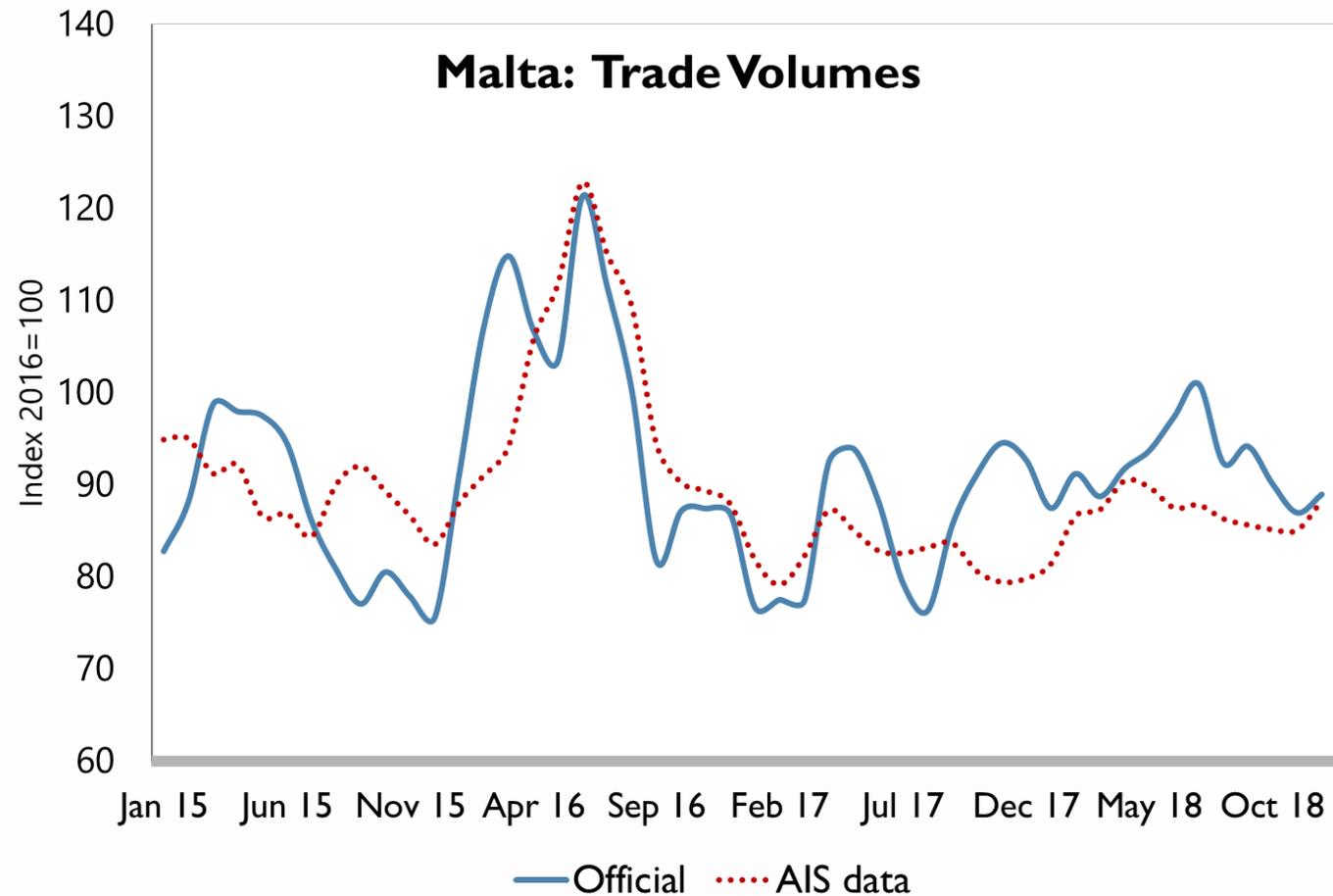
Sources: Eurostat, MarineTraffic, and authors' calculations.
Note: AIS=automatic identification system.

News-Based Sentiment Indicators

Huang, Simpson, Ulybina and Roitman, 2019 IMF Working Paper, forthcoming

Construct sentiment indices based on news as early warning indicators (EWIs) for financial crises

- Database of over 3 million news articles from the Financial Times from 1980 to 2019
- Semantic clustering techniques to capture sentiments

- Sentiment indices spike and/or trend up ahead of financial crises
- On average, general sentiment indices (e.g., positive) perform better than specific sentiment indices (e.g., fear)
- Consider a battery of different sentiments increases chances of detecting crises

News-Based Sentiment Indicators

Huang, Simpson, Ulybina and Roitman, 2019 IMF Working Paper, forthcoming

Semantic Clustering based on Computational Text Analysis (Word2Vec)
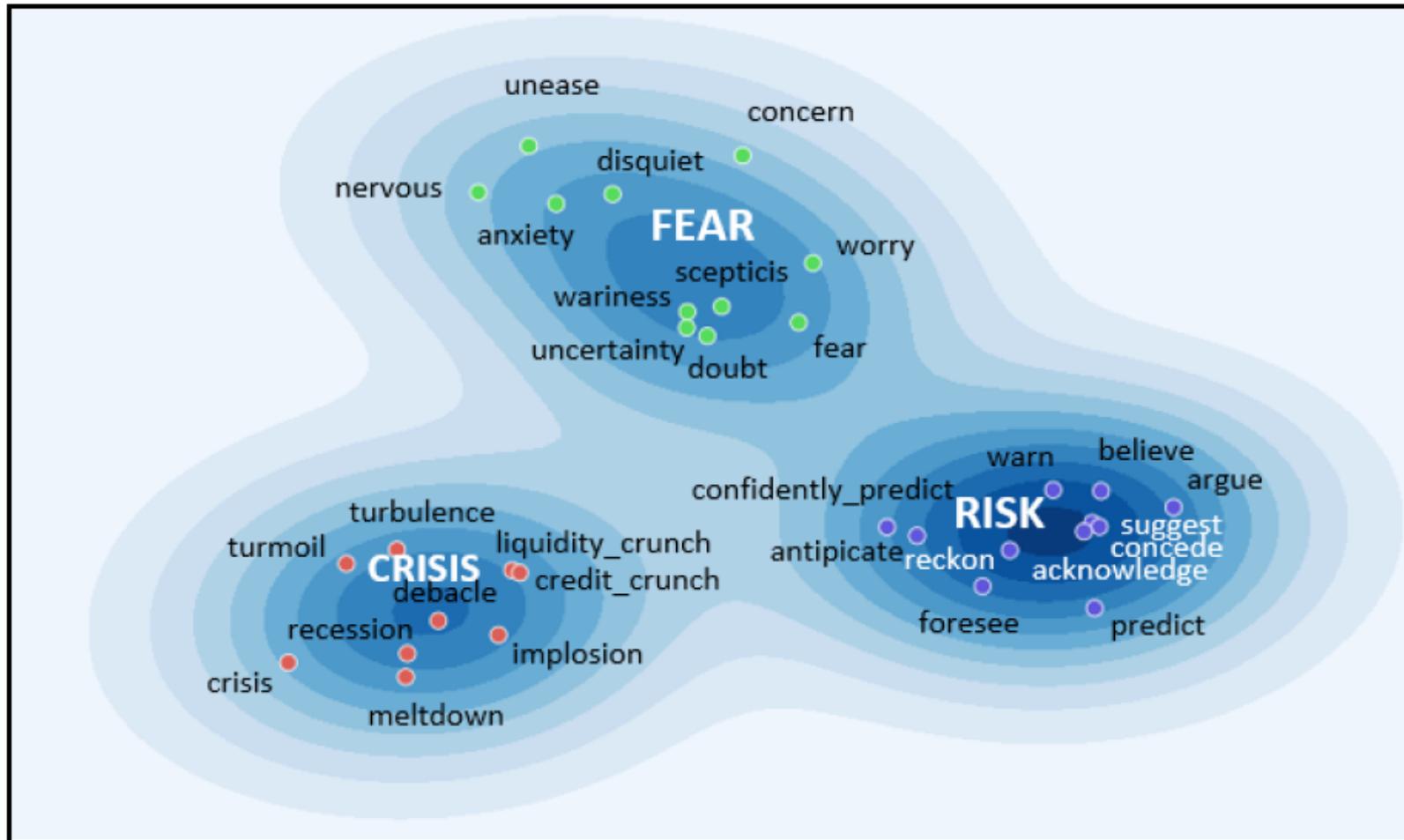
Machine Learning and Causality: The Impact of Financial Crises on Growth

Tiffin, 2019 IMF Working Paper, forthcoming

Use a 'causal' machine learning approach to estimate the impact of a financial crisis on output growth

- 46 macroeconomic and financial variables over 1985-2017 for 107 countries (both emerging markets and advanced economies)
- Shapley values used for interpreting the impact

- Average cumulative cost of a *hypothetical* crisis is estimated at 7.2 percentage points of growth over two years (consistent with previous estimates)
- Cost varies significantly from country to country, depending on individual characteristics and pre-crisis initial conditions

Machine Learning and Causality: The Impact of Financial Crises on Growth

*Tiffin, 2019 IMF Working Paper, forthcoming*

Figure 3. Distribution of Shapley Values
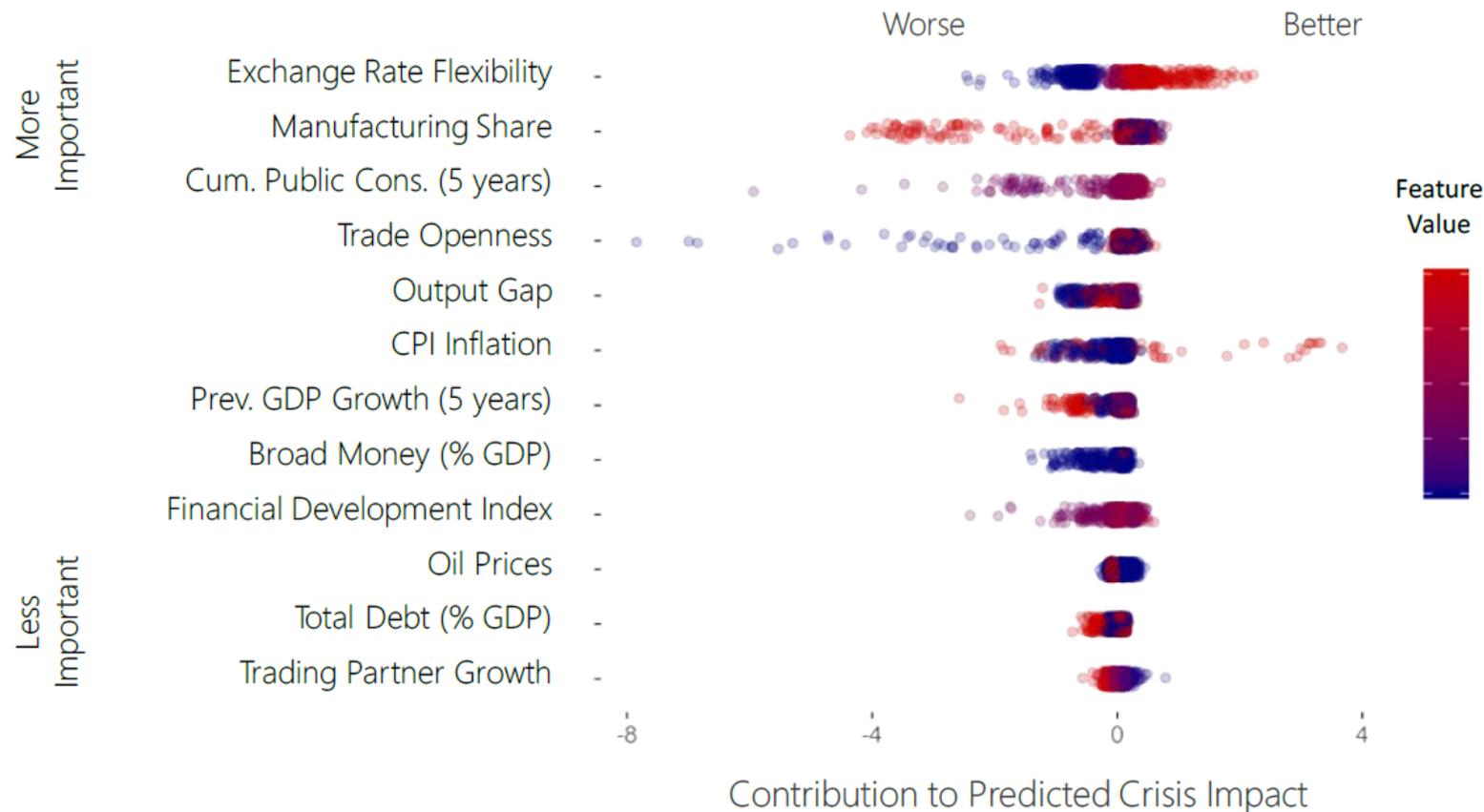
Study shows that 'Exchange-rate flexibility' and 'Manufacturing Share' as most important factors in shaping the impact of a financial crisis

Big Data and Machine Learning at the IMF

- Current Research Projects
- **Behind the Scene**
 - Information Sharing, Access, Infrastructure

Supporting the Use of Big Data in the IMF

Support the Use of Big Data

Build Internal Knowledge Sharing

Develop IT Infrastructure

Facilitate Data Access
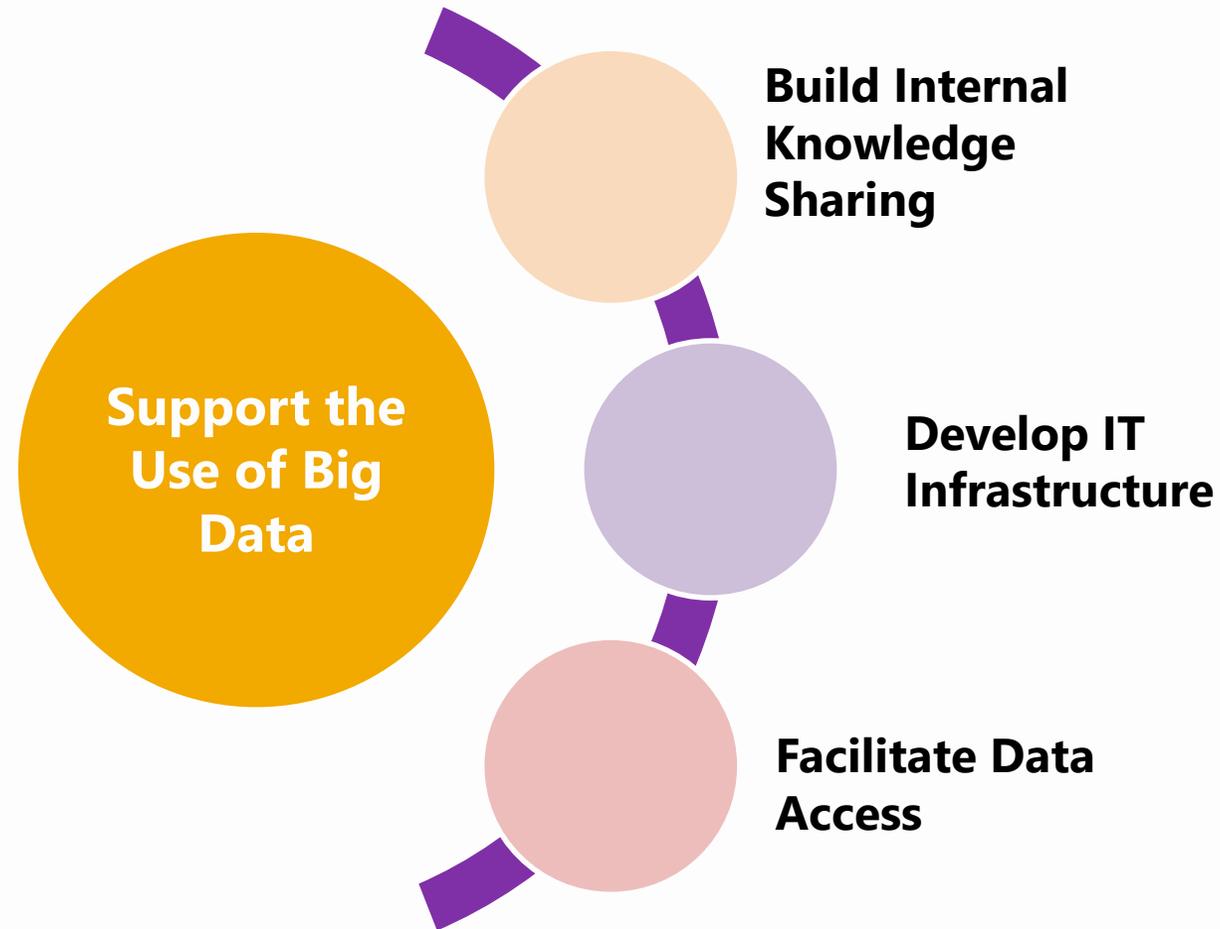
BigData@Fund: An Internal Community of Practice for Information Sharing and Collaboration

Support the Use of Big Data

Build Internal Knowledge Sharing

Sharing Knowledge and Exchange Technical Solutions

Leveraging Staff with Data Science Skills

Develop a Live Repository of Fund's Big Data Projects

Platform for Outreach to Peer Institutions

Big Data Lab: A Cloud-Based Platform to Support Storage and Analytics of Large Datasets

Projects

Flexibility

Scalability

Reproducibility

Deployment

DOMINO
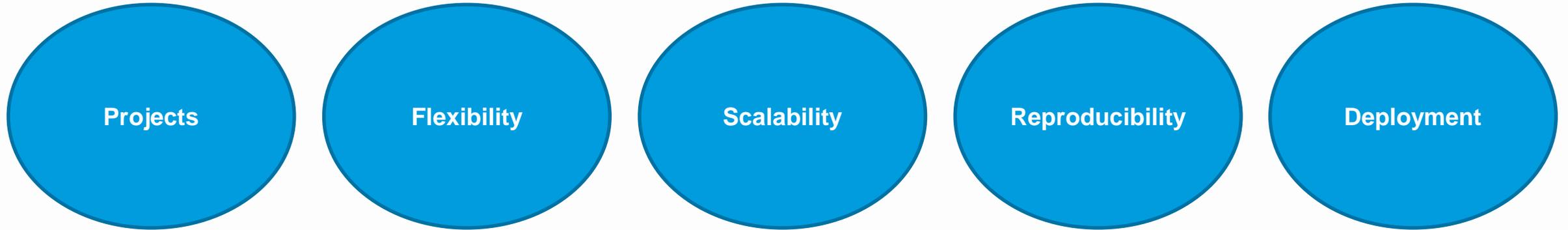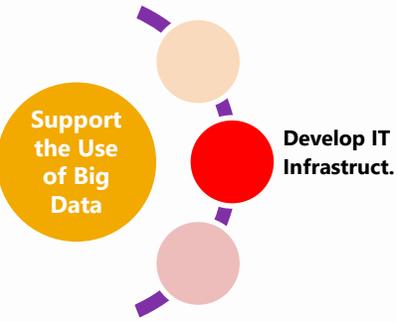
Proof of concept to offer Fund staff a cloud-based platform for data storage, code execution, and tool deployment for Big Data analytics
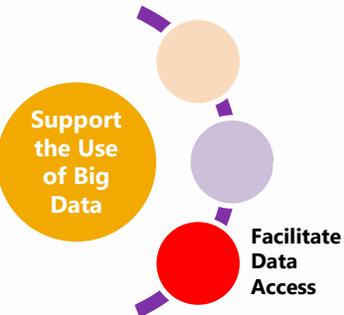
Data Collaboratives: A Consortium to Facilitate Data Sharing Agreements with Private Sector

o Key benefits for Consortium members and partners

  o Template data sharing framework agreement

  o Secure IT architecture to ingest, store, and access data

  o Repositories for derived data products and algorithms

Consortium Members

Data Partnerships Pipeline

Thank You!

For follow-up questions:
Marco Marini - mmarini@imf.org