

# Security, Privacy and Big Data Standards

ESR 2015 The Grand Hotel Malahide, Dublin, Ireland.

Ray Walshe Lead Editor  
Big Data Reference Architecture  
ISO IEC JTC1 WG9

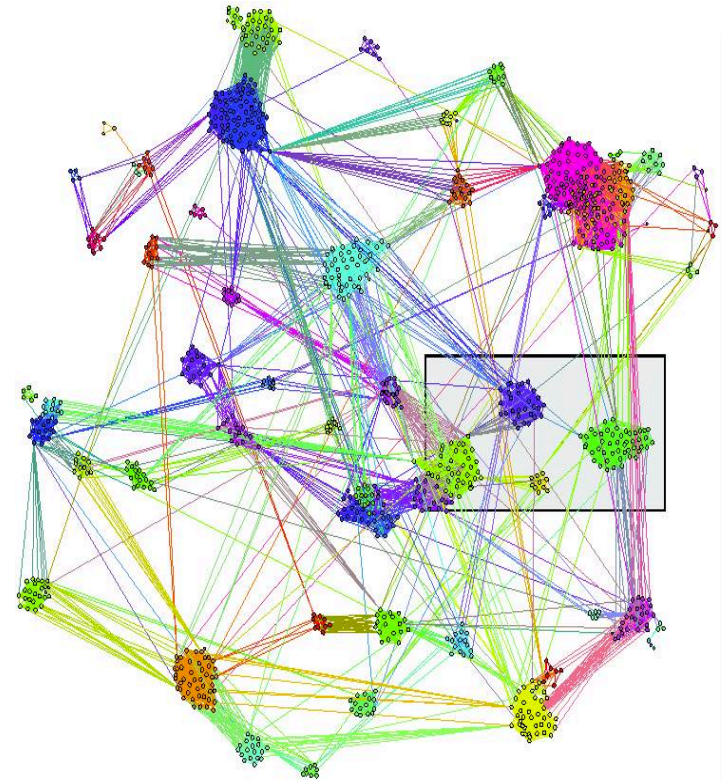
# Security, Privacy and Big Data Standards

- \* Security and Privacy issues arise in any distributed computing environment.
- \* Security and Privacy issues are exacerbated by Big Data



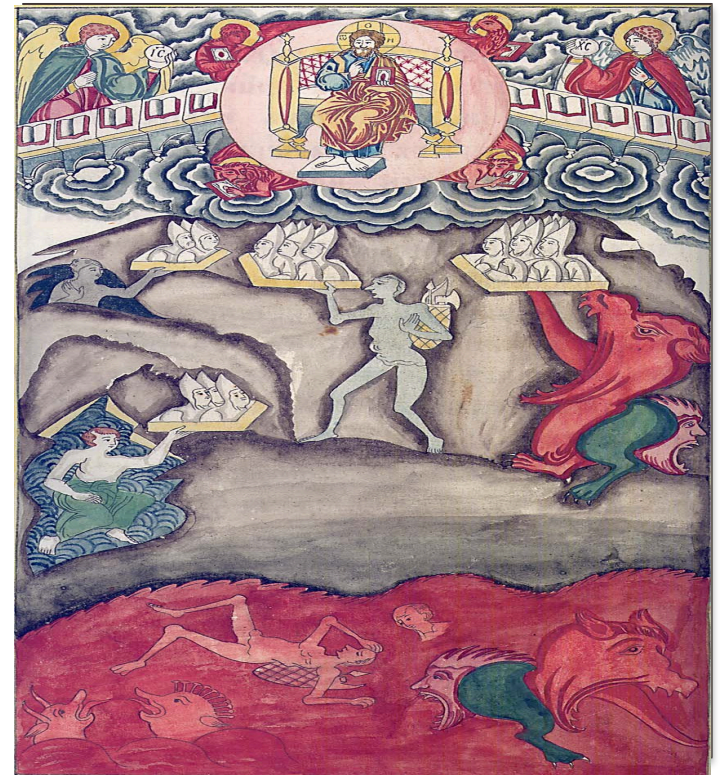
# Issues: Sources&Provenance

- \* Big Data comes from combining data from different sources
- \* Big Data necessitates storage in multi-tiered storage
- \* Big Data frequently moves across boundaries
- \* Metadata and Data Provenance



# Issues: Ownership

- \* Data Ownership
- \* Privileges conferred
- \* Ownership transparency
- \* Data Provenance facilitates traceability



MS 2623 Old believer Apocalypse. Russia, 1812

# Issues: Not Designed In

- \* Frameworks developed to support volume and velocity
- \* Malfunctioning computing nodes
- \* Partial infrastructure attacks
- \* Rogue nodes can be added



# Issues: Search

- \* Data search and selection can lead to privacy or security policy concerns
- \* A combination of user competency and system protections needed
- \* End-to-end information assurance practices



# Threats

- \* Systematic threat model analysis needed
- \* The threat model includes scenarios for confidentiality and integrity, provenance, Collusion attacks etc
- \* One positive is that analytics can also be performed to detect security breach events and fortify security.







# Policy

- \* US. Safe Harbor privacy principles inadequate
- \* Notice, choice, onward data transfer, security, data integrity, correction or deletion of data, effective enforcement of these guidelines.
- \* Big Data will further amplify the need for increased control



# People

- \* The human element
- \* security and privacy problems created through human agents will vary from benign to accidental to malicious.



# Regulatory environment will evolve

- \* Security and privacy measures must scale nonlinearly
- \* HIPAA guidance for personal health records.
- \* Full scope of risk poorly understood
- \* Different computer professionals will emerge



# Todo: Design, mechanise, protect, secure

- \* Design with security in mind
- \* Data aggregation and dissemination
- \* Personally Identifiable Information
- \* Anonymization and obfuscation
- \* Payload might be encrypted



# Why?

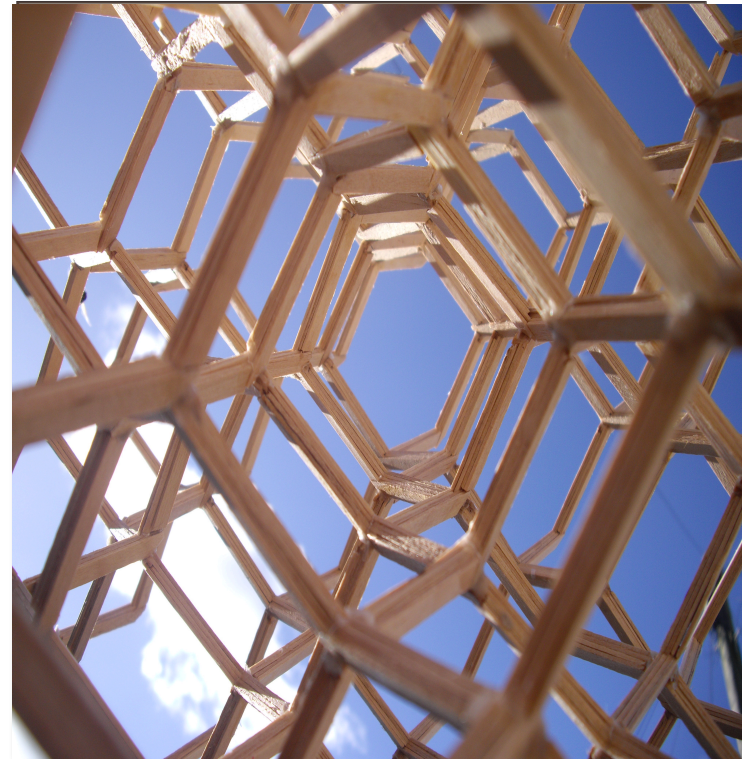
- \* Prosecution / Civil Actions Against the Company
- \* Customer Retention / Brand-Value
- \* Compliance
- \* Internationalisation



# Risk Management : Frameworks

## Use of Frameworks

- \* Incorporate Big Data Programs within RA/RM framework
- \* Incorporate Big Data Programs within DQM framework
- \* Ensure that DQM framework addresses intrinsic and contextual data quality factors



# Risk Management: ConsoliDATA

## Data Consolidation

- \* Collections not consolidated unless:
  - \* they satisfy threshold data quality tests
  - \* their purposes, quality, meanings of relevant data-items are compatible
  - \* policy constraints are respected



# Risk Management: Anonymous

## Effective Anonymisation

- \* Where sensitive data is involved
- \* anonymisation techniques are applied
- \* Data submitted to analysis is not re-identifiable





# Risk Management: Clean Up

## Data Scrubbing

- \* Only within context of the organisation's data quality assurance framework
- \* Involve external reference-points
- \* their accuracy and effectiveness are audited
- \* No decision-making on results unless threshold data quality criteria met



# Risk Management: Decisions

## Decision-Making

- \* Inferencing mechanisms not to be relied upon to make decisions
- \* When 'big data' is applied to decision-making:
  - \* the criteria of relevance, meaning, and transparency of decision mechanisms are all satisfied
  - \* the results are audited, including by testing against known instances
  - \* the outcomes are subjected to post-implementation assessment,





SDO/Consortium	Interests area on standardization	Main deliverables
ISO/IEC JTC 1/SC 32	Data management and interchange, including database languages, multimedia object management, metadata management, and e-Business.	e-Business standards, including role negotiation; metadata repositories, model specification, metamodel definitions; SQL; and object libraries and application packages built on (using) SQL.
ISO/IEC JTC 1/SC 38	Standardization for interoperable Distributed Application Platform and Services including Web Services, Service Oriented Architecture (SOA), and Cloud Computing	Cloud Data Management Interfaces, Open Virtualization Format, Web Services Interoperability
ITU-T SG13	Cloud computing for Big Data	Cloud computing based big data requirements, capabilities, and use cases.
W3C	Web and Semantic related standards for markup, structure, query, semantics, and interchange.	Multiple standards including ontology specification standards, data markup, query, access control, and interchange.
Open Geospatial Consortium	Geospatial related standards for the specification, structure, query, and processing of location related data.	Multiple standards related to the encoding, processing, query, and access control of geospatial data.
Organization for the Advancement of Structured Information Standards	Information access and exchange.	A set of protocols for interacting with structured data content such as OData ( <a href="https://www.oasis-open.org/standards#odatav4.0">https://www.oasis-open.org/standards#odatav4.0</a> ), standards for security, Cloud computing, SOA, Web services, the Smart Grid, electronic publishing, emergency management, and other areas
Transaction Processing Performance Council	Benchmarks for Big Data Systems	Specification of TPC Express, Benchmark™ for Hadoop system and the related kit
TM Forum	Enable enterprises, service providers and suppliers to continuously transform in order to succeed in the digital economy	Share experiences to solve critical business challenges including IT transformation, business process optimization, big data analytics, cloud management, and cyber security.

# ISO Standards

- \* ISO/IEC JTC 1/WG9 Big Data
- \* ISO/IEC JTC 1/WG10 IoT
- \* ISO/IEC JTC 1/SC 32, Data management and interchange
- \* ISO/IEC JTC 1/SC 38 Distributed Application Platforms and Services



# Other SDO's

- \* ITU-T SG13 Cloud computing based big data
- \* W3C standardization of Web technologies
- \* Organization for the Advancement of Structured Information Standards (OASIS) OASIS standards for Security, Cloud computing, SOA, etc
- \* TM Forum creating the tools, best practice guidance



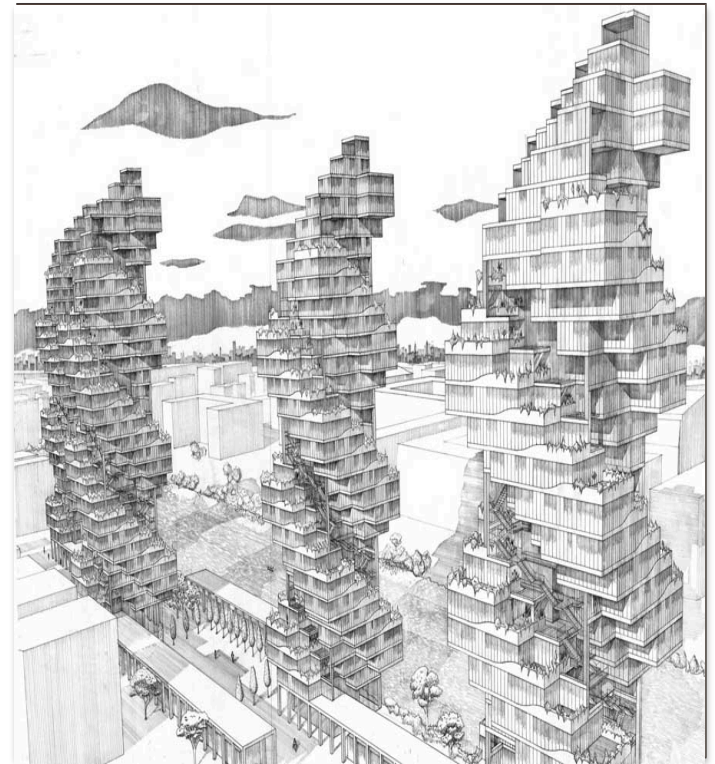
# ISO/IEC JTC1/WG9

ISO/IEC JTC 1/WG 9 was established by Resolution 28 during the 29th Plenary of the ISO/IEC JTC 1 in November 2014.



# WG9 Group

- \* WG9-BDWG met in the Computer Science School of Ciudad Real (Spain) to discuss the ISO/IEC 20546 (Big Data Overview and Vocabulary) and ISO/IEC 20547 (Big Data Reference Architecture) standards.





# WG9 Group

- \* Big Data experts USA, Japan, South Korea, Ireland, Germany, United Kingdom and Spain
- \* Working for different companies and academic organizations NIST, Oracle, Huawei, Microsoft, ETRI, InCadence Strategic Solutions, SAIC, Dublin City University and UCLM)



# ToR

Focus for JTC 1's Big Data standardization program

Develop foundational standards for Big Data throughout JTC 1 upon which other standards can be developed.

Develop other Big Data standards that build on the foundational standards

Identify gaps in Big Data standardization.

Develop and maintain liaisons with all relevant JTC 1 entities

Identify JTC 1 (and other organization) entities that are developing standards and related material that contribute to Big Data, and where appropriate, investigate ongoing and potential new work that contributes to Big Data.

Engage with the community grow the awareness of and encourage engagement in JTC 1 Big Data standardization efforts within JTC 1

# Current Work WG9 20547

## **ISO/IEC 20547** Information Technology – Big Data Reference Architecture (**5 Parts**)

**Scope:** This International Standard specifies the Big Data Reference Architecture (BDRA). The Reference Architecture includes the Big Data roles, activities, and functional components and their relationships.

# SCOPE

# 20547 Parts 1-3

## **Part 1: (TR) Framework and Application Process**

**Scope:** This technical report describes the framework of the Big Data Reference Architecture and the process for how a user of the standard can apply it to their particular problem domain.

## **Part 2: (TR) Use Cases and Derived Requirements \*\***

**Scope:** This technical report would decompose a set of contributed use cases into general Big Data Reference Architecture requirements.

## **Part 3: (IS) Reference Architecture \*\***

**Scope:** This International Standard specifies the Big Data Reference Architecture (BDRA). The Reference Architecture includes the Big Data roles, activities, and functional components and their relationships.

# 20547 Parts 4-5

## Part 4: (IS) Security and Privacy Fabric

**Scope:** This international standard specifies the underlying Security and Privacy fabric that applies to all aspects of the BDRA including the Big Data roles, activities, and functional components.

## Part 5: (TR) Standards Roadmap

**Scope:** This technical report will document Big Data relevant standards, both in existence and under development, along with priorities for future Big Data standards development based on gap analysis.

# ISO Big Data Editorial

Title	Editors	Co-Editors
ISO/IEC TR 20547-1, Information technology – Big Data Reference Architecture -- Part 1: <b>Framework and Application Process</b>	David BOYD (US)	Suwook HA (KR), Ray WALSHE (IE)
ISO/IEC TR 20547-2, Information technology – Big Data Reference Architecture - - Part 2: <b>Use Cases and Derived Requirements</b>	Ray WALSHE (IE)	Suwook HA (KR)
ISO/IEC 20547-3, Information technology -- Big Data Reference Architecture -- Part 3: <b>Reference Architecture</b>	Ray WALSHE (IE)	David BOYD (US)
ISO/IEC 20547-4, Information technology -- Big Data Reference Architecture – Part 4: <b>Security and Privacy Fabric</b>	Jacob DILLES (US)	
ISO/IEC TR 20547-5, Information technology – Big Data Reference Architecture -- Part 5: <b>Standards Roadmap</b>	David BOYD (US)	Toshihiro SUZUKI (JP), Abdellatif Benjelloun TOUIMI (UK)

