Machine-actionable web linking: datasets, services, metadata

An abstract specification for a hypermedia control to link to data or metadata services.

Working concept development document for discussion

S.M. Richard [steve.richard@azgs.az.gov](mailto:steve.richard@azgs.az.gov) October 23, 2013

# Introduction

This document outlines a scheme for properties that can be used to specify the behavior of a link in a hypermedia document such that machine agents can interpret the link and use it with minimal intervention by a human user. There are a variety of situations in which a machine client might acquire and process a listing of resources—the objective might be to make resources available to a user in their work environment, reconstruct an archived compound digital object, reconstruct a workspace, to generate data visualization, or to duplicate a workflow. The basic issue is how to assert associations (referred to as links) between resources for machine processing.

This issue impacts a variety of current activities considering use of Atom, GeoRSS, or various metadata formats to describe associations between resources, including (among others) the Open Geospatial Consortium [OWS context Standards Working Group](http://www.ogcnetwork.net/context), [data and service casting schemes](http://wiki.esipfed.org/index.php/Discovery_Cluster) being developed by the Earth Science Information Partners (ESIP) , energy industry ISO19115 metadata profile, the protocol for Web Description Resources (POWDER), USGS CGI Web Application Integration Framework group, linked data profiles being developed for JSON encoding (JSON-LD, <http://json-ld.org/>), and the Open Archives Initiative Object Reuse and Exchange specification.

The basic web architecture is designed to account for human-directed navigation of links to obtain resources that for the most part were intended for display and visual processing by human users. Implementation of REST-architecture, in which the links (hypermedia controls, affordances) in hypermedia resource representations must be self-explanatory in order to enable the hypermedia to drive application state [Fielding, web blog 2009]. With the increasing adoption of service-based architecture, linked data, and semantic web technology, machine interpretation and processing of resources is becoming an integral part of an evolving distributed computation system. Simply clicking on a link to see what you get does not work in this environment. Links between resources for machine-automated processing require additional information about the nature of the target resource, its capabilities, data structure, and content.

Software clients on the World Wide Web use the MIME type value of the content-type header parameter in the response to an ‘http GET’ to determine how to handle a response document (message). The handler has to know characteristics of the encoding scheme in the document to utilize the response content. [references to MIME type definition and registration rfc 6638]. This works for result documents that will be opened and understood by human users using common desktop applications. This proposal addresses a more complex situation: given a choice of several URIs to deference (typically using http framework, with implication that http header parameters may be involved), a software application (machine) must determine which one(s) exposes the representation or interface that the software can utilize. The solution proposed here is for link providers to supply some a priori knowledge of the link targets to help applications utilizing the links; this information is supplied as properties associated with the links in a resource representation.

## Use Scenarios

Here are a variety of scenarios framed in several contexts: a metadata record, a ‘service cast’ or ‘dataset cast’ document, an ‘OWS context’ (‘common operating picture’) document (atom xml or JSON), or an XML ‘data’ document (like GeoSciML). These use cases are all related to the issue of what information needs to be encoded to make a URI machine-actionable in a simple, useful way.

* A data citation provides a link to directly access a particular subset of some data set.
* A GeoSciML instance document contains URIs that specify terminological quantifiers for various property elements; a user interface must present these using labels intelligible to users of the data
* A WaterML instance document contains URIs that specify terminological quantifiers for various property elements; a data processing application is comparing these attribute values to data from another source and must assess concept similarity. An owl representation would be most useful.
* A WFS server is processing filter requests against a GeoSciML document with concept expansion on attribute values specified by URIs, and must determine the transitive closure of the concept in the containing concept scheme.
* A metadata record for a data granule links to a metadata record describing the collection that contains the granule
* A metadata description of a service resource links to metadata for datasets it serves.
* Metadata for a dataset contains actionable link/description of services providing the data so that a client can connect to the service and access data without human intervention.
* A service metadata record contains links to metadata for the datasets the service distributes or utilizes.
* An Atom feed describing an information resource provides links enabling a variety of human and machine interactions with the resource that access different representations and interfaces.
* A workflow description describes a chain of services; this requires “Hypermedia as the engine of application state” for a machine automated data-processing.

# Discussion

The internet is based on a layered is a stack of network protocols (see Open Systems Interconnection (OSI), see http://www.itu.int/rec/T-REC-X.200-199407-I/en). Each layer in the stack is designed to abstract some aspects of messaging between network clients, such that messages can be exchanged between agents operating in that layer with no need to know about lower layers in the stack. The application layer is the top layer of the stack and is concerned with communication between user applications. Operation of the [application layer](http://en.wikipedia.org/wiki/Application_Layer) typically is founded on protocol such as HTTP, FTP, or SMTP. Many applications, such as web browsers, e-mail clients, and file-transfer programs, require only the operations provided by these protocols. With the advent of more complex distributed applications various techniques have been developed and are in use to ‘wrap’ complex data bundles and operation invocations within messages transported via the basic application layer protocols. Open Geospatial Consortium WxS services, OpenDAP, and numerous ‘RESTful’ APIs are examples of this approach.

Software that automates workflows using machine actionable links needs are commonly presented with resource representations that may include a variety of links. In order to function correctly, the software must identify the links that will access resources that are useful to the application, or meet its processing requirements. Requirements for the REST concept of “Hypermedia as the engine of application state” should be noted as well. A software agent may only know how to parse CSV or NetCDF files, or might require an OGC WMS, WFS or WCS; perhaps it requires graphics encoded as SVG; it may require content encoded in particular xml schema, e.g. GeoSciML, WaterML, or a particular RDF vocabulary. A client might need to know the available options for representations of a requested resource. Content negotiation in http allows management of these options in simple cases, but the use of [MIME](http://en.wikipedia.org/wiki/Internet_media_type) types to specify details of xml schema, data structure, or vocabulary is problematic. These conventions become very application specific and MIME is intended to be a ‘standard’ that spans multiple applications and domains.

# Links vs. Identifiers

This proposal is about links on the World Wide Web. Links are machine actionable if a software application can parse the link and use the information there to access some resource necessary for its function. The term ‘identifier’ will be used here to denote a string that is intended to correspond to a specific, particular resource, and the term ‘link’ will be used to denote an object that has the purpose of accessing a resource (or representation of a resource) on the Web. On the World Wide Web, resources are accessed using URLs—resource locators. It is rapidly becoming a convention to use http URIs to identify both resources that are directly accessible electronically (e.g. a particular file), and those that may be physical or abstract for which only an electronic representation can be accessed via the Web.

# Protocol options

Protocol is used here to encompass considerations having to do with messaging between client and server—what requests are supported, how requests are encoded. The content of the service’s response is a representation of the requested resource, discussed in the next section. The internet functions on a series of [‘stacked’ protocols](http://en.wikipedia.org/wiki/Communications_protocol#Protocol_layering), starting from the hardware level, and working up to the level of file formats. The internet is defined by the stack of low-level protocols that allow machines to communicate—TCP and IP. This allows applications developed for use on the internet (like the World Wide Web) to be developed by defining higher level protocols transported using the lower level protocols. This specification that utilization of a URI-based link in the application layer can be specified with three properties:

Identifier protocol for URI—this determines the protocol for dereferencing the identifier and has implications for operations and options in that dereferencing. HTTP is the most common protocol; a variety of other identifier protocols (doi, issn, isbn, ark) use dereferencing schemes based on HTTP operation, but have separate name servers. The IETF URI specification mandates that “each URI begins with a scheme name” (section 1.1.1, IETF RPC-3986); thus explicit specification as a link attribute for links using URIs should be unnecessary.

Service type– this determines the operations, parameters etc. that may be used in construction of a request for a desired resource, operation, or application state. Service type identifiers should be version-specific. Example service types include OGC WMS, OGC WFS, OpenDAP, OGC CSW.

Service profile – In many cases, knowing the service type is not sufficient—there may be additional constraints in particular service instances that are standardized in a profile that needs to be understood in order to successfully use the service. These typically specify particular data schema, vocabularies, and data constraints. Example: 'INSPIRE Geology view WMS'.

# Media Types

MIME types were originally defined in the context of attachments to e-mail messages. As the world wide web has evolved, the concept has been generalized to Media Type. The definition of 'Media Type' has been purposefully left vague by the IEFT, left to the discretion of reviewer of RFC documents proposing new media registrations.

Ideally all clients and services would be utilizing a REST architecture in which the media type would contain information describing any application state transitions that might follow retrieval of the document. In actual practice, there will be services in the system based on various architectures, include component based remote-procedure calls, object-based URI schemes with CRUD operations, etc. For such services, specifying the hypermedia type of the response is insufficient guidance for an application inspecting a list of links to select a link that it can use to access the resource.

# Representation

Options for resource representation when dereferencing an identifier have traditionally been accounted for with MIME types and http content negotiation; this has worked fine for handling representations that can be mapped to user applications and resources that a human will interact with mostly by viewing on a computer screen. The evolution into machine-actionable links and linked data requires more explicit approaches to presenting choices to a client that are handled in the application before the http request is issued. Computer processing of representations returned in JSON, XML, CSV, NetCDF encoding may be possible to some degree with no a priori understanding of the content data schema, but things are much easier and more efficient if the client is working with a representation (encoding and schema) that it understands. Thus the client should be able to determine if there is a representation available that it can work with and how to request precisely that representation.

Metadata records that describe a resource intended for use by a software client may present a variety of distribution options (using distribution in the ISO19115 sense). This metadata may be in an Atom feed describing a dataset, a formal metadata record from a catalog search, an archive record for some aggregated resource (OAI Object Reuse and Exchange). A client processing the record needs to be able to identify the distribution / representation option it needs. For example the client might need a WMS that serves tiff images, a WFS that offers GeoSciML mapped features, or a CSW that uses the INSPRIRE ISO profile.

To explore what is being done to implement machine actionable links, a survey was made of several specifications that are in use (Table 1). The various approaches generally build on the html <Link> approach, with information encoded in one or more attributes on the link, or in the codelist/controlled vocabulary associated with these attributes. All utilize MIME type in some fashion, and generally in include some kind of ‘rel’ or ‘role’ attribute used to indicate the semantics of the link.

Table . Link parameter specifications reviewed for this analysis

| Specification | Link | Notes |
| --- | --- | --- |
| ESIP discovery | <http://wiki.esipfed.org/index.php/Discovery_Cast_Atom_Response_Format_v1.1> | Describes links to external information. These links include a relation and type. |
| ISO19115/ 19139 |  | formal metadata specification. CI\_OnlineResource element is used to encode machine actionable links |
| IETF Web Linking | <http://tools.ietf.org/html/rfc5988> | IETF-5988 specifies relation types for web links, defines a registry for type definitions to enable interoperability, and defines the use of the Link field in HTTP headers to encode links |
| IANA link type registry |  |  |
| ATOM | <http://tools.ietf.org/html/rfc4287> | An XML-based document format that describes lists of related information known as "feeds". Feeds are composed of a number of items, known as "entries", each with an extensible set of attached metadata. Defines link element with relation type, original vocabulary of 5 types extended in IANA link type registry |
| RDFa | http://rdfa.info/about/ | scheme to add attributes on xml elements to tag them with RDF encoded metadata, designed for use in XHTML web documents. These allow association of xml element values with URI’s for properties, datatypes, linked resources (only one, because they are xml attibutes), types, or identifiers |
| xlink |  | specification for attributes associated with links in xml documents |
| GML codeType |  | XML element scheme to associate a URI with a context, use for linking to concept resources. Similar to CodeList type in ISO 19139. |

Table . Link type relation vocabularies reviewed for this analysis

| Vocabulary | link | Notes |
| --- | --- | --- |
| ESIP link type |  | for data casting and service casting, to extend ATOM Link types |
| ISO19115-1 online function codes |  | for function property of CI\_Online Resource. From draft version of 19115-1 DIS. |
| IANA link relations | <http://www.iana.org/assignments/link-relations/link-relations.xml> | for rel attribute of Link element. extends link types defined in IETF-5988 |
| DataCite RelType | <http://schema.datacite.org/meta/kernel-2.2/doc/DataCite-MetadataKernel_v2.2.pdf> | relationships from a citation to a related resource |
| RDFa relation | <http://www.w3.org/TR/rdfa-syntax/#relValues> | RDF predicates for relationships between resources |
| Dublin Core Terms | <http://dublincore.org/documents/dcmi-terms/> | Compare terms in substitution group for dct:relation. |

# Results of compilation of specifications

Table 3 is a summary of parameters used as link properties in the various schema that were studied (see Table 1). These are grouped into related attributes color coded in the table. 1) A target URI (variously named ‘href’ or ‘linkage’), relationship type property (‘rel’ in most specs), a human-readable title or label for the link to use in GUI presentation, and a MIME type indicating the file format of the target are used in most of the specifications and are considered the core mandatory properties. 2) Additional useful information about the target is supplied by altTitle, description, media, length, and hreflang; these properties would be optional. 3) Information about the protocol, format, schema for the target are specified by function, protocol, serviceType and outputScheme. 4) The last 5 attributes in Table 3 (xlink:base, nofollow, show, actuate, noreferrer, prefetch) are switches that indicate suggested or required client behavior when a link is used, and are all optional. Sources of link relationship vocabularies are summarized in Table 2, and a compilation of the relationship types is presented in Table 4.

Table . Link properties that are in use. The first four are the most widely used and are considered the ‘core’ attributes. Abbreviations in the schemes column: xlink—W3C link specification; atom—IETF4287 Atom feed specification RFC; iso—ISO 19115/19139; 5988—IETF5988 web link RFC; esip—ESIP discovery cluster data and web casting discussion; ows—OpenGeospatial consortium OWS context discussion from Standards working group.

| Element | Scope note | Schemes that use |
| --- | --- | --- |
| 1. Basic information | | |
| targetURI | URI that identifies the resource that is the source of the link. This is generally an http URI, which will be dereferenced. The associated attributes provide guidance for client software to determine if it wants to dereference this identifier and what representations is can expect when it does. If identifier is not http, then the protocol property should indicate the scheme used. | xlink, atom, iso, 5988, esip, ows |
| rel | URI from IANA rel vocabulary for consistency with IETF5988. Semantics of link from global vocabulary for interoperability. Semantics in this context means calculatable (see discussion in Coyle, 2010 p. 19). Attribute value is list; best practice is to include one of the 5 original Atom link@rel values for interoperability. | xlink, atom, 5988, esip |
| title | free text to label link in GUI, used to describe the meaning of a link or resource in a human-readable fashion | xlink, atom, iso, 5988, owc |
| MIMEtype | MIME content type. Indicates low-level file format. Expects a MIME type (http://www.iana.org/assignments/media-types/application/in­dex.html). Intention is that if a type is listed here, it is known to be offered by the host that the href will dereference to | atom, 5988, esip, owc |
| 1. Additional information | | |
| altTitle | String that encodes title value in a different character set, and/or contain language information as per [RFC5987]. | 5988 |
| description | detailed text description of what the online resource is/does | iso |
| media | indicates intended destination medium or media for style information (see Le Hors et al., 1999, Section 6.13 http://www.w3.org/TR/html401 ). Example values include 'screen', 'tty', 'print', 'braille', 'aural'... Vocabulary appears to be related to type of device (including paper as a device...) that is intended target for resource representation. Default to 'screen', and it is anticipated that other values would be only rarely required. | 5988 |
| length | Indicates an advisory length of the linked content in octets; it is a hint about the content length of the representation returned when href identifier is dereferenced | atom |
| hreflang | describes the language of the resource pointed to by the href attribute. When used together with the rel="alternate", it implies a translated version of the entry. Multiple "hreflang" parameters on a single link-value indicate language options that may be indicated by the client. | atom, 5988 |
| 1. Function, format, schema of target | | |
| function | Controlled vocabulary—tells client why they’d use this link. Purpose property provides mechanism for more granular, application specific indication of link semantics. Example values: 'download', ‘browsing’, 'fileAccess'. Analogous to ISO19115 CI\_OnlineFunctionCode. See Table 5. Default is ‘download’. | iso |
| protocol | identifier protocol to be used e.g. http, ftp, dns, smb, nfs, smtp, pop. Also allow other identifier schemes that may not have implicit web behavior: ARK DOI EAN13 EISSN ISBN ISSN ISTC LISSN LSID UPC URN. See IETF registry at http://www.rfc-editor.org/rfcxx00.html. Protocol operating at the 'bottom' of the application layer of the OSI network protocol stack. Various other protocols are specific to particular applications, and are indicated using serviceType (and version) and the outputScheme. Default is http. | iso |
| serviceType | URI that identifies a service protocol. This specifies protocols for the network layer above http/ftp etc. Should be a URI that can dereference to some kind of service specification document. [e.g. CSW 2.0.2, WFS 1.1.1, WMS 1.3.0, OpenSearch 1.1, OpenDAP]. Proposal is that the type URI should indicate the specification version; one or more versions that may be requested can be explicitly indicated by multiple serviceType values. URI to identify serviceType and version should be defined by the service specification. | esip, owc, iso |
| outputScheme | profile for content of message retrieve by href URL; URI for xml schema or JSON scheme, other description of data structure and content. Clients look at this to pick the link that will get a representation they can use. This is the information scheme in the layers on top of the MIME type encoding; note that the same output scheme might be encoded using different MIME types, so the two are somewhat orthogonal. MIME types have been conflating this property with the lower level encoding (.vnd, +xml … stuff). | iso, esip |
| 1. Operational switches | | |
| xml:base | stem for relative URI in attributes, or for CURIEs | xml |
| nofollow | Indicates that the context’s original author or publisher does not endorse the link target. Optional, Boolean; default is ‘False’ | IANA |
| show | When link is to resource that is a component of the resource containing the link, indicates desired presentation of the ending resource on traversal from the starting resource. Value must be one of the values "new", "replace", "embed", "other", and "none". Used to assemble a resource 'by reference' to libraries of component parts. 'new' and 'replace' only make sense in the context of a window-based browsing application. | xlink |
| actuate | When link is to resource that is a component of the resource containing the link, indicates desired behavior when the containing resource is parsed or loaded into the client environment. Values: "onLoad", "onRequest", "other", and "none". Is ‘onLoadl same as ‘prefetch’? | actuate |
| noreferrer | Indicates that no referrer information is to be leaked when following the link | IANA |
| prefetch | Indicates that the link target should be preemptively cached | IANA |

Table . Compilation of relation type terms from analyzed specifications. Terms are grouped according to the scope of the relation; each attribute group is shaded with a different color and the groups are labeled with gray-shaded cells. Rel terms that specify link function are summarized separately in Table 5.

| Type | Subtype | Notes | Vocabularies |
| --- | --- | --- | --- |
| Access to resource representation and description | | | |
| current-version (iana:current) |  | link to current version of the context resource; MIMEtype, outputschema used to disambiguate different representations available. This is to allow advertising of specific representations. [syn: latest-version] | IANA |
| browseGraphic |  | link to low-resolution visualization of resource, used for determination of fitness for purpose | esip, iso |
| browsing |  | link is to web application that will allow user to explore the resource content | iso |
| documentation (iana:describedBy) |  | online information about the resource | esip, ISO, IANA, DataCite |
| metadata |  | link retrieves formal metadata record describing resource. outputScheme provides information to select metadata | esip, ISO, RDFa |
| monitor |  | Link to resource (feed) that can be used to monitor changes in the context resource. See http://tools.ietf.org/html/rfc5989. This relation type target URI apparently has to be a session Initiation Protocol (sip:) URI. | IANA |
|  | monitor-group | resource that can be used to monitor changes in a group of HTTP resources that includes the context resource. From rfc5989: The monitor-group URI corresponds only to an Resource List Server (RLS as defined in RFC 4662) and never an HTTP resource or fixed set of HTTP resources. This relation type target URI apparently has to be a session Initiation Protocol (sip:) URI. | IANA |
| contents |  | link to listing of the parts of the context resource. This listing is considered a kind of feed | IANA, RDFa |
| alternate |  | link is to a different representation (substitute, variant, version) of the context resource. [this appears to make sense only if the context resource is a particular representation, otherwise the alternates should be separate links to access the resource] | IANA, DataCite, RDFa, DCT |
|  | duplicate | link to a resource whose available representations are byte-for-byte identical with the corresponding representations of the context | IANA |
|  | successor-version | link to resource that supersedes the context resource in a version history | IANA, DataCite, DCT |
|  | predecessor-version | link to resource that precedes the context resource in a version history | IANA, dataCite |
|  | working-copy | link to a resource that is a revision draft for a successor resource | IANA |
|  | working-copy-of | link from a revision draft resource to the resource it is intended to supersede | IANA |
|  | isOriginalFormOf | link from the original version in version sequence to some successor version. | DataCite |
|  | isVersionOf | inverse link for isOriginalFormOf; link from a version to the original resource in a version sequence | DCT |
| icon |  | Link to icon resource representing the link context | IANA, RDFa |
| index |  | link to index resource for searching context resource(?) | IANA, RDFa |
| Alternate identifiers associated with resource | | | |
| bookmark |  | target of relationship is permanent link to use for bookmarking purposes. | IANA RDFa |
| canonical |  | target of relationship is preferred version of a set of URIs with highly similar content. It is intended to help search engines when the same or highly similar similar content is available at different URIs. | IANA |
| self |  | a link (URI) for the current context. E.g. use in search result listing to reproduce the search URI that produced the result. | IANA |
| Links between parts of a segmented resource | | | |
| start |  | first resource in the ordered collection of resources that contains the current context. | IANA, RDFa |
| first |  | to the first item in the ordered collection of resources that contains the current context. (see also start and top). | IANA, DataCite, RDFa |
| prev |  | to the previous item in the ordered collection of resources that contains the current context. (see also start and top). | IANA, DataCite, RDFa |
| last |  | last resource in the ordered collection of resources that contains the current context. | IANA, DataCite, RDFa |
| next |  | to the next item in the ordered collection of resources that contains the current context. (see also start and top). | IANA, DataCite, RDFa |
| up |  | link to a parent resource in a linked hierarchy of resources. | IANA, RDFa, DCT |
| hasPart |  | Generic type for links to parts of a resource | DCT |
|  | section | link to a section (part) of a collection of resource | IANA, RDFa |
|  | chapter | identifier for a chapter (part) within a resource | IANA, DataCite, RDFa |
|  | subsection | [how are chapters, sections, and subsections distinguished?] | IANA, RDFa |
| Links to resource specifying properties of the context resource | | | |
| copyright |  | Link to copyright statement that applies to the context resource | IANA, RDFa |
| license |  | Link to a resource that specifies licensing stipulations for use of context resource | IANA, RDFa |
| privacyPolicy |  | link to a Platform for Privacy Preferences (P3P) privacy scheme Policy Reference File. | RDFa |
| author |  | link is to Author. The author is the originating agent for a resource; this is a 'non-information' resource, thus there is some representation involved | IANA |
| tag |  | a literal value (string) to be associated with the context resoruce as a finding aid | IANA |
| conformsTo |  | dublin core relation type; explicit link to specification that resource conforms to. Normally this should be same as URI in outputScheme link parameter, but as a distinct relation element allows separate explicit link to conformance spec for target resource. Title of link should be same as outputScheme or ServiceType URI in a link rel. | DCT |
| hasFormat |  | if a type attribute is not a registered MIME type, or needs additional explanation, a separate hasFormat element with title={the type string} can be included. | DCT |
| requires |  | link to a resource to which the context resource has a dependency. This element must be contained in a metadata description for the resource that present the requirement. | DCT |
| via |  | link to a resource that is a source of information in the context resource; interpreted here to be generic term subsuming compilation, citation, referencing. Provenance. | IANA |
|  | compiles | Link to resource whose content has been incorporated into the context resource | DataCite |
|  | cites | link to a resource that is cited for some reason (evidence, authority, attribution) in the context resource. Distinction of Cite and Reference needs clarification | DataCite, RDFa |
|  | references | link to a resource that provided information used in the development of the context resource. Distinction of Cite and Reference needs clarification | DataCite, DCT |
| Link is to related resource | | | |
| event |  | link is to micro-article related to context resource [intention is somewhat unclear in esip discussions] | esip |
| related |  | link is to a resource that has some useful association to the context resource; generic link with essentially no semantics | IANA |
|  | enclosure | a related resource that is potentially large and might require special processing. [Not very useful because semantics are unclear] | IANA |
| collection |  | link gets Data Casting Collection. Function unclear; guess is a document that is a list of related resources. How is this different from esip feed? | esip |
|  | feed | link is to a related RSS or ATOM feed; esip restricts to feed of feeds… only necessary if use [serviceCast](http://sciflo.jpl.nasa.gov/scast/) and [dataCast](http://datacasting.jpl.nasa.gov/) as other collection subtypes | esip |
| replies |  | link is to a resource that responds in some way to assertions or information in the context resource | IANA |
| documents |  | link to a resource that is the subject of information in the context resource | DataCite |
| isCompiledBy |  | link to resource that incorporates context resource into a compiled resource | DataCite |
| appendix |  | link is to appendix resource | IANA, RDFa |
|  | isSupplementTo/ isSupplementedBy | subtyping is based on interpretation that an appendix is equivalent to a supplement | DataCite |
| archives |  | link to collection of resources of historical interest relative to the context resource | IANA |
|  | next-archive | Navigation through archive resource-- have to maintain relationship with what the archive context is… Intention is unclear | IANA |
|  | prev-archive | Navigation through archive resource-- have to maintain relationship with what the archive context is… Intention is unclear | IANA |
|  | version-history | a listing (feed) that enumerates ordered collection of all versions of the context resource | IANA |
| glossary |  | a glossary that defines terms used in the context resource | IANA, RDFa |
| help |  | link to a help resource that provides guidance on the use and interpretation of the context resource | IANA, RDFa |
| stylesheet |  | link to resource that provides instructions for presentation of the context resource | IANA, RDFa |
| isFormatOf |  | if resource is a format specification, this relation indicates link to resources that are examples of the format | DCT |
| isReferencedBy |  | generic type for inverse link to resource that uses information from and references (by link?) the context resource. Distinction of Cite and Reference needs clarification+C43 | DataCite, DCT |
| isCompiledBy |  | link to resource that incorporates context resource into a compiled resource | DataCite |
| isCitedBy |  | link to a resource that cites the context resource for some reason (evidence, authority, attribution). Distinction of Cite and Reference needs clarification | DataCite |
| isRequiredBy |  | inverse link for Requires--explicit link to resources that have a dependency on the context resource. | DCT |
| payment |  | this is considered a link function, not a relationship type; listed here because is IANA rel type. | IANA |
| replaces |  | link to a resource that the context resource is meant to supersede. See also 'predessor-version'. Distinction is that replace is used if resources are not a version sequence, for instance if a new specification is superseding some existing spec. | DCT |

Machine actionable links to resources that are not simple HTTP GETs for documents a browser will understand would require one or more of the Group 3 attributes in Table 3 (function, protocol, serviceType, and outputScheme). Table 5 is a summary of terms applicable to the function property of a link; these have been extracted from the relationship type compilation as a separate vocabulary. Protocol defaults to HTTP, and this will be the most common value. The protocol property could also be used to flag targetURI’s that are not standard web protocols, e.g. ISBN, DOI, ARK. For links to services with more complex behavior, protocol layers overlying the base (http, ftp) can be specified by serviceType, and if necessary an outputScheme applicable to the serviceType.

Table . Link relation terms that specify expected behavior when link is dereferenced

|  |  |  |  |
| --- | --- | --- | --- |
| Function | | | |
| download |  | link will retrieve data from web | ESIP, ISO |
|  | fileAccess | link is network file path specific to some local area network | iso |
| service |  | link is to service end point that provides access to resource through some interface | esip, IANA |
|  | search | link to online web application to search within the described resource; not clear if ISO intention is the same (or what it is!) | iso, IANA |
| upload |  | link endpoint will accept file upload (? POST, or get web-form that user interacts with to upload a file?) | iso |
| emailService |  | should be smtp or mailTo protocol type for user to interact with context contact agent | iso |
| offlineAccess |  | link to online instructions for requesting the resource from the provider | iso |
|  | order (iana:payment) | link to online order web application for obtaining the resource. Payment may be part of this process. | iso, IANA |
| hub |  | link to online web application that enables registration for notification of updates to the context (see monitor relation types…) | IANA |
| edit |  | link to a resource that can be used to edit the link's context | IANA |
| edit-media |  | link to a resource that can be used to edit media associated with the link's context | IANA |

## The 'profile' Link Relation Type

https://datatracker.ietf.org/doc/draft-wilde-profile-link/

This specification defines the 'profile' link relation type that allows resource representations to indicate that they are following one or more profiles. A profile is defined to not alter the semantics of the resource representation itself, but to allow clients to learn about additional semantics (constraints, conventions, extensions) tha t are associated with the resource representation, in addition to those defined by the media type and possibly other mechanisms.

see also http://microformats.org/wiki/html5-profile

# Need to handle template links

WADL, OpenSearch type approaches.

# What about xForms as a control paradigm for links?

# Proposed solution for machine actionable links

In order for software client to utilize a link, it first must understand the URI scheme for the link identifier and how to dereference that kind of URI. Table 6 lists elements that are considered most important for characterizing machine actionable links. This information could conceivably all be encoded in the MIME type, perhaps even utilizing some structured syntax to allow analysis of the MIME type string to extract some of these properties. It would lead to a massive proliferation of MIME types of increasing length and complexity. The solution favored here is to specify additional properties to link items.

Table . Proposed properties for useful machine-actionable links.

| **property** | **scope notes** |
| --- | --- |
| linkage (syn: href, targetURI) | URI that identifies the resource that is the target of the link. This is generally an http URI, which will be dereferenced. The associated attributes provide guidance for client software to determine if it wants to dereference this identifier and what representations is can expect when it does. |
| rel | Semantics of link. Semantics in this context means calculatable (see discussion in Coyle, 2010 p. 19). URI from IANA **rel** vocabulary should be included for consistency with IETF RFD-5988. Recommendation is to use the Terms not namespace qualified, following guidance in Atom Specification RFC-4287, section 4.2.7.2. |
| altRel | Other domain-specific terms specifying semantics of link, not from IANA registry. (Optional). Multiple values are separated by comma. Rel value string MUST be quoted if it contains a comma (","). |
| title | Free text to label link in user interfaces. Optional. The content of the "title" attribute is Language-Sensitive. Entities such as "&amp;" and "&lt;" represent their corresponding characters ("&" and "<", respectively), not markup. Link elements MAY have a title attribute. The "title" parameter MUST NOT appear more than once in a given link-value; occurrences after the first MUST be ignored by parsers. |
| type | MIME type of response. Specifies file format and optionally the native software application environment. Intention is that if a type is listed here, it is known to be offered by the host that the href accesses. Note that this is only a hint; for example, it does not override the Content-Type header of a HTTP response obtained by actually following the link. There MUST NOT be more than one type parameter in a link-value; occurrences after the first MUST be ignored by parsers. (default value text/html) |
| function | Term tells client why they’d use this link. (optional, default value ‘download’). Function property provides mechanism for more granular, application specific indication of link semantics. Example values: 'download', 'browsing', 'fileAccess', ‘service’. |
| protocol | Connection protocol to be used e.g. http, ftp, dns, smb, nfs, smtp, pop. See IETF registry at http://www.rfc-editor.org/rfcxx00.html. Protocol operating at the 'bottom' of the application layer of the OSI network protocol stack. Also allow other identifier schemes: ARK, DOI, EAN13, EISSN, ISBN, ISSN, ISTC, LISSN, LSID, UPC, URN, etc from IETF URI scheme registry; URIs using these other schemes are only dereferenceable if the client knows a priori how to process them. Protocols specific to particular applications that are layered on this base protocol are indicated using serviceType (and version) and the outputScheme. Optional, default is HTTP. |
| serviceType | URI that identifies a service protocol. This specifies protocols for network layer above http/ftp e.g. CSW, WFS, WMS, OpenSearch, OpenDAP, etc. Should be a URI that can dereference to some kind of service specification document. URI for serviceType may indicate a version, but if not, one or more versions that may be requested can be explicitly indicated. Syntax to identify versions is defined by the service specification. Optional. |
| outputScheme | profile for content of message retrieve by href URL; URI for xml schema or JSON scheme, other description of data structure and content? Need conventions for outputScheme vocabulary. Clients look at this to pick the link that will get a representation they can use; unknown outputSchemes can be ignored. This is the information scheme in the layers on top of the MIME type encoding and serviceType; note that the same output scheme might be encoded using different MIME types or accessed through different serviceTypes, so the two are somewhat orthogonal. Optional. |
| Other properties that may be useful (all optional) | |
| altTitle | String that encodes title value in a different character set, and/or contain language information as per [RFC5987]. |
| descriptionurl | detailed text description of what the online resource is/does. Since is not considered good practice to put extensive text in an element attribute, implement by reference with a url for an html description page. |
| media | Indicates intended destination medium or media for style information (see Le Hors et al., 1999, Section 6.13 http://www.w3.org/TR/html401 ). Example values include 'screen', 'tty', 'print', 'braille', 'aural'... Vocabulary appears to be related to type of device or material manifestation (including paper as a device...) that is intended target for resource representation. Default to 'screen', and it is anticipated that other values would be only rarely required. [debate point—is this necessary for links that are intended to operate on the web?] |
| length | Indicates an advisory length of the linked content in octets; it is a hint about the content length of the representation returned when linkage identifier is dereferenced |
| hreflang | describes the language of the resource pointed to by the linkage attribute. When used together with the rel="alternate", it implies a translated version of the entry. Multiple "hreflang" parameters on a single link-value indicate language options that may be indicated by the client. |
| behavior | A comma separated list of properties specifying behavior expected in client when link is actuated. See Table 7 for list of values. |

Table . Vocabulary for specifying link behavior property.

| Term | Explanation |
| --- | --- |
| nofollow | if property is specified, indicates that the context’s original author or publisher does not endorse the link target. |
| showreplace | equivalent to ‘show’ value for xlink:show property. |
| showembed | equivalent to ‘embed’ value for xlink:show property. |
| actuateOnLoad | equivalent to ‘onLoad’ value for xlink:actuate property. |
| actuateOnRequest | equivalent to ‘onRequest’ value for xlink:actuate property. |
| noreferrer | if property is specified, indicates that no referrer information is to be leaked when following the link |
| prefetch | if property is specified, indicates that the link target should be preemptively cached |

## Implementation—still waiting….

The Atom protocol allows attributes to be added to any element as long as they are not in the atom (<http://www.w3.org/2005/Atom>) namespace (Nottingham and Sayre, 2005, p. 7). An xml schema at <http://schemas.usgin.org/schemas/ExtLinkAttributes0.5.xsd> defines the attributes from Table 6 that extend the base attributes defined on the link element in the Atom Specification.

As a concrete example of the need for clearer conventions for specifying link properties, consider links in simple metadata. The Open Geospatial Consortium catalog service for the web requires all implementations to be able to provide a response conforming to the csw record xml schema (<http://schemas.open- gis.net/csw/2.0.2/CSW-discovery.xsd>). The content of this schema is essentially the Dublin Core and Dublin Core Term elements (<http://dublincore.org/documents/dcmi-terms/>). Metadata from the ESRI Geoportal (v1.0) CSW implementation encodes links to OGC service end points like this:

<dct:references scheme=” urn:x-esri:specification:ServiceType:ArcIMS:Metadata:Server “>http://url.to.server/wms?service=WMS&request= </dct:references>.

GeoNetwork opensource (OneGeology catalog hosted by BRGM) encodes the capabilities URL thus:

<dc:URI protocol="WWW:LINK-1.0-http--link" description="http://ogc.bgs.ac.uk/cgi-bin/BGS\_GSN\_Bedrock\_Geology/wms">[http://ogc.bgs.ac.uk/cgi-bin/BGS\_GSN\_Bedrock\_Geology/wms?</dc:URI](http://ogc.bgs.ac.uk/cgi-bin/BGS_GSN_Bedrock_Geology/wms?%3c/dc:URI)>

Note the dc:URI is not a valid element in the csw:record XML schema.

Other CSW servers may very well use different encodings. This is not interoperable—a web client cannot reliably parse even a ‘simple’ csw record to get the capabilities URL for an OGC service. The csw:record XML implementation of Dublin Core elements only allows a single attribute (scheme) on the relation element substitution group. It appears that the ESRI approach, using the <dct:references> element with a URI that identifies an OGC getCapabilities document would be a viable solution. The other extended link properties could be added in the <dct:references> element, but this would not be schema valid against the csw:record schema.

# References and related reading

Adida, Ben, Birbeck, Mark, McCarron, Shane, and Pemberton, Steven, editors, 2008-10-14, RDFa in XHTML: Syntax and Processing, A collection of attributes and processing rules for extending XHTML to support RDF: W3C Recommendation 14 October 2008, accessed at http://www.w3.org/TR/2008/REC-rdfa-syntax-20081014, 2011-09-30.

Booth, David, 2003-01-28, Four uses of a URL: Name, Concept, Web Location and Document instance: accessed at <http://www.w3.org/2002/11/dbooth-names/dbooth-names_clean.htm> (2010-02-15).

Brickley, Dan, 2000-09-08, Notes on RDF, Xlink as linked information systems: <http://www.w3.org/2000/02/rdf-xlink/> accessed 2010-02-07

Coyle, Karen, 2010, Understanding the Semantic Web: Bibliographic Data and Metadata: Library Technology Reports, v.46, American Library Association, 31 p., ISSN 0024-2586

Duerst, M., and Suignard, M., January 2005, Internationalized Resource Identifiers (IRIs): IETF RFC-3987, accessed at <http://www.ietf.org/rfc/rfc3987.txt> .

Fielding, R., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and Berners-Lee, T., 1999-06, Hypertext Transfer Protocol -- HTTP/1.1: RFC2616, accessed at <http://tools.ietf.org/html/rfc2616 2010-02-19>.

Fielding, Roy, 2008, REST APIs must be hypertext-driven: Untangled, Web BLOG, <http://roy.gbiv.com/untangled/2008/rest-apis-must-be-hypertext-driven> (accessed 2013-01-25).

Gregorio, J.C., and deHora, B., editors, 2007, The Atom Publishing Protocol: IETF RFC 5023.

Halpin, Harry, and Presutti, Presutti, 2009, An Ontology of Resources for Linked Data: LDOW 2009, April 20–24, 2009, Madrid, Spain, ACM 978-1-60558-487-4/09/04. <accessed at <http://events.linkeddata.org/ldow2009/papers/ldow2009_paper19.pdf>, 09/03/2011>

ISO19115 WG, in prep, ISO 19115-1 Geographic information — Metadata – Part 1: Fundamentals: ISO TC211 N3071

ISO19119, 2005-02-15, Geographic Information—Services.

ISO19119:2005/Amd.1, 2008-05-01, Geographic information — Services, AMENDMENT 1: Extensions of the service metadata model.

ITU, 1994, Information Technology – Open Systems Interconnection – Basic Reference Model: The Basic Model: ITU-T Recommendation X.200, 63 pages.

Le Hors, A., Raggett, D., and I. Jacobs, December 1999, HTML 4.01 Specification, W3C Recommendation REC-html401-19991224, <http://www.w3.org/TR/1999/REC-html401-19991224>. Latest version available at <http://www.w3.org/TR/html401>.

Linked Data for the Web, [http://www.rdfabout.com/intro/?section=8](%20%20http://www.rdfabout.com/intro/?section=8) accessed 2010-02-07

Mendelsohn, Noah, and Williams, Stuart, 2007-01-02, The use of Metadata in URIs: W3C TAG Finding, accessed at <http://www.w3.org/2001/tag/doc/metaDataInURI-31-20070102.html>.

Nottingham, M., IETF, 2010, Web Linking: IETF RFC 5988, ISSN:2070-1721.

Nottingham, M., and Sayre, R., eds., 2005, The Atom Syndication Format: IETF Network Working Group Request for Comments RFC-4287, accessed at <http://www.ietf.org/rfc/rfc4287.txt> .

Rosinger, David, and Tillman, Stan, eds., 2010-06-24, OGC® OWS-7 Information Sharing Engineering Report: Open Geospatial Consortium, Inc., document OGC 10-035r1, accessed at <https://portal.opengeospatial.org/files/?artifact_id=40031&version=1> (OCG members only).

Starr, Joan (head of working group) and DataCite Metadata Working Group, 2011-07, DataCite Metadata Schema for the Publication and Citation of Research Data, version 2.2: DOI:10.5438/0005, accessed at <http://schema.datacite.org/meta/kernel-2.2/doc/DataCite-MetadataKernel_v2.2.pdf> 2011-09-30.

van Kesteren, A., Glazman, D., Lie, H., and T. Celik, September 2009, "Media Queries", W3C Candidate Recommendation CR-css3-mediaqueries-20090915, <http://www.w3.org/TR/2009/CR-css3-mediaqueries-20090915/>.