Bigger Perspectives on Metadata — Structural Ideas for MDR/MFI Series
(SC32/WG2/N1336)

Presentation By

Frank Farance, Farance Inc.

+1 212 486 4700, **frank@farance.com**

Dan Gillman, Bureau of Labor Statistics

+1 202 691 7523, **Gillman.Daniel@bls.gov**

Summary of Presentation

- Exploring a broader perspective of metadata
- Helps understand structures of MDR/MFI
- First, explore smaller problem (datum)
- Second, show issues over time (lifecycle)
- Third, expand to fuller solution (add in services / processes)
- Fourth, relate to MDR/MFI structure

Overview of Interoperability

- What does it mean?
- Why do we care?
- Kinds of interoperability
- Interoperability scenarios
- Benefits of interoperability
- How do we recognize interoperability?
- Application to MDR/MFI structuring

Defining Interoperability

- Why define “interoperability”?
  - Starting point for discussion
  - Makes my presentation clearer
  - Not sure if everyone agrees with this definition
- Other definitions of “interoperability”
  - Broader definitions apply, but don’t clarify
    - e.g., interoperability of machine parts
  - Some definitions are too narrow
    - e.g., focus on particular techniques / technologies

Defining Interoperability

<u>interoperability</u>: capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units

- ISO/IEC 2382-1, Information Technology — Vocabulary — Part 1: Fundamental Terms
  - Contains definition of "interoperability"
  - Particular to information technology
  - Developed by JTC1 (JTC1's scope *<u>is</u>* Information Technology standards)
  - Useful for the present discussion

Defining Interoperability

- Exploring the definition of “interoperability”
  - “capability ...”
    - interoperability is a kind of capability
  - “to communicate, ...”
    - successful communication is a kind of interoperability
  - “execute programs, or ...”
    - consistent/compatible execution is a kind of interoperability
  - “transfer data ...”
    - consistent interpretation of data is a kind of interoperability

Defining Interoperability

- Exploring the definition of "interoperability" ...
  - "among various functional units ..."
    - a "functional unit" is a system, subsystem, component, etc.
    - interoperability is a capability that is perceived across multiple functional units
  - "in a manner that requires the user ..."
    - a "user" can be a human, a computer, or combination
  - "to have little or no knowledge of the unique characteristics of those units"
    - don't need to know how the systems are implemented
    - there is an implied or explicit "specification" whose scope includes certain requirements, and the satisfaction of those requirements implies a particular degree of interoperability

Defining Interoperability

- JTC1 definition of “interoperability” is good enough
  - Useful for providing a framework for discussion
  - Useful for addressing technical issues
  - Useful for technical solutions
- In other words:
  - “good enough” is good enough

Overview of Interoperability

- What does it mean?
- Why do we care?
- Kinds of interoperability
- Interoperability scenarios
- Benefits of interoperability
- How do we recognize interoperability?
- Application to MDR/MFI structuring

Why Do We Care About Interoperability?

- Systems, services, data, etc. all need to connect to other systems, services, data, etc.
- Without interoperability, nothing would connect
  - Computer parts would not connect
  - Computers wouldn't "talk" to each other
  - Systems wouldn't "talk" to each other
  - Data wouldn't work with other data
  - and so on
- Interoperability is essential

Overview of Interoperability

- What does it mean?
- Why do we care?
- Kinds of interoperability
- Interoperability scenarios
- Benefits of interoperability
- How do we recognize interoperability?
- Application to MDR/MFI structuring

Kinds of Interoperability

- Examples within Information Technology
  - CD/DVDs: physical media, bit streams, file formats, file names, metadata, services (read/write)
  - Programming Languages: syntax/semantics, datatypes, character set, application programming interface (API), application binary interface (ABI), libraries, services
  - Security: hash functions, authentication methods, authorization methods, privacy techniques, confidentiality techniques, data integrity techniques
  - Networking: decomposition framework (ISO OSI layers), layer N services, layer N protocols
  - Data: datatypes, data exchange (codings, APIs, protocols), encodings, codes, metadata

Digression On Important Concepts (Farance/Gillman Research)

- datum: designation of a value
  - a datum is a kind of designation (in the terminological sense)
  - associates a signifier to a concept, a special kind of concept called a “value”
- value: concept that includes a notion of equality
  - any concept that includes/implies an “IsEqual()” function is a candidate as a *value* kind of concept
  - e.g., numbers, character strings

Digression On Important Concepts (Farance/Gillman Research)

- metadata: data that defines/describes objects
  - also known as “descriptive data”
  - data itself is a kind of object (data describing other data)
  - metadata can describe non-data objects, such as books (e.g., Dublin Core)
  - metadata is a kind of data that exists in the presence of a descriptive relationship
    - data without a connection to the object (object description) is merely data
    - a data element description (called a “data element” for short in 11179) is a kind of metadata

Digression On Important Concepts (Farance/Gillman Research)

- data element: object that can hold a datum
  - e.g., data elements store data
- data terminology: terminology that contains values
  - a terminology that includes value kind-of concepts
  - e.g., an 11179 value domain — signifiers (called "permissible values" in 11179) that designate value kind-of concepts (called "permissible value meanings" in 11179)

NOTE: Ongoing research by Farance/Gillman

Kinds of Interoperability

- Data interoperability:
  - important for data exchange / data interchange
  - meaning of data (e.g., broad, narrow, industry-specific) is agreed upon
  - satisfies characteristics of “interoperability”:
    - capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units

Kinds of Interoperability

- No single variety/specification of “quantity”
- Thus, metadata is essential for describing different variants of “quantity”
  - ISO/IEC 11179 is an example of one kind of descriptive data
- Thus, if the implied specification for interoperability is *“we’re exchanging quantities”*, then metadata is essential for interoperability
- The broad specification “exchanging quantities” is, essentially, blind data exchange
  - An important business/technical goal for organizations without established data exchange agreements

Kinds of Interoperability

- Metadata interoperability:
  - standardized descriptive data
    - standardized descriptive characteristics
    - standardized properties of those characteristics
    - standardized meanings of the property values
- ISO/IEC 11179 provides
  - common data semantics
    - standardized descriptive characteristics
    - standardized properties of those characteristics
    - standardized meanings of the property values
- Other standards may be layered
  - e.g., ISO/IEC 11404 standard for datatypes

Overview of Interoperability

- What does it mean?
- Why do we care?
- Kinds of interoperability
- Interoperability scenarios
- Benefits of interoperability
- How do we recognize interoperability?
- Application to MDR/MFI structuring

Interoperability Scenarios

- Degrees of conformity:
  - variety control = 1: conforming
  - variety control = 2: e.g.
    - "strictly conforming" — highest degree of interoperability
    - "conforming" — possibly, less interoperability
  - variety control = ∞:
    - interoperability specification is focused upon agreement of metadata (descriptive data)
- Degrees of conformity are independent of conformity roles
  - e.g., "strictly conforming API environment"
  - e.g., "conforming data instance"

Interoperability Scenarios

- Data interoperability conformity roles:
  - For codings (e.g., XML, ASN.1, etc.)
    - data instance
    - data writer/export
    - data reader/import (needs to conform to mandatory *and* optional elements)
    - data repository (needs to conform to mandatory *and* optional elements)
  - For APIs (C, Java, PHP, etc.)
    - application conformity (e.g., program conforms to spec)
    - environment conformity (e.g., implements library calls)
  - For protocols, dependent upon protocol nature:
    - client role for client-server protocols
    - server role for client-server protocols
    - peer role for peer-to-peer protocols

Overview of Interoperability

- What does it mean?
- Why do we care?
- Kinds of interoperability
- Interoperability scenarios
- Benefits of interoperability
- How do we recognize interoperability?
- Application to MDR/MFI structuring

Benefits of Interoperability

- Reduces technical risk
  - Can build systems, data, etc. towards the requirements/recommendations of a specification
  - Data exchange complexity is reduced from N^2 to N because each “handshake” need not be explored
- Reduces business risk
- Facilitates
  - Current capabilities: data interchange is possible, to a particular degree
  - Future capabilities: new kinds of data requires additional descriptive data
    - Reuse existing metadata infrastructure

Overview of Interoperability

- What does it mean?
- Why do we care?
- Kinds of interoperability
- Interoperability scenarios
- Benefits of interoperability
- How do we recognize interoperability?
- Application to MDR/MFI structuring

Techniques for Measuring Data Interoperability

- Step #1: Reformulate data, data elements, datatypes, classes, object classes, etc. according to terminological principles (ISO 704)
  – Express data as designations with concept relationships
  – Example: a data element contains the value 17 might be reformulated as “On 2007-07-09 at 10:00 on Roosevelt Island, NY, USA, the temperature, in Celsius, measured by thermometer XYZ, to 0.1 degrees precision, with accuracy 0.2, is: 17”

Techniques for Measuring Data Interoperability

- Step #2: Express reformulation as series, lattice, etc. of derived concepts:
  - Example: “On 2007-07-09 at 10:00 on Roosevelt Island, NY, USA, the temperature, in Celsius, measured by thermometer XYZ, to 0.1 degrees precision, with accuracy 0.2, is: 17” describes:
    - a kind of 17
    - a kind of temperature
    - a kind of temporal extent
    - a kind of spatial extent
    - a kind of precision
    - a kind of accuracy
    - a kind of measuring instrument
    - a kind of measuring technique
    - and so on ...

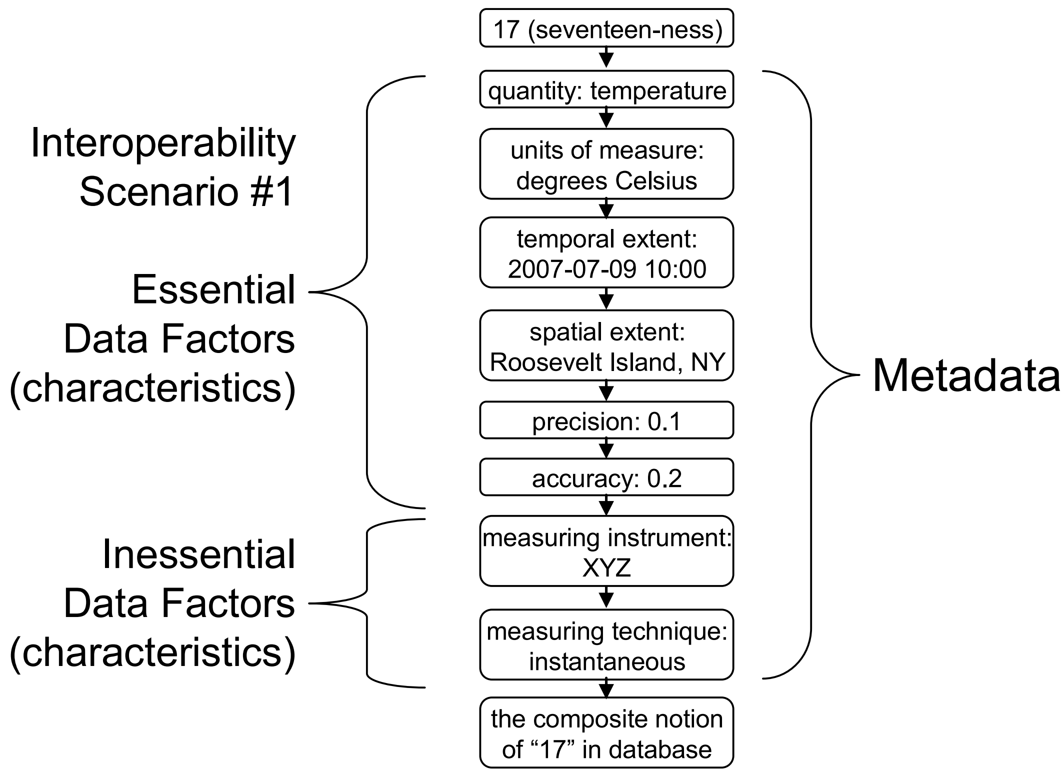
Techniques for Measuring Data Interoperability

- Step #3: Determine which characteristics are essential
  - Same as the terminological process for developing good definitions
  - Partition into essential and inessential characteristics
- Data interoperability is dependent upon essential characteristics

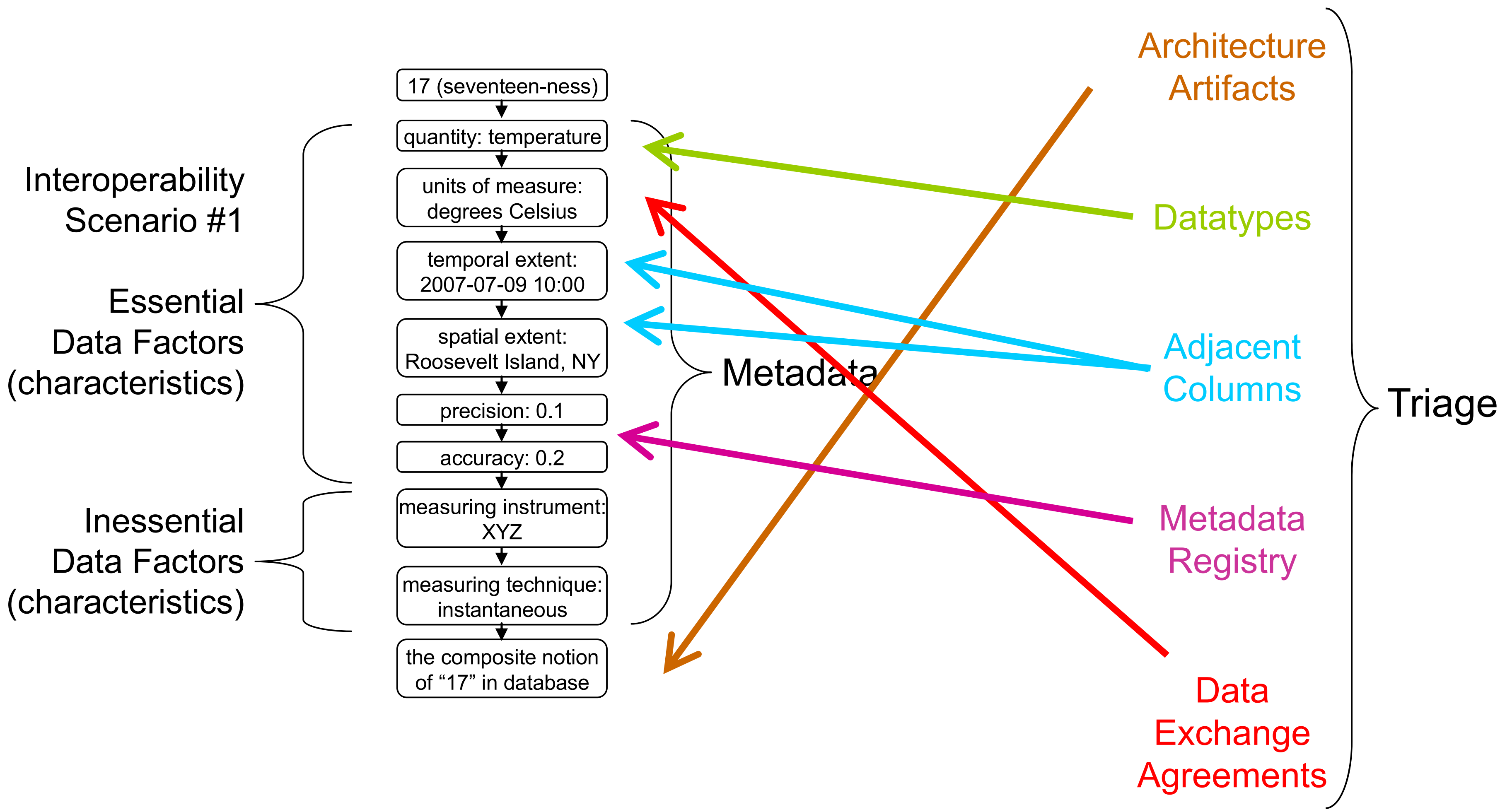
Techniques for Measuring Data Interoperability

- Example #1:
  - Example: “On 2007-07-09 at 10:00 on Roosevelt Island, NY, USA, the temperature, in Celsius, measured by thermometer XYZ, to 0.1 degrees precision, with accuracy 0.2, is: 17” describes:
    - a kind of 17 — quantity is essential
    - a kind of temperature — dimension is essential
    - a kind of temporal extent — interval dimension is essential
    - a kind of spatial extent — interval dimension is essential
    - a kind of precision — computation limitation is essential
    - a kind of accuracy — data quality criterion is essential
    - a kind of measuring instrument — instrument is inessential
    - a kind of measuring technique — technique is inessential

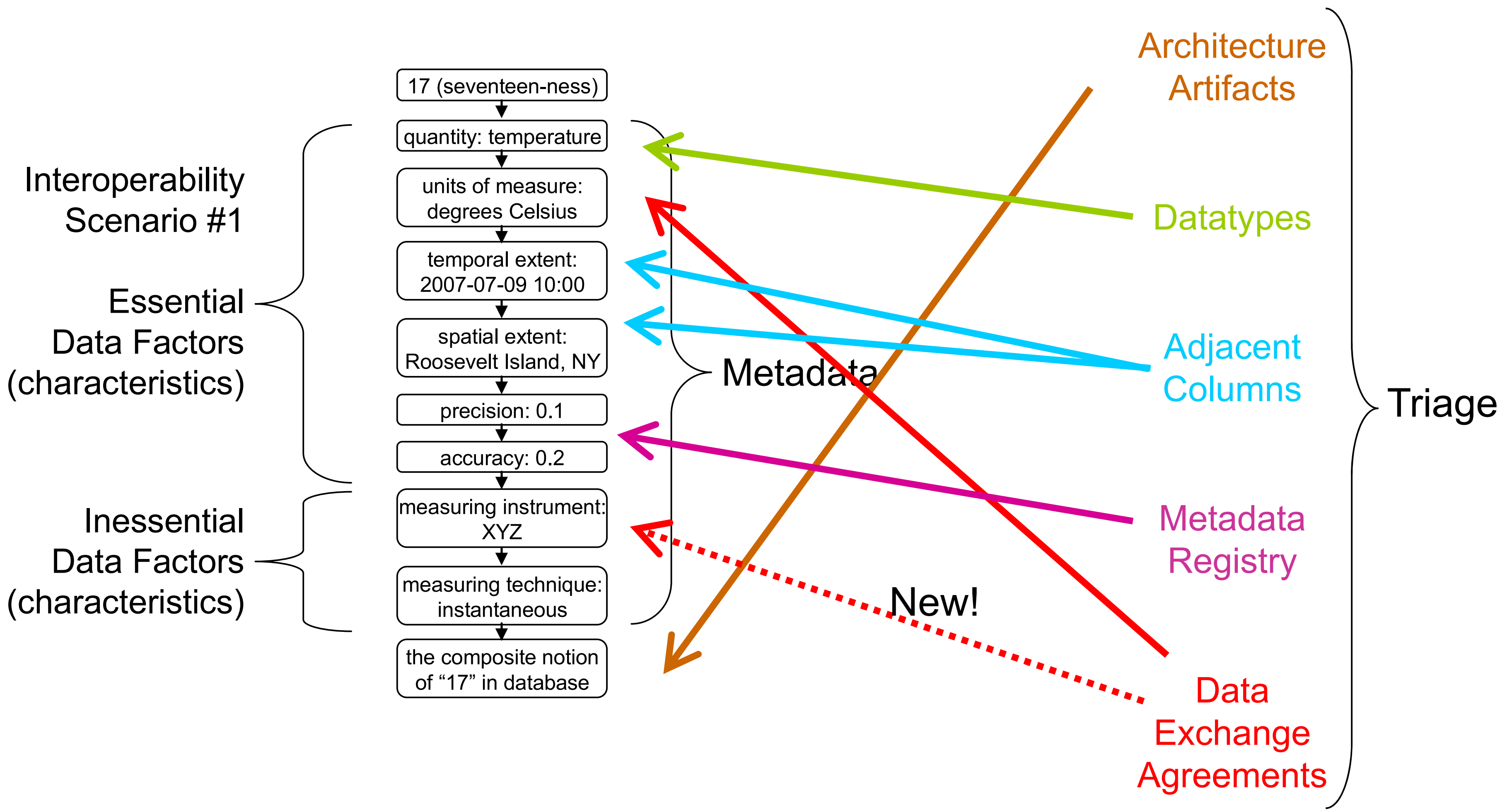
Techniques for Measuring Data Interoperability



Software Development Lifecycle Isues (Triaging)



Software Development Lifecycle Isues (Triaging -- Changes over time)



Techniques for Measuring Data Interoperability

- Thus, the interoperability specification concerns: number, temperature, temporal-spatial extent, precision, accuracy, but NOT measuring instrument, measuring technique
- Likewise another interoperability specification might include measuring instrument and measuring technique, but not spatial extent

Techniques for Measuring Data Interoperability

- By using well-defined characteristics and well-defined derivations, the metadata *affords* a well-defined computation model for examining data interoperability
- Characteristics may be structured linearly, hierarchically, or in some other group
- Organization of characteristics is optimized for re-use

Other Kinds of Data-Related Description

- Above discussion only shows concept derivation
- Data has computational descriptions, too
  - Characterizing operations
  - Data-like objects
  - Services
  - Processes (e.g., object-oriented objects)

Overview of Interoperability

- What does it mean?
- Why do we care?
- Kinds of interoperability
- Interoperability scenarios
- Benefits of interoperability
- How do we recognize interoperability?
- Application to MDR/MFI structuring

Fundamental Aspects of Description

- Object description (in several categories)
- Identifiers (for naming)
- References (for re-use)
- Grouping (for re-use)
- Relationships (for connections)
- Organization processes (e.g., biz process)

Fundamental Aspects of Description Applied to MDR/MFI

- Object description (in several categories)
  - Relationship systems [MDR part 2]
  - Data semantics [MDR part 3]
  - Terminological features [MDR part 4]
  - Names/identifiers [MDR part 5]
  - Registration process/data [MDR part 6]
  - Instance descriptions [MDR/MFI ???]
  - Mapping equivalences [MFI part 2]
  - Service Descriptions [MFI part 7]
  - Process Descriptions [MFI part 5]
  - RGPS as standardized relationships [MFI part?]
  - Computational Descriptions [MDR/MFI part?]

Fundamental Aspects of Description Applied to MDR/MFI

- MDR Part 1:
  - Core features, including
    - Tuples (identifier / kind / value)
    - References
    - Grouping
    - Relationship recording
- MFI Part 1:
  - Core features, including
    - Mappings
    - Relationships

Fundamental Aspects of Description Applied to MDR/MFI

- Conclusions
  - Descriptive data (metadata) applies broadly
  - Use standardized methods for attributes
  - Very very simple typing for data
  - Consider standardized relationship types
  - Need to support re-use/grouping
  - Presentation (standard parts) structure should match conformity boundaries

More Information

- Contact:
  - Frank Farance, Farance Inc.,
    +1 212 486 4700
    frank@farance.com
  - Dan Gillman, Bureau of Labor Statistics
    +1 212 486 4700
    Gillman.Daniel@bls.gov