

OBSERVATION We plan to convey the vocabulary in both JSON-LD and Turtle formats.

ISSUE: Does Dataset Usage Vocabulary change for open/closed data here?

<http://www.w3.org/2013/dwbp/track/issues/153>

## Abstract

The Dataset Usage Vocabulary (DUV) can illustrate how datasets are used by humans and machines as well as describe consumer feedback in the form of user experiences and citation references. Datasets as defined by DCAT are *a collection of data, published or curated by a single agent, and available for access or download in one or more formats*. From this perspective a dataset is representative of a logical container used to organize physical data.

By specifying a number of foundational concepts used to collect dataset consumer feedback and experiences, APIs can be written to support collaboration across the web by structurally publishing consumer opinions and experiences, and provide a means for data consumers and producers advertise and search for published open dataset usage.

<http://www.w3.org/2013/dwbp/track/issues/154>

ISSUE: Dataset is being strictly defined along the lines of the DCAT definition. It is fairly common that organizations or scientific domains refer to a dataset as a collection of data. How do we accommodate the need to provide feedback on a collection of data?

For instance if I am performing a simulation of climate change and sharing the results with a collaborator they will be mainly providing feedback on the entire collection of datasets and not each dataset individually.

If we use `dcat:Dataset` and `dcat:Catalog` without breaking the DCAT model then we basically have to duplicate all properties in the data model. I'd like to avoid this.

What I propose is that we have:

- 1) a `duv:Collection` class as you suggested Bernadette
- 2) make a parent class for `dcat:Dataset` and `duv:Collection` that would serve as a "inode"
- 3) Call this parent class `duv:Data` (or something else - please suggest)
- 4) All data usage and feedback would relate to `duv:Data`.

Also we could consider: [schema.org](http://schema.org) tends to use a parent class called `CreativeWork` that carries many of the aspects of data usage. `CreativeWork` (<https://schema.org/CreativeWork>)

) acts as a parent class of anything on the web including DCAT. It also provides mechanisms for citation, discussion, etc. [http://schema.org/docs/schema\\_org\\_rdfa.htm](http://schema.org/docs/schema_org_rdfa.htm)

## Status of This Document

*This section describes the status of this document at the time of its publication. Other documents may supersede this document. A list of current W3C publications and the latest revision of this technical report can be found in the [W3C technical reports index](#) at <http://www.w3.org/TR/>.*

This is a draft document which may be merged into another document or eventually make its way into being a standalone Working Draft.

This document was published by the [Data on the Web Best Practices Working Group](#) as an Editor's Draft. If you wish to make comments regarding this document, please send them to [public-dwbp-wg@w3.org](mailto:public-dwbp-wg@w3.org) ([subscribe](#), [archives](#)). All comments are welcome.

Publication as an Editor's Draft does not imply endorsement by the W3C Membership. This is a draft document and may be updated, replaced or obsoleted by other documents at any time. It is inappropriate to cite this document as other than work in progress.

This document was produced by a group operating under the [5 February 2004 W3C Patent Policy](#). W3C maintains a [public list of any patent disclosures](#) made in connection with the deliverables of the group; that page also includes instructions for disclosing a patent. An individual who has actual knowledge of a patent which the individual believes contains [Essential Claim\(s\)](#) must disclose the information in accordance with [section 6 of the W3C Patent Policy](#).

This document is governed by the [1 August 2014 W3C Process Document](#).

## Table of Contents

- 1. [Audience](#)
- 2. [Introduction](#)
- 3. [Scenarios](#)
- 4. [Related Work](#)
- 5. [Data Usage at a Glance](#)
- 6. [Classes and Properties](#)
- 7. [Acknowledgements](#)
- 8. [The Data Usage Vocabulary](#)

### 1. Audience

The DUV is intended for data producers and publishers interested in tracking, sharing, and persisting consumer dataset usage. It is also intended for collaborators who require an exchange medium to advertise and interactively convey dataset usage.

ISSUE: Is the target audience scoped appropriately? The audience has been scoped to provide guidance to support producers, publishers desiring to collect consumer feedback and convey consumer usage information. From a consumer perspective the vocabulary is exploring usage from a collaborator perspective.

<http://www.w3.org/2013/dwbp/track/issues/155>

## 2. Introduction

This vocabulary is meant to fill a niche that helps standardize the way web published dataset usage be conveyed and shared. At this time is no clear standard way to describe dataset usage on the web. Without a means to systematically describe dataset usage, searching and conveying techniques are application specific and discovery and collaboration across the Web is more difficult.

This vocabulary also recommends and requires data publishers to provide a mechanism of receiving data usage from data consumers in the form of feedback, citation and data correction.

<http://www.w3.org/2013/dwbp/track/issues/156>

ISSUE: Should we include background material that describes the current approaches to data usage related efforts? DCAT, Schema.org, Dublin Core, Force-11 Citations

- <https://www.force11.org/datacitation>
- <http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/data-citation-and-linking>
- <http://dublincore.org/documents/dc-citation-guidelines/>

What about efforts that are not open, but might serve as models? Do we get into possible trouble with proprietary models? E.g. Stack Overflow has an incredible following supporting interactions between software developers, sharing code, ideas etc.

<http://stackoverflow.com/help/searching> . Github last May is in the process of developing a feedback model: <https://github.com/blog/1843-diversity-and-feedback-at-github> .

Laufer also provided SIOC for social interaction <http://rdfs.org/sioc/spec/>

The scope of DUV is defined by the Data on the Web Best Practices (DWBP) Use Case document based on the data usage requirements about datasets. These requirements include: citing data on the Web, tracking the usage of data, sharing feedback and rating data. These requirements were derived from fourteen real world use cases examples provided in the use case document.

### 3. Scenarios

<http://www.w3.org/2013/dwbp/track/issues/157>

ISSUE: What possible undocumented use cases are still outstanding and need to be considered for illustrative purposes?

ISSUE: I'm not sure if any of our use cases cover this, but I did find the Twitter feedback forms covers feedback on violations. <https://support.twitter.com/forms> . We should also take into account the privacy interest group, are there aspects of usage <http://www.w3.org/2012/dnt-ws/report>

ISSUE: Need to examine the stories from the social working group to determine if there either might be new use cases or illustrative stories we could consider using. Since the last release of the Use Case document, the data usage team has been in contact with the Social Web Working group <https://www.w3.org/wiki/Socialwg>

<NOTE From Ann Bassetti>

Yes, definitely interesting! I realize I should have included this link to the User Stories: [https://www.w3.org/wiki/Socialwg/Social\\_API/User\\_stories](https://www.w3.org/wiki/Socialwg/Social_API/User_stories) . The WG is in the middle of voting on that initial set of stories and doesn't want them changed during the voting. But we anticipate there are gaps, so have set up a new page to collect new ideas: [https://www.w3.org/wiki/Socialwg/Social\\_API/More\\_user\\_stories](https://www.w3.org/wiki/Socialwg/Social_API/More_user_stories)

If, reading through the first set, you think the attributes of scientific collaboration aren't represented, I encourage you to add one or more User Stories to the 2<sup>nd</sup> page. (Note that you can login to the wiki using your usual W3C access credentials.)

Feel free, also, to send a note to the SocWeb WG mailing list ([public-socialweb@w3.org](mailto:public-socialweb@w3.org) ) to explain your logic, or ask questions, if you feel like that would be helpful.

ISSUE: Is there a WCAG req we need to consider for Data Usage. I recommend we leave this action open for discussion. I don't believe have any DWBP use cases that reflect WCAG needs. <http://www.w3.org/WAI/intro/wcag> . In terms of 115, if it's going to remain open, can you suggest a concrete way to resolve it? Something like:

1. review existing use-cases for WCAG reqs (this is what we're doing for quality vocab)
2. add a new use-case that specifically outlines WCAG req, and/or
3. add to list of data usage requirements currently not covered by the data usage vocabulary

E.g. Feedback on inaccessible website: <http://www.w3.org/WAI/users/inaccessible> , See How People with Disabilities Use the Web: <http://www.w3.org/WAI/intro/people-use-web/stories>

Scenarios used in this document are used to illustrate data usage concepts and are adapted to appeal to a wide audience.

OBSERVATION: Sumit based on our conversation yesterday about your need for a mechanism, it occurred to me that we really need to nail down different usage patterns. Could they be:

1. Publisher wants to get feedback from consumers so he provides services, apis or forms to track feedback.
2. Publisher wants to provide README instructions about the dataset format, advice to consumers for using the data etc.
3. Collaborators working on a research effort want to notify other collaborators about current activities, they exchange their experiences to other as they work. These experiences can be exchanged through a subscribe publication system.
4. A user wants to search and discover tools that consume a particular kind of dataset. To do this the user wants to correlate feedback extracted from Github, Open Source, and other publicly available code repositories with published data.

#### 4.3.1 Requirements for Data Usage

##### R-Citable

*It should be possible to cite data on the Web*

<http://www.w3.org/2013/dwbp/track/issues/158>

ISSUE: Data citation on web can be interpreted in many different ways. It is proposed that citation be constrained as a means providing feedback or instruction from a data usage perspective only.

ISSUE: Should the citation vocabulary classes be comprehensive attempting to accommodate many different approaches or should the classes be considered open where citation classes can serve as a handle to customized approaches?

Motivation: [ASO](#), [GS1Digital](#), [LATimes](#), [LusTRE](#), [RDESC](#), [SharePSI \(](#), [Share-PSI Emergency Albania \)](#), [UKOpenResearchForum](#)

Use Case	R-Citable Benefits
----------	--------------------

<a href="#">OpenExperimentalFieldStudies</a>	Various experiments and fields studies are performed to generate data which is used in computational models and bigger studies. Capability to capture all the citations of the published data can justify the efforts used in publishing. Citation information can be used to identify all the user communities interested in data source.
<a href="#">LATimes</a>	On 27 March 2014, the LA Times published a story <i>Women earn 83 cents for every \$1 men earn in L.A. city government</i> . It was based on an Infographic released by LA's City Controller, Ron Galperin. This report could only cite data portal of all the resource. It could not cite to exact dataset because tool long URI. <b>It raises the question of granularity of dataset citation.</b>
<a href="#">RDESC</a>	RDESC is a data curator so it uses data from different sources. But this usage is not communicated to data publishers because of lack of such mechanism provided by publishers.

### R-TrackDataUsage

*It should be possible to track the usage of data.* Capability of tracking data usage can help enhancing reputation of the datasets. Records of data usage shows all the successful outcome of the data usage and all the entities associated with it such as the person, organisation, application, research projects that has used these datasets. It increases trust in the data. It also provides provence about how data versions over the time.

**Motivation:** [ASO](#), [LandPortal](#), [LusTRE](#), [OpenExperimentalFieldStudies](#), [RDESC](#), [UKOpenResearchForum](#)

Use Case	R-TrackDataUsage Benefits
<a href="#">ASO</a>	Data is used in decision making process by Water Reservoir Managers. Capability to track usage of data will lead to identification of all the decisions and policy changes made by authorities based on this data. It will also list applications, tools and frameworks suitable for analysis of this kind of data.
<a href="#">LandPortal</a>	Data is used in Research; Policy Making, Journalism; Development;

	Investments; Governance; Food security; Poverty; Gender issues. Usage tracking will help in assessing the impact of published data.
<a href="#">LusTRE</a>	Data is put in public for reuse and reference in nature conservation activities. Information about use of this data will determined impact of this framework. Usage of this data MUST lead to future publications of less heterogenous data and more and more used of standardised thesauri.
<a href="#">OpenExperimentalFieldStudies</a>	Data is used in computational models and studies. Capabilities to track usage of data will enable data publishers to identify all the users communities making use of this data. It will also identify combined use of multiple datasets in one big study. This will identify related datasets which can be recommended to future users.
<a href="#">RDESC</a>	Data is published in Linked Data Format for discovery and recommendations of related datasets. Capability to keep track of its usage will list all the tools and application suitable to be used with this data. Because RDESC is not data publisher but more of a data facilitator, usage tracking will identify highly search dataset and the trends in the temporal, spatial and domain specific search queries.
<a href="#">UKOpenResearchForum</a>	Data is published with “intelligent openness” to support research projects. Capability to track data usage will provide adequate acknowledgement to data originator.

## R-UsageFeedback

*Data consumers should have a way of sharing feedback and rating data.* User feedback is important to address data quality concerns about published dataset. Different users may have different experience with the same dataset so it is important to capture the context in which data was used and the profile of the user who uses it. R-UsageFeedback should also provide a way to communicate suggested corrections and update to the datasets by the users back to data publisher. Data publishers should have a review mechanism to incorporate submitted corrections.

<http://www.w3.org/2013/dwbp/track/issues/159>

**ISSUE:** Feedback can be about other qualified or unqualified opinions written in natural language. Might PROV-O be beneficial here to track exchanges between users?

**OBSERVATION:** Data curation is a common type of feedback mechanism to suggest changes. It could be in the form of natural language, data structures, or a combination of natural language and data structures.

OBSERVATION: Here there are several proposed intersections with the DWBP data quality vocabulary.

**Motivation:** [ASO](#), [DadosGovBr](#), [DataUsage](#), [LusTRE](#), [OKFNTransport](#), [OpenExperimentalFieldStudies](#), [SharePSI \( Share-PSI Feedback, Share-PSI Feedback 2, Share-PSI France \)](#), [RDESC](#)

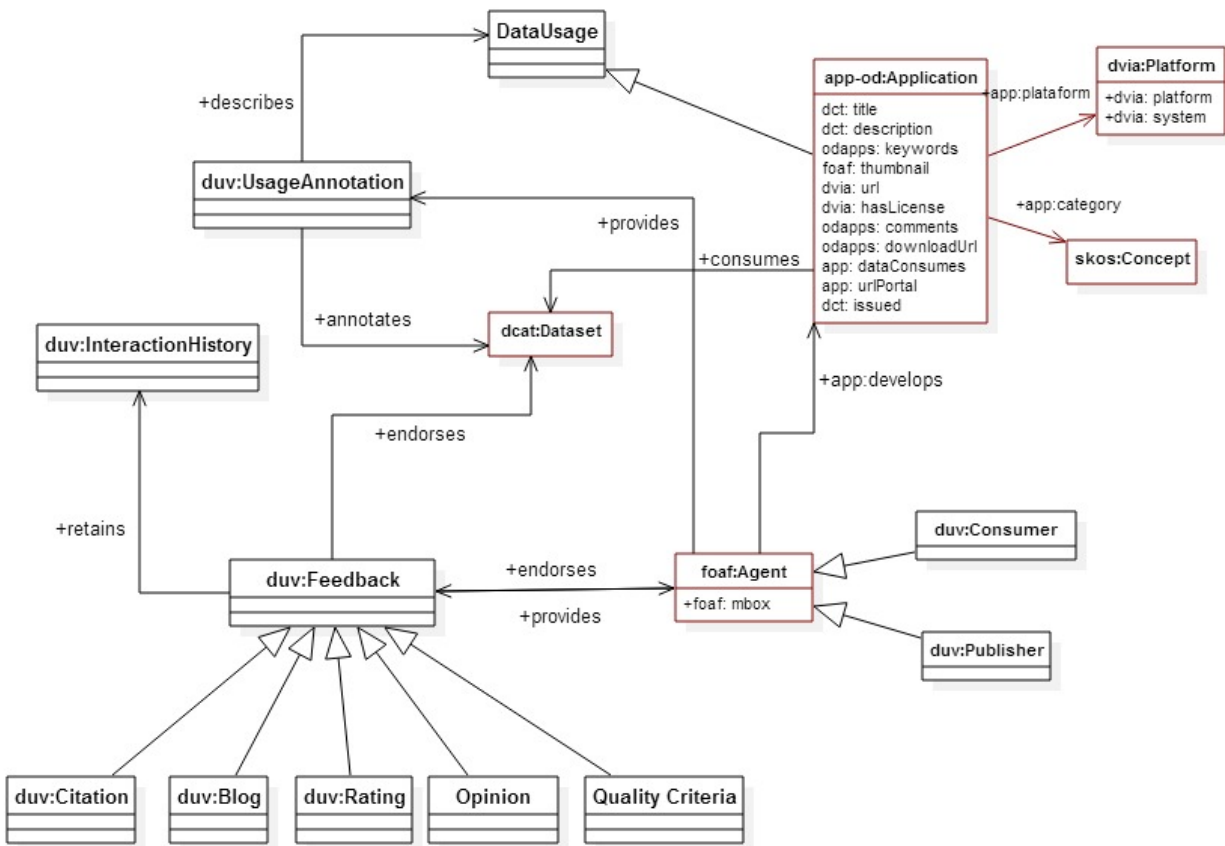
Use Case	R-UsageFeedback Benefits
<a href="#">ASO</a>	Data grows rapidly each year. User feedback can reports issues of data completeness and correctness.
<a href="#">DadosGovBr</a>	Data came from various publishers. As a catalog, the site has faced several challenges, one of them was to integrate the various technologies and formulas used by publishers to provide datasets in the portal. User feedback can provided usabilities of those technologies and formulas. User feedback can be used to crowdsource discrepancies in the vocabularies used to describe datasets.
<a href="#">DataUsage</a>	
<a href="#">LusTRE</a>	Data multilingualism is one of the challenge for this use case. User feedback can be used to crowdsource multilingual text alignment.
<a href="#">OKFNTransport</a>	
<a href="#">OpenExperimentalFieldStudies</a>	Data is used in computational models and studies. User feedback can be used to identify good quality data required for good quality research. completeness, time resolution and usability can be captured using user feedback.
<a href="#">RDESC</a>	RDESC curate different data source and publish metadata in Linked Data Format. User feedback is useful to assess metadata quality. Availability of the source datasets, Correctness of persistent URI, Correctness of the concepts defined in RDESC such as FOAF Agents, Organizations, Physical Properties and Usability of the search interface can be captured in user feedback.

## 4. Related Work

In the development of DUV, many state of the art vocabularies were identified and leveraged including DCAT (<http://www.w3.org/TR/vocab-dcat/>), and the forthcoming Data Quality vocabulary to track feedback in the form of metrics and ratings.

**OBSERVATION:** The relationship between Data Quality and Data Usage is that Data Quality is more about being an annotation describing a dataset and can exist independently of Data Usage. Data Usage “uses” as a tool to describe a usage activity.

## 5. Data Usage at a Glance



**ISSUE:** Is Opinion more of a data usage concept as opposed to a data quality concept?

**ISSUE:** Data usage is parent class for the Application, however there are other types of data usage such as Standard Operating Procedures (SOP), README, or instructions? Could Annotation support this?

**OPERATION:** We are missing `duv:Producer` class.

## 6. Classes and Properties

### 6.1 Classes

#### [foaf:Agent](#)

**Definition:** A person or possibly software.

**Subclass:** A class representing a person performing data usage activities.

**Usage Note:** This class will be used to annotate consumers, producers and publishers either sharing usage information or providing feedback.

#### [app-od:Application](#)

**Definition:** Software application that consumes a dataset.

**Subclass:** `duv:DataUsage`

**Usage Note:** This class represents an application that consumes a dataset.

#### [duv:Blog](#)

**Definition:** The blog class used to capture consumer feedback in natural language form.

**Subclass:** `duv:Feedback`

**Usage Notes:** This class represents a type of feedback.

**Status:** **draft**

## duv:Citation

**Definition:** The citation class used to capture dataset feedback in the form of published material

**Subclass:** `duv:Feedback`

**Usage Notes:** This class represents published material as a type of feedback.

**Status:** **draft**

## duv:Collection

**Definition:** A collection of datasets.

**Subclass:** [dcat:Catalog](#)

**Usage Note:** .

**Source:** <http://www.w3.org/TR/vocab-dcat/>

## skos:Concept

**Definition:** The class `skos:Concept` is the class of SKOS concepts.

**Usage Notes:** A SKOS concept can be viewed as an idea or notion; a unit of thought. However, what constitutes a unit of thought is subjective, and this definition is meant to be suggestive, rather than restrictive.

**Source:** <http://www.w3.org/TR/skos-reference/>

## duv:Consumer

**Definition:** A person or possibly software.

**Subclass:** `foaf:Agent`

**Definition:** This class represents an agent in the role of a data consumer or collaborator.

**Usage Note:** This class will be used to annotate consumers, producers and publishers either sharing usage information or providing feedback.

**Status:** draft

## dcat:Dataset

**Definition:** A collection of data, published or curated by a single agent, and available for access or download in one or more formats.

**Usage:** This class represents the actual dataset as published by the dataset publisher. In cases where a distinction between the actual dataset and its entry in the catalog is necessary (because metadata such as modification date and maintainer might differ), the [\*catalog record\*](#) class can be used for the latter.

**Subclass:** [dctype:Dataset](#)

**Source:** <http://www.w3.org/TR/vocab-dcat/>

## duv:DataUsage

**Definition:** The DUV data usage class used to advertise instructions, notes, and applications that consume dataset.

**Usage Notes:** This class represents recommendations for different data usages available for a given dataset. It could either be provided by the data publisher for advertising data usage or as a mechanism for communities to notify fellow collaborators about usage possibilities.

**Status:** draft

## duv:Feedback

**Definition:** The feedback class used to capture consumer feedback in a variety of forms.

**Usage Notes:** This class is the parent class used to collect feedback in the form of discrete metrics, information, and quality assessments.

**Status:** draft

## duv:InteractionHistory

**Definition:** The interaction history class used to capture a discussion, or feedback history. This is most likely going to be expanded using provenance classes as appropriate.

**Usage Notes:** This class is the parent class used to collect feedback history in the form of discrete metrics, information, and quality assessments.

**Status:** draft

## duv:Opinion

**Definition:** The opinion class represents a qualified opinion used to assess, the dataset or other feedback.

**Subclass:** duv:Feedback.

**Usage Notes:** This class represents a type of feedback.

**Status:** draft Note: DATA QUALITY AND GRANULARITY PLACEHOLDER

## dvia:Platform

**Definition:** Platform that supports software application.

**Subclass:**

**Usage Note:** This class represents a supported platform used by a software application.

Note that the duv:Feedback class is closely tied to ratings, data criteria, and qualified opinions. The definitions (including domain and range) of terms outside the dcat namespace are provided here only for convenience and must not be considered normative. The authoritative definitions of these terms are in the corresponding specifications: [DC11], [FOAF], [RDF-SCHEMA], [SKOS-REFERENCE], [xmlschema-2] and [vcard-rdf].

## duv:Producer

**Definition:** A person or possibly software that produced a dataset

**Subclass:** foaf:Agent

**Definition:** This class represents an agent in the role of a data producer that provides this to a data publisher. In some, perhaps many cases the data producer and publisher will be the same.

**Status:** draft

## duv:Publisher

**Definition:** A person or possibly software responsible for publishing data on the web.

**Subclass:** foaf:Agent

**Definition:** This class represents an agent in the role of a data publisher that makes data accessible to consumers. In some, perhaps many cases the data producer and publisher will be the same.

**Status:** draft

## duv:QualityCriteria

**Definition:** The quality criteria class represents a quality measures used to assess characteristics of the data.

**Subclass:** duv:Feedback.

**Usage Notes:** This class represents a type of feedback.

**Status:** draft Note: DATA QUALITY AND GRANULARITY PLACEHOLDER

## duv:Rating

**Definition:** The rating class represents a type of metric used to assess, the dataset or other feedback.

**Subclass:** duv:Feedback.

**Usage Notes:** This class represents a type of feedback.

**Status:** draft Note: DATA QUALITY AND GRANULARITY PLACEHOLDER

## duv:UsageAnnotation

**Definition:** TBD

**Usage Notes:** TBD

**Status:** draft

## Acknowledgements

TBD

## 8. The Data Usage Vocabulary

TBD

### Important Links

<http://www.w3.org/TR/ld-bp/#vocabulary-creation>

<http://w3c.github.io/data-shapes/data-shapes-primer/no-class-templates>