

From Data Lakes to Data Hubs: How to make Data Semantic!

Prof Renato Iannella
Adjunct Professor, The University of Hong Kong
Lead Enterprise Data Architect, Airservices Australia



1

ICOM6046-L5-Expert-Address.key - 15 October 2019

Overview

- ✱ **The move towards Data Lakes**
- ✱ **Enabling maturity of data capabilities**
- ✱ **The Data Lake maturity model**
- ✱ **Data Hubs**
- ✱ **Industry Directions and Trends**
- ✱ **Semantics**

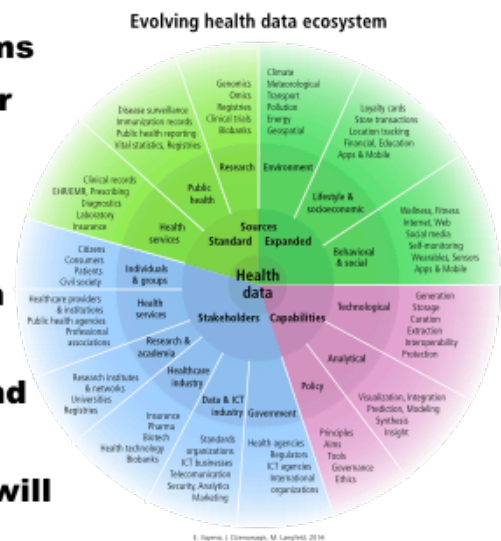


2

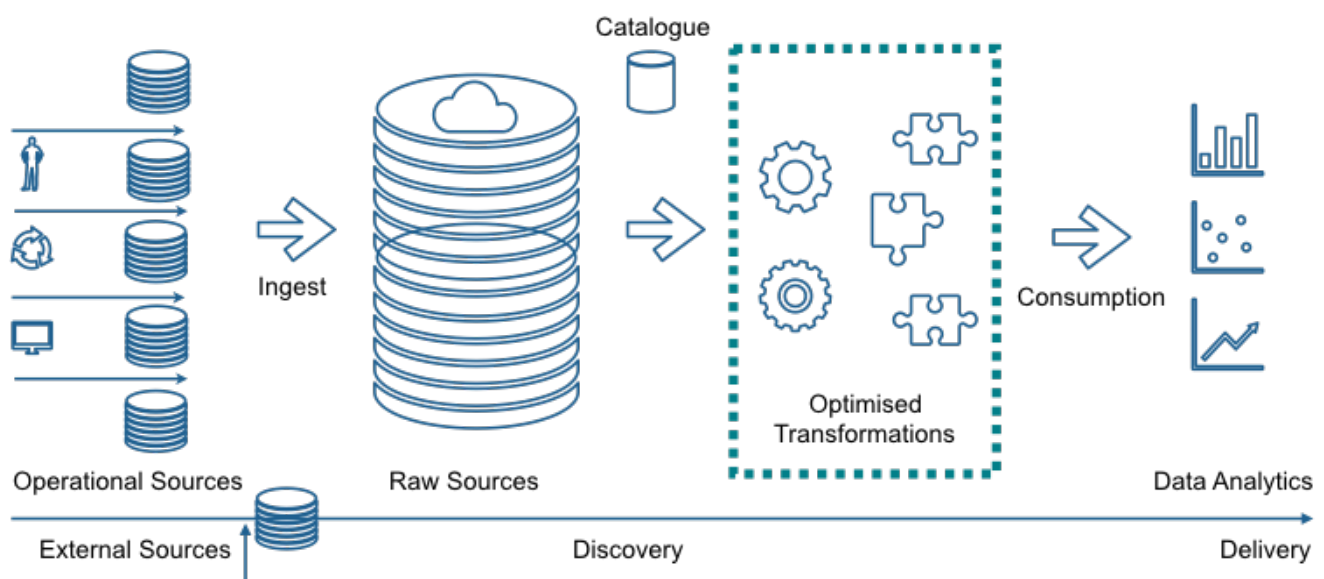
ICOM6046-L5-Expert-Address.key - 15 October 2019

The move to Data Lakes...

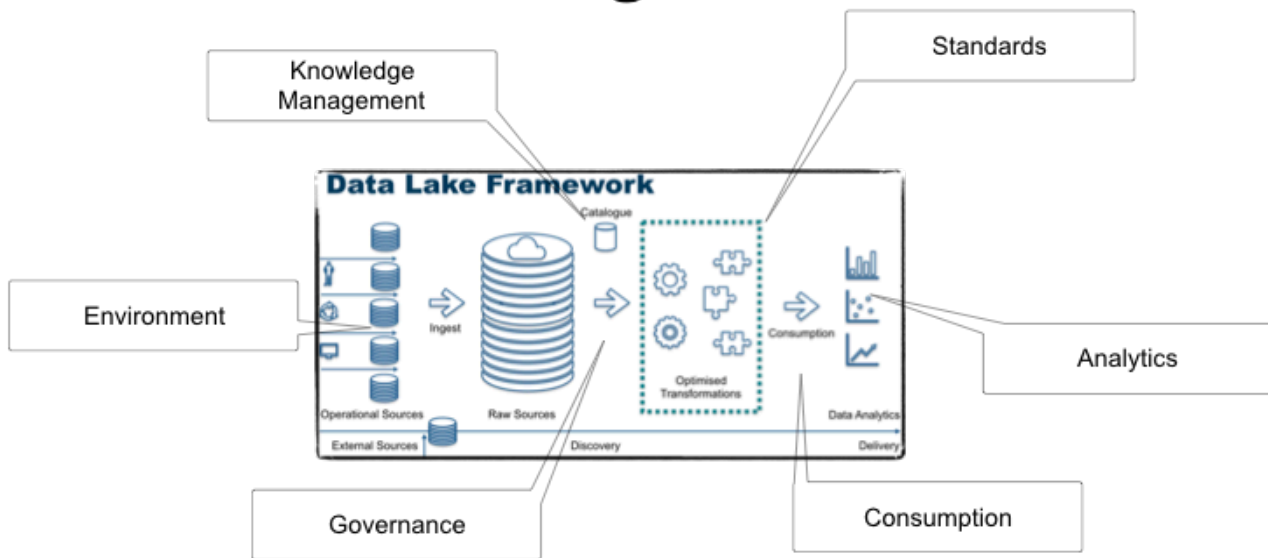
- * **Industry sectors have large data ecosystems**
- * **They collect significant amounts of data for specific operational purposes**
 - * **But don't share very well or reuse to maximise benefits**
- * **Data Lakes are about the potential for such data to be collected and optimised**
- * **Enables a reference platform for sharing and advanced data analytics**
- * **The design and governance of Data Lakes will be constantly changing and maturing...**



Data Lake Framework



Information Management



5

ICOM6046-L5-Expert-Address.key - 15 October 2019

Information Maturity Capabilities

- ✳ **Knowledge** - this capability is the collective knowledge of information across the enterprise
- ✳ **Standards** - this capability is the alignment of information to conform to applicable standards
- ✳ **Consumption** - this capability is the sharing of information across the enterprise to consumption points
- ✳ **Analytics** - this capability is the integration of information across the enterprise to support common objectives
- ✳ **Governance** - this capability is the control, protection, and assurances of information across the enterprise
- ✳ **Environment** - this capability is the platform for information platforms and services



Source: Gartner, 2017

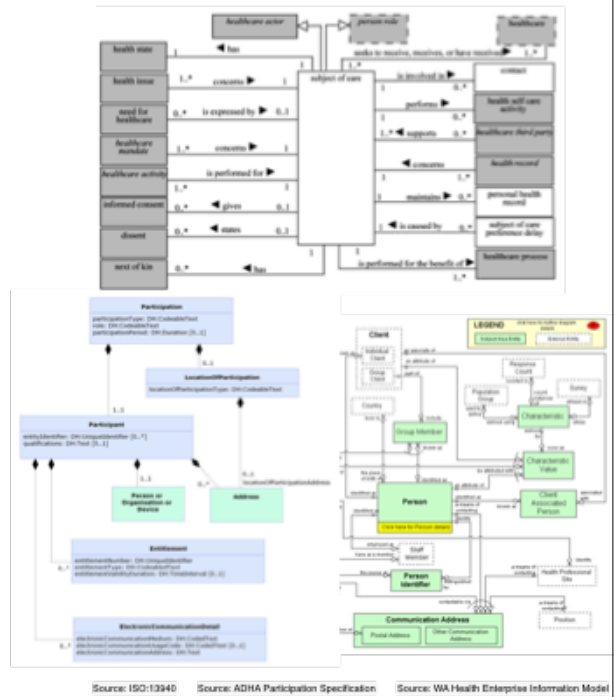
Source: <https://trishamplane.com/wp-content/uploads/2014/10/CMMS-Levels.jpg>

6

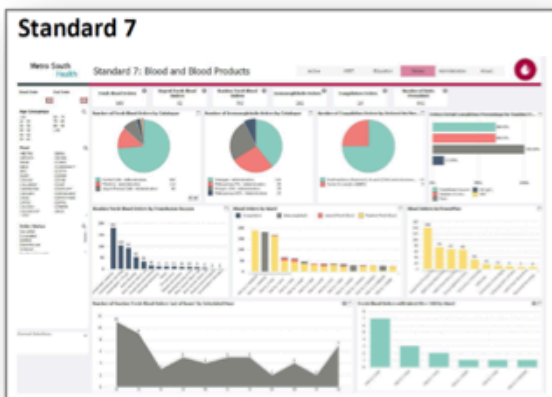
ICOM6046-L5-Expert-Address.key - 15 October 2019

Maturity - Standards

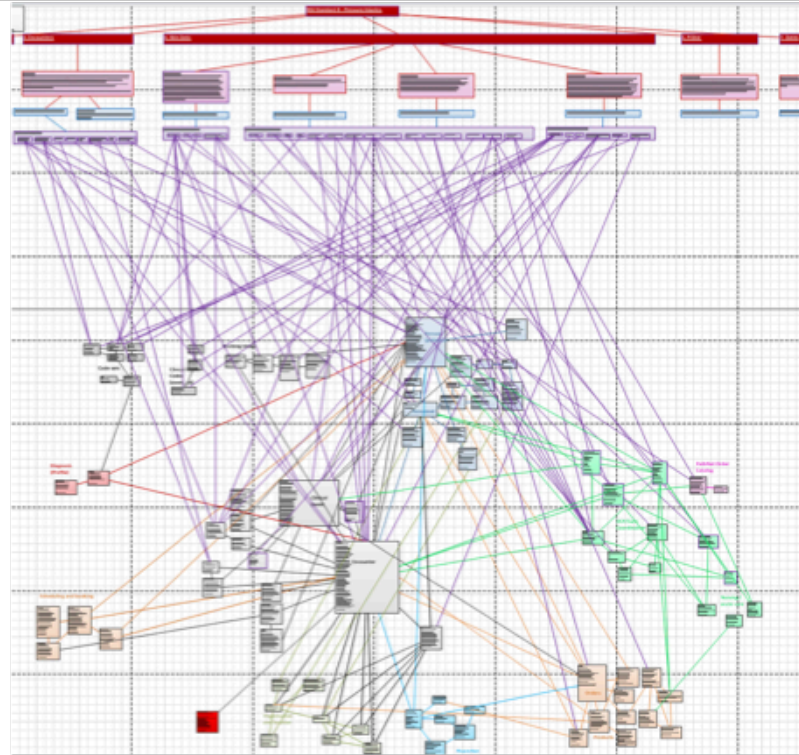
- * **Common views of data entities**
 - * **Shared healthcare concepts**
- * **Data Model Standards**
 - * **ADHA, HL7, FHIR, ISO, Jurisdictions...**
- * **Terminology, Reference Data**
 - * **SNOMED, AMT, ICD...**
- * **Standard KPIs**
 - * **ACHS, Clinical Quality, HACs...**
- * **Identity**
 - * **Local Patient Identifiers**
 - * **IHI, HPI-I, PHI-O...**



Maturity - Data Analytics

[illegible]

Data Analytics



Source: eHealth Queensland, Benson Choy

9

ICOM6046-L5-Expert-Address.key - 15 October 2019

Maturity - Data Governance

- ✳ **Business Glossaries for Policy development and analysis**
- ✳ **Business Rules Management for compliance and exceptions**
- ✳ **Business semantics (models, dictionaries, ontologies) to enable platform understanding and inter-relationships**
- ✳ **Data Catalogues for knowledge management and discovery**
- ✳ **Data Quality for monitoring and measuring data maturity**
- ✳ **Data Lineage for data provenance and transformations**
- ✳ **Data Impact Analysis to trace change impacts and dependencies**
- ✳ **Information Policies model and representation for integration in enterprise applications**



10

ICOM6046-L5-Expert-Address.key - 15 October 2019

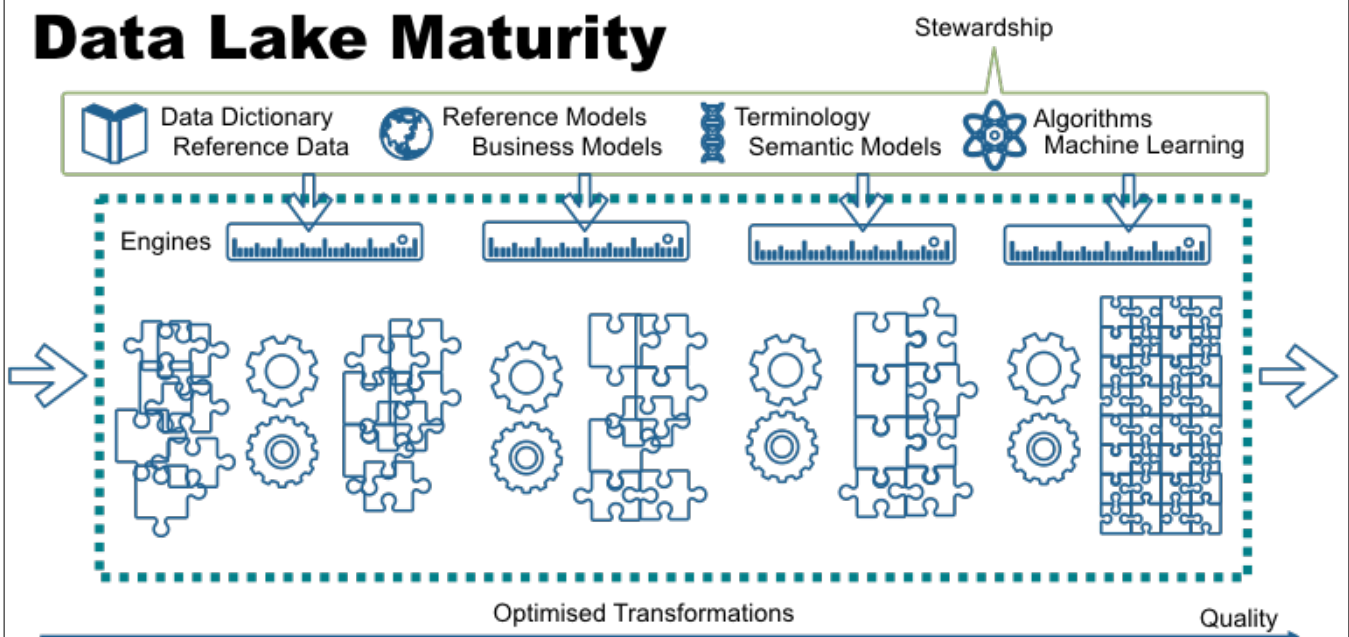
The Data Lake Maturity Journey



11

ICOM6046-L5-Expert-Address.key - 15 October 2019

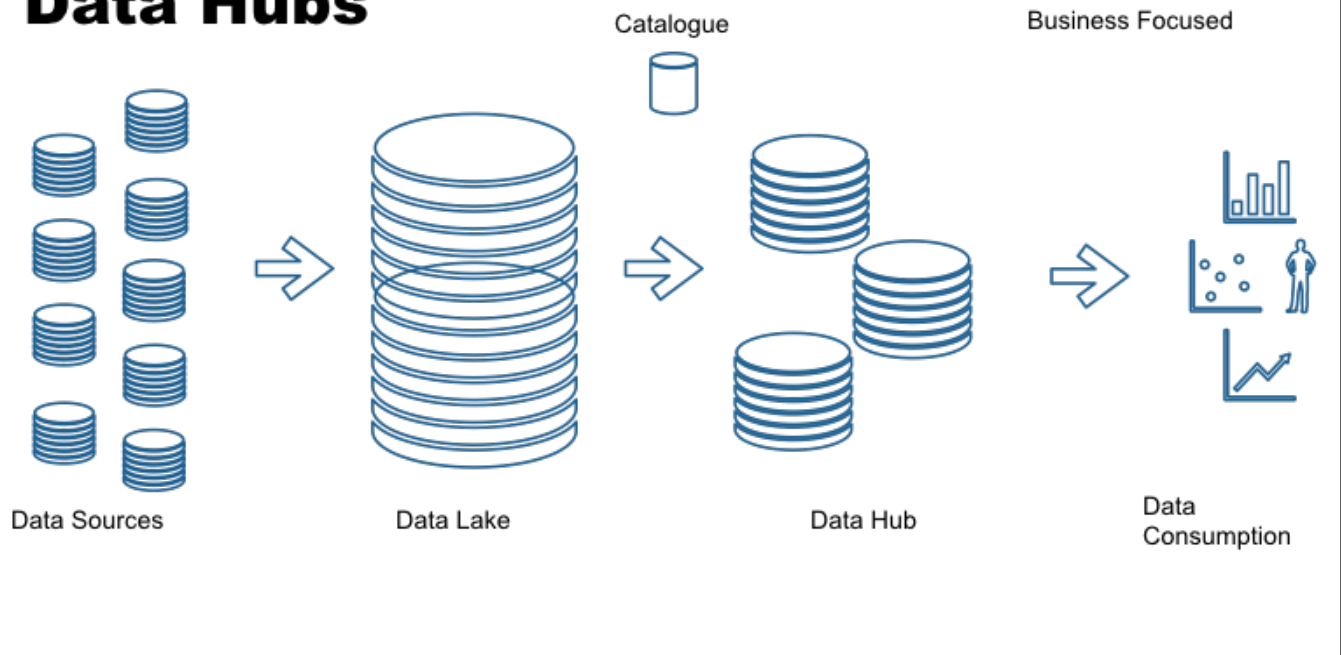
Data Lake Maturity



12

ICOM6046-L5-Expert-Address.key - 15 October 2019

Data Hubs



13

ICOM6046-L5-Expert-Address.key - 15 October 2019

Data Hub Strategy

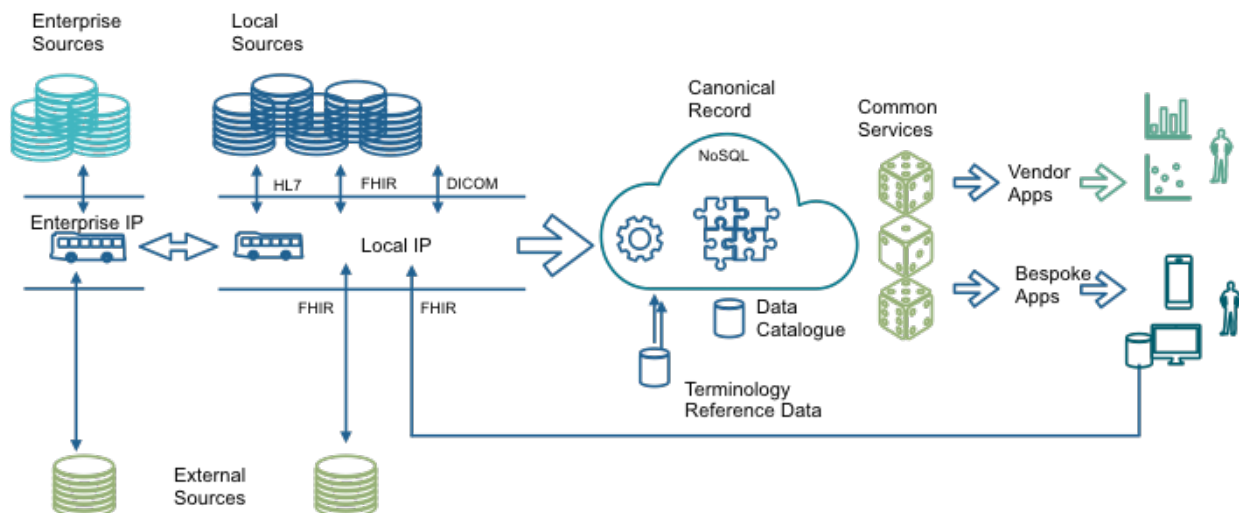
- ✳ **A data hub strategy completes a governance and sharing architecture and drives integration**
- ✳ **Develop a conceptual data hub strategy by looking at the needs of mediation in support of the most shared data**
 - ✳ **Extrapolate this down to logical and implementable layers**
- ✳ **A data hub is a node on an enterprise network (graph model) with explicit and clear relationships between hubs**
 - ✳ **Can be initially centralised and later distributed**
- ✳ **Develop a persistence model reconciling the needs of availability, timeliness and access to operational costs of storage and delivery**



14

ICOM6046-L5-Expert-Address.key - 15 October 2019

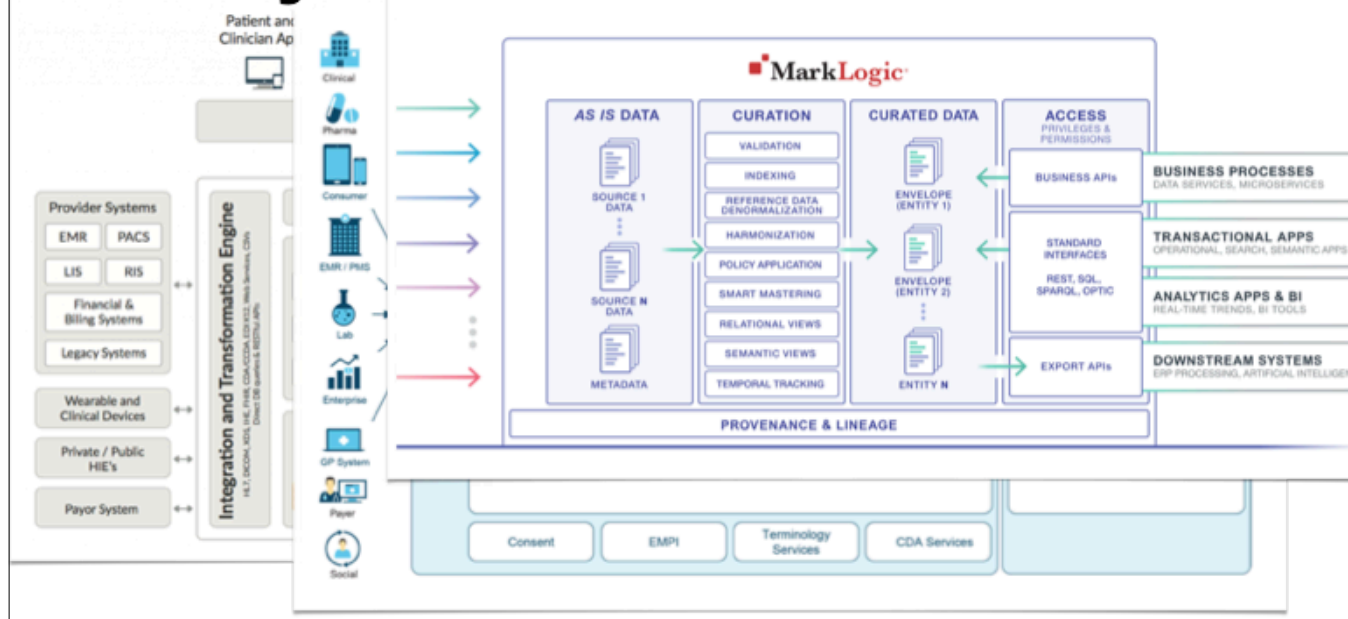
Health: Data Platform



15

ICOM6046-L5-Expert-Address.key - 15 October 2019

Industry Direction



16

ICOM6046-L5-Expert-Address.key - 15 October 2019

Future Trends

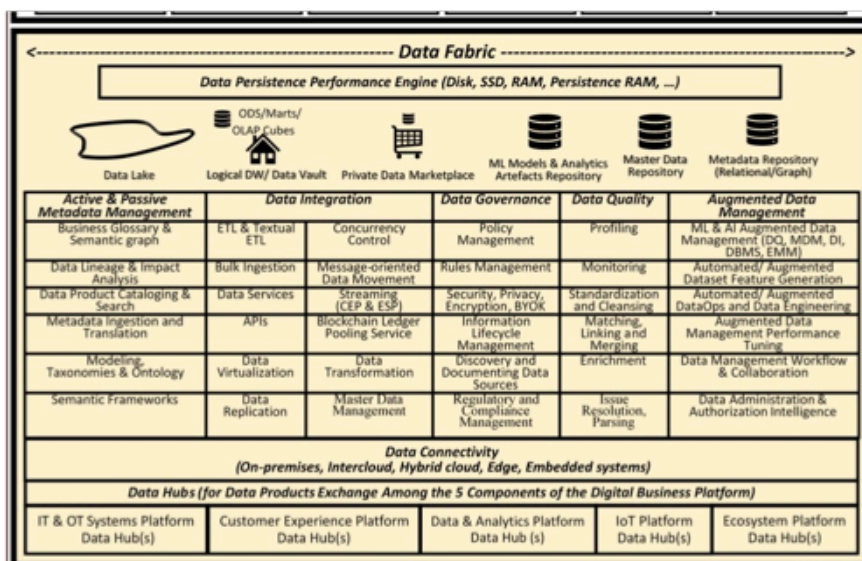
Gartner



17

ICOM6046-L5-Expert-Address.key - 15 October 2019

Data Fabric



<https://www.linkedin.com/pulse/part-2-3-how-modernize-your-data-analytics-platform-sha/>

18

ICOM6046-L5-Expert-Address.key - 15 October 2019

Graph

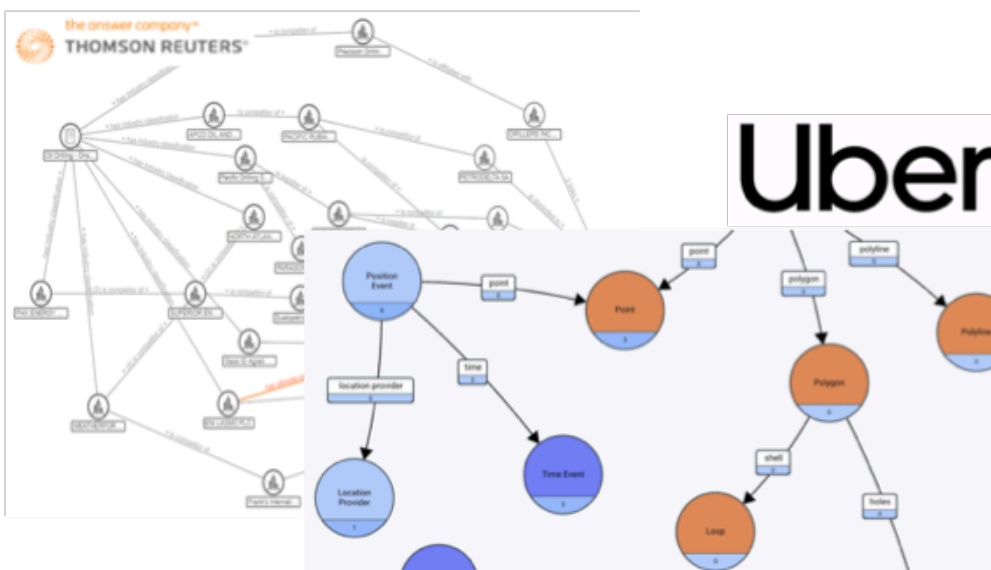
- ✱ **Explore data the way people think**
 - ✱ **Relationships between concepts and entities are exposed and transparent**
- ✱ **Complex network investigations**
 - ✱ **Data storage for complex traversal operations**
- ✱ **Scalable (cf relational)**
- ✱ **Emergent knowledge networks**
 - ✱ **Creates a human-comprehensible data views of connected ontologies**
 - ✱ **Linking of diverse heterogeneous data**



19

ICOM6046-L5-Expert-Address.key - 15 October 2019

Graph - Industry

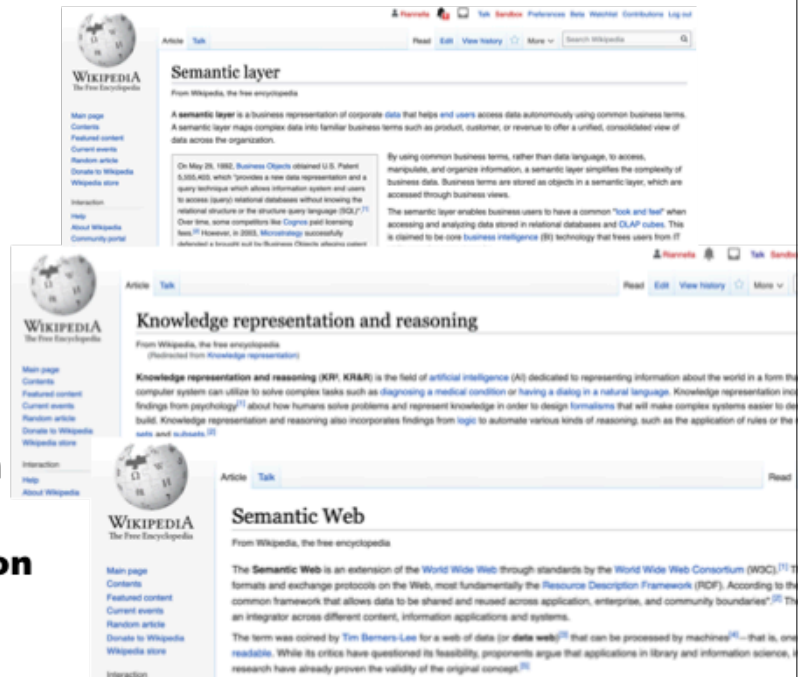


20

ICOM6046-L5-Expert-Address.key - 15 October 2019

Semantics

- ✱ **What does Semantics mean (in our industry)**
- ✱ **The “semantic layer”?**
- ✱ **Example**
 - ✱ **The tag “A380” means “Airbus A380”**
- ✱ **Difference response from the Semantic Web and Knowledge Representation community**

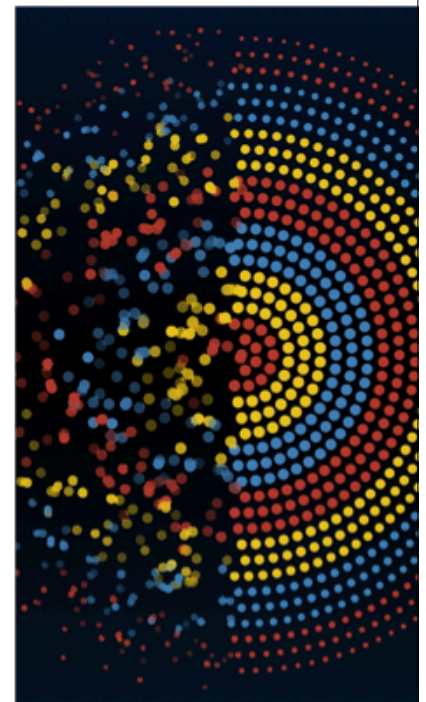


21

ICOM6046-L5-Expert-Address.key - 15 October 2019

Rough Guide to Semantics

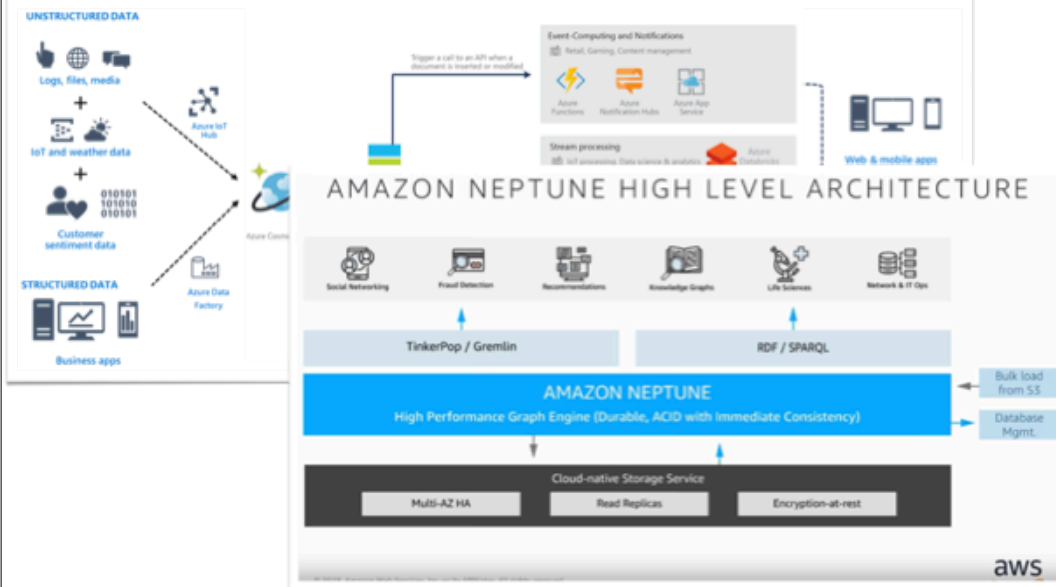
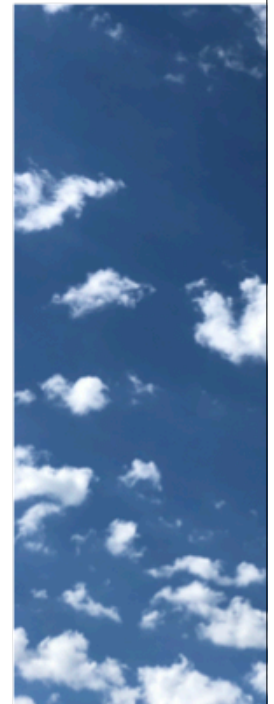
- ✱ **Types**
 - ✱ **Is an A380 an Aircraft?**
- ✱ **Properties first-class**
 - ✱ **How is it related to the A340? Cathay Pacific?**
- ✱ **Sameness**
 - ✱ **Is it the same as the A380neo?**
- ✱ **Uniqueness**
 - ✱ **Which individual has #454322**
- ✱ **Graph model**
 - ✱ **Is it represented as triples**
- ✱ **Reasoning (Inference)**
 - ✱ **Flew from BNE to HKG ->> travelled by Aircraft**



22

ICOM6046-L5-Expert-Address.key - 15 October 2019

Cloud Semantic Hubs



23

ICOM6046-L5-Expert-Address.key - 15 October 2019

Summary

- ✱ **The Data Lake provides a platform for maturing enterprise data analytics capabilities**
- ✱ **Data Hubs provide a focus on business requirements**
- ✱ **Prioritise business cases that can show direct benefit from Semantic Hubs**
- ✱ **Develop an Enterprise Information Architecture**
 - ✱ **Prepare for a continuous platform and services evolution**
- ✱ **Develop a maturity Roadmap for your “data fabric”**



24

ICOM6046-L5-Expert-Address.key - 15 October 2019

Questions? 有什么问题吗

