HAR Vocabulary Specification Draft 0.03

Preparatory draft to come to a Namespace Document 30 September 2017

IRI:

This version:

Authors:

Contributors: Niels Vandekeybus, Aad Versteden, Jonathan Langhens, Johan Delaure

---

0.1. Abstract

The Http ARchive (HAR-) format is used to register and exchange the web browser's interaction with a site. The aim of current vocabulary is to describe a HAR file such that it can be used and reused in an automated way. The description involves linking the HAR-specific descriptors to terms described in adopted vocabularies.

0.2. Status of This Document

The aim is to establish a stable core of classes and properties that allows the interoperable use of HAR files.
The HAR vocabulary specification is produced as part of the Big Data Europe project.

0.3. Table of Contents

1. Introduction

1.1. Context of this specification

A vocabulary is defined as a dictionary of terms, each of which is either a class or a property. With a vocabulary we aim to settle a common framework that allows data to be shared and reused across applications, enterprises and communities.

Different vocabularies define different sets of classes and properties. As the aim is to improve interoperability, vocabulary definitions should be unique and not overlap with those in other vocabularies. A good vocabulary reuses as much as possible definitions of existing and adopted vocabularies .

The vocabulary description serves as a “namespace document” that describes the classes and properties that constitute the vocabulary. The vocabulary is identified by the namespace URI. Revisions and extensions of the vocabulary are conducted through edits to this document.

The maturity of the vocabulary will be a function of the stability of the individual vocabulary terms and its adoption.

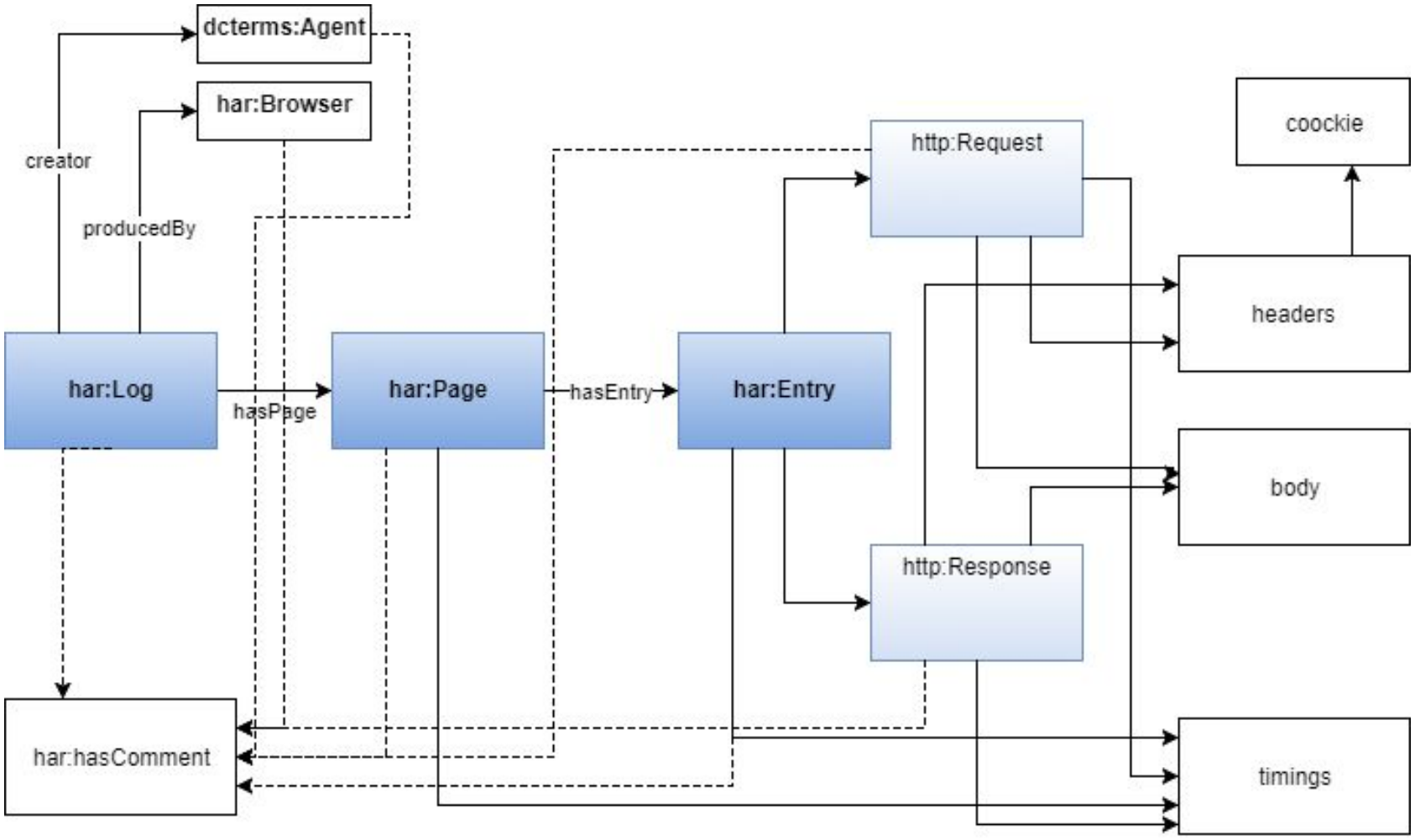
This document provides a first approach to the definition of an ontology for the HAR specification. **To provide transparency to the options and suggestions considered, considerations made and decisions taken, notes are included at this stage.** These notes are only relevant in the current process, they will not be retained in a more final version of the vocabulary.

1.2. Aim of HAR Vocabulary specification

The Http ARchive (HAR-) format is used to log, capture and store the client/server conversation in a web browser's interaction with a site. The data collected by monitoring HTTP communication includes detailed performance data and is archived in JSON format.

In order to share, use and reuse HAR files, it is a mandatory condition that there is a common understanding of the semantics used to describe HAR files.
This vocabulary describes HAR-files up to the level they are compliant with the HAR specification as described on following URL:
https://dvcs.w3.org/hg/webperf/raw-file/tip/specs/HAR/Overview.html#sec-object-types-log

1.3. HAR Namespace overview



1.4. External Vocabulary References

The description of the terms in the provided 'dictionary' often make reference to classes and properties defined in other vocabularies. This section provides a placeholder reference for any mention of externally defined terms.

@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
@prefix rdfs: <http://www.w3.org/2001/XMLSchema#>
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>
@prefix owl: <http://www.w3.org/2002/07/owl#>
@prefix dc: <http://purl.org/dc/elements/1.1/>
@prefix dcterms:http://dublincore.org/2012/06/14/dcterms
@prefix dpedia: <:http://dbpedia.org>
@prefix foaf: <http://xmlns.com/foaf/0.1/>
@prefix http: <http://www.w3.org/2011/http#>
@prefix time: <https://www.w3.org/2006/time#>
@prefix proton: <http://www.ontotext.com/proton/protonext#>
@prefix nfo: <http://oscaf.sourceforge.net/nfo.html>

For terms unique or new defined for the HAR vocabulary, we project a temporary URL defined as:
http://www.w3.org/2018/har

2. List of elements in HAR

2.1. HAR_File

The instances of this class are the actual HAR files.

Class	
HAR Spec:	HAR file: The actual HAR file that contains the har log. The HAR files has specific technical properties as being UTF-8 encoded, having the ".har" extension and having the HAR internal structure as specified by the HAR specification.
Ontology:	http://oscaf.sourceforge.net/nfo.html
Resource URI:	nfo:FileDataObject http://oscaf.sourceforge.net/nfo.html#nfo:WebDataObject
URI semantics:	(Nepomuk File Ontology (NFO) is one of the fundamental parts of NIE. It deals with files and other desktop resources. Files are understood as sequences of bytes stored in a Filesystem or on a Network.) A resource containing a finite sequence of bytes with arbitrary information, that is available to a computer program and is usually based on some kind of durable storage. A file is durable in the sense that it remains available for programs to use after the current program has finished.

2.3. HAR_filename

The Har file has a filename.

Property	
HAR Spec:	File name of the Har file
Ontology:	http://oscaf.sourceforge.net/nfo.html#nfo:fileName
Resource URI:	http://oscaf.sourceforge.net/nfo.html#nfo:fileName
URI semantics:	Name of the file, together with the extension

2.4. Log

Within the file, this is the root.

Class	
HAR Spec	This object represents the root of the exported data. This object MUST be present and its name MUST be "log".
HAR properties:	Req: "version"|string|Version number of the format. Req: "creator"|object|An object of type creator that contains the name and version information of the log creator application. Opt: "browser"|object|An object of type browser that contains the name and version information of the user agent.

	Opt: "pages"|array|An array of objects of type page, each representing one exported (tracked) page. Leave out this field if the application does not support grouping by pages. Req:"entries"|array|An array of objects of type entry, each representing one exported (tracked) HTTP request. Opt."comment"|string| A comment provided by the user or the application.
Ontology:	http://www.w3.org/2018/har
Resource URI:	har:Log
URI semantics:	Identifies a log or trace of a HTTP client/server conversation. The log is in the HAR-format.

2.5. Agent

Class	
HAR Spec	This object contains information about the log creator application
HAR properties:	Req:”name|string|The name of the application that created the log. Req:”version”|string|Its version Opt:”comment”|string|:a comment of the user or app.
Ontology:	http://dublincore.org/documents/2012/06/14/dcmi-terms/
Resource URI:	dcterms:Agent
URI	A resource that acts or has the power to act..

semantics:	

2.6. log created

Property	
HAR Spec	This object contains information about the log creator application
HAR properties:	Req:”name|string|The name of the application that created the log. Req:”version”|string|Its version Opt:”comment”|string|:a comment of the user or app.
Ontology:	http://dublincore.org/documents/2012/06/14/dcmi-terms/
Resource URI:	dcterms:creator
URI semantics:	An entity primarily responsible for making the resource.Examples of a Creator include a person, an organization, or a service.

2.7. Browser

Class	
HAR Spec**:**	This object contains information about the browser that created the log
HAR properties:	Req:”name|string|The name of the browser Req:”version”|string|Its version Opt:”comment”|string|:a comment of the user or browser.

Ontology	http://www.w3.org/2018/har
Resource URI:	har:Browser
URI semantics:	Software application used to navigate content that is formatted as for the World Wide Web and using the communication protocols of the internet.
SubClass of	http://purl.org/dc/dcmitype/Software
	NOTE
	It might be advisable to develop a separate vocabulary that describes the domain of browsers and their capacity.

2.8. producedby

Property	
HAR Spec	This object contains information about the log creator application
HAR properties:	Req:”name|string|The name of the application that created the log. Req:”version”|string|Its version Opt:”comment”|string|:a comment of the user or app.
Ontology:	http://vocab.getty.edu/ontology
Resource URI:	http://vocab.getty.edu/ontology#aat2212_produced_by
URI semantics:	things - [are] produced by - things/[equipment]. Example: stereoscopic photographs are produced by stereoscopic cameras; punchwork is produced by punches (marking tools)

2.9. browser:version

Property	
HAR Spec	Version of the browser
Ontology:	http://open.vocab.org/
Resource URI:	http://open.vocab.org/terms/versionnumber
URI semantics:	http://open.vocab.org/terms/versionnumber
Further detail:	Property to encode the version of things – simple intent, to record the sequence of items

2.10. Entries

Class	
HAR Term:	entries
HAR use:	An array of objects of type entry, each representing one exported (tracked) HTTP request. An entry stores additional information on each HTTPrequest.
HAR properties:	1. pageref [string, unique, optional] - Reference to the parent page. Leave out this field if the application does not support

	grouping by pages. 2. startedDateTime [string] - Date and time stamp of the request start (ISO 8601 - YYYY-MM-DDThh:mm:ss.sTZD). 3. time [number] - Total elapsed time of the request in milliseconds. This is the sum of all timings available in the timings object (i.e. not including -1 values) . 4. request [object] - Detailed info about the request. 5. response [object] - Detailed info about the response. 6. cache [object] - Info about cache usage. 7. timings [object] - Detailed timing info about request/response round trip. 8. serverIPAddress [string, optional] (new in 1.2) - IP address of the server that was connected (result of DNS resolution). 9. connection [string, optional] (new in 1.2) - Unique ID of the parent TCP/IP connection, can be the client port number. Note that a port number doesn't have to be unique identifier in cases where the port is shared for more connections. If the port isn't available for the application, any other unique connection ID can be used instead (e.g. connection index). Leave out this field if the application doesn't support this info. 10. comment [string, optional] (new in 1.2) - A comment provided by the user or the application.
Ontology	http://www.w3.org/2018/har
Resource URI:	har:Entry http://www.w3.org/2018/har/Entry
URI semantics:	Collection of metadata regarding the traced HTTP conversation.

2.11. page

Class	
HAR Term:	pages
HAR use:	This object represents list of exported pages
HAR properties:	1. *startedDateTime [string]* - Date and time stamp for the beginning of the page load (ISO 8601 - YYYY-MM-DDThh:mm:ss.sTZD, e.g. 2009-07-24T19:20:30.45+01:00). 2. *id [string]* - Unique identifier of a page within the . Entries use it to refer the parent page. 3. *title [string]* - Page title. 4. *pageTimings[object]* - Detailed timing info about page load. 5. *comment [string, optional]* (new in 1.2) - A comment provided by the user or the application. 6. onContentLoad [number, optional] - Content of the page loaded. Number of milliseconds since page load started (page.startedDateTime). 7. Use -1 if the timing does not apply to the current request. onLoad [number,optional] - Page is loaded (onLoad event fired). Number of milliseconds since page load started (page.startedDateTime). Use -1 if the timing does not apply to the current request. 8. comment [string, optional] (new in 1.2) - A comment provided by the user or the application.
Ontology:	http://www.w3.org/2018/har

Resource URI:	http:Response http://www.w3.org/2011/http#Response
URI semantics:	https://www.w3.org/TR/HTTP-in-RDF10/#RequestClass

2.12. request

Class	
HAR Term:	request
HAR use:	This object contains detailed info about performed request.
HAR properties:	1. method [string] - Request method (GET, POST, ...). 2. url [string] - Absolute URL of the request (fragments are not included). 3. httpVersion [string] - Request HTTP Version. 4. cookies [array] - List of cookie objects. 5. headers [array] - List of header objects. 6. queryString [array] - List of query parameter objects. 7. postData [object, optional] - Posted data info. 8. headersSize [number] - Total number of bytes from the start of the HTTP request message until (and including) the double CRLF before the body. Set to -1 if the info is not available. 9. bodySize [number] - Size of the request body (POST data payload) in bytes. Set to -1 if the info is not available. 10. comment [string, optional] (new in 1.2) - A comment provided by

	the user or the application.
Ontology:	https://www.w3.org/TR/HTTP-in-RDF10
Resource URI:	http:Request http://www.w3.org/2011/http#Request
URI semantics:	https://www.w3.org/TR/HTTP-in-RDF10/#RequestClass

2.13. response

Class	
HAR Term:	response
HAR use:	This object contains detailed info about the response.
HAR properties:	method [string] - Request method (GET, POST, ...). url [string] - Absolute URL of the request (fragments are not included). httpVersion [string] - Request HTTP Version. cookies [array] - List of cookie objects. headers [array] - List of header objects. queryString [array] - List of query parameter objects. postData [object, optional] - Posted data info. headersSize [number] - Total number of bytes from the start of the HTTP request message until (and including) the double CRLF before the body. Set to -1 if the info is not available. bodySize [number] - Size of the request body (POST data payload) in bytes. Set to -1 if the info is not available.

	comment [string, optional] (new in 1.2) - A comment provided by the user or the application.
Ontology:	https://www.w3.org/TR/HTTP-in-RDF10
Resource URI:	http:Response http://www.w3.org/2011/http#Response
URI semantics:	https://www.w3.org/TR/HTTP-in-RDF10/#RequestClass

3. Detailed HAR elements and vocabulary

See Spreadsheet

<end of document>