Workshop
Asset Reporting Format (ARF) and Asset Identification

Adam Halbardier
Booz Allen Hamilton
National Institute of Standards and Technology (NIST)

John Wunder
MITRE Corporation

6th Annual IT Security Automation Conference

What is ARF and Asset Identification

- What is Asset Identification
  - NIST Interagency Report (IR) 7693
  - A specification governing the method and format to identify and represent assets
- What is ARF
  - NIST Interagency Report (IR) 7694
  - A specification governing the formatting of reports about assets
  - Defines how tools should report on information about assets

Agenda

- Asset Identification Issues
- ARF Use Cases and Relationships
- Timeline and Ways to Participate

Agenda

- Asset Identification Issues
- ARF Use Cases and Relationships
- Timeline and Ways to Participate

Asset Identification

How do you associate information about an asset with the asset itself?

Asset Identification

Or,

Asset Identification

How do you uniquely identify an asset and represent that identification?

Use Cases

- Reporting
  - E.g. assessments, remediations, events
- Tasking
  - E.g. assessments, remediations
- Contextual Information
  - E.g. owning organization, location, network, etc.
- Federation of asset databases

What do you get?

- Correlation of data across the management domain, including from varying…
  - Sensor types
  - Timeframes
  - Result types
  - Vendors

Are we there yet?

- Automated security specifications use varying mechanisms to identify assets
  - **Incompatible** specifications
  - **Inconsistent** implementations
  - **Incomplete** information

How can we get there?

- Single specification to identify assets
- May be used by specification authors as identification elements
  - OVAL
  - XCCDF
  - OCIL
  - Digital event reporting
  - Remediation

How it works

Assets may be identified using a combination of

zero to many **canonical identifiers** and/or

some set of **identifying information**

Canonical Identifiers

- Many tools assign identifiers to assets they manage
- Assets may be identified using an **assigned identification element** in the context of a **namespace**
- Ex:
  - Namespace: VendorProduct1
  - Identifier: Asset3544

Identifying Information

- Sometimes, assigned identifiers are unavailable or not shared
- But, some information that is **collectable** or **discoverable** about an asset is available
  - Devices: hostname, IPv4 address, MAC address
  - People: Full name, location, organization
  - Organizations: Name, type
- Some amount of certainty of an accurate identification

How it works

Assets may be identified using a combination of

zero to many **canonical identifiers** and/or

some set of **identifying information**

Examples

Canonical IDs:
- Asset1234@MITRE

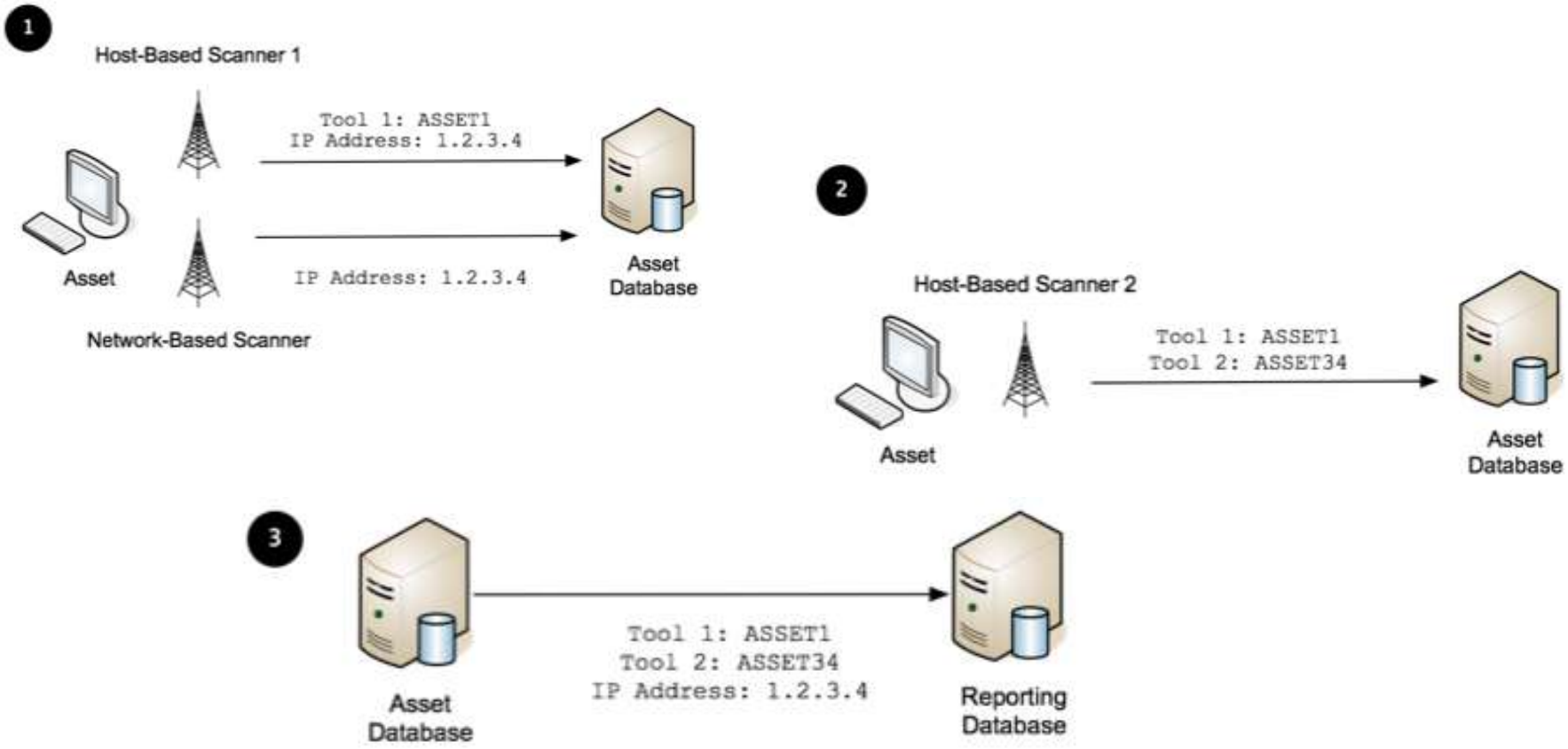
Canonical IDs:
- Asset1234@Tool1
- Asset4321@Tool2

Canonical IDs:
- Asset1234@Tool1
- Asset4321@Tool2

Identifying Information:
- IPv4: 1.2.3.4
- Hostname: mm123123

Identifying Information:
- IPv4: 1.2.3.4
- Hostname: mm123123

Sample Usage (Reporting)



What's an asset?

- Device
- Person
- Organization
- Network
- System
- Software
- Circuit

Problem: Enumerations vs. Open-Ended Values

- Enumeration: canonical set of possible values
  - Greater compatibility
- Open-Ended Value: any valid value
  - Greater flexibility
- Controlled Vocabulary: namespaced set of allowable values
  - Trade-off between compatibility and flexibility

Case Study: Organization Type

- Enumeration
  - Government (NIST)
  - For-Profit Corporation (Booz-Allen Hamilton)
  - Non-Profit (MITRE)
- Open-Ended Value
  - “Government”
  - “Federal Government”
  - “Federally Funded Research and Development Corporation”

Controlled Vocabulary

- Namespaced set of allowed values
  - Core namespace to meet common use cases
  - Extension namespaces can be created ad-hoc to meet emerging use cases
- Easier to change, prevents inconsistencies
- But harder to validate and manage
- Trade-Off: how open are values, how often do they change?

Problem: Schema Incompatibilities

- AI imports xAL to do addresses and GML to do geolocations
  - GML and xAL import independent Xlink schemas that define the same thing (this is bad)
  - AI, by itself, currently has a conflict
- Specifications importing AI introduce additional incompatibilities if they include Xlink
  - E.g. ARF

Possible Solutions

- Only rely on either GML or xAL
  - Solves immediate problem
  - But specifications relying on AI might need to import the other
- Don’t rely on either
  - Solves permanent problem (AI is not importing troublesome schemas)
  - At expense of reuse
- Other technical solutions?
  - XML Catalog

Proposal

- ???
- GML is easily replaced by a simple custom implementation (point and radius)
- xAL is more powerful and harder to replace
  - Are there any other options for international addresses?
  - Or should we roll our own?
  - Or keep xAL and warn specifications that want to include it

Problem: Data Source of Identifying Information

- At times, the data source for collected information matters
  - ie network scanners may report different IP addresses depending on where they are on the network
- At Developer Days, it was suggested that identifying information is not valuable unless it is tied to collecting sensor

Possible Solutions

- Tag each piece of identifying information with originating sensor
  - Optional or required?
  - Include canonical IDs as well?
  - How do you handle correlated data?
- More robust data element for handling data source
- Assume identifying information is sourced by the immediate data source
- Assume identifying information is unsourced

Proposal

- Assume the originating source is always the immediate source
  - Is is workable?

Agenda

- Asset Identification Issues
- ARF Use Cases and Relationships
- Timeline and Ways to Participate

Purpose of ARF

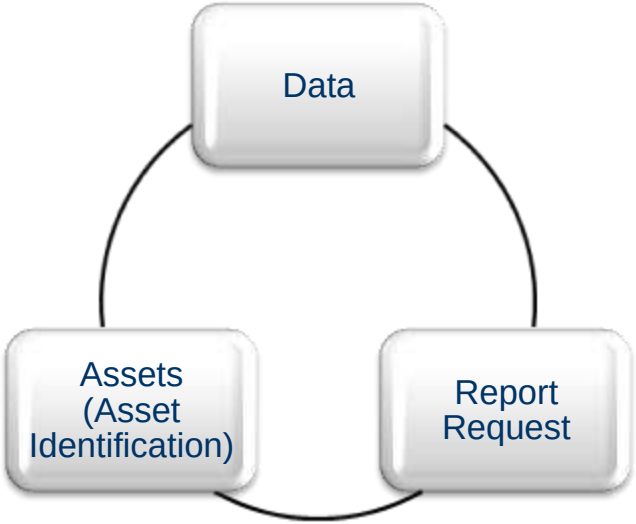
- Define a data model to house data about:
  - Assets
  - Asset identification information
  - Requests for asset information
  - The relationships between the components above
- Define a specification to report about assets in support of numerous use cases in government and industry at various levels of detail

Purpose of ARF (con't)

- Enable asset report correlation
  - Leverage the Asset Identification specification to identify the subjects of reports enabling different reports about the same assets to be correlated across and enterprise

Scope of ARF

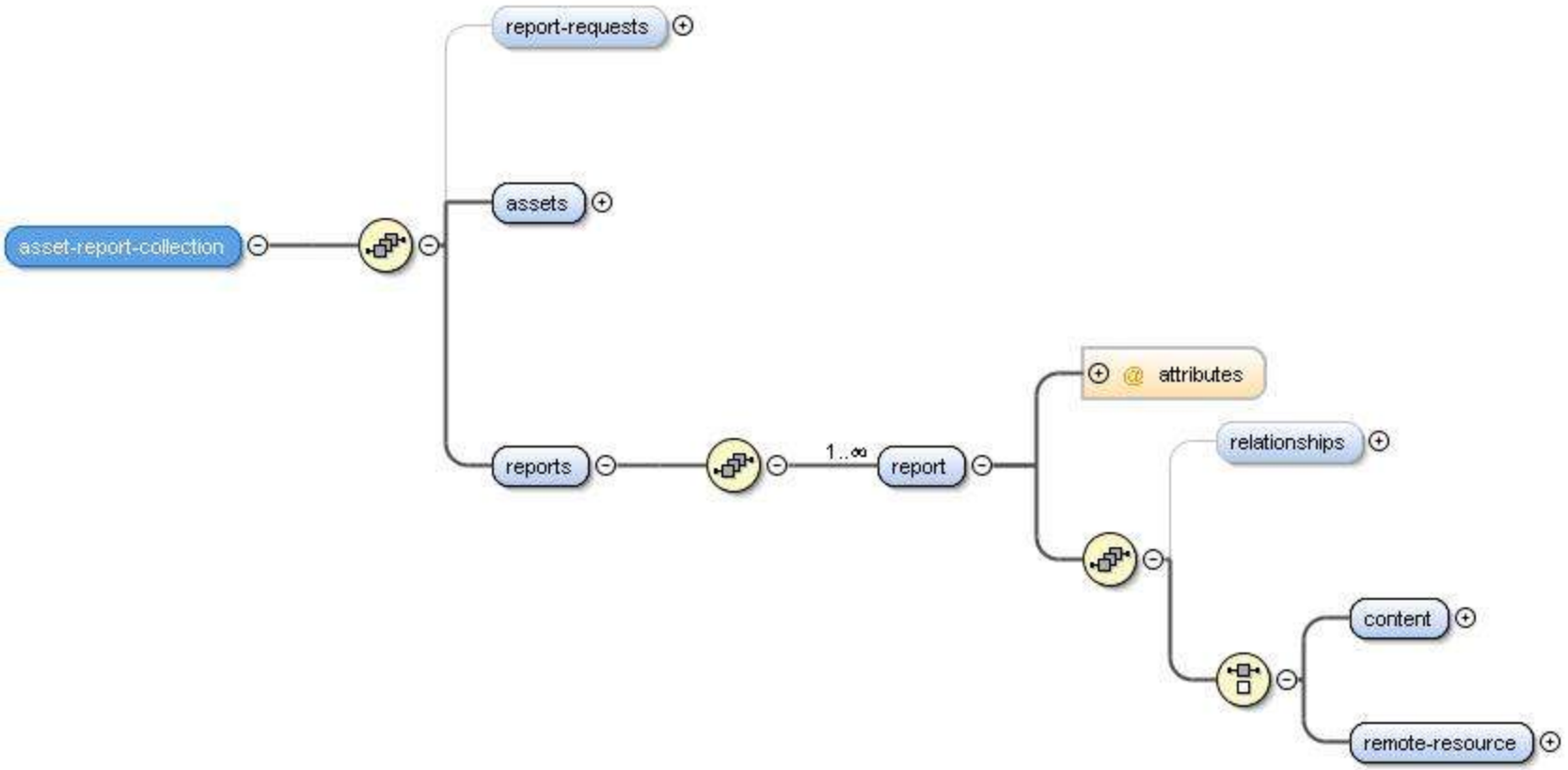
- Define the report transport data model
- Define the relationships between asset report components, while leaving the low-level data models to other specifications



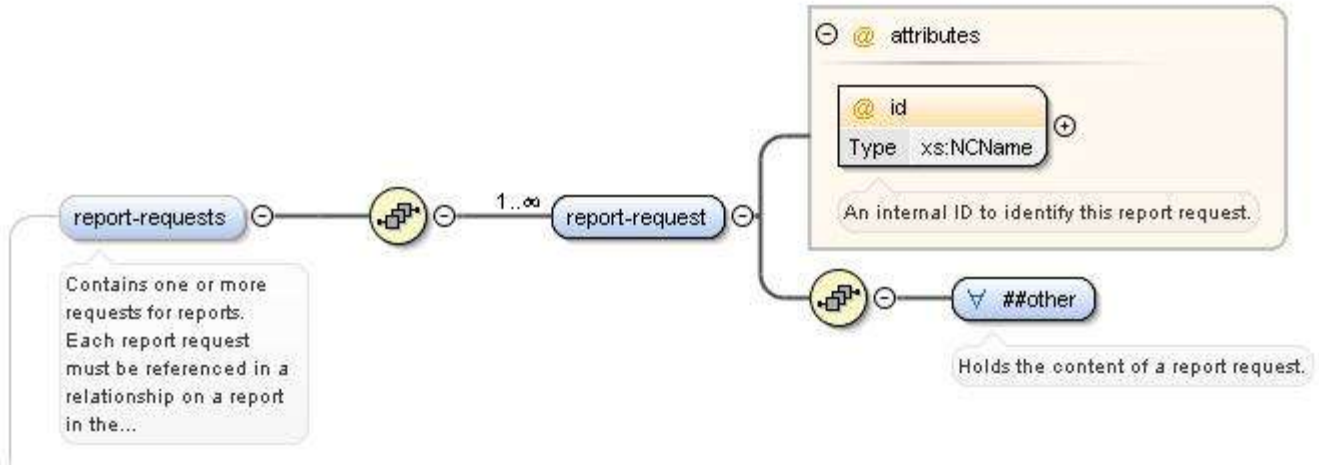
High-level Requirements

- Must be able to:
  - associate one or more assets with arbitrary payloads
  - define explicit relationships between payloads and assets
  - combine multiple ARF reports into a single ARF report
  - define reusable sets of data
  - reference data external to the ARF report

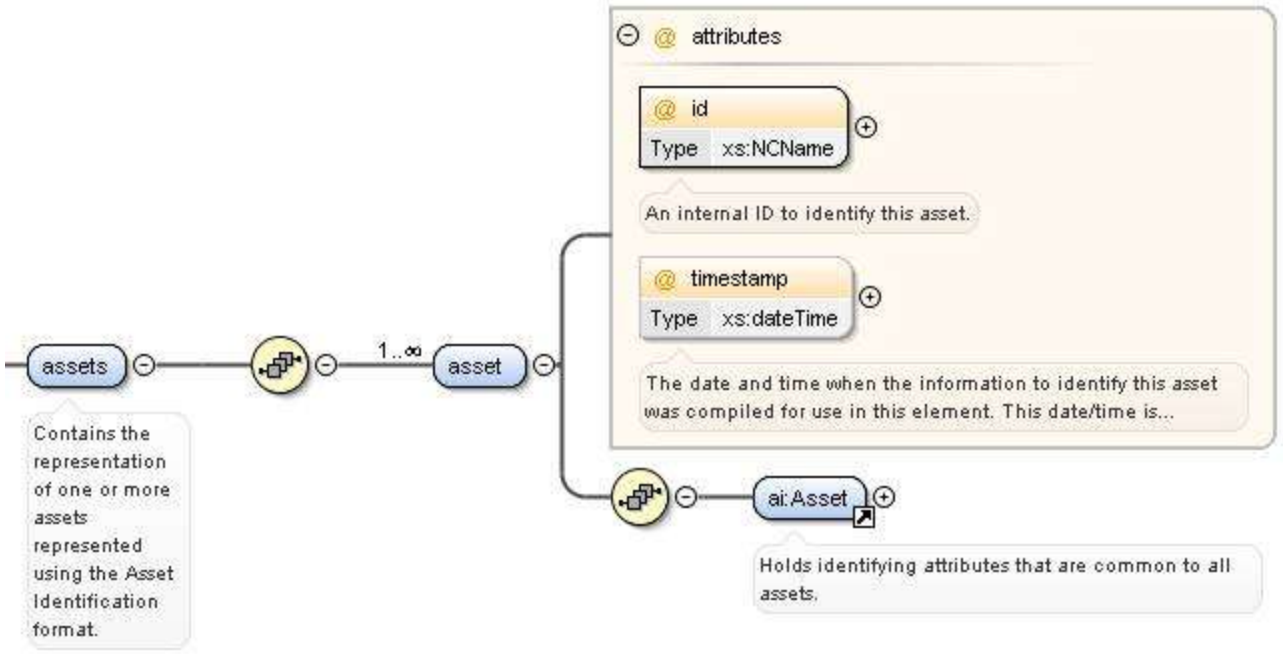
Data Model


report-requests
assets
asset-report-collection
reports
1..∞
report
attributes
relationships
content
remote-resource

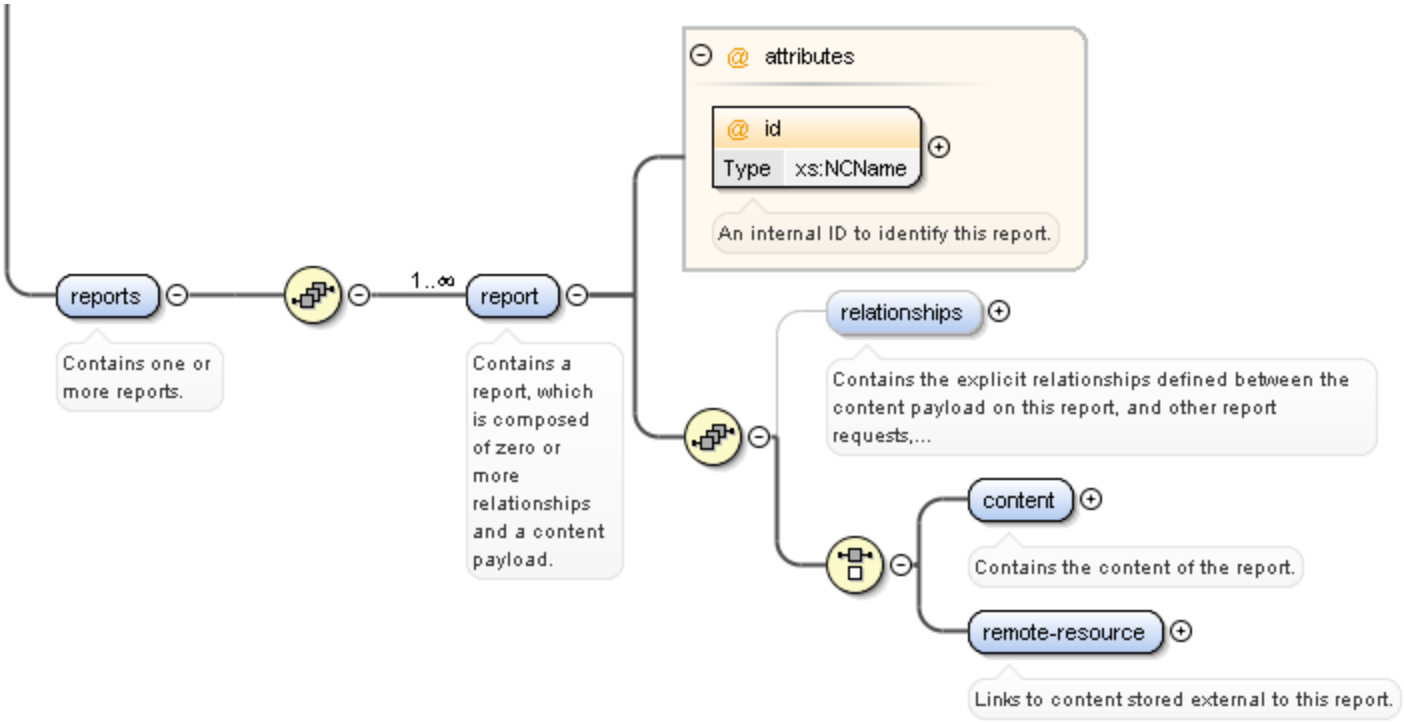
Data Model – report-request



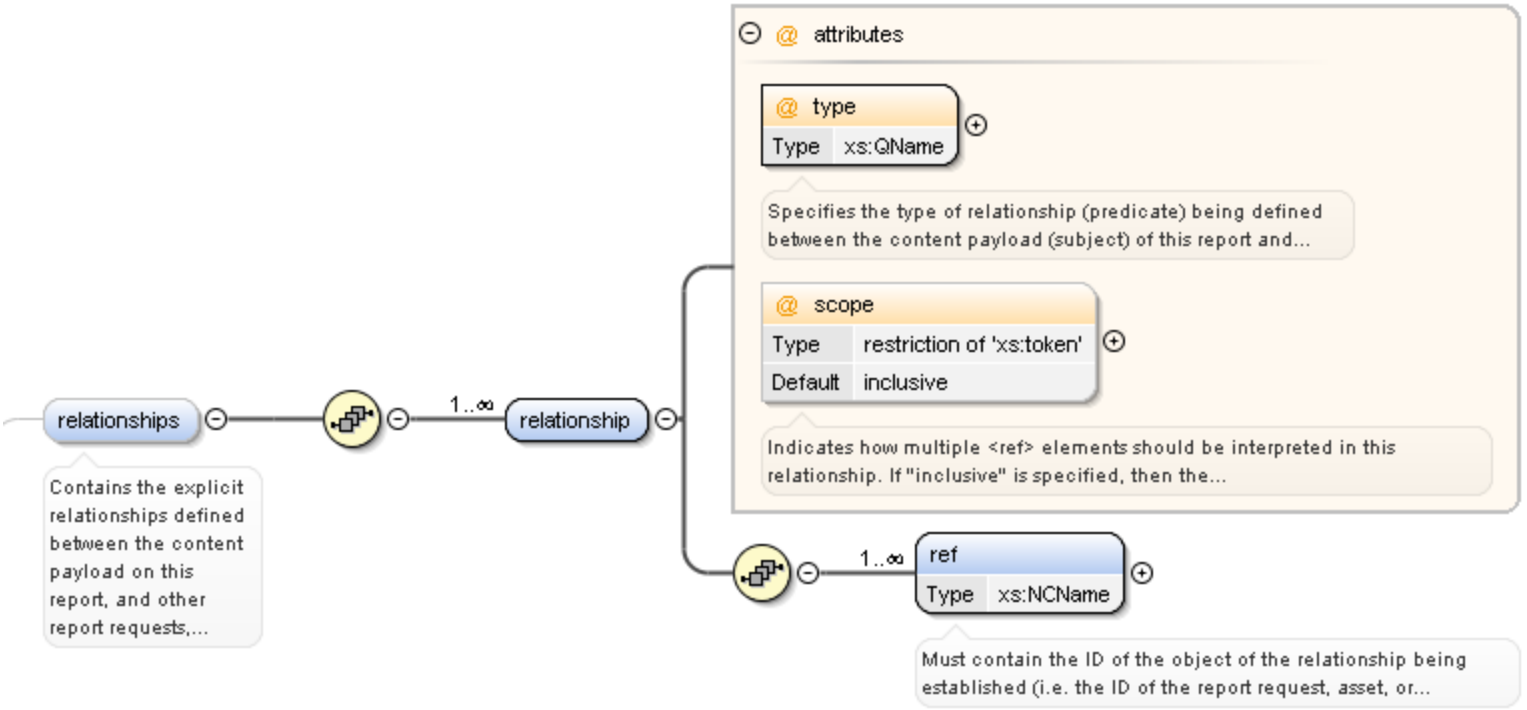
Data Model - asset



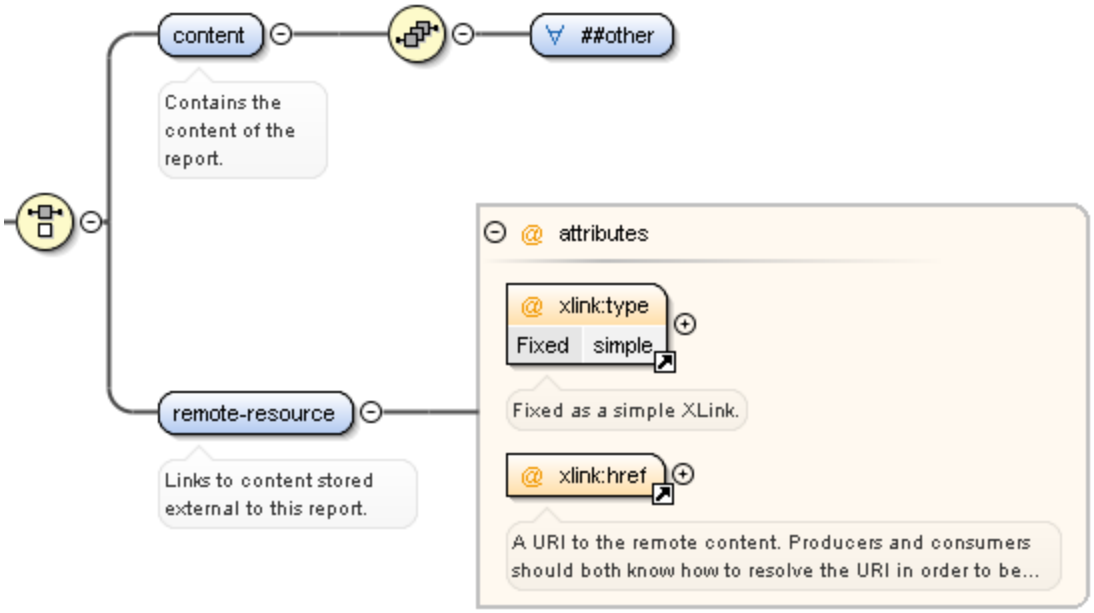
Data Model - report



Data Model – relationship



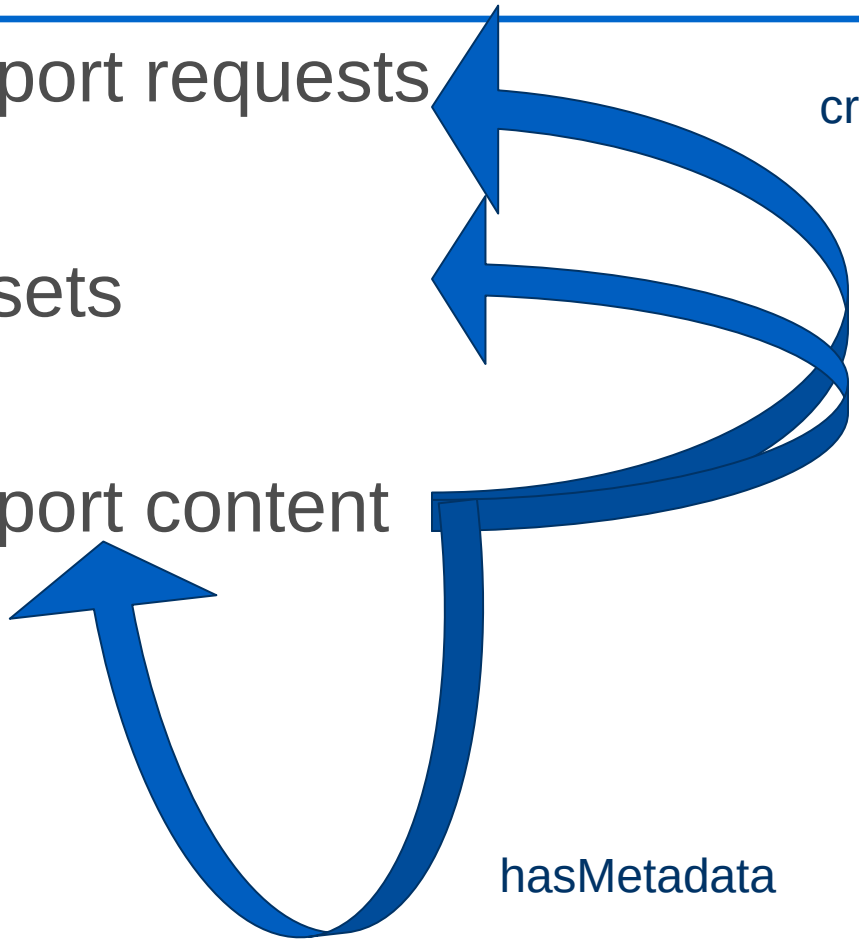
Data Model – content



Data Model Suggestions?

Objects to Be Related

- Report requests
- Assets
- Report content



createdFor

isAbout
hasSource
retrievedFrom
createdBy
emittedBy

hasMetadata

http://scap.nist.gov/vocabulary/arf/relationships/1.0#

Are We Missing Anything?

- createdFor -> report-request
- isAbout -> asset
- hasSource -> asset
- retrievedFrom -> asset
- createdBy -> asset
- emittedBy -> asset
- hasMetadata -> report

Use cases

- Compliance Assessment
- Vulnerability Management

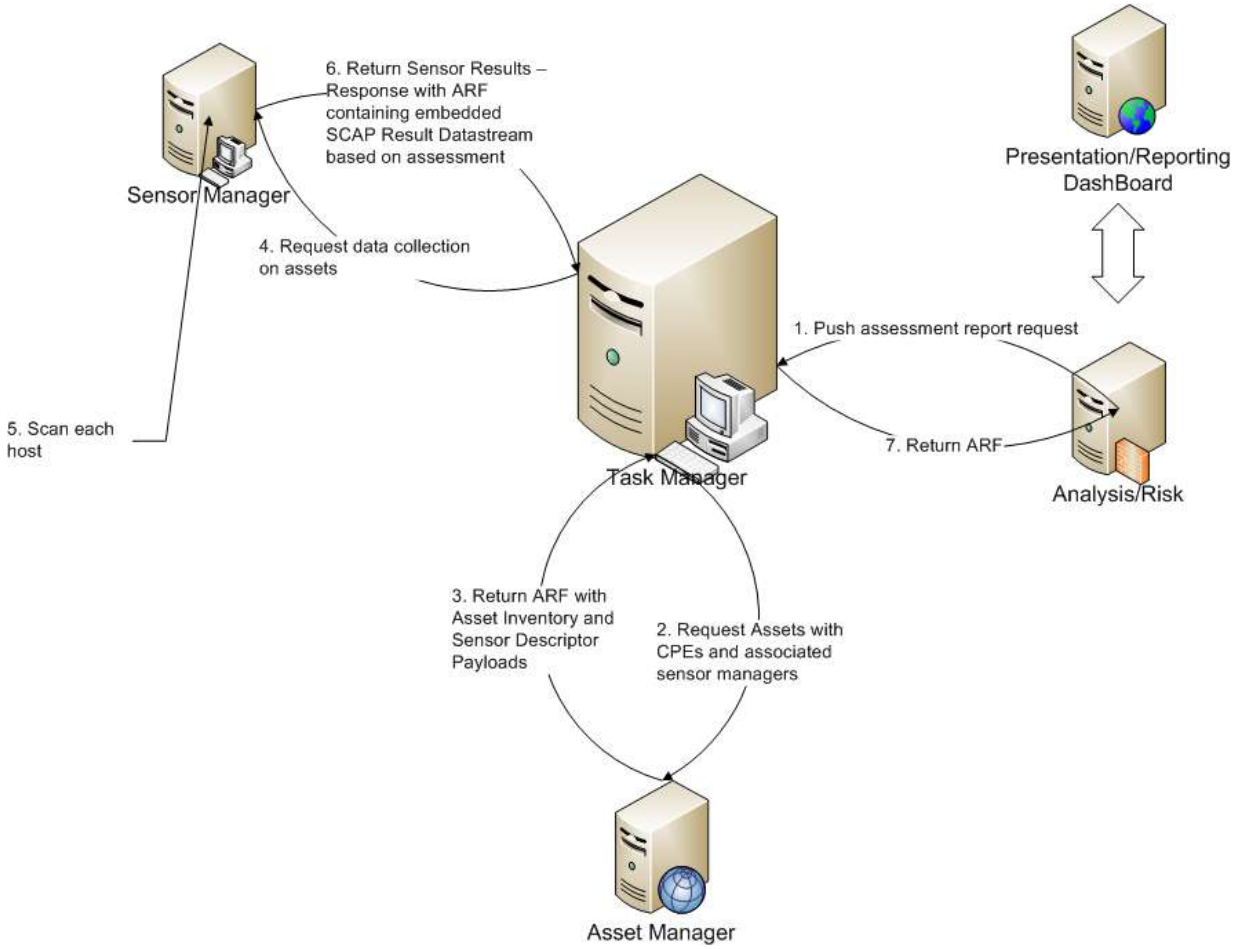
} Assessment

- Asset Discovery and Inventory Management

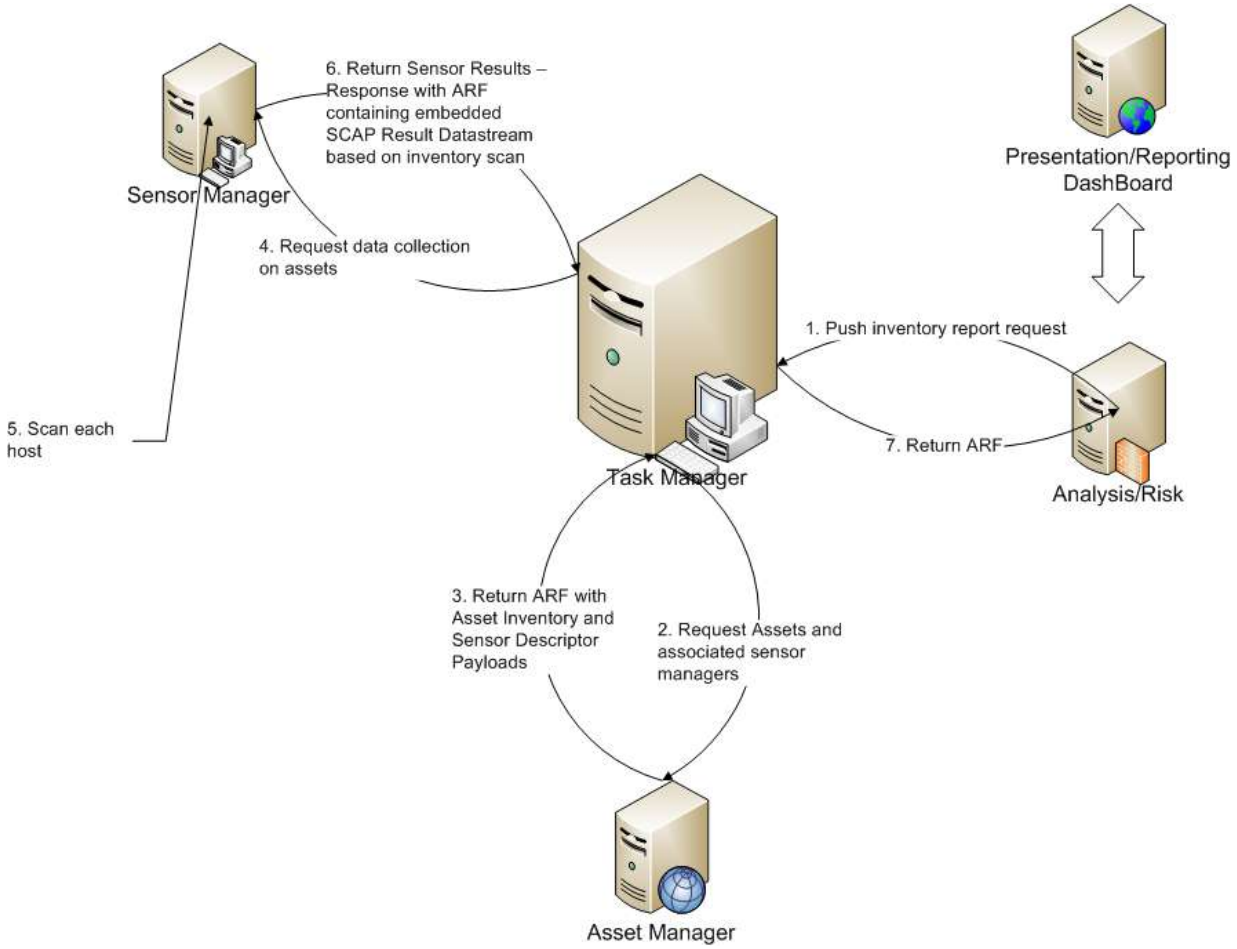
} Inventory

- Digital Event Analysis

Assessment Use Case



Inventory Use Case



Thoughts?

Agenda

- Asset Identification Issues
- ARF Use Cases and Relationships
- Timeline and Ways to Participate

Timeline

- After workshop, changes will be incorporated into ARF and Asset Identification and drafts will be released
- Drafts will enter NIST 30 day public review period
- Specifications final in Winter 2010
- Inclusion in SCAP 1.2

Get Involved

- Contact any member of the working group
  - Adam Halbardier – adam.halbardier@nist.gov
  - John Wunder – jwunder@mitre.org
  - Dave Waltermire – dave.waltermire@nist.gov
  - Mark Johnson – mark.johnson@nist.gov
- Email to emerging-specs@nist.gov
- Ask about getting involved in the working group
- Submit comments on NIST IR 7693 and 7694

Questions & Answers / Feedback

John Wunder
MITRE Corporation
jwunder@mitre.org
(781) 271-4602

Adam Halbardier
Booz Allen Hamilton
Supporting National Institute of Standards and Technology (NIST)
adam.halbardier@nist.gov
(310) 297-5444