PRAGMA BIG DATA PANEL

SATOSHI SEKIGUCHI
(AIST)

Abe Administration Statement　（安倍内閣宣言）

『世界最先端ＩＴ国家創造』宣言 – June, 2013
- ▶ Statement on "Forging the World-Leading IT Nation"

Basic Philosophy　（基本理念）
- ▶ Japan in Need of Revitalization　（閉塞打破、再生日本）
- ▶ Toward the Achievement of the World-Class IT-Driven Society　（世界最高水準IT利活用）
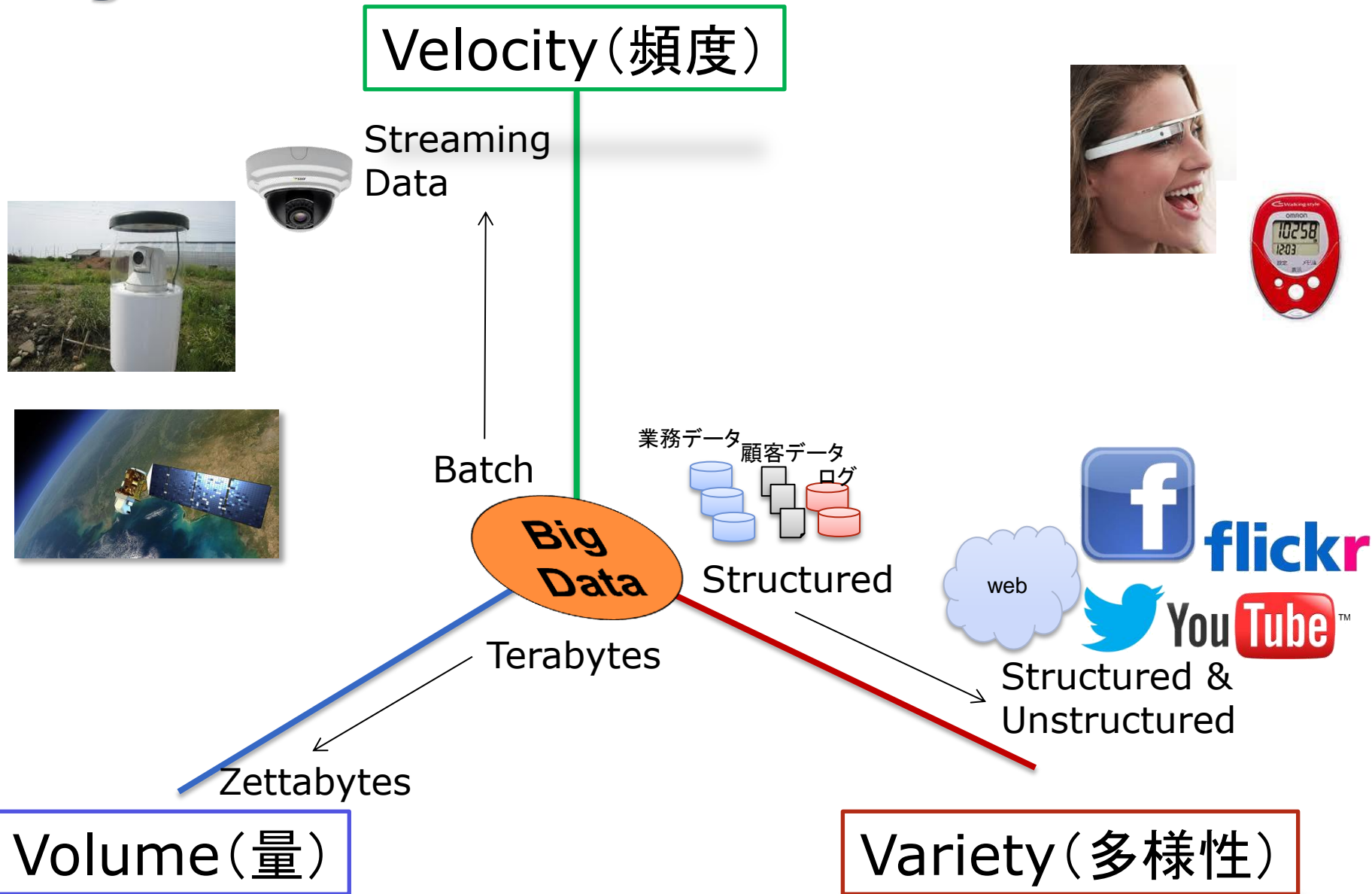
Abe Administration Statement （安倍内閣宣言）

Our Goals for the Future Society

1. A society that enhances the creation of innovative technologies and new services and the growth of their providers

    - 増強革新的新産業・新服務的創作&全産業成長促進社会

2. The most natural disaster-proof society in the world, in order to ensure a safe and healthy lifestyle

    - 健康/安心/快適生活 & 社会抗災能力/安全的世界

3. A society which guarantees that public services are available on a "one-stop" basis for anyone, anytime and any where
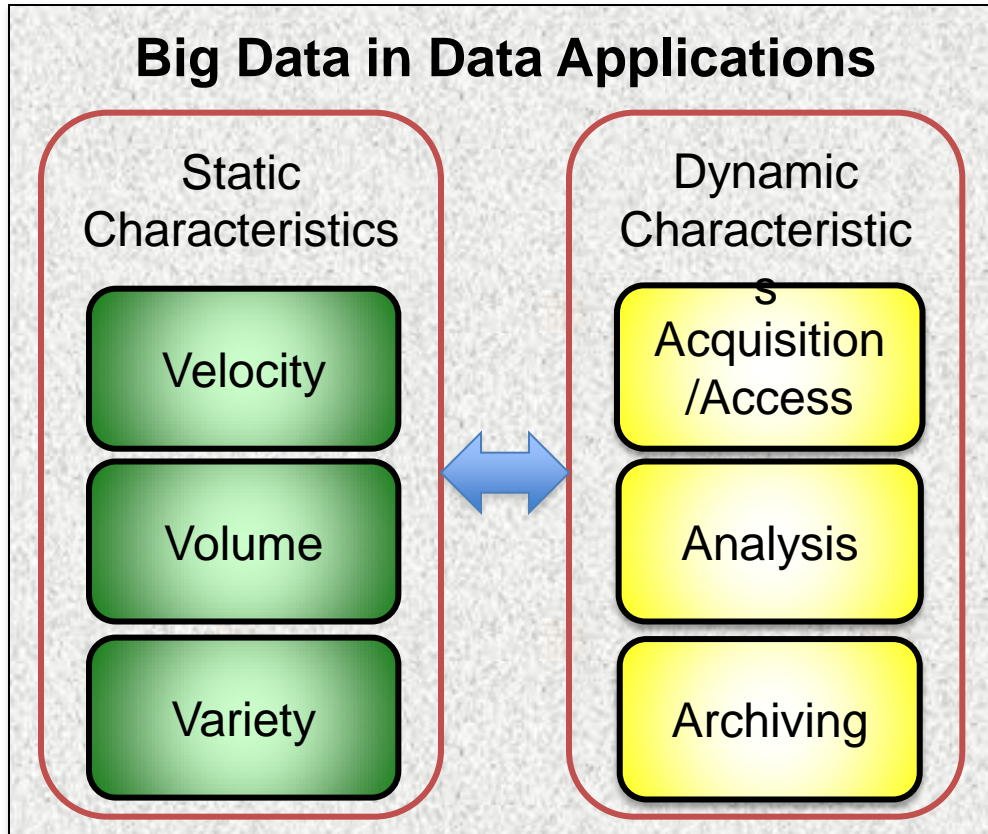
独立行政法人 産業技術総合研究所

Efforts to Make to Fulfill the Goals

1. Creation of innovative technologies…

    a. Promotion to make the best use of both Open data and Big data

        • *Open data* for public, *Big data* for commercial use

2. The most natural disaster-proof society in the world to ensure a safe and healthy lifestyle

    a. Realization of a healthy and long-lived society through appropriate community-based health care, nursing-care support, and health promotion

3. One-stop public services

    a. Innovation of the government information system across both local and national government

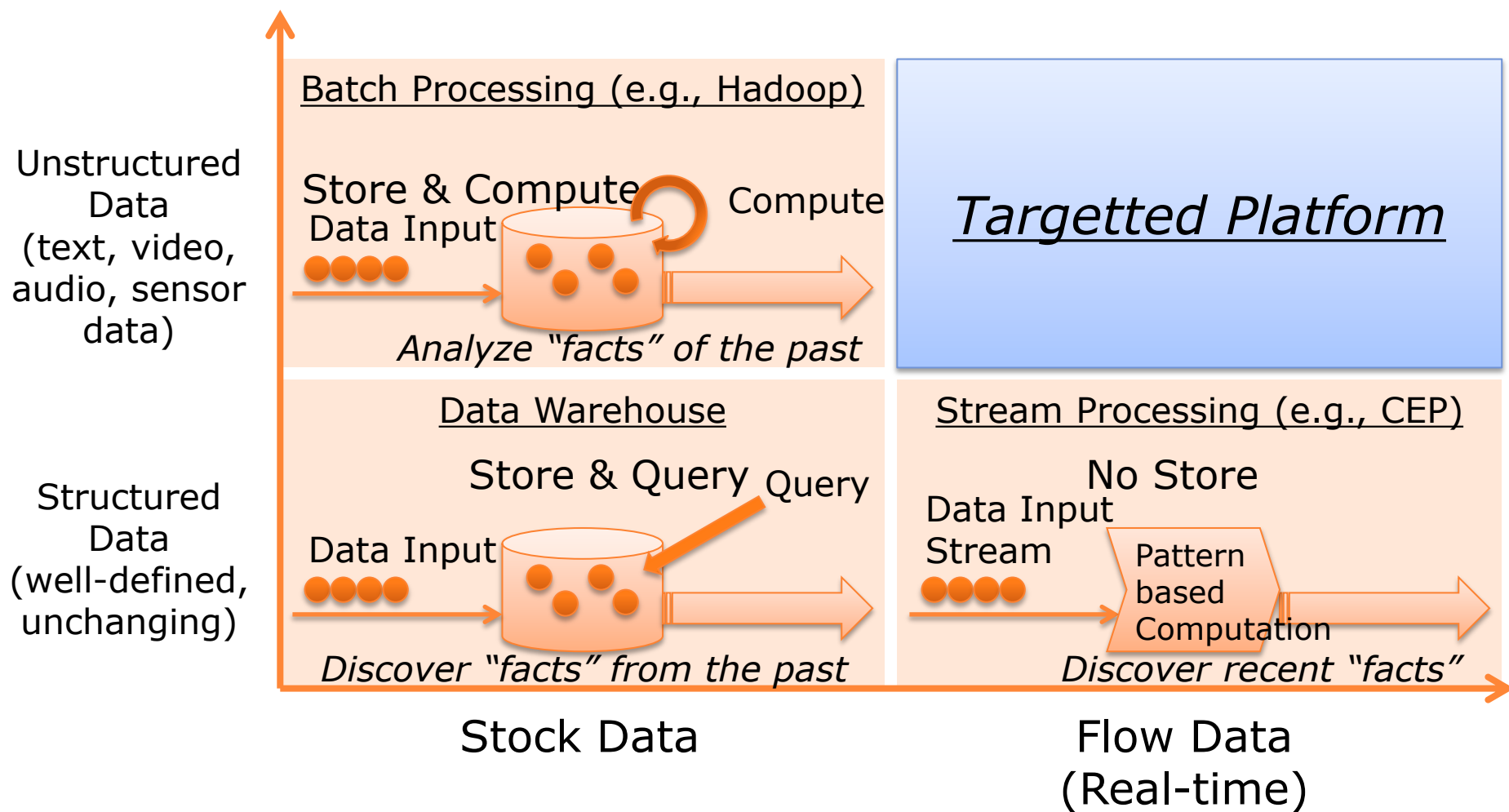   • "Cloud First" Concept for the system cost reduction

Integrating "Big Data" with Visualization/Human Computer Interfaces

Big Data in Data Applications

Static Characteristics		Dynamic Characteristics
Velocity		Acquisition /Access
Volume	⟷	Analysis
Variety		Archiving

- Data of all types are experiencing exponential growth
- To be useful, data must be acquired/accessed, analyzed, and archived

- Potential opportunities in
  - *Geoscience (sensor data)*
  - *Life sciences (drug discovery)*
  - *Manufacturing (cost reduction)*
  - *Meteorology/environment (climate modeling/prediction)*
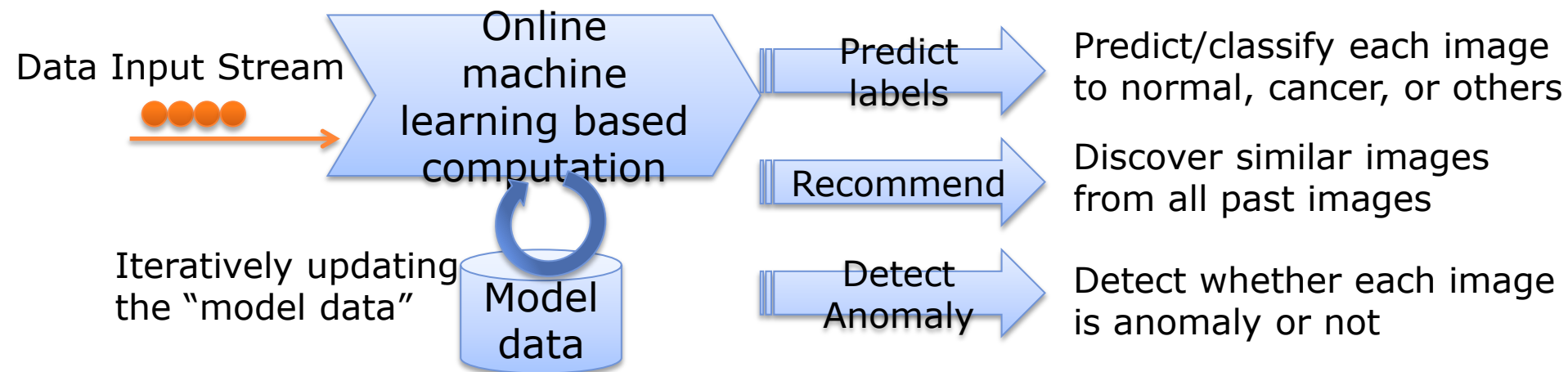- Requires a multidisciplinary approach

Existing Big Data Solutions

Batch Processing (e.g., Hadoop)

Store & Compute

Data Input

Compute

Analyze "facts" of the past

Targetted Platform

Data Warehouse

Store & Query Query

Data Input

Discover "facts" from the past

Stream Processing (e.g., CEP)

No Store

Data Input Stream

Pattern based Computation

Discover recent "facts"

Unstructured Data (text, video, audio, sensor data)

Structured Data (well-defined, unchanging)
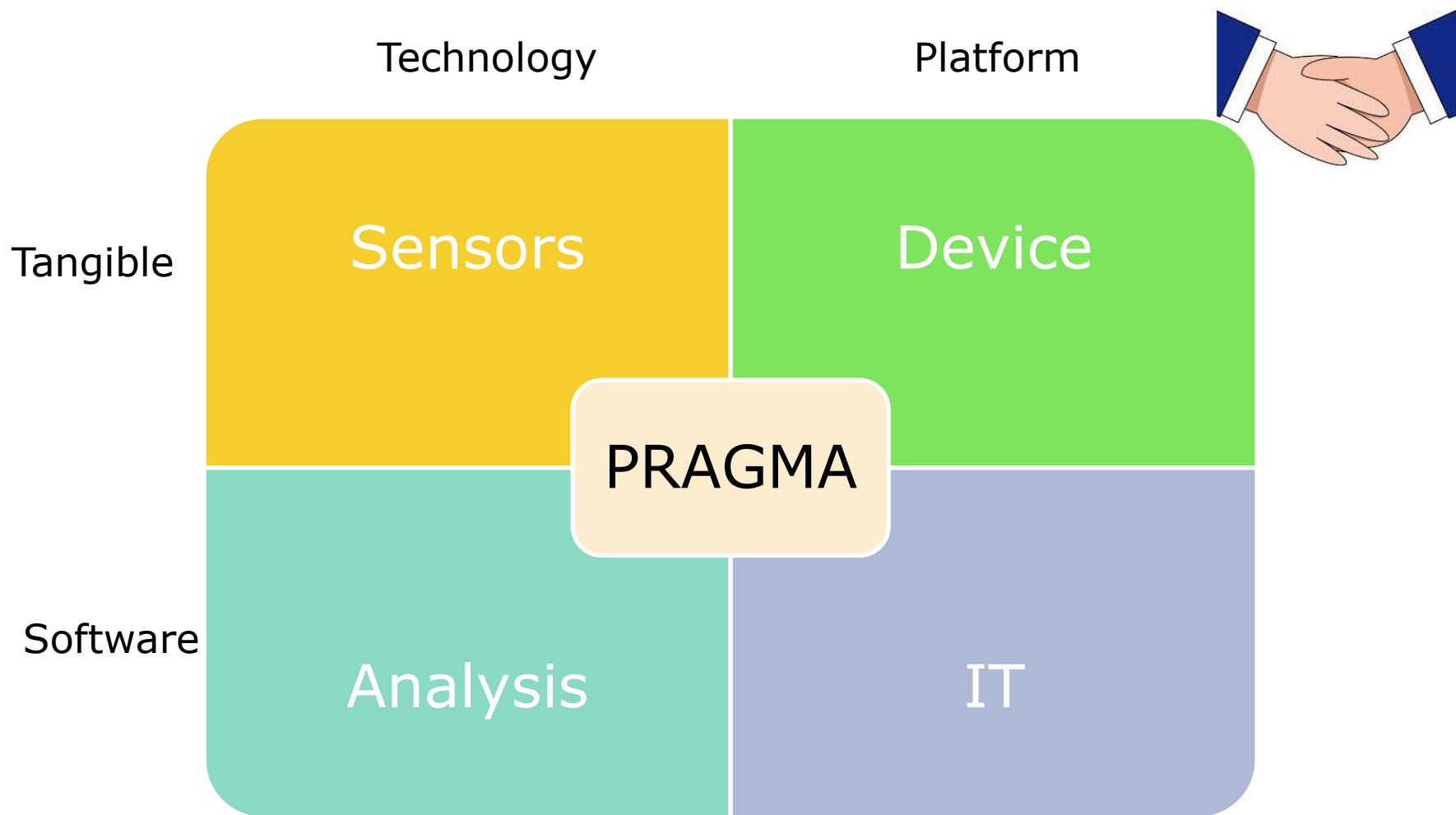
Stock Data

Flow Data (Real-time)

Real-time Analytics Platform for Big Data

- The platform must have a highly scalable online machine learning system
  - ▶ Continuously captures incoming streamed data
  - ▶ And performs deep analytics using machine learning algorithms, e.g., label prediction, recommendation, anomaly detection, etc.
  - ▶ Up to 10K real-time events can be processed in a second

  Discover "facts" from the past on real-time +
  Predict "future" using prior knowledge

Data Input Stream

Online machine learning based computation

Predict labels → Predict/classify each image to normal, cancer, or others

Recommend → Discover similar images from all past images

Detect Anomaly → Detect whether each image is anomaly or not

Iteratively updating the "model data"

Model data

Research Data Alianse

- Grid Reunited

- Call for idea & proposals

  - ▶ RDA 4th plenary meeting

  - ▶ September, 2014 Amsterdam