

Email Ontology Tutorial

Dave Salmen

Bill Mandrick

Data Tactics Corporation



Email Ontology



Introduction

- Email Domain Ontology - Purpose
- Email Domain Ontology - Scope
- Ontology Basics
- Where To Start?
- Ontology Levels and Re-Use
- Warm-up Exercise – Email Standards Ontology
- Email Ontology - Base Classes
- Where To Stop?
- Email Ontology - Information Content - Classes
- Email Ontology – Definitions
- Email Ontology - Information Content – Properties
- Email Information Content Extraction – Empirical Results

Purpose

- Email Domain Ontology
 - To illustrate detailed steps of the ontology creation methodology using a domain of information artifacts that is familiar to a wide audience.
 - Intelligence Community work often involves working with content across the spectrum of information artifacts

Repeatable Process for Ontology Development

Inputs 	Activities 	Outputs
Subject Matter Expertise User Requirements Authoritative Sources/Definitions Salient Databases	I Scope the Domain SME Interviews Survey Sources Identify Baseline Terms Establish Metrics	Domain Definition Initial List of Terms Metrics Statement
SME Feedback Taxonomies Folksonomies Indexes	II Create Iterative Lexicon Decompose Baseline Terms Create Ontological Definitions Identify Relations	Iterative List of Terms List of Relations Versioned Domain Lexicon
Iterative List of Terms List of Relations Versioned Domain Lexicon	III Create Initial Ontology Extend from Upper Ontology Relate Entities and Events Employ Tool (e.g. TBC)	Versioned owl file Graphic Depictions SME Update Briefing
Metrics Versioned owl file SME Update Briefing	IV Revise Ontology SME Review Review Metrics Revision Iterations	Revised OWL File Revised Briefings Revised Domain Lexicon
Revised OWL File Revised Briefings Revised Domain Lexicon	V Publish Ontology Post to Repository Post Change Request Process Conduct Briefings	Domain Lexicon Versioned OWL File Lessons Learned Executive Briefings Change Request Process

Scope

- Detailed terms from RFC 5322 - Internet Message Format and related RFC documents
- Core terms from related to Multipurpose Internet Email Extension (MIME) from RFC 2045, RFC 2046, RFC 2047
- Core terms for email network protocols:
 - POP - RFC 1939,
 - IMAP - RFC 3501
 - SMTP – RFC 5321
- Additional consideration given to terms from the JSR919 - JavaMail API Specification 1.5 and the Java email parsing library implementation

Reply Reply All Forward

Information Artifact: E-Mail



Blowback: The Unintended Consequences of Hezbollah's Role in Syria // CFC Complex Coverage Publication

CFCPublications@cimicweb.org

IBE: Computer Display or Printed Paper Copy

ICE: Subject Line

To: CFCPublications@cimicweb.org

Cc: robin.barnett@cimicweb.org

ICE: To Address

Attachments: 20130915_Hezbollah_Final.pdf (996 KB)

ICE: PDF in an Attachment Role

Monday, September 16, 2013 1:30 AM

The NATO Allied Command Operations (ACO) Civil-Military Fusion Centre (CFC) is pleased to release another publication titled:

Blowback: The Unintended Consequences of Hezbollah's Role in Syria

Manage your subscriptions by clicking here: <https://www.cimicweb.org/layouts/Profile/Manage.aspx>

Access all CFC reports on our website <https://www.cimicweb.org>

We appreciate your feedback. Email us at cfcpublications@cimicweb.org

On behalf of CDR Joergen Jessen,

Robin Michelle Barnett, CFC Executive Editor

Norfolk, Virginia, USA 23508

T + 1 757 683 6466

robin.barnett@cimicweb.org

IQE: Color Scheme, Font, Resolution...

ICE: URI

Creating the Email Domain Ontology

- Classes
 - Sub-Classes
- Properties
 - Domain/Range
 - Property type
 - ObjectProperty, DatatypeProperty, AnnotationProperty
 - Sub-Properties
- Instances
- Ontology Level and Ontology Re-Use

Email Domain Ontology

Where To Start?

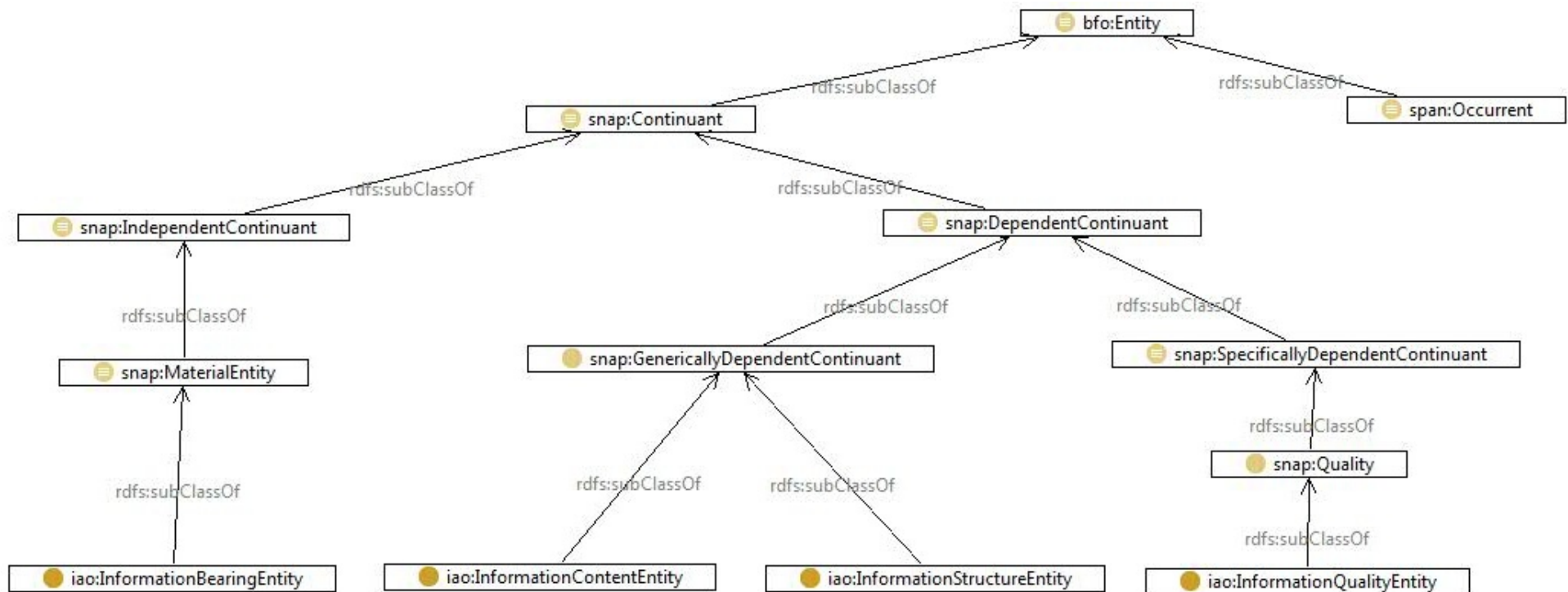
- Modular Ontology Construction
 - Ontology Levels
 - Downward Population
 - Ontology Re-use
- Email Domain Expertise

Ontology Levels and Ontology Re-use

- ULOs
 - Basic Formal Ontology (BFO)
 - Relationship Ontology (RO)
- MLOs and LLOs
 - Information Artifact Ontology (IAO)
 - Email Domain Ontology
 - Contact Ontology
 - Computer Network Ontology (CNO)
 - Software Ontology (SWO)

Basic Formal Ontology (BFO)

Information Artifact Ontology (IAO)



Where To Start (continued)?

- Email Domain Expertise
 - Email Related Internet Standards
 - IETF - Request for Comments (RFC)
 - Internet Message Format
 - » Multipurpose Internet Mail Extensions (MIME)
 - Post Office Protocol
 - Internet Message Access Protocol
 - Simple Mail Transfer Protocol
 - Java Specification Request (JSR)
 - JavaMail API Specification
 - Email Parsing - Empirical Results

Warm-up Exercise

Email Standards Ontology

- Email Related Standards
 - IETF Request for Comments (RFC)
 - RFC 5322 - Internet Message Format
 - RFC 2045, RFC 2046, RFC 2047 – MIME Extensions
 - RFC 1939 – Post Office Protocol
 - RFC 2060 – Internet Message Access Protocol
 - RFC 5321 – Simple Message Transfer Protocol
 - Java Specification Request (JSR)
 - JSR 919 JavaMail API Specification

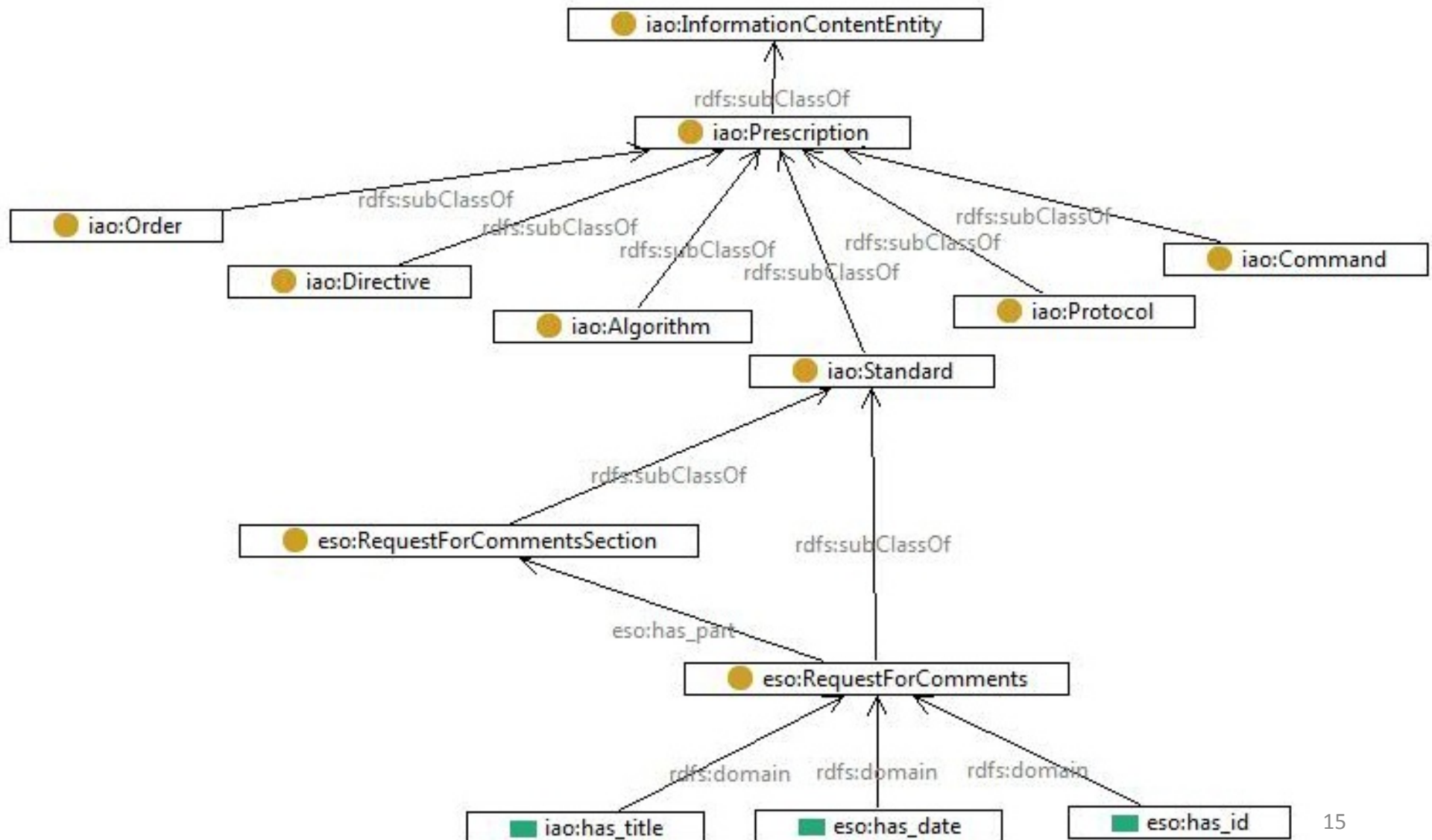
Email Standards Ontology

- RFC instances ontology diagram from:
 - rfc-0.3.xlsx
 - RFC5322 - Internet Message Format
 - RFC1939 - Post Office Protocol – Version 3
 - RFC2060 - Internet Message Access Protocol – Version 4rev1
 - RFC5321 – Simple Message Transfer Protocol
 - JSR919 – JavaMail API Design Specification Version 1.5

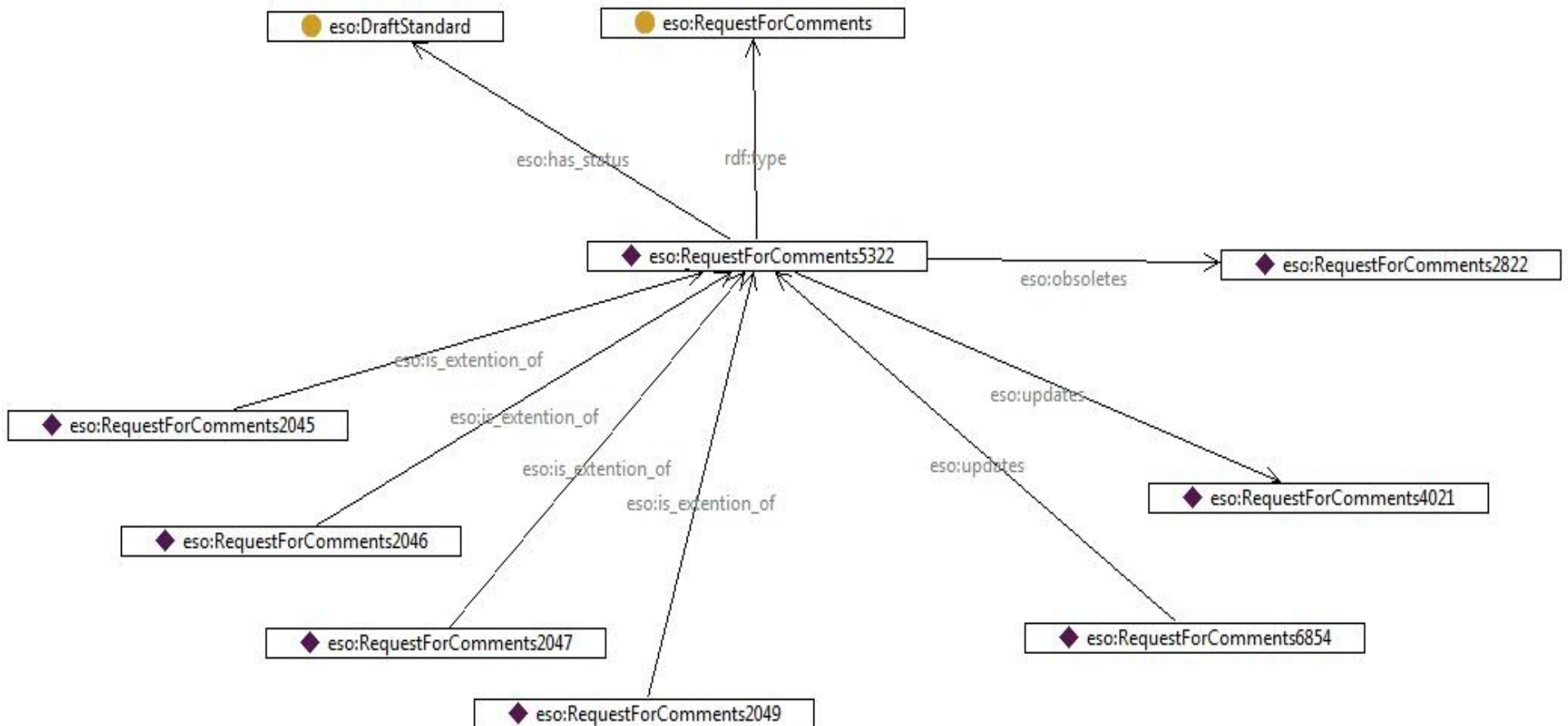
Email Standards Ontology

RFC	Title	Category	Status	Date	Relationships
RFC5322	Internet Message Format	Standards Track	Draft Standard	October 2008	obsoletes RFC2822, updates RFC4021
RFC2822	Internet Message Format	Standards Track	Proposed Standard	April 2001	obsoletes RFC822
RFC822	Standard for the Format of ARPA Internet Text Messages	Standards Track	Internet Standard	August 3, 1982	
RFC4021	Registration of Mail and MIME Header Fields	Standards Track	Proposed Standard	March 2005	
RFC6854	Simple Authentication Schemes for the Asynchronous Layer Coding (ALC) and NACK-Oriented Reliable Multicast (NORM) Protocols	Standards Track	Proposed Standard	March 2013	updates RFC5322
RFC2045	Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies	Standards Track	Draft Standard	November 1996	extends RFC5322, obsoletes RFC1521, obsoletes RFC1522, obsoletes RFC1590
RFC2046	Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types	Standards Track	Draft Standard	November 1996	extends RFC5322, obsoletes RFC1521, obsoletes RFC1522, obsoletes RFC1590
RFC2047	MIME (Multipurpose Internet Mail Extensions) Part Three: Message Header Extensions for Non-ASCII Text	Standards Track	Draft Standard	November 1996	extends RFC5322, obsoletes RFC1521, obsoletes RFC1522, obsoletes RFC1590
RFC2049	Multipurpose Internet Mail Extensions (MIME) Part Five: Conformance Criteria and Examples	Standards Track	Draft Standard	November 1996	extends RFC5322, obsoletes RFC1521, obsoletes RFC1522, obsoletes RFC1590
RFC2184	MIME Parameter Value and Encoded Word Extensions: Character Sets, Languages, and Continuations	Standards Track	Proposed Standard	August 1997	updates RFC2045, updates RFC2047, updates RFC2183
RFC2231	MIME Parameter Value and Encoded Word Extensions: Character Sets, Languages, and Continuations	Standards Track	Proposed Standard	November 1997	obsoletes RFC2184, updates RFC2045, updates RFC2047, updates RFC2183
RFC5335	Internationalized Email Headers	Standards Track	Experimental	September 2008	updates RFC2045, updates RFC2822
RFC6532	Internationalized Email Headers	Standards Track	Draft Standard	February 2012	updates RFC2045, obsoletes RFC5335
RFC2646	The Text/Plain Format Parameter	Standards Track	Proposed Standard	August 1999	updates RFC2046
RFC3676	The Text/Plain Format and DelSp Parameters	Standards Track	Proposed Standard	February 2004	obsoletes RFC2646
RFC3798	Message Disposition Notification	Standards Track	Draft Standard	May 2004	updates RFC2046, updates RFC3461, obsoletes RFC2298
RFC5147	URI Fragment Identifiers for the text/plain Media Type	Standards Track	Proposed Standard	April 2008	updates RFC2046
RFC6657	Update to MIME regarding "charset" Parameter Handling in Textual Media Types	Standards Track	Proposed Standard	July 2012	updates RFC2046
RFC2298	An Extensible Message Format for Message Disposition Notifications	Standards Track	Proposed Standard	March 1998	
RFC5337	Internationalized Delivery Status and Disposition Notifications	Experimental	Experimental	September 2008	updates RFC3461, updates 3464, updates 3798 obsoletes RFC5337, updates RFC3461, updates RFC3464, updates RFC3798,
RFC6533	Internationalized Delivery Status and Disposition Notifications	Standards Track	Proposed Standard	February 2012	updates RFC6522
RFC3461	Simple Mail Transfer Protocol (SMTP) Service Extension for Delivery Status Notifications (DNSs)	Standards Track	Draft Standard	January 2003	obsoletes RFC1891
RFC3464	An Extensible Message Format for Delivery Status Notifications	Standards Track	Draft Standard	January 2003	obsoletes RFC1984

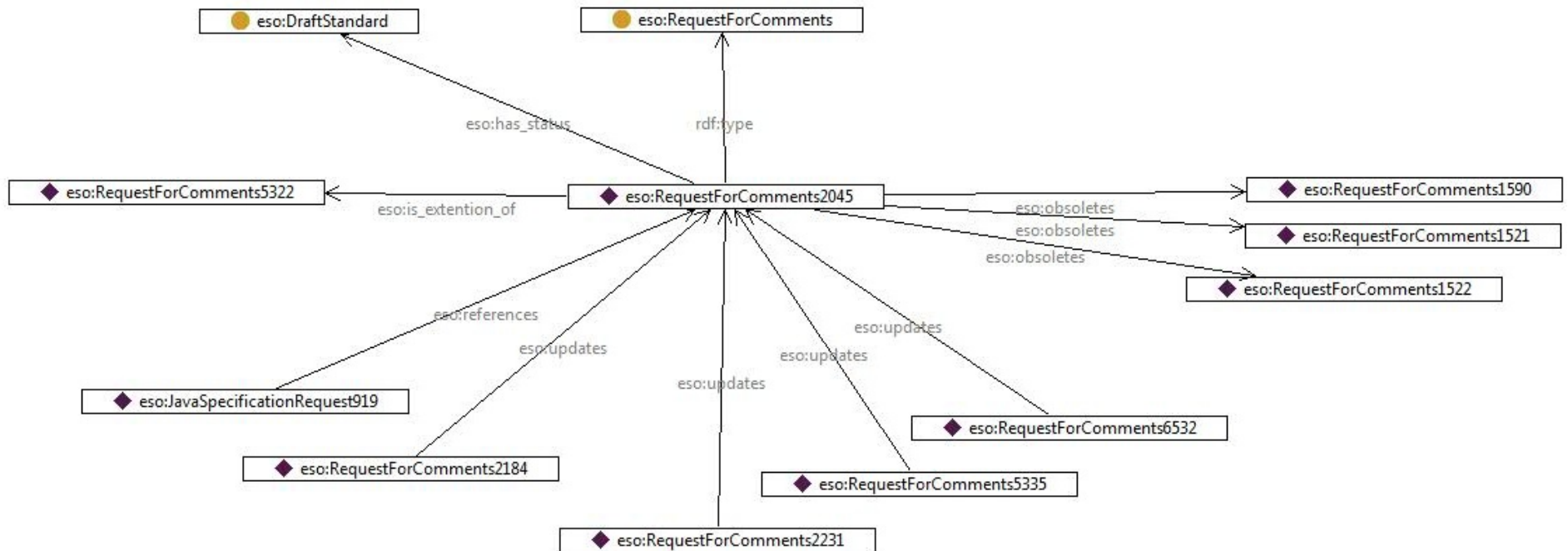
Email Standards Ontology



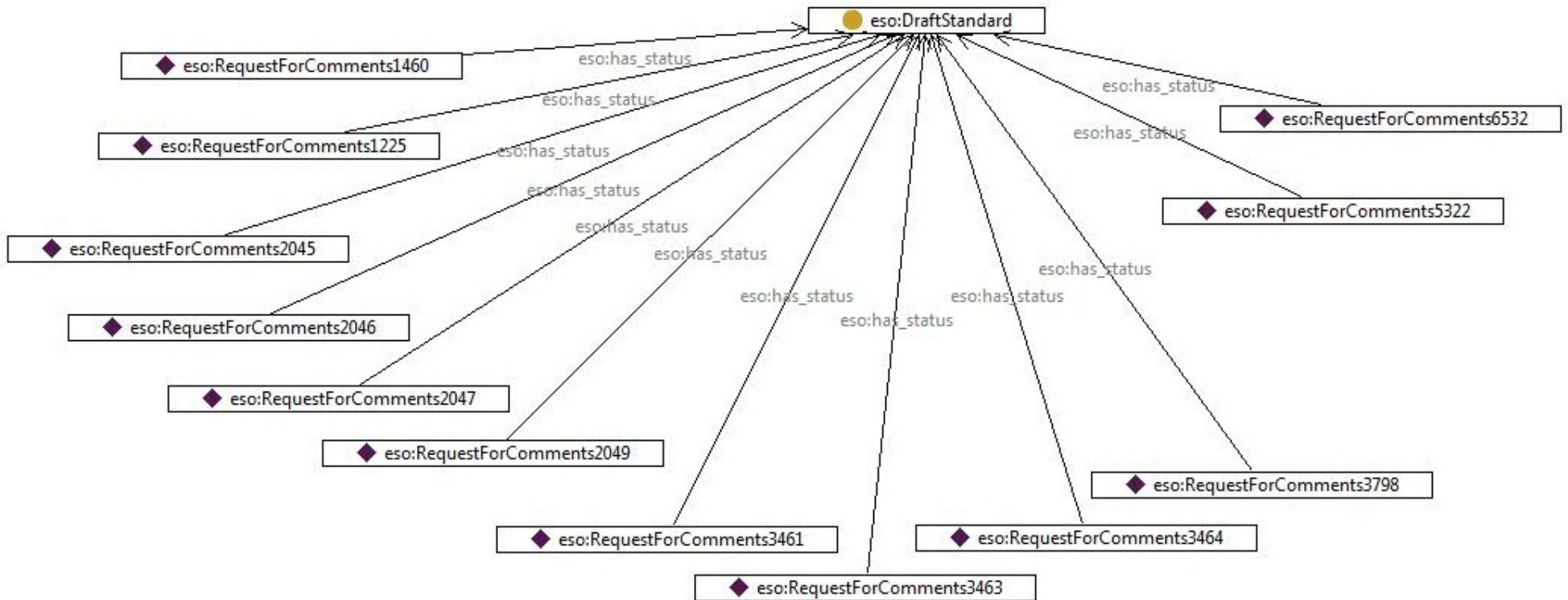
RFC 5322 Internet Message Format



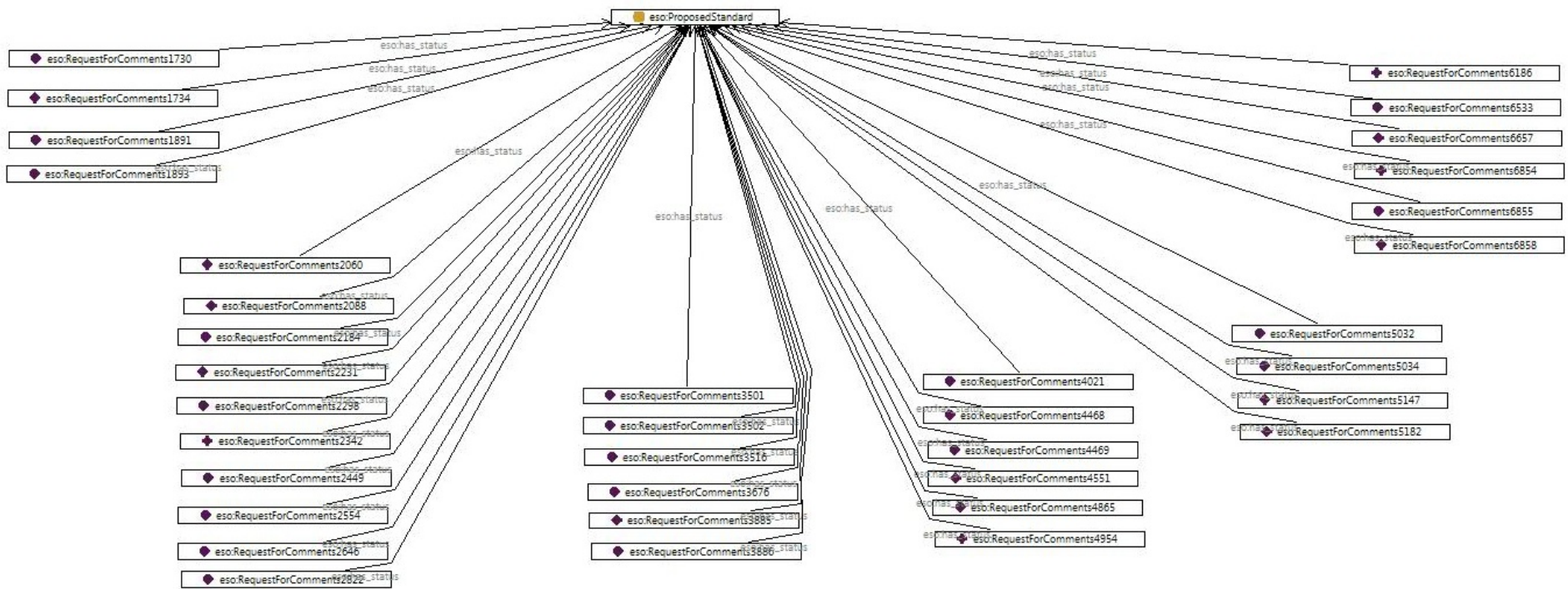
RFC 2045 MIME Extension



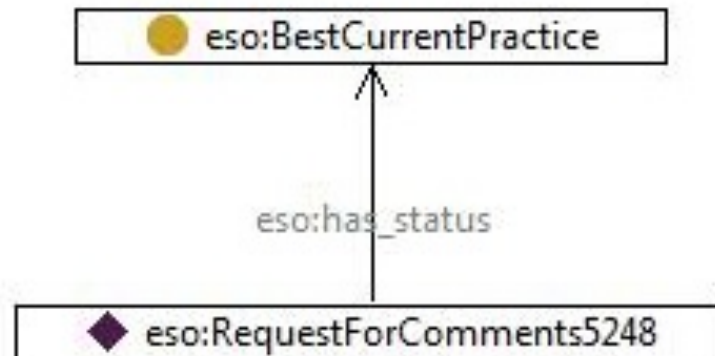
RFC Draft Standard



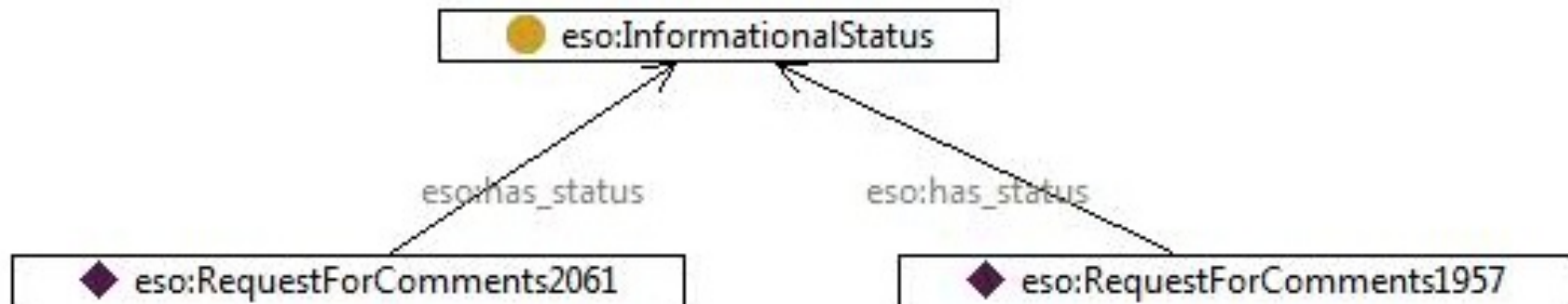
RFC Proposed Standard



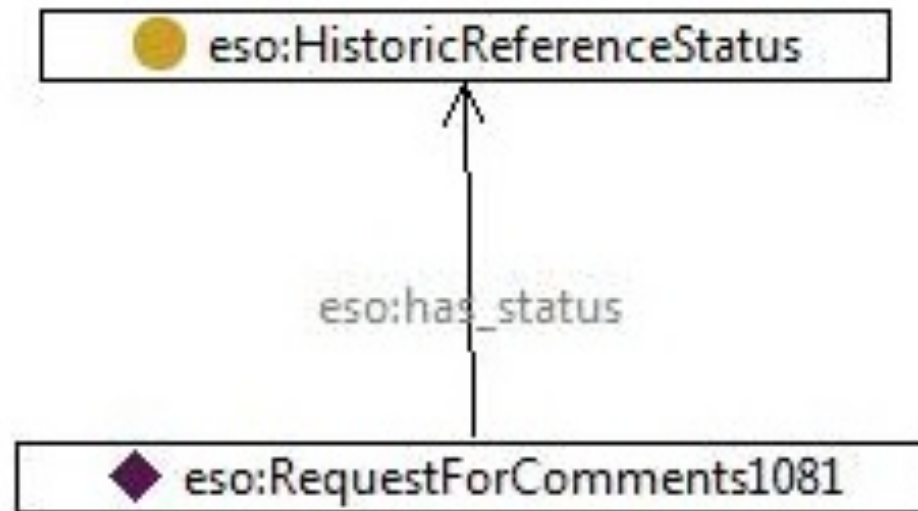
RFC Best Current Practice



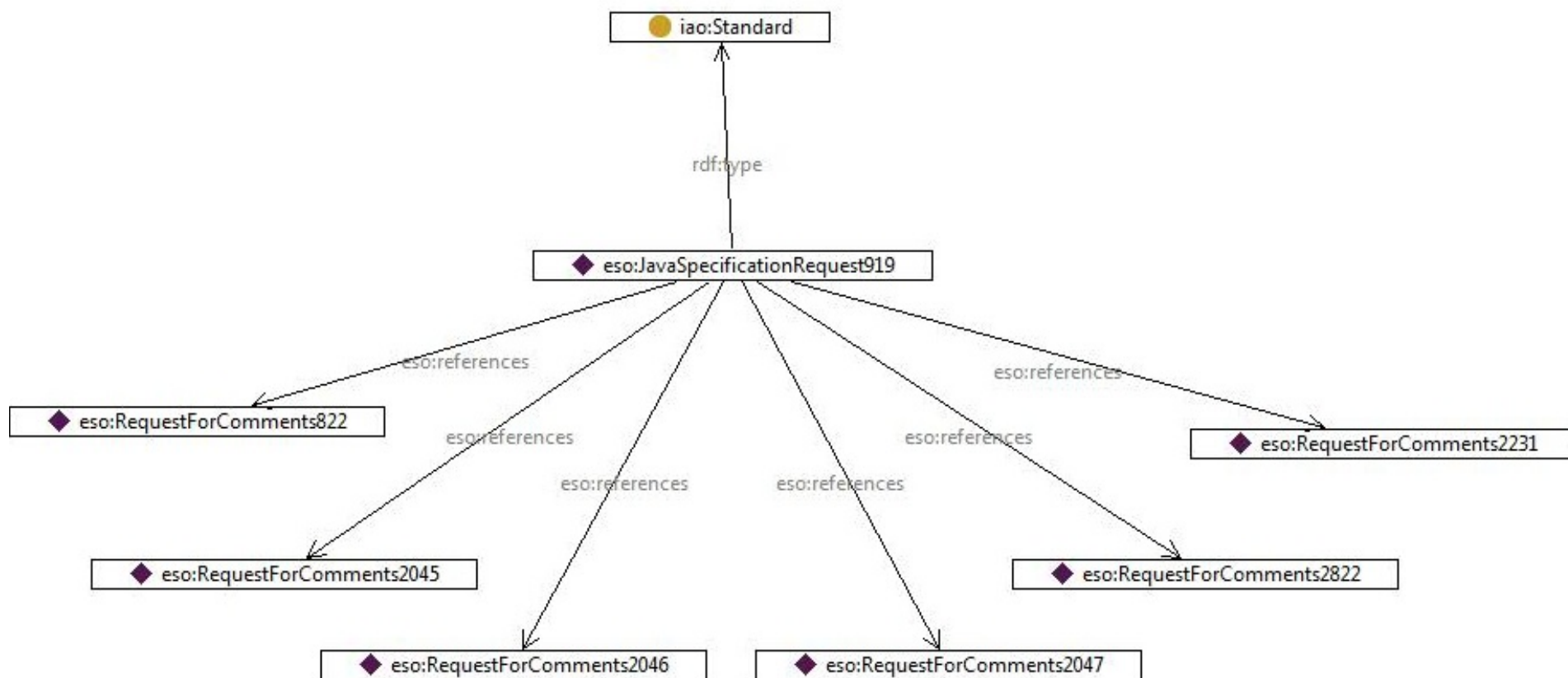
RFC Informational Status



RFC Historic Reference



JSR 919 JavaMail API Specification RFC 5322 vs RFC 2822, RFC 822



Email Ontology

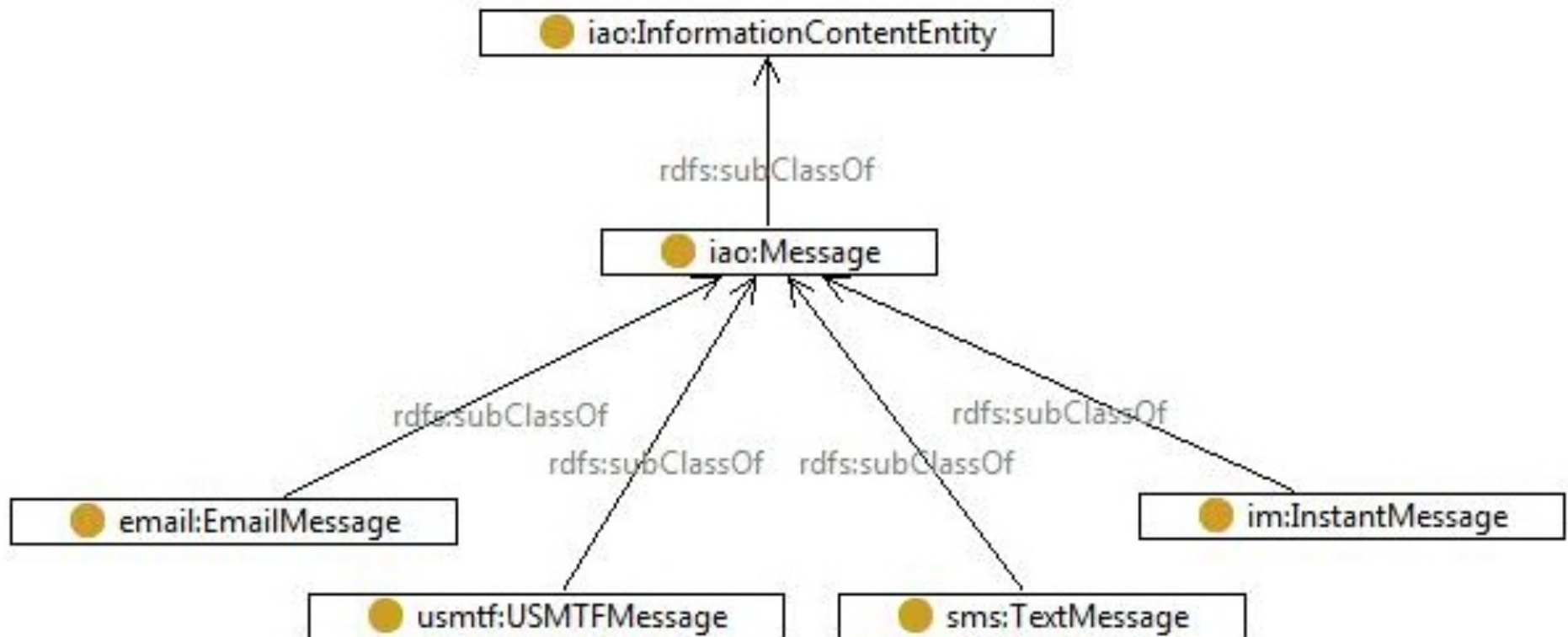
Base Classes

- EmailMessage
 - Email
 - Message
 - InternetMessage
 - EmailMessage
 - InternetEmailMessage
 - ElectronicMailMessage

Other Possible Message Domain Ontologies

- Simple Message System (SMS)
 - Text Message
- Instant Message (IM)
 - Instant Message
- United States Message Text Format (MIL-STD-6040)
 - USMTF Message

Other Message Types



Email Ontology

Base Classes (Continued)

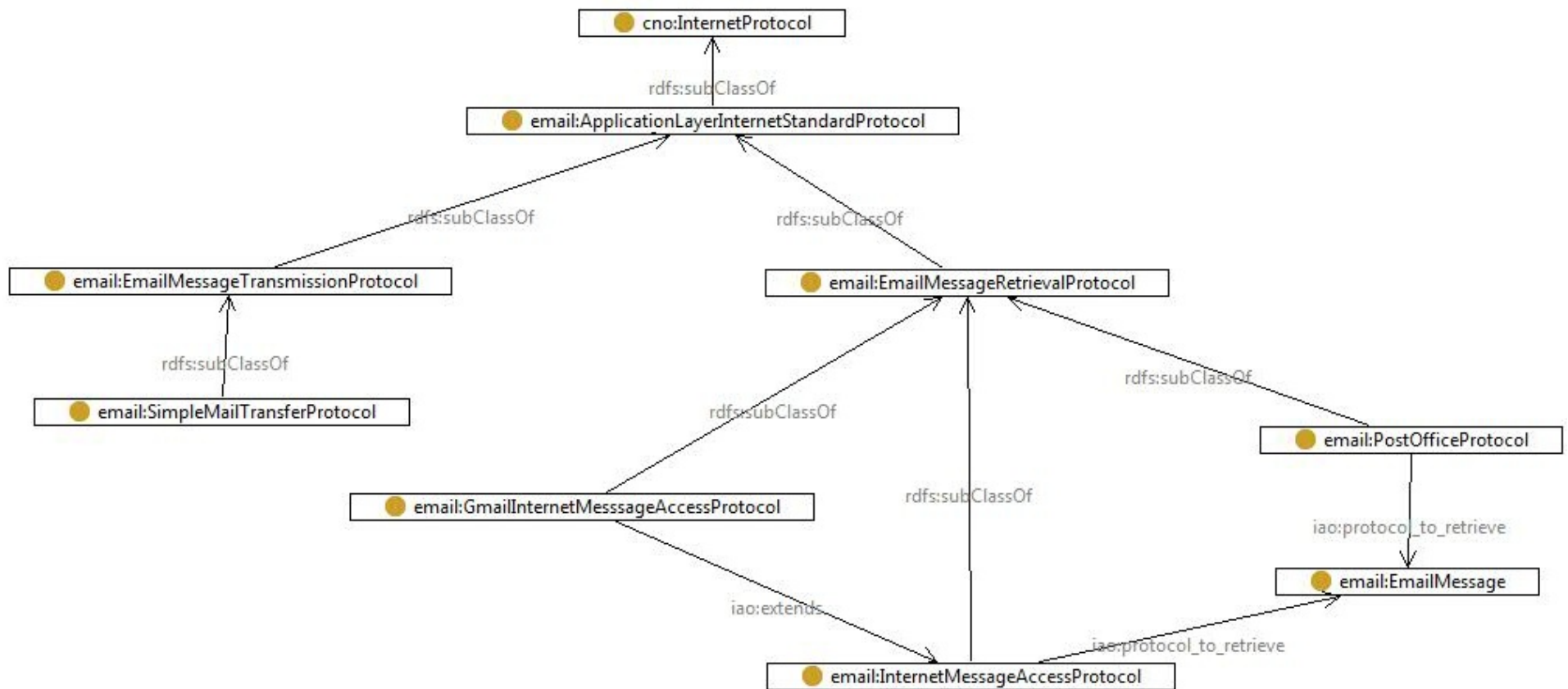
- InternetProtocol **
- ApplicationLayerInternetStandardProtocol **
- EmailMessageRetrievalProtocol
- EmailMessageTransmissionProtocol
- PostOfficeProtocol (POP)
- InternetMessageAccessProtocol (IMAP)
- GmailIMAP (GIMAP)
- SimpleMessageTransferProtocol (SMTP)

Where To Stop?

- Limited by domain ontology scope definition
- Crossing boundary into another domain
- No further decomposition

Email Ontology

Base Classes



Email Ontology

Information Content Classes

- EmailMessage
- MessageHeader
- MessageBody
- MultipartBody
- ContentType
- ContentTypeParameter
- EmailContact
- EmailAddress
- EmailMessageIdentifier

Ontology Definitions

Text Definitions and Logical Definitions

Class Form 

Name: email:EmailMessage

Annotations

dc:source

rdfs:comment

rdfs:label

Class Axioms

rdfs:subClassOf

owl:equivalentClass

owl:disjointWith

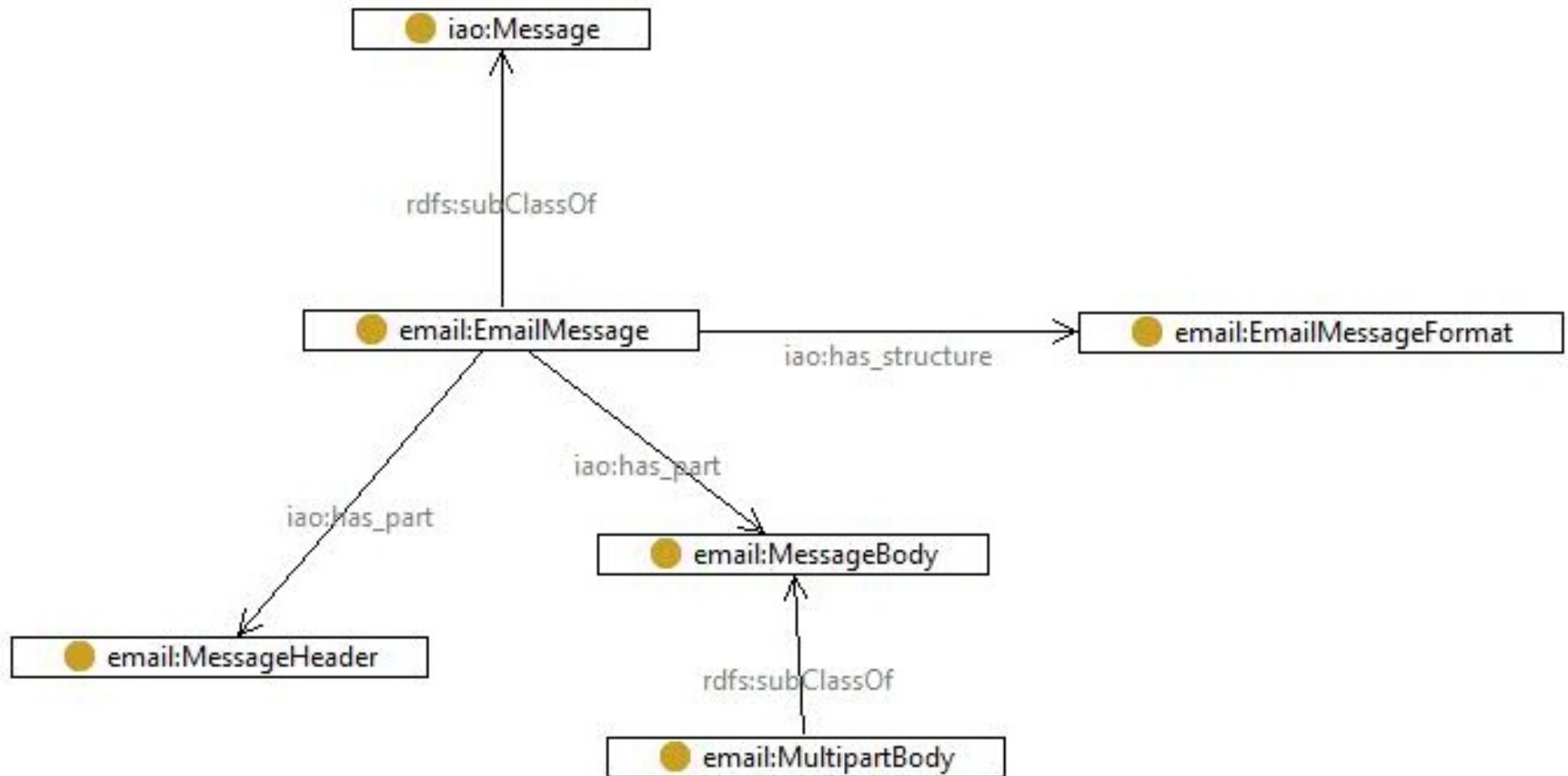
owl:hasKey

Other Properties

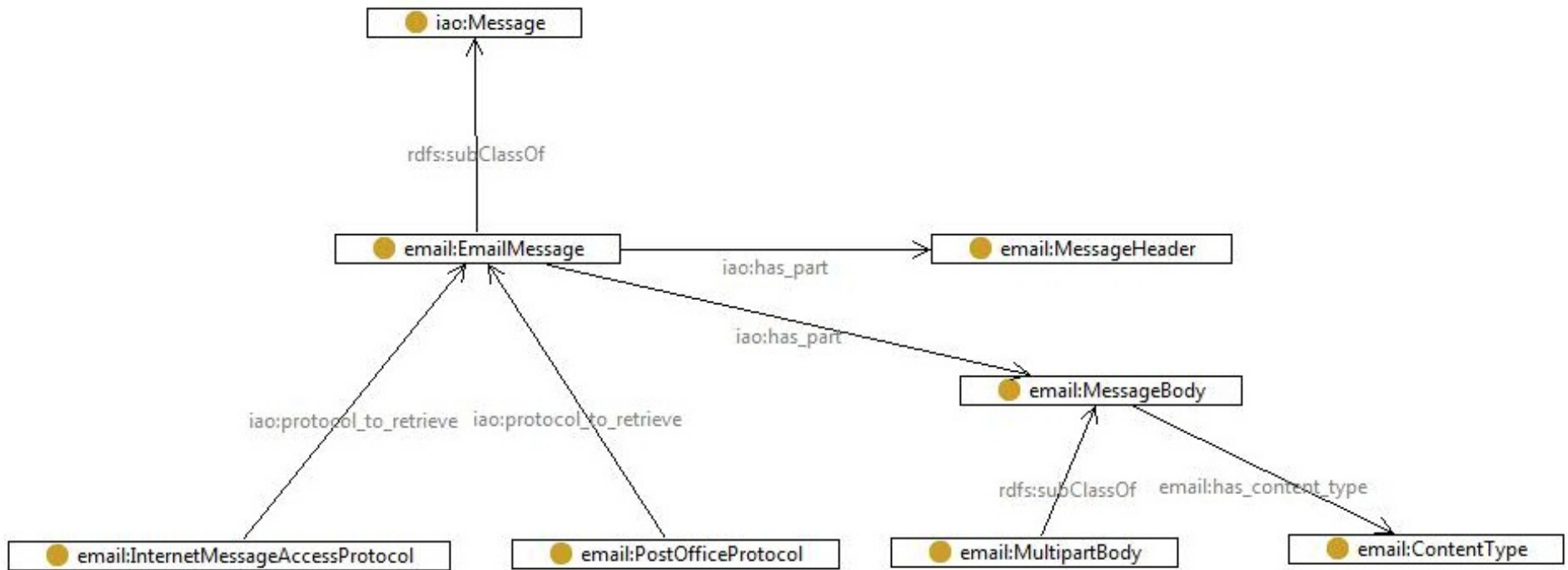
iao:has_part

iao:has_structure

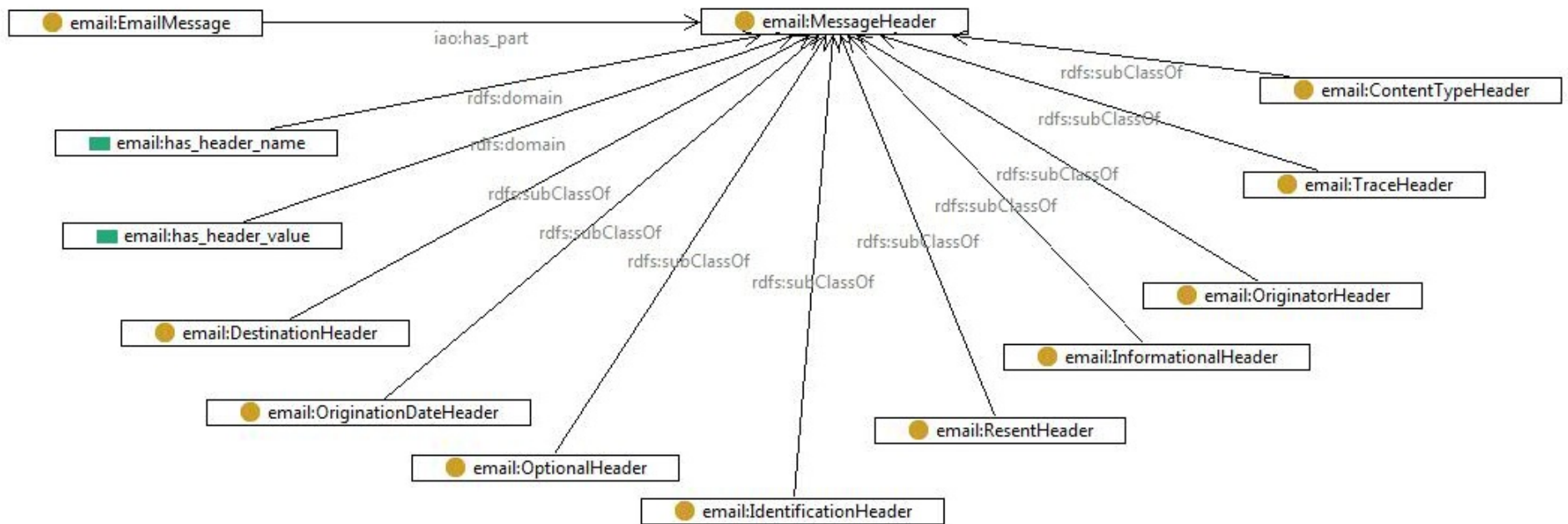
EmailMessage



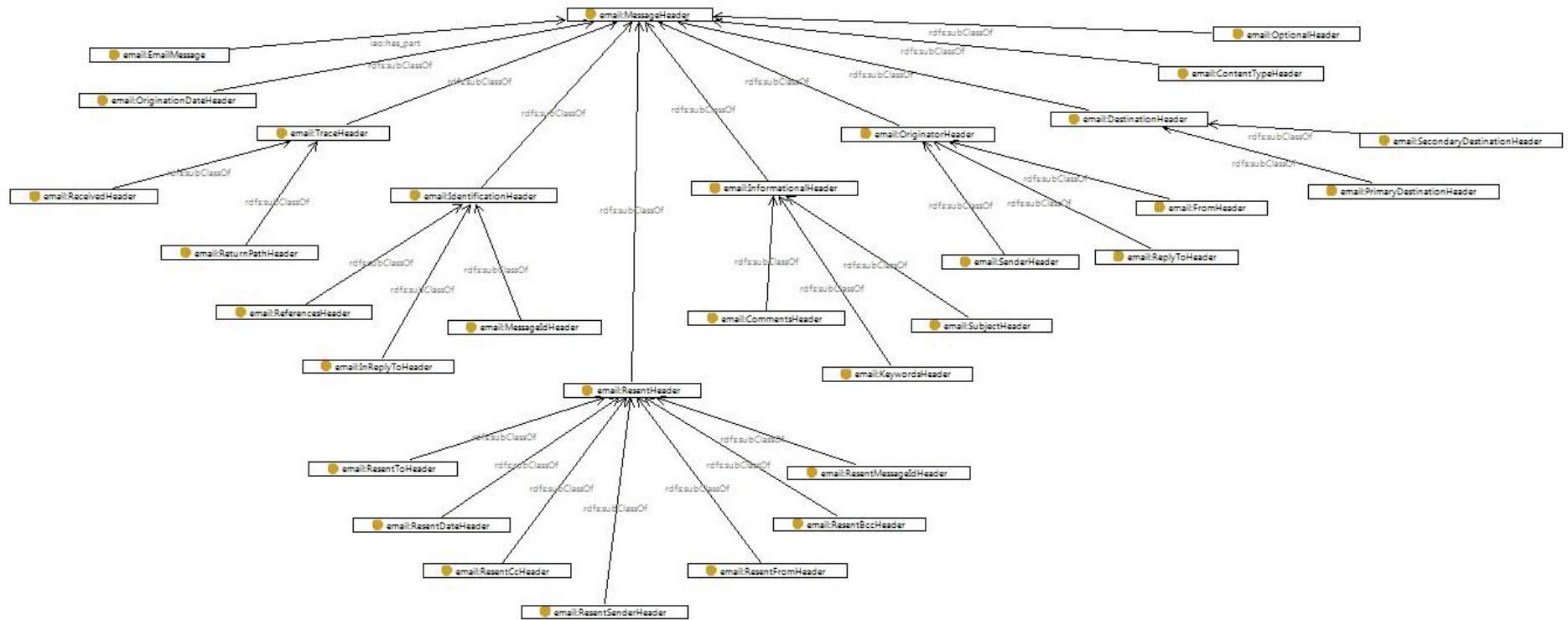
EmailMessage



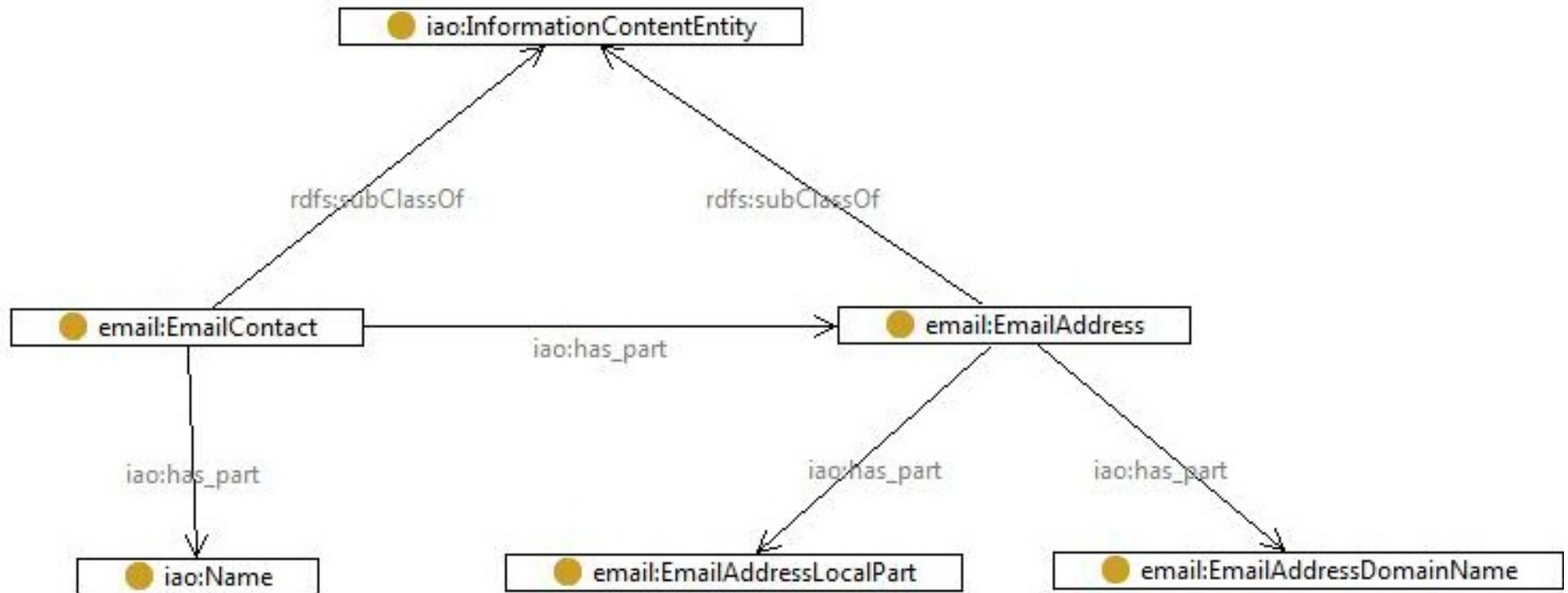
MessageHeader



MessageHeader



EmailContact

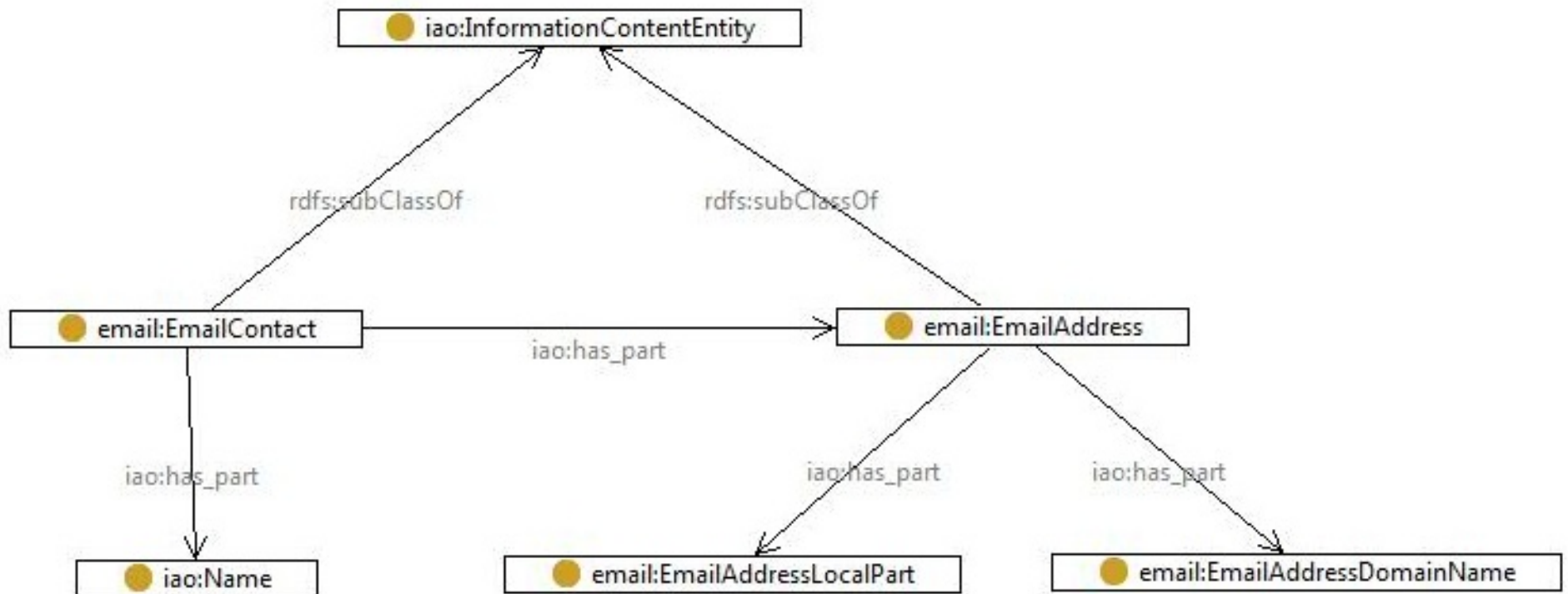


EmailMessage /EmailContact

Property Hierarchy Approach

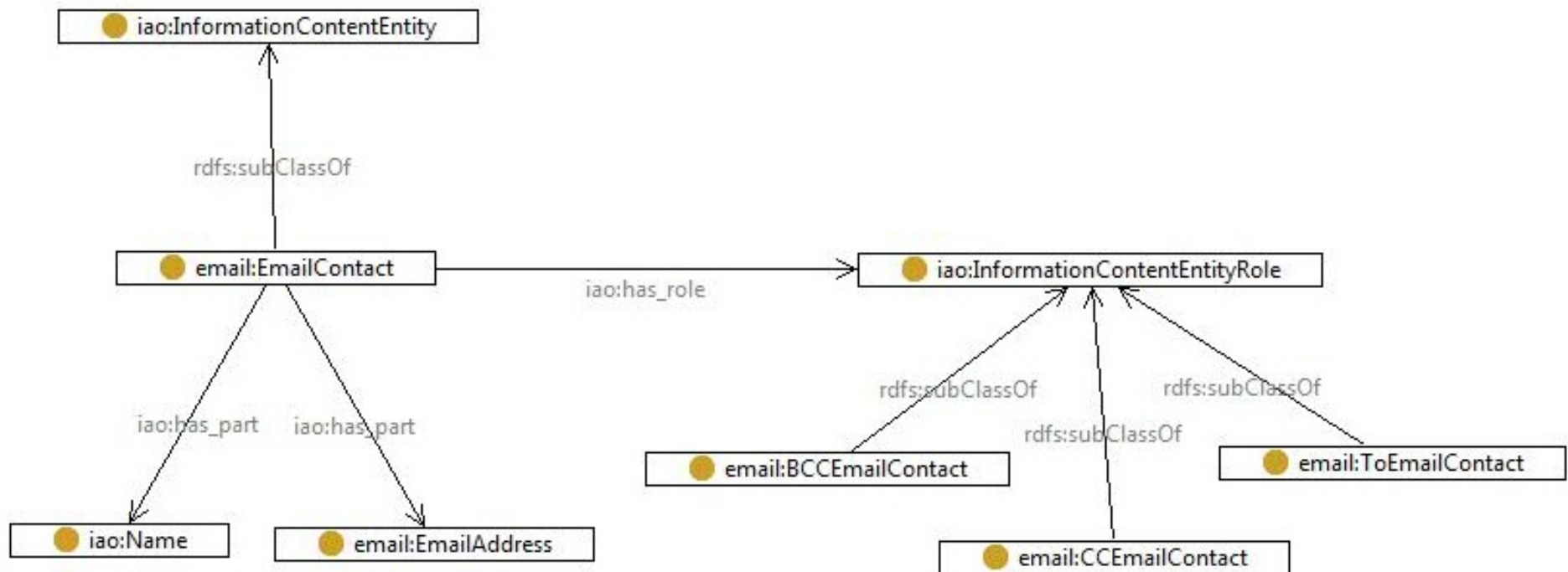
- has_email_contact
 - has_originator_email_contact
 - has_from_email_contact
 - has_sender_email_contact
 - has_replay_to_email_contact
 - has_destination_email_contact
 - has_primary_destination_email_contact
 - has_to_email_contact
 - has_secondary_destination_email_contact
 - has_cc_email_contact
 - has_bcc_email_contact

EmailAddress

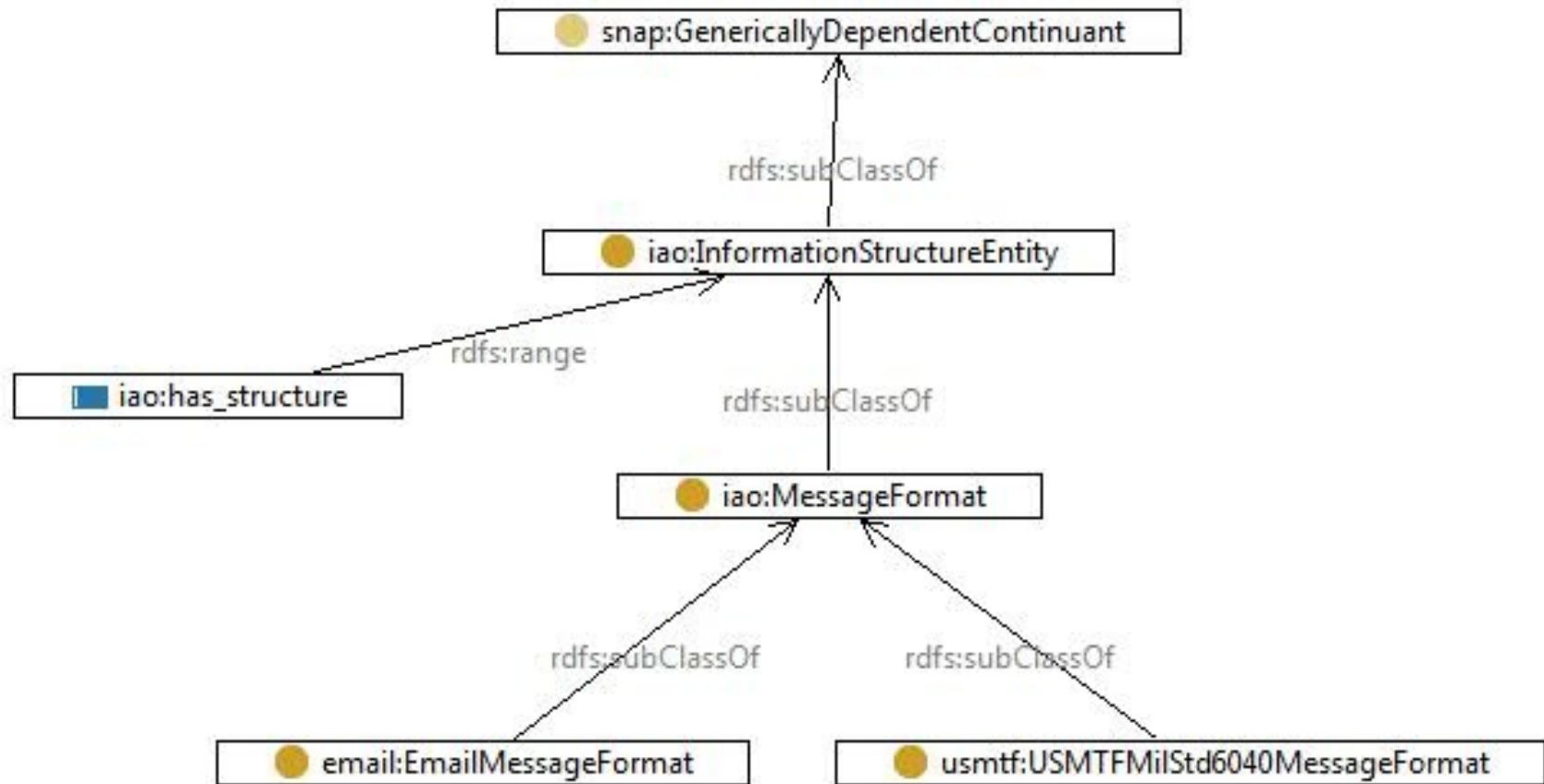


EmailContact

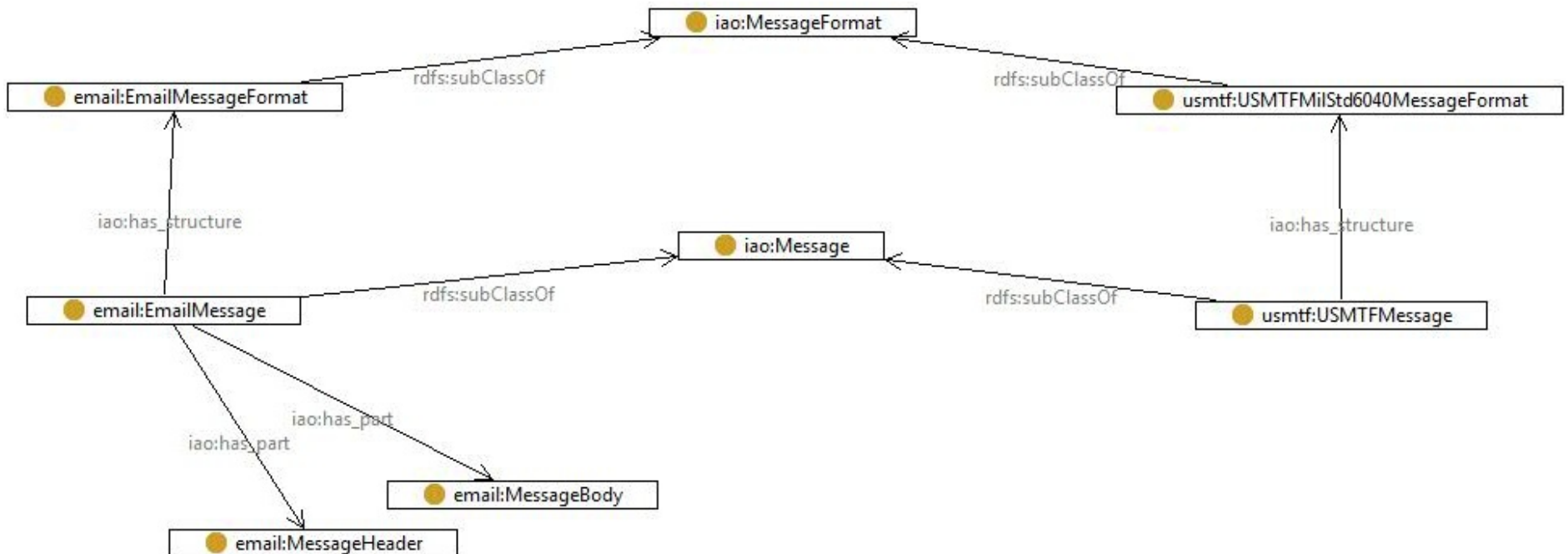
Roles versus Properties



MessageFormat



MessageFormat (continued)



Email Information Content Extraction

Empirical Email Parsing Results

- Extract email information content using JavaMail 1.5.0 library
 - Approximately 80,000 emails spanning > 2 years
 - Email message headers – header type distribution
 - Email message body – content type distribution
 - Email messages with multipart body
 - Body part content type distribution

Email Parsing Results

- Email Messages: 83,897
 - Message Headers: 2,217,060
 - Unique Message Header Names: 1,025
 - Message Body
 - Unique Content Types: 13
 - With Multipart Body: 53,079
 - Body Parts: 106,760
 - Unique Content Types: 87

Email Parsing Results

Message Body Content Types

Content Type	Count
application/octet-stream	1
application/pkcs7-mime	8
application/x-pkcs7-mime	22
message/rfc822	1
multipart/alternative	43922
multipart/mixed	7583
multipart/related	916
multipart/report	61
multipart/signed	596
text/calendar	189
text/html	14634
text/plain	15962
NULL (empty charset param)	2

Email Parsing Results

- Full statistics
 - email_parse_0.1.xlsx