JournalList.net

Specification for `trust.txt` file and underlying system

Version 1.5
of a Reference Document
Originally Published May 20th, 2020.

Updated version approved July 11th, 2024

Abstract **3**
Introduction **3**
Scope **4**
Provenance **4**
Notices **4**
Document Status 4
Commenting 5
Conduct 5
Statement of Openness 5
Agreement 5
License 5
Other Documents 5
Disclaimer 6
Description of roles **6**
The Specification **7**
Access method 7
File Format 8
File Content 9
File Methods 10
Variable Declaration Records 10
Expiration 10
Limits 10
Where to put it 11
Note about base URI for Publishers on other platforms 11
Note on reliability of signal 11
Note on importance of network effects 12
Examples of Files **12**
Reviewers **14**
References **15**

Abstract

This document standardizes the use of a `trust.txt` file and an underlying system to systematically publish the connection between website Publishers and Associations those Publishers choose to make.

Introduction

The concept of *trust* in journalism and all over the web has been under assault, both from governments trying to destabilize democracy and profiteers who seek financial gain from exploiting the reputation that legitimate journalists have so carefully built over the centuries.

While there are many worthwhile efforts to build up a network of trust, some of which are as old as the publishers themselves, those efforts are largely invisible to search engines, platforms, advertisers, researchers and others. They exist in the offline world as associations, cooperatives, and other affiliations based on commonalities, but modern consumers of current websites don't get all of the benefit of the work that is already being done.

This concept seeks to make those offline networks of trust visible online.

The concept of a `trust.txt` file borrows heavily from two previous very successful efforts improving the overall experience of the internet: `robots.txt` and `ads.txt`. With both, website publishers are able to create a small and very manageable file that they have full control over that helps platforms and advertisers improve the overall ecosystem, and thereby the experience for users. So it will be with `trust.txt`.

Scope

First, It is *not* in the scope of this document to define, for example, what is "truth" or to say who is or is not a journalist.

This is a technical document providing a method for any web publisher or group of publishers to produce and post a small text file on their own website indicating how they chose to associate and detailing what related URLs they publish on, what they choose to disclose about policies regarding things such as generative Artificial Intelligence, and doing so in a way that can be uniformly visible to platforms, advertisers, and other interested parties.

A key attribute is that the file is posted to the web serving system of the Publisher, thus proving that the controllers of that website created the file.

Provenance

This project is a stand-alone specification from JournalList, but much of the groundwork for it came from the Certified Content Coalition, a now-dormant organization. The CCC was an outgrowth of the Innovator In Residence program that was part of the UpRamp suite of initiatives at CableLabs, the research arm of the cable and broadband industry based in Colorado.

Notices

Document Status

This section describes the status of this document as it exists in current form.

Other documents may supersede this document. Contact JournalList staff or the JournalList to be sure you are reviewing the current version.

Status: ~~In-Progress~~ ~~Draft~~ Released ~~Closed~~

This is a released 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 a work-in-progress.

Commenting

Please make all comments by email to JournalList staff. Please refer to the page number and heading for each comment.

Conduct

JournalList is a standalone organization, but in general it operates along guidelines developed by other standards organizations, including the W3C, the IEEE, and AFNOR. All comments will be reviewed by JournalList staff and if any do not meet the general guidelines put forth by other standards organizations in the opinion of the JournalList staff, those comments will be deleted. Membership privileges can be lost

through persistent violation of the fundamental principles of operation or disregard of standards of conduct.

Statement of Openness

For the development of standards, openness and due process are essential.

Openness requires that any person who has, or could be reasonably expected to have an interest, and who meets the requirements of these procedures, will have a right to participate by:

a) Attending Working Group meetings (in person or electronically),
b) Becoming a member of the Working Group,
c) Becoming an officer of the Working Group,
d) Expressing a position and its basis, and,
e) Having that position considered.

Agreement

This technical specification is the result of a cooperative effort undertaken at the direction of JournalList.net, Inc.. You may download, copy, distribute, and reference the documents herein only for the purpose of developing products or services in accordance with such documents, and educational use.

License

Other Documents

This document may contain references to other documents not owned or controlled by JournalList. Use and understanding of this document may require access to such other documents. Designing, manufacturing, distributing, using, selling, or servicing products, or providing services, based on this document may require intellectual property licenses from third parties for technology referenced in this document. To the extent this document contains or refers to documents of third parties, you agree to abide by the terms of any licenses associated with such third-party documents, including open source licenses, if any.

Disclaimer

This document is furnished on an "AS IS" basis and neither JournalList Inc. nor its members provides any representation or warranty, express or implied, regarding the accuracy, completeness, non-infringement, or fitness for a particular purpose of this document, or any document referenced herein. Any use or reliance on the information or opinion in this document is at the risk of the user, and JournalList and its members shall not be liable for any damage or injury incurred by any person arising out of the completeness, accuracy, or utility of any information or opinion contained in the document.

JournalList reserves the right to revise this document for any reason including, but not limited to, changes in laws, regulations, or standards promulgated by various entities, technology advances, or changes in equipment design, or operating procedures described, or referred to, herein.

This document is not to be construed to suggest that any company modify or change any of its products or procedures, nor does this document represent a commitment by JournalList or any of its members to purchase any service or endorse any content whether or not it meets the characteristics described in the document. Unless granted in a separate written agreement from JournalList, nothing contained herein shall be construed to confer any license or right to any intellectual property. This document is not to be construed as an endorsement of any content, product or company or as the adoption or promulgation of any guidelines, standards, or recommendations.

Description of roles

For this document, the following words describe specific roles. While these roles are broadly described, there is an implied trust relationship between organizations that fall into these respective roles. This trust relationship is typically based on a legal agreement executed by the respective parties (for example, a membership agreement or a purchase agreement).

Publisher — Any organization or individual who has control of a website that is generally available to the public on the internet. If part of that site is behind a paywall, that is allowed as long as some information is available at the root domain.

Association — Any group of Publishers, as defined by that group. This may include traditional state-level associations, buying collectives, titles held by one owner, news-sharing organizations, etc. An Association will typically have a membership agreement that a Publisher will need to execute to be considered a member of the Association.

Control — Allowable variables in the file include `control` and `controlledby`. This allows, for example, an ownership group to express in the `trust.txt` file control over online publications, and vice versa. So, The New York Times owns and has control over thewirecutter.com. The Times is a member of the Associated Press, but is *not* controlled by the AP.

Vendor — Any organization that sells products or services to a Customer.

Customer	Any Publisher or Association that purchases products or services from a Vendor. A Customer will typically execute a purchase agreement with the Vendor for products or services provided.
Data Consumer	Any organization or individual who ingests data from any `trust.txt` file. This may take the form of a crawler, a robot, an agent, or any automated script. It may also include human users who want to look at the file using a browser.

This document applies to services that provide resources that clients can access through URIs as defined in RFC3986. For example, in the context of HTTP, a browser is a client that displays the content of a web page.

Crawlers are automated clients. Search engines for instance have crawlers to recursively traverse links for indexing. This specification is not a form of access authorization.

The Specification

This Reference Document specifies a format for encoding instructions in a plain-text file available to Data Consumers. Robots may retrieve these instructions before visiting other URLs on the site. Owners of those robots may use the data to learn about Associations and other affiliations of a web Publisher. The use of that data is completely up to the consumer of that data.

Access method

The use of the “Well Known Uniform Resource Identifiers” is recommended.

(For more information, please see this document: https://tools.ietf.org/html/rfc8615).

All of the instructions in that standard should be followed. In short, in addition to a path to the file residing at the root domain, an additional path or redirect should be included to the file that has the format: `foo.com/.well-known/trust.txt`.

Specifically, publishers should post the "`trust.txt`" file on their root domain with a redirect from "`/.well-known/trust.txt`" to "`/trust.txt`" for backward compatibility. For the purposes of this document the "root domain" is defined as the "public suffix" plus one string in the name. Robots should incorporate Public Suffix list to derive the root domain. It applies to all subdomains.

The declarations must be accessible via HTTP and/or HTTPS from the website that the instructions are to be applied to under a standard relative path on the server host: "`/.well-known/trust.txt`" and HTTP request header containing "`Content-Type: text/plain`". It may be advisable to additionally use "`Content-Type: text/plain; charset=utf-8`" to signal UTF8 support.

It is also advisable to prefer HTTPS connections over HTTP when crawling `trust.txt` files. In any case where data is available at an HTTPS and an HTTP connection for the same URL, the data from HTTPS should be preferred.

If the server response indicates Success (HTTP 2xx Status Code) the Data Consumer is advised to read the content, parse it, and use the declarations.

If the server response indicates an HTTP/HTTPS redirect (301, 302, 307 status codes), the Data Consumer should follow the redirect and consume the data as authoritative for the source of the redirect, if and only if the redirect is within scope of the original root domain as defined above. Up to three redirects are valid as long as each redirect location remains within the original root domain. For example an HTTP to HTTPS redirect within the same root domain is valid.

Any other redirect should be interpreted as an error and ignored.

If the server response indicates the resource is restricted (HTTP 401) the Data Consumer should seek direct contact with the site for authorization keys or clarification. Lacking direct contact, the Data Consumer should assume no declarations are being made under this system.

If the server response indicates the resource does not exist (HTTP Status Code 404), the Data Consumer can assume no declarations exist. For any other HTTP error encountered for a URL which the crawler previously found data, the Data Consumer should assume that previous declarations by the Publisher are no longer valid.

If the `trust.txt` file is unreachable due to server or network errors, this means the file is undefined and the Data Consumer can assume no declarations exist. For example, in the context of HTTP, an unreachable `trust.txt` has a response code in the 500-599

range. For other undefined status codes, the Data Consumer should assume the file does not exist.

Trust URI

A Publisher can place a trust URI of the form “trust://<domain>!” in the HTML of the social network pages they control (identified by the “social=” entries in their trust.txt file). This will advertise their trust.txt file in their social network pages. The corresponding trust.txt file can be retrieved by replacing the “trust://” scheme with “https://”, by removing the trailing character “!”, and by appending “/.well-known/trust.txt” to the path. For example, if the Trust URI is “trust://example.com!” the resulting trust.txt file URL is “https://example.com/.well-known/trust.txt”

When visiting a page with a trust URI, a Data Consumer can fetch the corresponding trust.txt and verify that this page is listed in the retrieved trust file, therefore confirming that the Publisher controlling the origin domain also controls the referenced “social” URI.

See the section on Trust URI Placement Guidelines for information on where to place the Trust URI on various social media platforms.

File Format

The instructions are encoded as a formatted plain text object, described here.

The format logically consists of a non-empty set of records, separated by blank lines, returns, line-feeds or end-of-line command. The records consist of a set of lines of the form:

```
<variable> "=" <value>
```

Comments are allowed anywhere in the file, and consist of optional whitespace, followed by a comment character '#' followed by the comment, terminated by the end-of-line.

File Content

Not all of the lines here need to be used, but all of them are available. All Fields can be used more than once with the exception of `controlledby` and `datatrainingallowed`, which should be used only once. For instance, an Association will in nearly all cases have many members. Each one will get its own line in the file. All Data Consumers should store all valid data with each URI.

Note that there is no distinction in the file between Publishers and Associations. This is by design. An organization may both have members, and be a member of other Associations.

Field	Valid Data	Notes
`member`	URL	Included here will be the URL for a member of any Association. One line for each member.
`belongto`	URL	This is the place to list an Association or other organization that a Publisher may belong to. One line for each organization.
`control`	URL	A domain directly controlled by one entity. For use by ownership groups or other similar organizational units. One line for each organization controlled.
`controlledby`	URL	Domain of owner or other controlling entity. There should be only one listing for the controlling organization.
`social`	URI	Any social media account directly controlled by the Publisher.
`vendor`	URL	Included here will be the URI for a Vendor to any Association or Publisher. One line for each Vendor.
`customer`	URL	Included here will be the URI for a Customer to any Vendor. One

		line for each Customer.
`disclosure`	Directory on base URI	If a Publisher has, for example, an ethics policy, it can publish the URI for that.
`contact`	Contact information that can be in any form, including physical or email addresses, a URI, etc.	As part of full transparency, Publishers or Associations may want to associate contact data so that people who are part of Data Consumer organizations can make contact with questions.
`datatrainingallowed`	“yes” or “no”	This is a directive to any scraper from an AI, a large language model, or any other tool designed to collect data from the site of the publisher to be used in forms other than referring users to the site of origin. A “yes” reply means that tools can scrape the data without restriction. A “no” means that a tool can not do that without a legally binding contract in place before collecting any data.

File Methods

Variable Declaration Records

Any line containing a pattern of <VARIABLE>=<VALUE> should be interpreted as a variable declaration and the crawler or robot should store the data associated with the root domain.

The <VARIABLE> is a string identifier without internal whitespace. The only supported separator is the equals sign ‘=’. The <VALUE> is an open string that may contain arbitrary data.

The declaration line is terminated by the end-of-line marker. The Data Consumer should liberally interpret CR, CRLF, etc., as a line separator. For human readability it is recommended that variables be declared at the end of the file, but this is not a strict requirement and should not be assumed.

Expiration

Data Consumers may independently store files, but if they do it is recommended that they regularly verify their own cache. Standard HTTP cache-control mechanisms can be used by both origin server and robots to influence the caching of the `trust.txt` file. Specifically consumers and replicators should take note of `HTTP Expires` header set by the origin server.

Limits

Crawlers may impose a parsing limit, but it is recommended that the limit be at least 500 kibibytes (KiB).

Where to put it

As detailed above, in the `well-known` directory, and in the top-level directory of your web server for backward compatibility.

When a robot looks for the "`/trust.txt`" file for URL, it strips the path component from the URL (everything from the first single slash), and puts "`/trust.txt`" in its place.

For example, for "`http://www.example.com/news/index.html`," it will remove the "`/news/index.html`" and replace it with "`/trust.txt`", and will end up with "`http://www.example.com/trust.txt`".

So, as a web site owner you need to put it in the right place on your web server for that resulting URL to work. Usually that is the same place where you put your web site's main "index.html" welcome page. Where exactly that is, and how to put the file there, depends on your web server software.

Remember to use all lower case for the filename: "`trust.txt`", not "`Trust.TXT`."

Note about base URI for Publishers on other platforms

There are some Publishers who work on a platform for which they do not control the base URI, for example a video Publisher working exclusively on YouTube. In that case, the Publisher would be advised to create a single-page website. The traditional

public-facing homepage of that site can be just a link to the YouTube page. Then a `trust.txt` page can be placed on that page.

Another example is a local, independent newsroom that is part of a chain that uses one URL, for example www.bizjournals.com. In that case it will be up to the Publisher to have all `trust.txt` information in one file, or to set up individual URIs for each publication. Either one is acceptable.

Note on reliability of signals

Search engines and social networks never reveal exactly what goes into their algorithms for determining search engine results or placement in a social feed. To reveal that would be an invitation to fraudulent manipulation. That said, they clearly rely on "signals" from pages, and from the way those pages are referenced and used. The existence of this Reference Document is designed in large part to create a new and useful signal.

That said, while the intention of the `trust.txt` system is to increase the trust of legitimate Publishers, the existence of a file on a site should not be regarded *a priori* as a signal of trust on its own. Also, the lack of a `trust.txt` file should not on its own be regarded as a negative indicator.

The platforms and social networks must weigh for themselves the trustworthiness of any individual Publisher or Association.

The goal of this Reference Document and the underlying framework is to make the inference of trust by affiliation much more accessible and scalable.

Note on importance of network effects

While every organization publishes a `trust.txt` file completely at its own discretion, the importance of networked connections is vital to make the signal valuable to algorithms assessing the value of the information in the file.

In other words, if a publisher says that it belongs to an association, but that association does not publish a `trust.txt` file confirming that the publisher is indeed a member, the strength of that signal may be lost, or even become negative. If a publisher feels participation in an association is a positive signal, that organization should strongly encourage the association to publish its own `trust.txt` file.

Examples of Files

These examples (created by JournalList, so data is not accurate) are examples of files that would be generated by individual organizations that would be placed on their own URL and controlled by them.

This the file that might be created by a Publication, *The Durango Herald*, of Colorado:

```
#Durango Herald trust.txt file from Ballantine Communications Inc.
belongto=https://coloradopressassociation.com
belongto=https://www.ap.org/
belongto=https://www.journallist.net/
control=http://www.adventurepro.us/
control=http://www.directoryplus.com/
control=http://www.doradomagazine.com/
control=http://www.dgomag.com/
control=http://the-journal.com/
control=http://pinerivertimes.com/
datatrainingallowed=no
social=https://facebook.com/TheDurangoHerald
social=https://twitter.com/durangoherald
social=https://instagram.com/durango_herald
social=https://www.youtube.com/channel/UCSfC3ozxDs8aOVDaMnaUAQA
contact=https://durangoherald.com/contact_us/staff
```

This is the file that might be created by a Publication that is owned by the owner of *The Durango Herald*, but does not have any other association memberships. This example, taken from the *Herald*'s file, is called *Adventure Pro Magazine*:

```
#Adventure Pro Magazine trust.txt file from Ballantine Communications Inc.
controlledby=http://www.durangoherald.com/
datatrainingallowed=no
social=https://www.facebook.com/AdventureProMag
social=https://twitter.com/AdventureProMag
social=https://www.instagram.com/adventurepromagazine/
social=https://www.youtube.com/channel/UCm0EL3_uRC6BFBtCud8mw7Q
social=https://flipboard.com/@AdventurePro
contact=https://adventurepro.us/about/
```

The Durango Herald could place the following trust URI on its X (Twitter) page (e.g., in the bio of its "durangoherald" account): "trust://durangoherald.com!". A Data Consumer visiting the durangoherald X/Twitter account would then be able to verify that this X/Twitter account is listed in the Durango Herald's trust.txt file.

This is the (shortened) file that might be created by an Association, The Colorado Press Association:

```
#CPA trust.txt file
belongto=http://newspapers.org/
belongto=https://www.nammembers.com/
belongto=https://coloradofoic.org/
belongto=https://www.journallist.net/
member=http://www.akronnewsreporter.com/
member=http://www.alamosanews.com/
member=http://www.theflume.com/
member=http://arvadapress.com/
member=http://www.aspendailynews.com/
member=http://www.aspentimes.com/
member=http://www.hightimbertimes.com/
member=http://www.aurorasentinel.com/
datatrainingallowed=yes
social=https://www.facebook.com/coloradopressassociation/
social=https://twitter.com/ColoradoPress
social=https://www.linkedin.com/company/colorado-press-association/
social=https://www.youtube.com/channel/UCDHXPIQtH1ze7UM3aT8ivKA/
contact=https://coloradopressassociation.com/contact/
```

This is the (shortened) file that might be created by the Associated Press:

```
#Associated Press trust.txt file
belongto=https://iptc.org/
belongto=https://journallist.net/

member=https://www.hearst.com/
member=https://scripps.com/
member=https://www.jsonline.com/
member=https://www.swiftcom.com/
member=https://www.spokesman.com/
member=https://www.nytimes.com/
member=https://www.ogdennews.com/

social=https://www.facebook.com/APNews
social=https://www.instagram.com/apnews/
social=https://twitter.com/ap
social=https://www.linkedin.com/company/associated-press
```

```
social=https://www.youtube.com/ap

contact=https://www.ap.org/contact-us/
```

Trust URI Placement Guidelines

The trust.txt specification is platform agnostic. In order to improve interoperability one SHOULD follow these guidelines when creating Trust URI entries on the following platforms.

- **Facebook**: Trust URI `https://www.facebook.com/<accountname>` on the Facebook account page: in the account page's Intro field.
- **GitHub**: Trust URI `https://github.com/<accountname>` on the account page: in the GitHub account page's Bio field.
- **Instagram**: Trust URI `https://www.instagram.com/<accountname>` on the account page: in the Instagram account page's Bio field.
- **LinkedIn**: Trust URI `https://www.linkedin.com/<type>/@<accountname>/`, where `<type>` is `in` (for individuals), `school` or `company`, on the LinkedIn account page: for individuals - in the LinkedIn account page's About field; for schools and companies - in the account page's Overview section. This will appear on the About page (`<url>/about/`) of the account.
- **Medium**: Trust URI either the default `https://medium.com/@<accountname>/` form or the subdomain `https://<accountname>.medium.com` form (depending on the owner's Medium account settings) on the account page: preferably in the Medium account page's Bio field, optionally in the Medium About page's description.
- **Rumble**: Trust URI `https://rumble.com/c/<accountname>` for individuals, in the user's Rumble channel Description field (per channel). This will appear on the About page (`<url>/about/`) of the account.
- **Telegram**: Trust URI `https://t.me/<accountname>` for accounts - in the Telegram account bio field; for channels - in the Telegram channel description field.
- **Threads**: Trust URI `https:///www.threads.net/@<accountname>` in the Threads account page's Bio field.
- **TikTok**: Trust URI `https://www.tiktok.com/@<accountname>` on the TikTok account page, in the account page's Bio field.

- **Vimeo**: Trust URI `https://vimeo.com/c/<accountname>` on the Vimeo account page, in the account page's Bio field.
- **X (Twitter)**: Trust URI `https://twitter.com/c/<accountname>` on the X (Twitter) account page, in the account page's Bio field.
- **YouTube**: Trust URI `https://youtube.com/@<accountname>` on the YouTube account page, in the About page's description. This will appear on the About page (`<url>/about/`) of the account.

Reviewers

Many thanks to these people who reviewed this document. (Reviewing does not imply endorsement):

Claire Wardle, First Draft News; Ralph Brown; John Daniszewski, Heather Edwards, Associated Press; Tom Brand, NAFB; Justin Sasso, Colo. Association of Broadcasters, Mickey Osterreicher, NPPA; Bill Skeet and Cedar Milazzo, Trustie; Sandro Hawke, W3C fellow; Jill Fraschman, Colo. Press Association; Connie Moon Sehat, NewsQ/Credibility Coalition; Gabriel Altay, Kensho; Sean La Roque-Doherty lawyer, writer and IEEE P.7011 participant; Andres Rodriguez, Instituto Tecnológico de Buenos Aires (ITBA) and IEEE P.7011 participant; Brendan Quinn, IPTC; Scott Cunningham original ads.txt advisor; Ed Bice, Scott Hale and Megan Marrelli, Meedan; Jason Kint, Chris Pedigo, Digital Content Next; Kati London, Microsoft.

UPDATE of additional reviewers for Version 1.3: The board of JournalList, including Scott Yates, Claire Wardle, Ralph Brown, Randy Picht, and Susan Kantor. Other reviewers of the changes included Brendan Quinn, IPTC; Thad Swiderski, eTypeServices; Michael W. Kearney, Ph.D., AI & Digital Media Expert; Kenny Katzgrau, CEO, BroadStreet Ads; Laura Ellis and Chris Needham, BBC.

References

RFC 9309 Robots. Exclusion Protocol https://www.rfc-editor.org/rfc/rfc9309.html

Robots.txt about page https://www.robotstxt.org/robotstxt.html

Ads.txt
https://iabtechlab.com/wp-content/uploads/2019/03/IAB-OpenRTB-Ads.txt-Public-Spec-1.0.2.pdf

W3C recommendations on robots.txt
https://www.w3.org/TR/html4/appendix/notes.html#h-B.4.1.1

IETF RFC 8615 - Well-Known Uniform Resource Identifiers (URIs)
-https://datatracker.ietf.org/doc/html/rfc8615

IANA's list of registered Well-Known URIs
https://www.iana.org/assignments/well-known-uris/well-known-uris.xhtml

Schema.org

International Press Telecommunications Council
https://iptc.org/standards/