Retinal Image and Health Metadata Specification v0.2

March 2011

DRAFT

1.0 Introduction

1.1 Objectives

This document defines a data format for retinal image and health metadata, intended to promote a standard way to associate health metadata with retinal images. The initial focus of this specification is to define the data format for a community database of retinal images. The eventual goal is for this specification to evolve to an open community standard that facilitates sharing of datasets of retinal images and health metadata across the research, clinical, and citizen community.

1.2 Background

This specification was initiated by the EyesFirst research project. EyesFirst is an open source research project that aims to apply advanced digital signal processing and data analysis techniques to retinal images to detect early signs of multiple diseases. The open source nature of this project promotes:

- Open Standards: a standard means to share retinal images and health metadata that is accessible to everyone.
- Open Data: retinal images and health metadata is freely available through an open database under flexible licensing terms.
- Open Source: the research results are publically available for others to study, use, and improve.
- Open Access: transparency during all phases of research.

This specification is intended to be a starting point toward the development of an open standard for retinal image and health metadata. To promote open data, EyesFirst is establishing an open source, self-service database of retinal images (DORI). DORI will make publically available, with minimal usage restrictions, clinical and research datasets of retinal images annotated with health metadata. In the near-term, the goal is for DORI to become a reference dataset of retinal images that can be used to, among other things, evaluate automated retinal diagnostics. As DORI grows over time with multiple samples per subject, the reference datasets within DORI will evolve to form the basis for longitudinal studies to investigate novel retinal predictors for multiple diseases. Through open standards and open data, EyesFirst will promote the development of open source image processing algorithms and data analysis tools.

2.0 Data Design

The retinal image and health metadata specification is organized as a set of primary categories, each consisting of one or more information object types. The information object types are associated with each other through an anonymous and opaque unique patient identifier, instantiated within the patient information object.

2.1 Primary Categories

The retinal image and health metadata specification consists of five primary categories of information objects (outlined in Table 1):

1. Patient
2. Demographics
3. Personal Medical History
4. Family Medical History
5. Ocular Images

Existing medical imaging and electronic health record standards (as listed in Table 1) were leveraged in defining information object types, where applicable. While certain standards are used as a basis for information objects, it does not preclude the use of other standards provided there are reasonable mappings to information object types defined in this specification. Appendix A provides a more detailed definition of each information object type within each category.

Primary Categories		
Category	**Information Object Type**	**Relevant Standards**
Patient	Patient	None Identified
Demographics	Demographics	HITSP C83 – CDA Content Module v1.1
Personal Medical History	Conditions	HITSP C83 – CDA Content Module v1.1
	Results	HITSP C83 – CDA Content Module v1.1
	Medications	HITSP C83 – CDA Content Module v1.1
	Allergies	HITSP C83 – CDA Content Module v1.1
	Procedures	HITSP C83 – CDA Content Module v1.1
Family Medical History	Family Medical History	HITSP C83 – CDA Content Module v1.1
Ocular Images	Ophthalmic Tomography	DICOM Supplement 110
	Ophthalmic Photography	DICOM Supplement 91
	Clinical Interpretation	None Identified

Table 1. Primary Modules

2.2 Patient Identifiers

Since content published to DORI will be made publically accessible, the metadata defined in this document is explicitly de-identified to comply with HIPAA privacy regulations.[1] As the patient identifier used by the data publisher may contain personally identifiable information, an opaque unique identifier is generated. This patient identifier is associated with each patient record and is used to identify multiple records for a single patient and to anonymously track the patient over time. As illustrated in Figure 1, the opaque unique patient identifier is based on the concatenation of existing information, including:

- Publishing organization's identifier
- Publishing clinician's identifier
- Original patient identifier used by the publisher
- Patient's year of birth.

This information is put through a hash function to generate a single, unique, and opaque identifier that is exposed to data consumers. Use of multiple input variables for the hash function allows for greater entropy of the hash function.

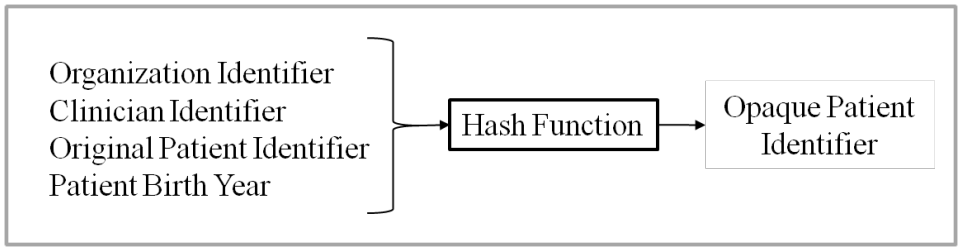

Figure 1. Generation of a unique, opaque patient identifier

2.3 DORI Record Identifiers

A Uniform Resource Identifier (URI) uniquely identifies DORI records. A DORI record URI has the following components:

- Domain Name
- Patient Identifier
- Information Object Type
- Instance Document Label

The URI then looks like the following
http://eyesfirst.org/patient_identifier/information_object_type/instance_document_name

This URI is used to uniquely reference individual records within DORI. It can be used to cluster records based on a single patient and/or a particular type of document. For example, the URL can associate multiple health records taken at different locations and different points of time to a single patient. Instance documents are instantiated as a

[1] Summary of the HIPAA Privacy Rule,
http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/privacysummary.pdf

particular type of information object (which is categorized into one of five primary categories) and provides another way to cluster records.

3.0 Example Usage Scenario

This scenario illustrates how a clinical site (hospital, private practice) publishes data to DORI. Prior to registering participants to publish their data into DORI, the clinical site needs to map the data format used internally by their Electronic Health Record (EHR) system to the metadata specification defined in this document such that DORI records may be generated automatically. Furthermore, each data publisher must register their respective clinical site and set of clinicians in DORI, resulting in an organization ID and clinician IDs. The diagrams below illustrate the process to register new patients (Figure 2), baseline medical information (Figure 3) and updated medical information (Figure 4).

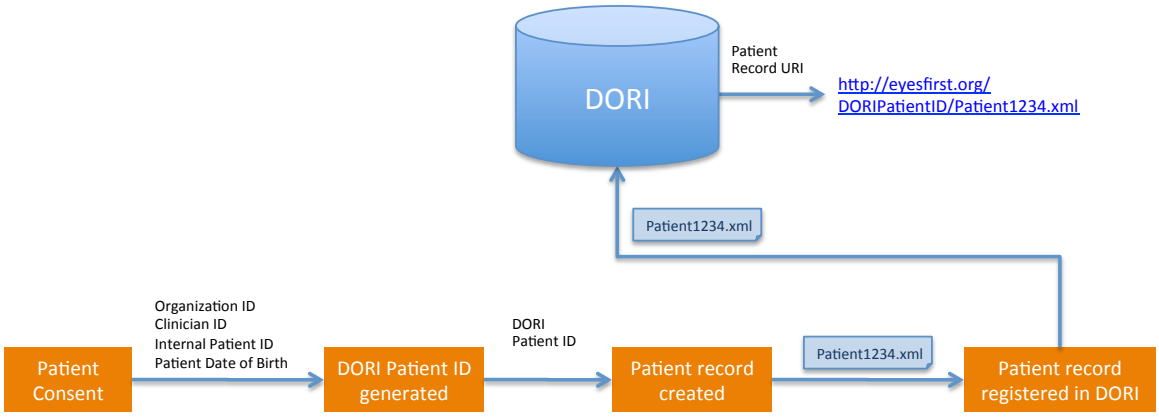

Figure 2. New Patient Registration Process

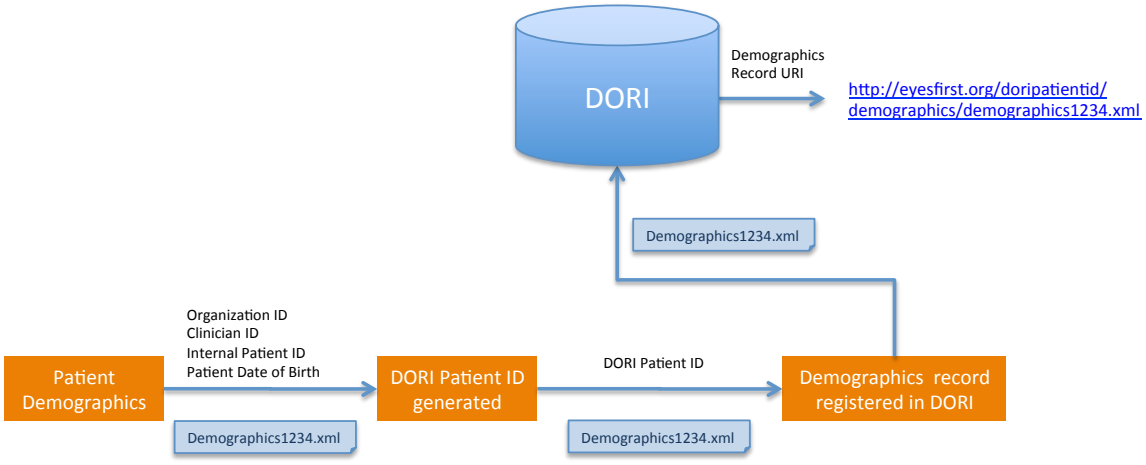

Figure 3. Baseline Medical Information Registration Process

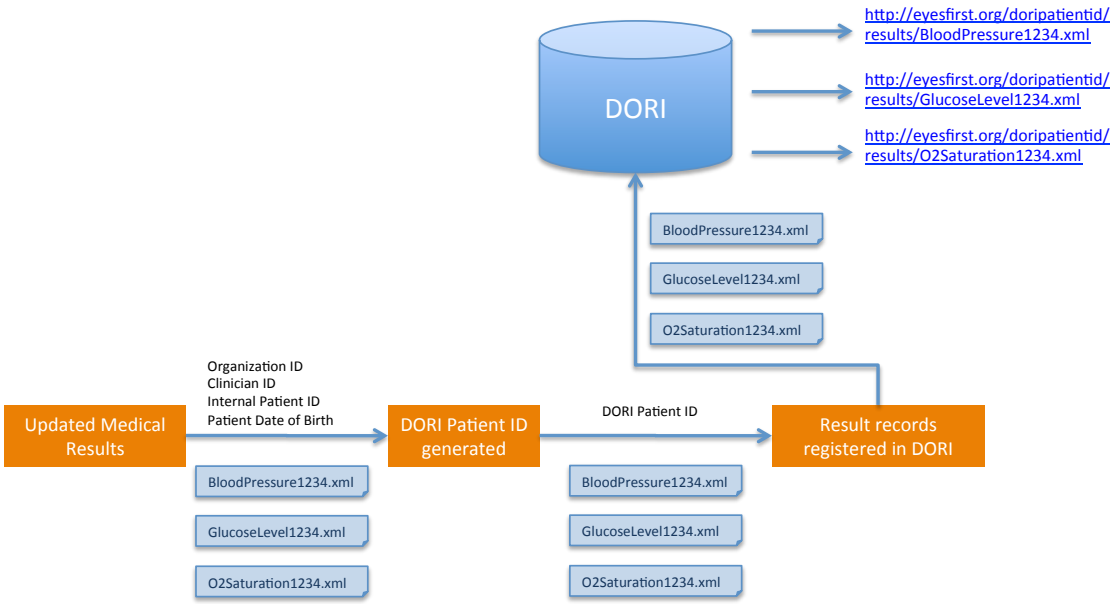

Figure 4. Updated Medical Information Registration Process

4.0 References

HITSP/C80: HITSP Clinical Document and Message Terminology Component, http://wiki.hitsp.org/docs/C80/C80-1.html

HITSP/C83: HITSP CDA Content Modules Component, http://wiki.hitsp.org/docs/C83/C83-1.html

DICOM Part 3: Information Object Definitions, PS 3.3-2009, ftp://medical.nema.org/medical/dicom/2009/09_03pu3.pdf

DICOM Supplement 91: Ophthalmic Photography Image SOP Classes, ftp://medical.nema.org/medical/dicom/final/sup91_ft2.pdf

DICOM Supplement 110: Ophthalmic Tomography Image Storage SOP Class, ftp://medical.nema.org/medical/dicom/final/sup110_ft4.pdf

Appendix A: Information Object Type Definitions

This section defines the information object types, organized by the primary categories. The obligation field with each attribute indicates whether the attribute is required or optional. In certain cases, the attribute is required only when an instance of the particular information object exists. The comments field identifies any particular constraints associated with the attribute and/or references applicable standards for further information.

Patient

Patient			
Name	**Definition**	**Obligation**	**Comments**
Patient Identifier	An opaque identifier that uniquely identifies the individual to which the exchange refers.	Mandatory	
Comments	Any additional free-text comments.	Optional	
Timestamp	The date and time that this patient is registered in DORI.	Mandatory	

Demographics

Demographics			
Name	**Definition**	**Obligation**	**Comments**
Gender	Gender is used to refer to administrative sex rather than biological sex and therefore should easily be classified into female and male. It is included in the exchange for purposes of linking to insurance information and other patient identification linkages and the value chosen by the patient should reflect the information under which any insurance or financial information will be filed, as well as the same information given to other healthcare providers, institutions or health data exchange networks.	Optional	C83-[DE-1.06-1] Gender SHALL be coded as specified in HITSP/C80 Section 2.2.1.2.1.2 V3 Administrative Gender
Year of Birth	The year of the birth of the individual to which this Exchange refers. The date of birth is typically a key patient identifier variable and used to enable computation of age at the effective date of any other data element. It is assumed to be fixed throughout the patient's lifetime.	Optional	
Place of Birth	The state and country in which the individual was born	Optional	
Place of Residence	The current state and country in	Optional	

	which the individual resides		
Year of Death	The year this individual died.	Optional	
Cause of Death	The cause of this individual's death.	Optional	
Comments	Any additional free-text comments.	Optional	
Timestamp	The date and time that this information was last updated.	Mandatory if demographic information is specified.	

Personal Medical History

The personal medical history category consists of five information object types:

1. Condition
2. Result
3. Procedure
4. Medication
5. Allergy

Condition			
Name	**Definition**	**Obligation**	**Comments**
Condition Name	This is a text description of the problem suffered	Optional	
Condition Date	This is the range of time of which the problem was active for the patient or subject	Mandatory if condition information is specified.	
Condition Type	This is a fixed value indicating the level of medical judgment used to determine the existence of a problem	Optional	C83-[DE-7.02-1] The problem type SHALL be coded as specified in HITSP/C80 Section 2.2.3.1.2 Problem Type
Condition Code	This value is a code describing the problem according to a specific vocabulary of problems	Optional	C83-[DE-7.04-1] The problem SHALL be coded as specified in HITSP/C80 Section 2.2.3.1.1 Problem
Comments	Any additional free-text comments.	Optional	
Timestamp	The date and time that this information was last updated.	Mandatory if demographic information is specified.	

Result			
Attribute	**Definition**	**Obligation**	**Comments**
Result Date/Time	The biologically relevant date/time for the observation	Mandatory if result information is specified.	
Result Type	A coded representation of the observation performed	Optional	C83-[DE-15.03-1] Result Type SHOULD be selected from LOINC (codeSystem 2.16.840.1.113883.6.1) or SNOMED CT (codeSystem 2.16.840.1.113883.6.96) C83-[DE-15.03-2] Result Type for laboratory results SHOULD be coded as specified in HITSP/C80 Section 2.2.3.6.1 Laboratory Observations.
Result Status	Status for this observation, e.g., complete, preliminary	Optional	
Result Value	The value of the result, including units of measure if applicable	Optional	
Result Interpretation	An abbreviated interpretation of the observation, e.g., normal, abnormal, high, etc	Optional	
Result Reference Range	Reference range(s) for the observation	Optional	
Comments	Any additional free-text comments.	Optional	
Timestamp	The date and time that this information was last updated.	Mandatory if demographic information is specified.	

Procedure			
Attribute	**Definition**	**Obligation**	**Comments**
Procedure Type	This is a coded value describing the type of the Procedure	Optional	
Procedure Description	Free text describing the Procedure	Optional	
Procedure Date / Time	The date and time of the Procedure, including duration if pertinent	Mandatory if result information is	

		specified.	
Comments	Any additional free-text comments.	Optional	
Timestamp	The date and time that this information was last updated.	Mandatory if demographic information is specified.	

Medication			
Attribute	**Definition**	**Obligation**	**Comments**
Type of Medication	A classification based on how the medication is marketed (e.g., prescription, over the counter drug)	Optional	C83-[DE-8.19-1] The type of medication SHALL be coded as specified in HITSP/C80 Section 2.2.3.3.5 Medication Type.
Coded Product Name	A code describing the product from a controlled vocabulary	Optional	C83-[DE-8.13-1] The coded product name SHALL be coded as specified in HITSP/C80 Section 2.2.3.3.8 Medication Clinical Drug Names. C83-[DE-8.13-2] When only the class of the drug is known (e.g., Beta Blocker or Sulfa Drug), it SHALL be coded as specified in HITSP/C80 Section 2.2.3.3.9 Medication Drug Class. C83-[DE-8.13-3] When only the medication ingredient name is know, the coded product name MAY be coded as specified in HITSP/C80 Section 2.2.3.3.11 Ingredient Name
Status of Medication	If the medication is Active, Discharged, Chronic, Acute, etc	Optional	C83-[DE-8.20-1] The medication status MAY be recorded using the CCD Medication Status observation using the value set defined in the CCD
Dose	A Sig Component: the amount of the product to be given. This may be a known, measurable unit (e.g., milliliters), an administration unit	Optional	C83-[DE-8.08-1] Units MAY be present when needed. If present it SHALL be coded as

	(e.g., tablet), or an amount of active ingredient (e.g., 250 mg). May define a variable dose, dose range or dose options based upon identified criteria (see Dose Indicator)		specified in HITSP/C80 Section 2.2.3.6.6 Units of Measurement C83-[DE-8.08-2] When the coded product or brand name describes the strength or concentration of the medication, and the dosing is in administration units (e.g., 1 tablet, 2 capsules), units SHOULD contain the preferred name of the presentation units within braces { } using the units of presentation from the NCI Thesaurus
Frequency	A Sig Component: defines how often the medication is to be administered as events per unit of time. Often expressed as the number of times per day (e.g., four times a day), but may also include event-related information (e.g., 1 hour before meals, in the morning, at bedtime). Complimentary to Interval, although equivalent expressions may have different implications (e.g., every 8 hours versus 3 times a day)	Optional	
Indication	A Sig Component: The medical condition or problem intended to be addressed by the ordered product. For example: for chest pain, for pain, for high blood pressure	Optional	C83-[DE-8.21-1] The indication SHALL be coded as specified in HITSP/C80 Section 2.2.3.1.1 Problem
Start Date	The date the medication was started	Mandatory if medication information is specified.	
Stop Date	The date the medication was/will be stopped	Mandatory if medication information is specified.	
Reaction	Any noted intended or unintended effects of the product. For example: full body rash, nausea, rash resolved	Optional	
Comments	Any additional free-text	Optional	

	comments.		
Timestamp	The date and time that this information was last updated.	Mandatory if demographic information is specified.	

Allergy			
Attribute	**Definition**	**Obligation**	**Comments**
Adverse Event Date	This is a date that expresses when this particular allergy or intolerance was known to be active for the patient	Optional	
Adverse Event Type	Describes the type of product and intolerance suffered by the patient. The type of product shall be classified with respect to whether the adverse event occurs in relationship with a medication, food, or environmental or other product. The adverse event should also be classified more specifically as an allergy, non-allergy intolerance, or just adverse reaction if that level of detail is not known	Optional	C83-[DE-6.02-1] The vocabulary used for adverse event types SHALL be coded as specified in HITSP/C80 Section 2.2.3.4.2 Allergy/Adverse Event Type
Product Description	This is the name or other description of the product or agent that causes the intolerance	Optional	
Product Coded	This value is a code describing the product	Optional	C83-[DE-6.04-1] Food and substance allergies SHALL be coded as specified in HITSP/C80 Section 2.2.3.3.11 Ingredient Name C83-[DE-6.04-2] Allergies to a class of medication SHALL be coded as specified in HITSP/C80 Section 2.2.3.3.9 Medication Drug Class C83-[DE-6.04-3] Allergies to a specific medication SHALL be coded as specified in HITSP/C80 Section 2.2.3.3.7 Medication Brand

			Name or HITSP/C80 Section 2.2.3.3.8 Medication Clinical Drug Names.
Reaction Description	This is the reaction that may be caused by the product or agent	Optional	
Reaction Coded	This value is a code describing the reaction	Optional	C83-[DE-6.06-1] The reaction SHALL be coded as specified in HITSP/C80 Section 2.2.3.4.1 Allergy/Adverse Event (Reaction)
Severity Description	This is a description of the level of severity of the allergy or intolerance	Optional	
Severity Coded	This value is a code describing the level severity of the allergy or intolerance	Optional	C83-[DE-6.08-1] The terminology used for severity of the adverse event SHALL be coded as specified in HITSP/C80 Section 2.2.3.4.3 Allergy/Adverse Event Severity
Alert Status	Whether the allergy is active or not.	Optional	
Comments	Any additional free-text comments on the result.	Optional	
Timestamp	The date and time that this information was last updated.	Mandatory if demographic information is specified.	

Family Medical History

Family Medical History			
Attribute	**Definition**	**Obligation**	**Comments**
Family Member Relationship	Relationship of Family Member to Patient or other Family Member. Record information on relatives including 1st and 2nd degree, such as: - Mother - Siblings - Children - Aunts/uncles - Cousins - Grandchildren - Nieces/Nephews This data element also allows the	Optional	C83-[DE-18.04-1] The Family Member Relationship (to Patient) SHALL be coded as specified in HITSP/C80 Section 2.2.1.2.5 Family Relationship Type

	recording of the natural father and mother, and the adoptive status. Consanguinity can be determined by including relationships between consanguineous individuals. Note: In order to record information for 3rd degree relatives and beyond, implementations can provide recursive entries. For example, the first entry could state a "grandmother", then the next entry could state the "grandmother's mother" which would be a 3rd degree relative. This enables any degree on relatives.		
Family Member Gender	Gender is used to refer to administrative sex rather than biological sex and therefore should easily be classified into female and male. It is included in the exchange for purposes of linking to insurance information and other patient identification linkages and the value chosen by the patient should reflect the information under which any insurance or financial information will be filed, as well as the same information given to other healthcare providers, institutions or health data exchange networks	Optional	C83-[DE-1.06-1] Gender SHALL be coded as specified in HITSP/C80 Section 2.2.1.2.1.2 V3 Administrative Gender
Family Member Year of Birth	Year of Birth of Family Member	Optional	
Family Member Year of Death	The age (real or approximate) of the family member at death	Optional	
Family Member Cause of Death	An indicator that a particular problem was the cause of death of the family member	Optional	
Family Member Condition Name	Condition is the generic term used in the model to designate conditions, problems, diagnoses, etc.	Optional	C83-[DE-18.12-1] The problem type SHALL be coded as specified in HITSP/C80 Section 2.2.3.1.2 Problem Type. C83-[DE-18.12-2] The problem SHALL be coded as specified in HITSP/C80 Section 2.2.3.1.1 Problem
Family Member Condition Date	This is the range of time of which the problem was active for the	Optional	

	family member.		
Family Member Condition Type	This is a fixed value indicating the level of medical judgment used to determine the existence of a problem	Optional	C83-[DE-7.02-1] The problem type SHALL be coded as specified in HITSP/C80 Section 2.2.3.1.2 Problem Type
Family Member Condition Code	This value is a code describing the problem according to a specific vocabulary of problems	Optional	C83-[DE-7.04-1] The problem SHALL be coded as specified in HITSP/C80 Section 2.2.3.1.1 Problem
Comments	Any additional free-text comments on the condition.	Optional	
Timestamp	The date and time that this information was last updated.	Mandatory if demographic information is specified.	

Ocular Imaging

Ocular imaging primarily consists of information objects defined by the Digital Imaging and Communications in Medicine (DICOM) standard. Rather than duplicate the DICOM standard in this document, the reader is referred to appropriate DICOM supplements depending on the type of ocular images they are tagging, as defined below. When constructing a DICOM record, any personally identifiable information (based on HIPAA privacy regulations) should be omitted.

- Optical Coherence Tomography: *Supplement 110 – Ophthalmic Tomography Image Storage SOP Class*
- Fundus Photography: *Supplement 91 – Ophthalmic Photography Image SOP Classes*

The Clinical Interpretation information object is defined as follows.

Clinical Interpretation			
Attribute	**Definition**	**Obligation**	**Comments**
Timestamp	The date and time that this interpretation was recorded.	Mandatory if clinical interpretation information is specified.	
Reviewer Identifier	An identifier for the individual that provided this interpretation of the retinal images.	Mandatory if clinical interpretation information is specified.	
Image Identifier	An identifier used to reference the	Mandatory if	

	image relevant to the clinical interpretation.	clinical interpretation information is specified.	
Uniform Resource Identifier (URI)	URI for the vocabulary used to tag the images.	Optional	
Name	Name of attribute used to tag the images with corresponding vocabulary URI.	Optional	
Value	Value of attribute used to tag the images with corresponding vocabulary URI.	Optional	
Probability	The degree of certainty that the finding is what the reviewer believes it to be.	Optional	
Comments	Any additional comments provided by the reviewer.	Optional	