

# Big Data + Analytics: *Making Analytics as Services*

*Wo Chang*

Digital Data Advisor

NIST Big Data Public Working Group, Co-Chair

ISO/IEC JTC 1/SC42/WG2 Big Data, Convenor

IEEE BDGMM, Chair

wchang@nist.gov

May 26, 2020



# Standards Roadmap Subgroup

## NIST Big Data Interoperability Framework (NBDIF)

[https://bigdatawg.nist.gov/V3\\_output\\_docs.php](https://bigdatawg.nist.gov/V3_output_docs.php)

**NIST SP1500-1:  
Definitions**

**NIST SP1500-2:  
Taxonomies**

**NIST SP1500-3: Use  
Cases &  
Requirements**

**NIST SP1500-4:  
Security & Privacy**

**NIST SP1500-5:  
Architecture Survey –  
White Paper**

**NIST SP1500-6:  
Reference  
Architecture**

**NIST SP1500-7:  
Standards Roadmap**

**NIST SP1500-8:  
Reference  
Architecture Interface**

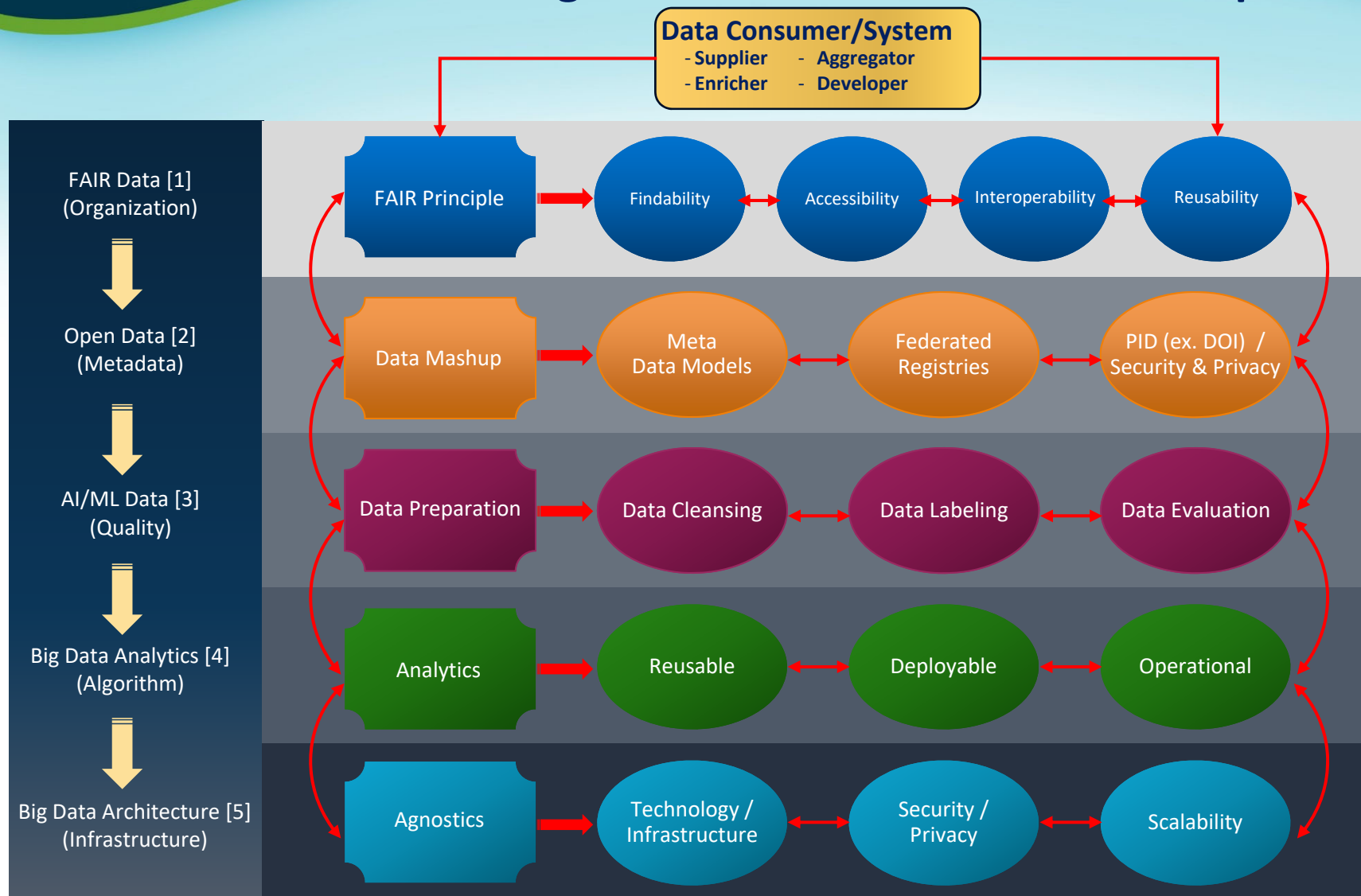
**NIST SP1500-9:  
Adoption &  
Modernization**

# ISO/IEC JTC 1/SC 42(AI)/WG2 Big Data Standards Activities

- Members: 220+ from 24 NBs: Australia, Austria, Belgium, Canada, China, Denmark, Finland, France, Germany, India, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Netherlands, Norway, Russian Federation, Singapore, Sweden, United Arab Emirates, UK, US
- ISO/IEC Projects

NBDIF Volumes	ISO/IEC Documents (publication date)
Vol. 1: Big Data Definitions	20546 Big data Overview and Vocabulary (Feb. 2019)
--	20547-1 Big Data Framework and Application Process (April 2020)
Vol. 2: Big Data Taxonomies	<i>Skip</i>
Vol. 3: Big Data Use Cases and Requirements	20547-2 Big Data Use Cases and Derived Requirements (April 2018)
Vol. 4: Big Data Security and Privacy	20547-4 Big Data Security and Privacy (Oct. 2020)
Vol. 5: Big Data Architecture Survey White Paper	<i>Skip</i>
Vol. 6: Big Data Reference Architecture	20547-3 Big Data Reference Architecture (March 2020)
Vol. 7: Big Data Standards Roadmap	20547-5 Big Data Standards Roadmap (April 2018)
Vol. 8: Big Data Reference Architecture Interfaces (new)	<i>Explore</i>
Vol. 9: Big Data Adoption and Modernization (new)	<i>Explore</i>
<b>Next Step: Explore Big Data + Analytics – Making Analytics as Services</b>	Plus other "data" related projects such as <ul style="list-style-type: none"> <li>• Process Management Framework for big data analytics</li> <li>• A series of data quality for analytics and machine learning</li> <li>• Data exploration</li> </ul>

# NIST Big Data Activities and Roadmap

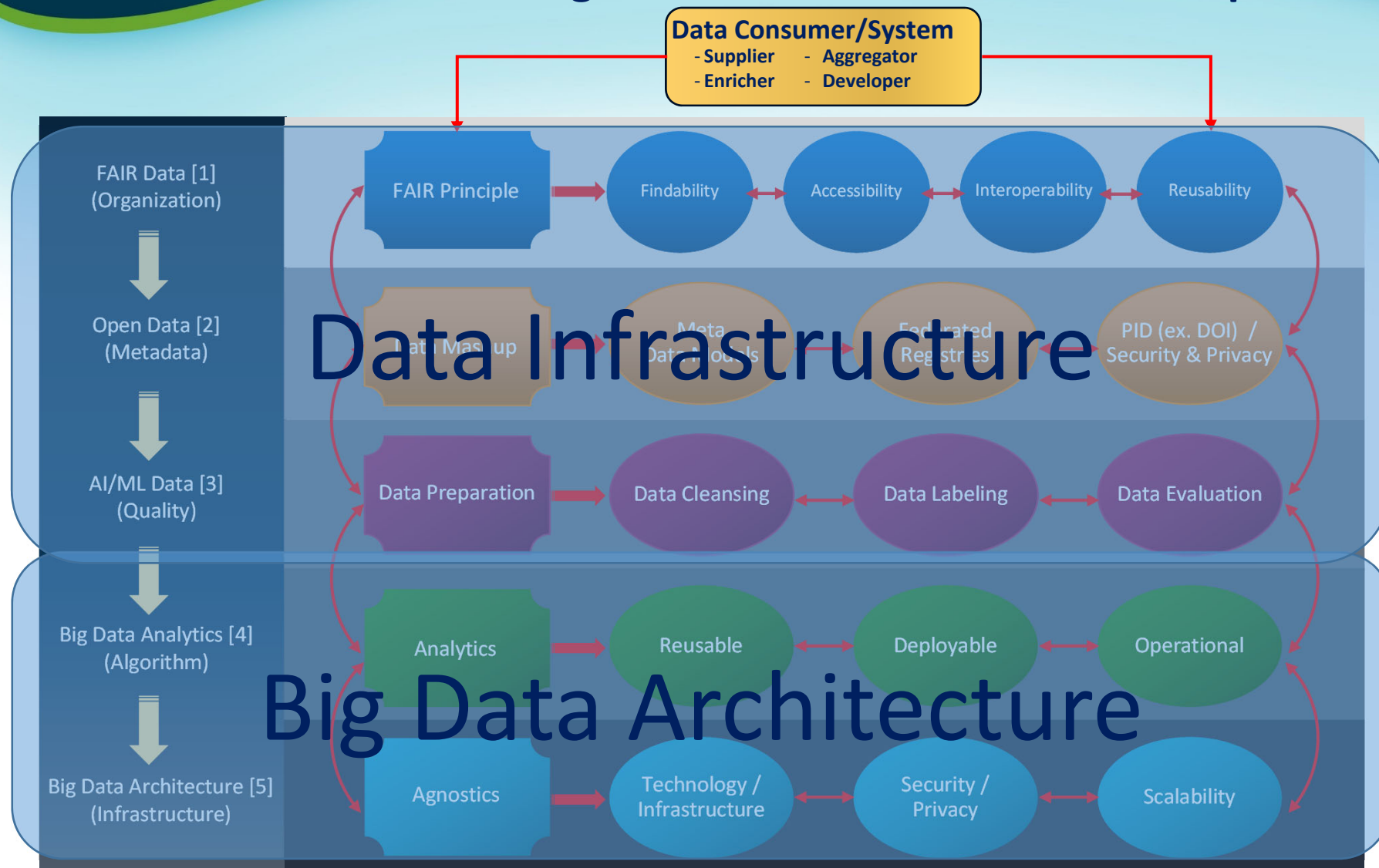


[1]: RDA GO FAIR IG, [2]: IEEE BDGMM, [3]: ISO/IEC JTC 1/SC 42, [4][5]: NIST Big Data PWG

**Big Data + Analytics: Making Analytics as Services, Wo Chang, NIST/ITL, May 28, 2020**



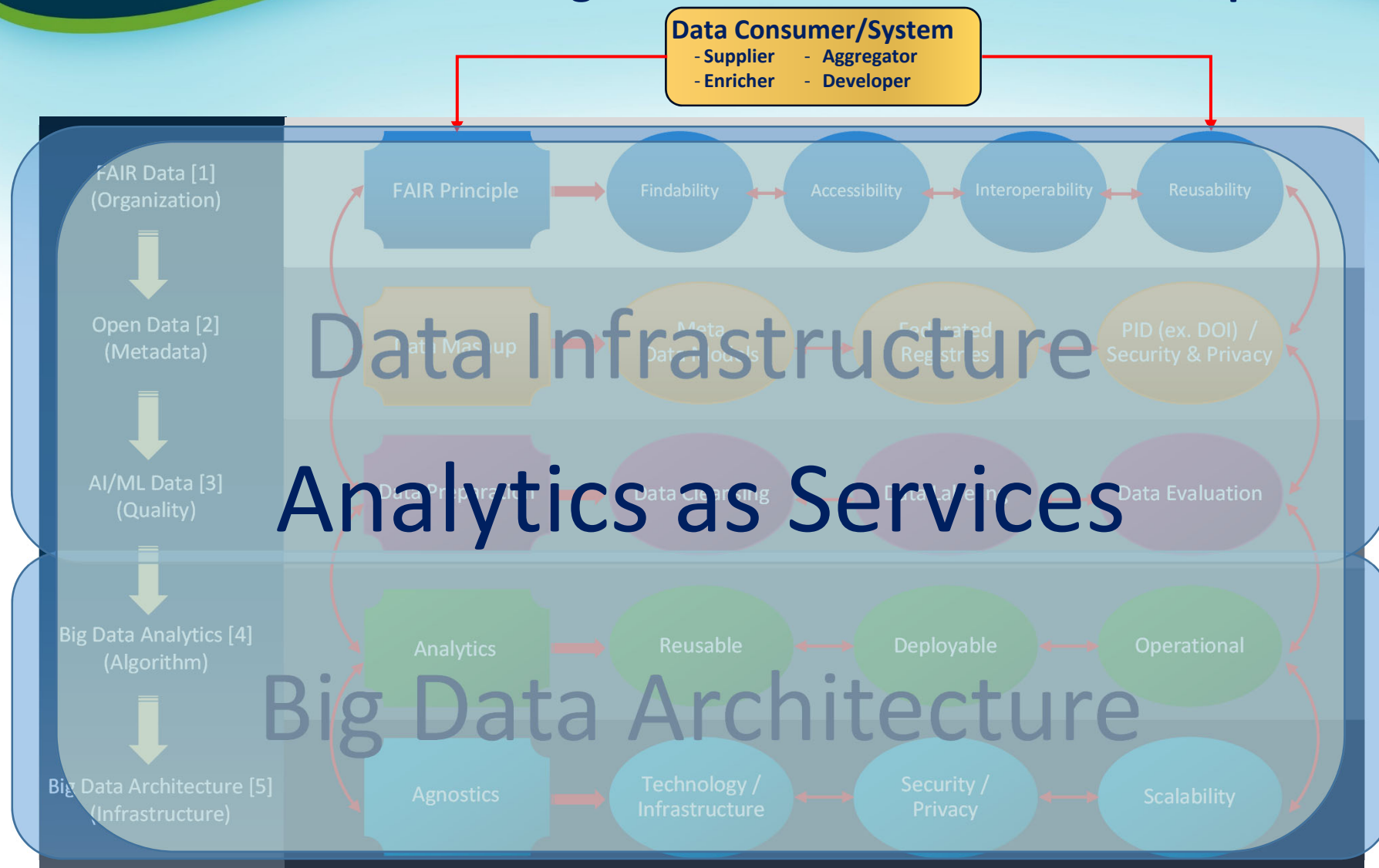
## NIST Big Data Activities and Roadmap



[1]: RDA GO FAIR IG, [2]: IEEE BDGMM, [3]: ISO/IEC JTC 1/SC 42, [4][5]: NIST Big Data PWG

*Big Data + Analytics: Making Analytics as Services, Wo Chang, NIST/ITL, May 28, 2020*

## NIST Big Data Activities and Roadmap



[1]: RDA GO FAIR IG, [2]: IEEE BDGMM, [3]: ISO/IEC JTC 1/SC 42, [4][5]: NIST Big Data PWG

**Big Data + Analytics: Making Analytics as Services, Wo Chang, NIST/ITL, May 28, 2020**

## Links

- [1] FAIR Data (<https://www.rd-alliance.org/groups/go-fair-ig>)
- [2] Open Data (<https://ieeesa.io/BDGMM>)
- [3] AI/ML Data (<https://www.iso.org/committee/6794475.html>)
- [4] Big Data Analytics (<https://bigdatawg.nist.gov/>)
- [5] Big Data Architecture (<https://bigdatawg.nist.gov/>)

# NBD-PWG Next Step

## Motivation and Industry Trends

With Big Data's compound annual growth rate at **61** percent and its ever-increasing deluge of information in the mainstream, the collective sum of world data will grow from [33 zettabytes \(ZB, 10<sup>21</sup>\) in 2018 to 175 ZB by 2025](#). The presence of such a rich source of information requires a massive analysis that can effectively bring about much insight and knowledge discovery.

### Programming Libs



Others...

### ML Frameworks



Others...

### Analytics Services



Others...



# NBD-PWG Next Step

## *Explore Big Data + Analytics – Making Analytics as Services*

NBD-PWG is exploring how to extend NBDIF for packaging scalable analytics as services to meet the challenges of so much information. These services would be reusable, deployable, and operational for Big Data, High Performance Computing, and AI machine learning (ML) and deep learning (DL) applications, regardless of the underlying computing environment.

### Topics for discussion include:

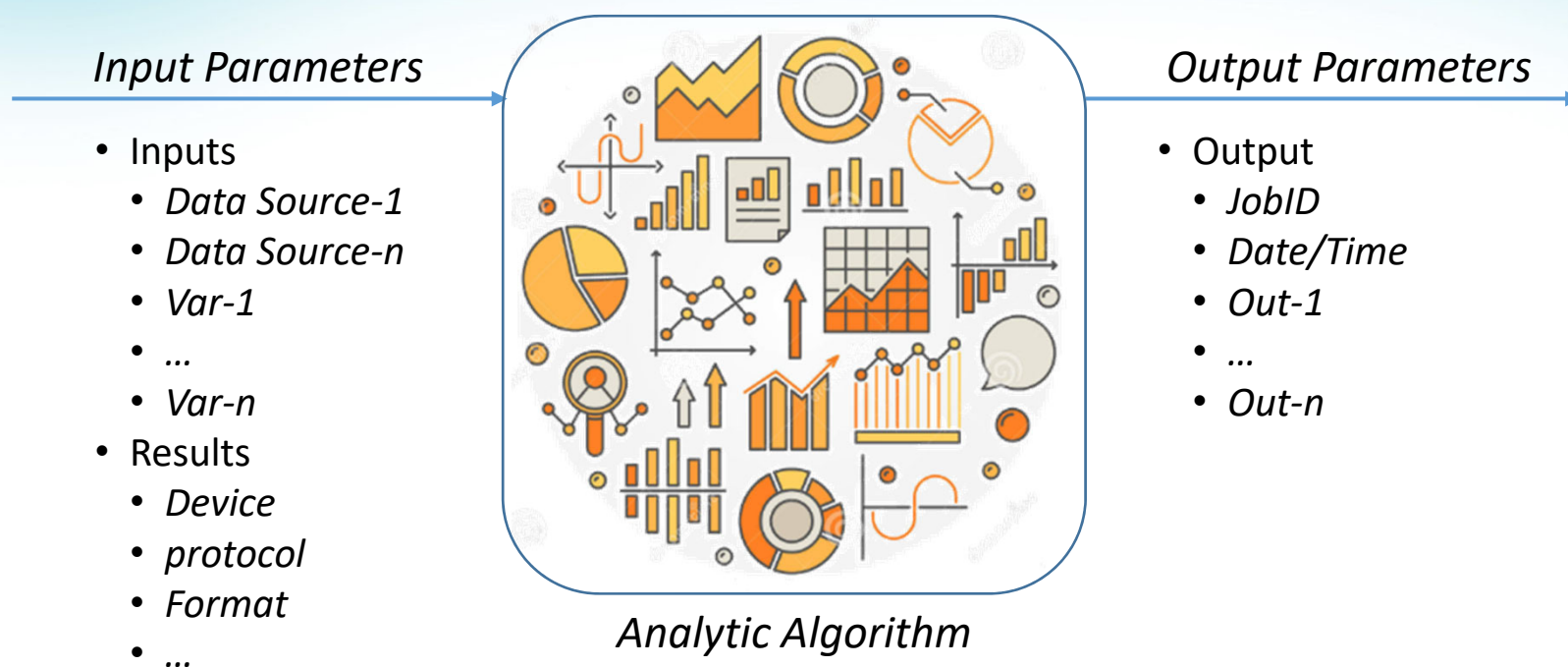
- ❖ **Exploration:** Determine and document the level of interest from industry, government, and academia in extending the NBDIF to develop scalable analytics as services that are reusable, deployable, and operational, regardless of the underlying computing environment.
- ❖ **Key Focus Areas**
  1. Compile and organize use cases, analytic services from traditional statistical, AI/ML/DL, and emerging analytics application domains; identify and document technical requirements.
  2. Package analytic algorithms with well-defined input and output parameters as service payloads that can be reusable, deployable, and operational across multi-cores, CPUs, and GPU computing platforms.
  3. Encapsulate service payload with well-defined format, interface, and end-to-end access control for open and secure computing environment.
  4. Establish federated registries to locate and consume analytics services with persistent identifiers across organizations.
  5. Provide resource management for application orchestration and workflow between processes.

## Focus 1 - Compile Analytic Service Use Cases

*Big Data + Analytics: Making Analytics as Services, Wo Chang, NIST/ITL, May 26, 2020*

# NBD-PWG Next Step: *Making Analytics as Services*

## *Focus 2 - Package Analytic Algorithms with well-defined I/O Parameters*



Explore the following:

- ModelOps (Operating Model for DevOps in AI): <https://www.modelop.com/modelops/>
- Portable Format for Analytics (PFS): <http://dmg.org/pfa>
- XML Markup: Predictive Model Markup Language [https://en.wikipedia.org/wiki/Predictive\\_Model\\_Markup\\_Language](https://en.wikipedia.org/wiki/Predictive_Model_Markup_Language)

# NBD-PWG Next Step: *Making Analytics as Services*

## *Focus 3 - Encapsulate Analytic Service as Payload*

Enable Big Data analytics services/tools for *interoperability, portability, reusability, and extensibility*.

Practical Aspect: Analytics services/tools can be **reusable, deployable, and operational** (max. use of resources) on any of Big Data, HPC, machine/deep learning, etc. computing environment.



Potential ISO/IEC standard specification



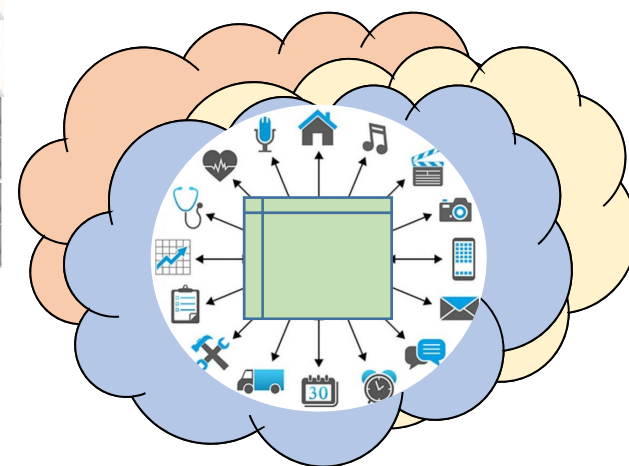
Laptop



Server



Data Center  
Many CPUs/Cores/GPUs/  
FPGAs/ASICs/accelerators

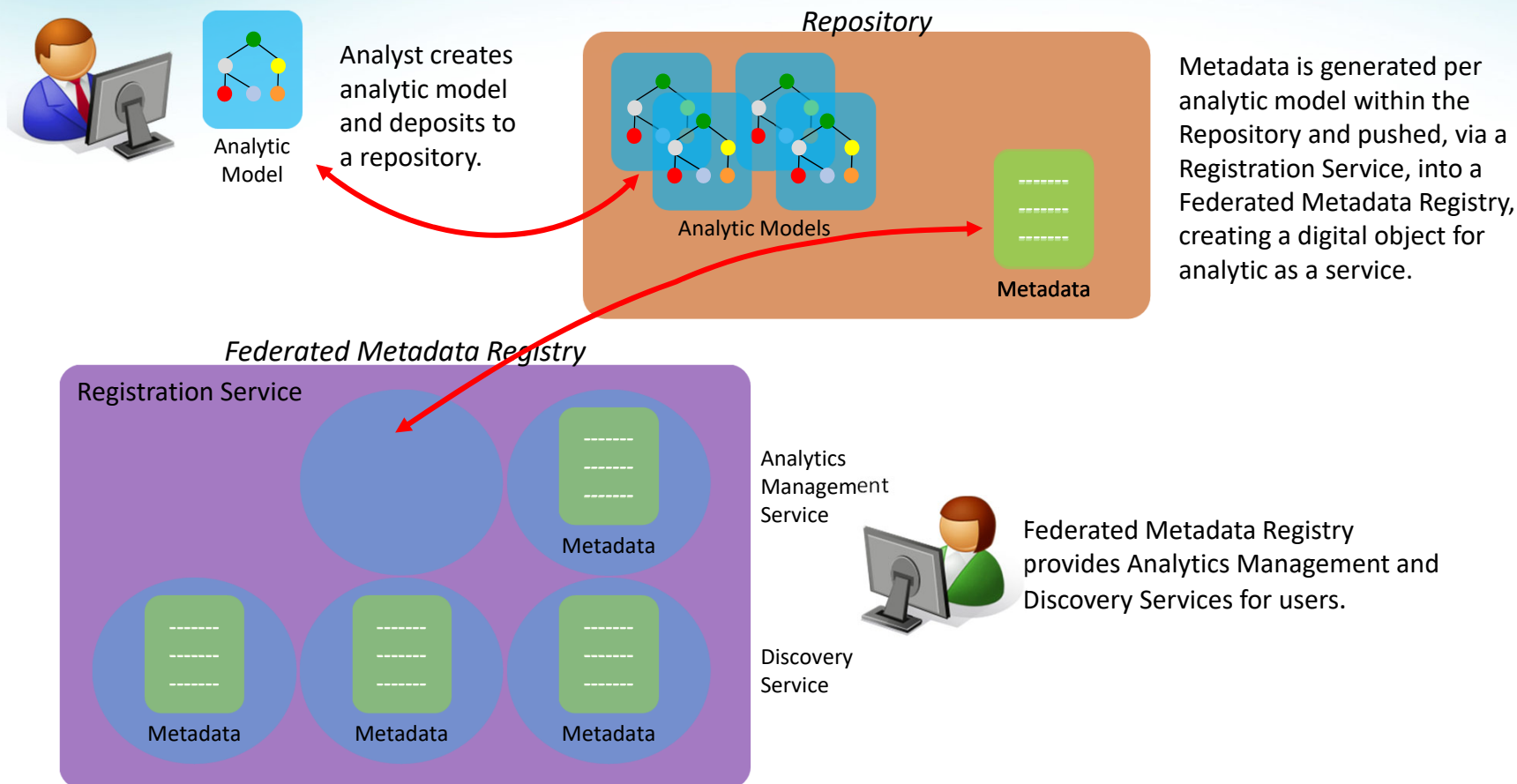


Cloud/Multi-clouds



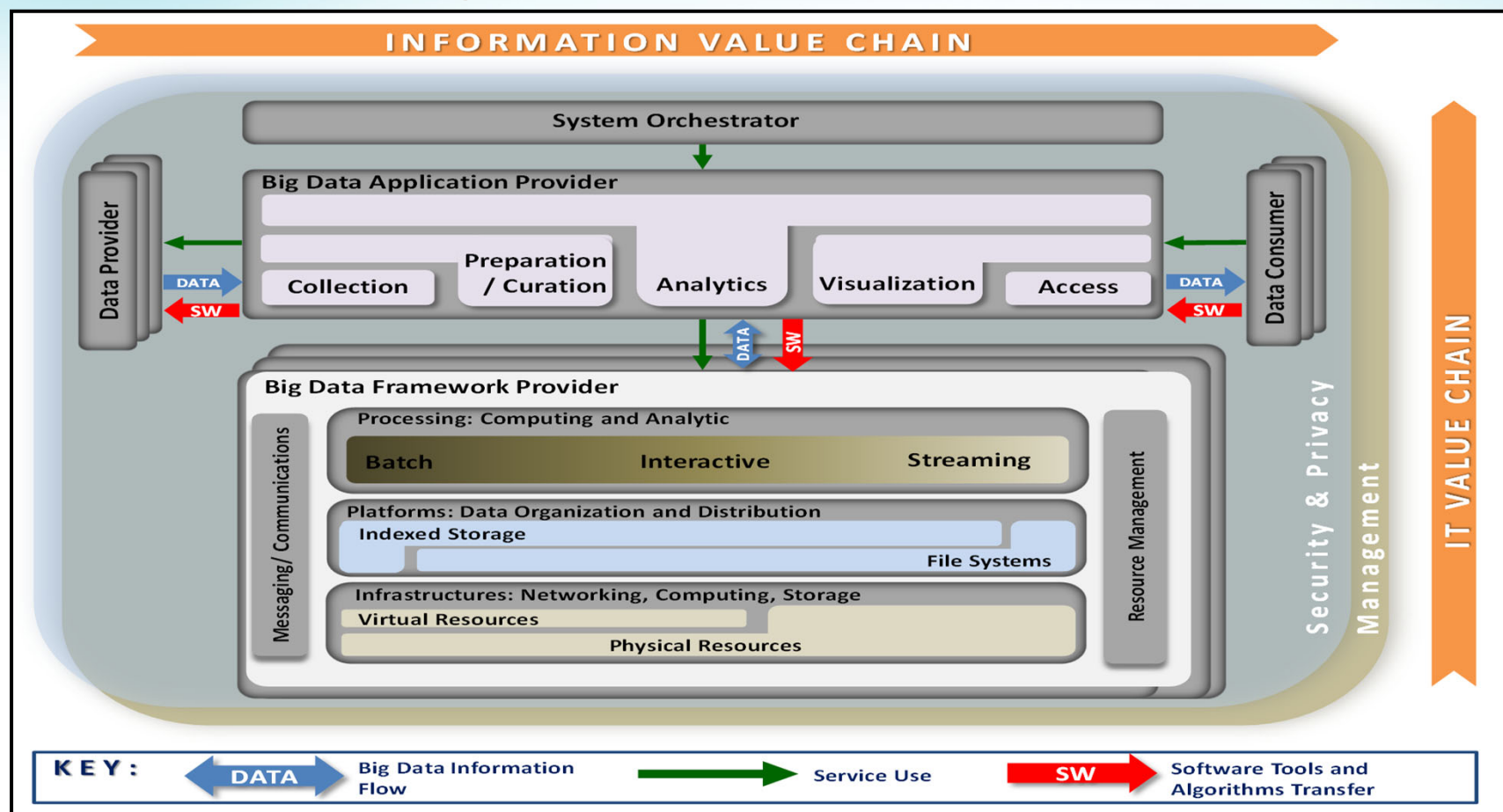
# NBD-PWG Next Step: *Making Analytics as Services*

## *Focus 4 – Federated Analytic Services Registry*



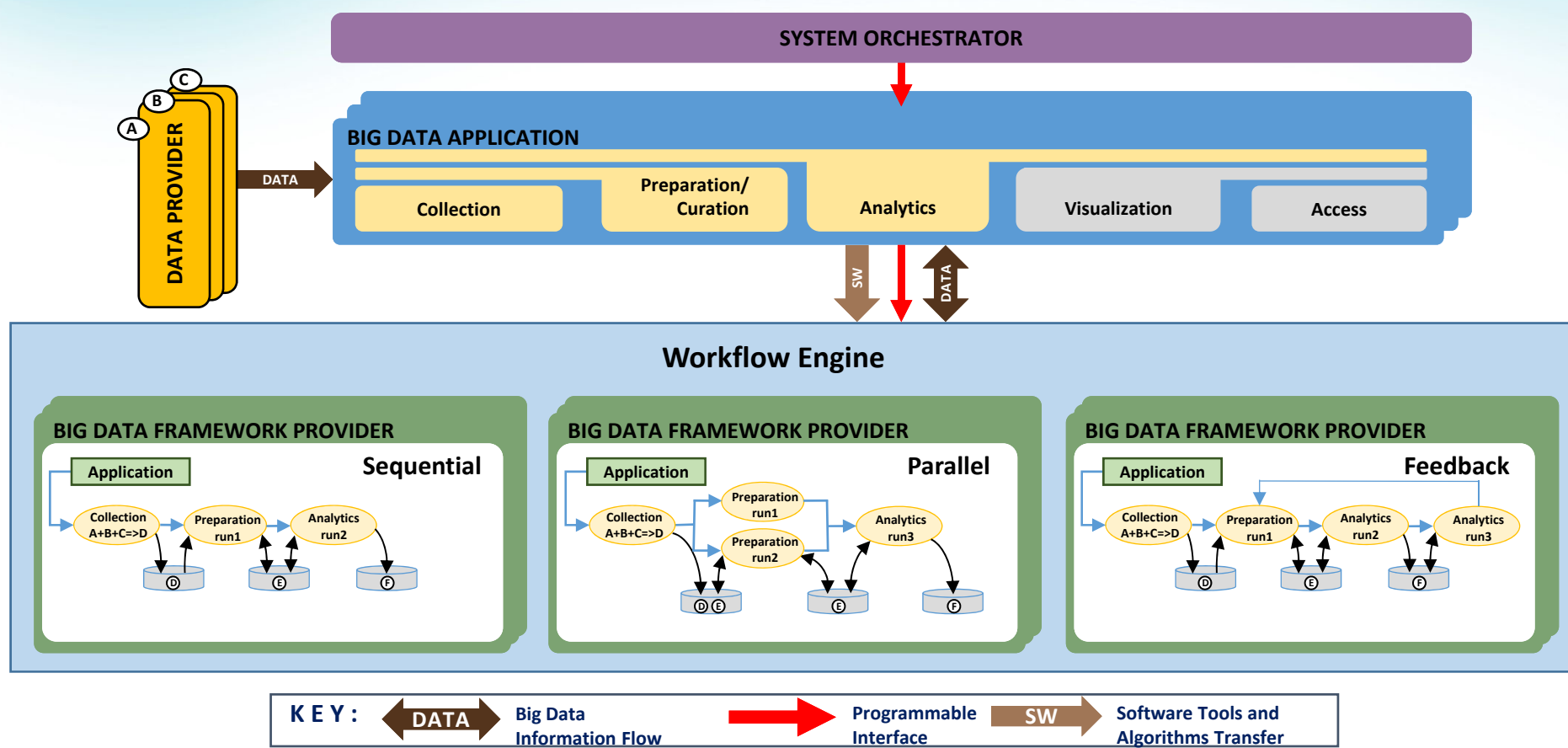
# Reference Architecture Subgroup

## NBDIF Vol. 6: NIST Big Data Reference Architecture (NBD-RA)

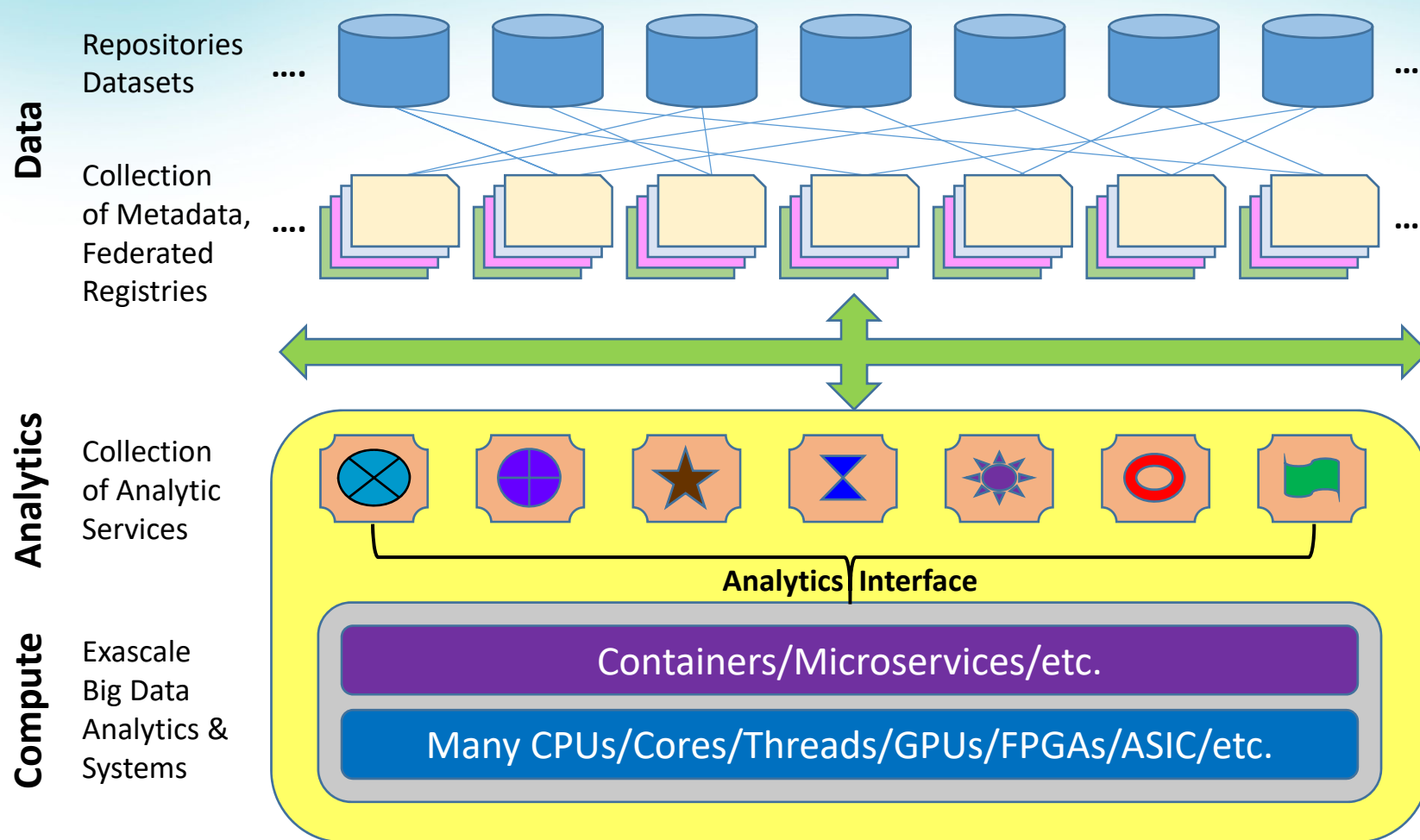


# NBD-PWG Next Step: *Making Analytics as Services*

## *Focus 5: Resource Management – Orchestration and Workflow*



## Beyond: Enable Convergence of Data + Analytics + Compute





# *Questions ?*

Please contact: [wchang@nist.gov](mailto:wchang@nist.gov)