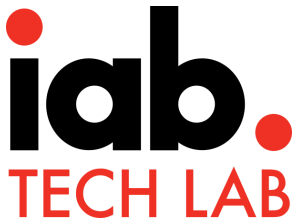

IAB Tech Lab
Ads.txt Specification Version 1.0

Final

June 27, 2017

About Ads.txt

The Ads.txt Specification was developed in the spring of 2017. This document is the final Ads.txt Specification Version 1.0; a peer-reviewed standard developed with the support of the OpenRTB working group. This document is available at https://iabtechlab.com/ads-txt.

About IAB Tech Lab

The IAB Technology Laboratory is a nonprofit research and development consortium charged with producing and helping companies implement global industry technical standards and solutions. The goal of the Tech Lab is to reduce friction associated with the digital advertising and marketing supply chain while contributing to the safe growth of an industry.

The IAB Tech Lab spearheads the development of technical standards, creates and maintains a code library to assist in rapid, cost-effective implementation of IAB standards, and establishes a test platform for companies to evaluate the compatibility of their technology solutions with IAB standards, which for 18 years have been the foundation for interoperability and profitable growth in the digital advertising supply chain. Further details about the IAB Technology Lab can be found at: https://iabtechlab.com/

Authors:
Neal Richter, Consulting Data Scientist, Hebbian Labs, and IAB Tech Lab OpenRTB Co-Chair
George Levitte, Product Manager, Google

Other Significant Contributions Include:
Per Bjorke, Senior Product Manager, Google; Drew Bradstock SVP, Product, Index Exchange; Jim Butler VP Engineering, Publisher Platforms, AOL; Andrew Casale President & CEO, Index Exchange; Sam Cox, Group Product Manager, AdX BuySide and Policy, Google; Allen Dove CTO, SpotX; Alanna Gombert, SVP, Technology & Ad Operations, IAB, General Manager, IAB Tech Lab; Tamer Hassan, CTO, White Ops; Dan Kaminsky, Chief Scientist and Founder, White Ops; Curt Larson, VP Product, Sharethrough; Pierre Nicolas, Product Manager, Criteo; Rachel Nyswander Thomas, Vice President, Operations and Policy, Trustworthy Accountability Group (TAG); Shree Madhavapeddi, Product Manager, Google; Brian O'Kelley, CEO, AppNexus; Bill Simmons, CTO, DataXu; Jud Spencer, Principal Lead Software Engineer, The Trade Desk; Scott Spencer, Director of Sustainable Ads, Google; Sam Tingleff VP, Buyer Engineering, Rubicon Project; Ian Trider Director, RTB Platform Operations, Centro; Mike Zaneis, CEO, Trustworthy Accountability Group (TAG)

IAB Tech Lab Contact:
Jennifer Derke, Director of Product, Programmatic & Data, IAB Tech Lab
openRTB@iabtechlab.com

Table of Contents

1. ABSTRACT

As part of a broader effort to eliminate the ability to profit from counterfeit inventory in the open digital advertising ecosystem, provide a mechanism to enable content owners to declare who is authorized to sell their inventory.

2. INTRODUCTION

For brevity, we'll assume readers are already familiar with the problem of fraud in ad tech and its vast scale [1][2][3]. Fraud can come in various forms, here we are concentrating on the form wherein ad inventory is being offered to buyers with a misrepresented label and account during the real-time bidding process. Typically the domain of the webpage, or the ID of the mobile app has been falsified to look like a site or app they do not have authorization to sell.

Here, we propose a new standard to enable content owners to explicitly declare a set of advertising systems and resellers who are authorized to sell their inventory. This will enable buyers to acquire advertising space through safe supply chains via authorized entities.

2.1 CHANGE LOG

Version	Date	Changes
1.0	June 27, 2017	First Version

3. SPECIFICATION

This memo specifies a mechanism for publisher content distributors to publicly declare their authorized advertising systems and identifiers within those systems. It also describes the format for encoding the instructions to be consumed by advertising systems and their customers. Advertising systems should retrieve these declarations before buying or selling advertising claiming to be on the website.

This specification is specifically inspired by the robots.txt standard [5][6]. A key attribute is that the file is posted to the web serving system of the content, thus proving that the website authored the file. We refer the reader to various other advertising API specifications such as

IAB Tech Lab's OpenRTB [7] and Google's AdX API [8] for real-time ad space sales and IAB Tech Lab's OpenDirect [9] for non real-time sales.

3.1 ACCESS METHOD

Publishers should post the "/ads.txt" file on their root domain and not on subdomains.

The declarations must be accessible via HTTP and/or HTTPS [2] from the website that the instructions are to be applied to under a standard relative path on the server host: "/ads.txt" and 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 fetching ads.txt files.

For convenience we will refer to this resource as the "/ads.txt" file, though the resource need in fact not originate from a file-system.

If the server response indicates Success (HTTP 2xx Status Code,) the advertising system must read the content, parse it, and utilize the declarations.

Other HTTP status codes such as HTTP 302 redirects are not supported, a future version may address delegating authority to a third-party web server to publish the data.

If the server response indicates the resource is restricted (HTTP 401) the advertising system should seek direct contact with the site for authorization keys or clarification.

If the server response indicates the resource does not exist (HTTP Status Code 404), the advertising system can assume no declarations exist and that no advertising system is unauthorized to buy and sell ads on the website.

3.2 FILE FORMAT

The instructions are encoded as a formatted plain text object, described here. A complete description of the syntax of this format is given in section 3.4 below.

The format logically consists of:

- A non-empty set of records, separated by line breaks. The records consist of a set of lines of the form:

  <FIELD #1>, <FIELD #2>, <FIELD #3>, <FIELD #4>

- Lines starting with # symbol are considered comments and are ignored.

Refer to section 3.4 for more syntax details.

3.3 THE RECORD

The following defines the contents within each field. We refer to the IAB OpenRTB [7] and IAB OpenDirect [9] specs as needed.

FIELD	NAME	DESCRIPTION
Field #1	Domain name of the advertising system	(Required) The canonical domain name of the SSP, Exchange, Header Wrapper, etc system that bidders connect to. This may be the operational domain of the system, if that is different than the parent corporate domain, to facilitate WHOIS and reverse IP lookups to establish clear ownership of the delegate system. Ideally the SSP or Exchange publishes a document detailing what domain name to use.
Field #2	Publisher's Account ID	(Required) The identifier associated with the seller or reseller account within the advertising system in field #1. This must contain the same value used in transactions (i.e. OpenRTB bid requests) in the field specified by the SSP/exchange. Typically, in OpenRTB, this is publisher.id. For OpenDirect it is typically the publisher's organization ID.
Field #3	Type of Account/ Relationship	(Required) An enumeration of the type of account. A value of 'DIRECT' indicates that the Publisher (content owner) directly controls the account indicated in field #2 on the system in field #1. This tends to mean a direct business contract between the Publisher and the advertising system. A value of 'RESELLER' indicates that the Publisher has authorized another entity to control the account indicated in field #2 and resell their ad space via the system in field #1. Other types may be added in the future. Note that this field should be treated as case insensitive when interpreting the data.
Field #4	Certification Authority ID	(Optional) An ID that uniquely identifies the advertising system within a certification authority (this ID maps to the entity listed in field #1). A current certification authority is the Trustworthy Accountability Group (aka TAG), and the TAGID would be included here [11].

Note that if a parent company is operating multiple distinct SSP/Exchanges, the domain in field #1 should refer to the domain of the RTB connection that the bidder is receiving bid requests from.

3.4 SYNTAX DEFINITION

3.4.1 THE RECORD

The core syntax is a comma separated format with three defined fields and one record per line.

The consumer systems should ignore any sequence of whitespace or tabs. If the data is obviously corrupted or malformed the contents of the file should be ignored. No field should contain tabs, commas or whitespace, otherwise it should be escaped with URL encoding [13].

Individual records are separated by an end-of-line marker. The consumer systems should liberally interpret CR, CRLF etc as a record separator.

The allowed identifiers in field #1 and by definition assumed to be valid DNS domain names obeying RFC 1123 [10], associated errata for RFC 1123 or subsuming RFCs.

Identifiers in field #2 can be strings or integers. For reference OpenRTB's publisher.id [14] is a string field.

3.4.2 COMMENTS

Comment are denoted by the character "#". Any line containing "#" should inform the data consumer to ignore the data after the "#" character to the end of the line.

3.4.3 EXTENSION FIELDS

Extension fields are allowed by implementers and their consumers as long as they utilize a distinct final separator field like ";" before adding extension data to each record. Extensions should be placed after the comment character.

3.5 EXPIRATION

Consuming systems of /ads.txt should cache the files, but if they do they must periodically verify the cached copy is fresh before using its contents.

Standard HTTP cache-control mechanisms can be used by both origin server and robots to influence the caching of the /ads.txt file. Specifically consumers and replicators should take note of HTTP Expires header set by the origin server.

If no cache-control directives are present consuming systems should default to an expiry of 7 days.

4. EXAMPLES

As defined above there are three required fields. The optional certification authority ID field is included in some of the examples.

4.1 SINGLE SYSTEM DIRECT

The first example is a website with only one authorized system that is directly controlled/operated by the website owner.

http://example.com/ads.txt

```
greenadexchange.com, XF7342, DIRECT, 5jyxf8k54
```

4.2 SINGLE SYSTEM RESELLER

The second example is a website with only one authorized system that is operated by a separate company doing resale of inventory. Their advertising system has not been independently certified, so no the optional fourth field is omitted.

http://example.com/ads.txt

```
redssp.com, 57013, RESELLER
```

4.3 MULTIPLE SYSTEMS AND RESELLERS

The third example is a website with multiple authorized systems and multiple resellers. Some of their authorized advertising systems are independently certified and have an ID issued.

http://example.com/ads.txt

```
# Ads.txt file for example.com:
greenadexchange.com, 12345, DIRECT, d75815a79
silverssp.com, 9675, RESELLER, f496211
blueadexchange.com, XF436, DIRECT
orangeexchange.com, 45678, RESELLER
silverssp.com, ABE679, RESELLER
```

5. IMPLEMENTER'S NOTES

5.1 VERSION

This is version 1.0 of the specification and every attempt will be made to make future versions backward compatible if possible.

5.2 GUIDANCE BY PARTY

5.2.1 SSP/EXCHANGE

SSPs and exchanges should decide which canonical domain they wish to be used in field #1. They should make documentation available to publishers and DSPs. Publisher-facing documentation should indicate how publishers can retrieve the appropriate ID for field #2. DSP-facing documentation should indicate which field in bid requests should be used by DSPs for checking against the ads.txt file.

Ideally, implementing SSPs/exchanges should provide a tool to generate the exact lines for a publisher to place in the ads.txt file. SSPs/exchanges should also consider crawling publishers' domains and notifying publishers (e.g. a warning in the publisher dashboard, e-mail, etc.) of the absence of an ads.txt file or the absence of appropriate declarations in the file.

5.2.2 DSP

DSPs should consult documentation provided by SSPs/exchanges as to the canonical domain used by the exchange (field #1) and the appropriate field in bid requests to be checked against ads.txt (field #2).

5.2.3 PUBLISHERS

Publishers should consult documentation provided by SSPs/exchanges as to the canonical domain used by the exchange (field #1) and the appropriate ID to place in field #2.

5.3 INTEROPERABILITY

Implementors should pay particular attention to the robustness in parsing of the /ads.txt file. It is expected that the /ads.txt files are created with automated systems or manual platform-specific text editors consumers of the data should be liberal in accepting files with different end-of-line conventions, specifically CR and LF in addition to CRLF and varying whitespace or field separation characters.

5.4 SECURITY

The /ads.txt declarations are retrieved and applied in separate, possibly unauthenticated HTTP transactions, and it is possible that one server can impersonate another or otherwise intercept a request for /ads.txt, and provide a consuming system with false information.

If this is a worry then the website owner should redirect unsecure http requests to https requests for the /ads.txt file.

5.5 SUBDOMAINS

Subdomains are currently not supported. This is an open item that will be discussed in future iterations of the spec.

When writing spiders, implementers should request the /ads.txt from the domains that are driving significant requests for advertising. Publishers should post the /ads.txt file on their root domain and not on subdomains. The crawler should strip the subdomains when creating the crawler's URL list. The public suffix list [12] should be utilized in implementing subdomain stripping.

5.6 ADS.TXT CRAWLERS

An reference implementation of an ads.txt data crawler can be found here on github [15].

Crawlers that may want to crawl publisher content beyond ads.txt and read and display publisher content, advertisements, and connected metadata may do so if not prohibited by the instructions in robots.txt.

6. SCOPE AND FUTURE DIRECTIONS

6.1 SCOPE

Scope of this initial version of this standard is to define a mechanism to define authorized sellers for web content from the perspective of the domain owner, for the purpose of addressing some of the fraud scenarios related to counterfeit inventory.

6.2 OPEN ISSUES

Open issues to be considered for resolution in a future version of the spec should be brought to attention via contacting openrtb@iabtechlab.com or on the openrtb-dev@googlegroups.com mailing list.

6.3 FUTURE DIRECTIONS

Future directions include covering mobile apps and other non-web environments, allowed ad formats, subdomains and delegating authority to a third party.

7. ACKNOWLEDGEMENTS

The authors would like to thank the original authors of the robots.txt [5][6] file for providing inspiration. We would also like to thank numerous people within the IAB Tech Lab, TAG and multiple companies for their comments on the early drafts and supporting the initiative.

8. REFERENCES

1. https://techcrunch.com/2016/01/06/the-8-2-billion-adtech-fraud-problem-that-everyone-is-ignoring/
2. http://adage.com/article/digital/ana-report-7-2-billion-lost-ad-fraud-2015/302201/
3. http://boingboing.net/2016/12/21/methbot-a-3m-5mday-video-a.html
4. https://www.emarketer.com/Article/Ad-Industrys-Focus-on-Fraud-Has-Intensified/1014430
5. http://www.robotstxt.org/norobots-rfc.txt
6. https://en.wikipedia.org/wiki/Robots_exclusion_standard
7. https://www.iab.com/guidelines/real-time-bidding-rtb-project/
8. https://developers.google.com/ad-exchange/rtb/downloads
9. https://www.iab.com/guidelines/iab-opendirect-specification/
10. https://tools.ietf.org/html/rfc1123
11. https://www.tagtoday.net
12. https://publicsuffix.org/
13. https://www.w3schools.com/tags/ref_urlencode.asp
14. http://www.iab.com/wp-content/uploads/2016/11/OpenRTB-API-Specification-Version-2-5-DRAFT_2016-11.pdf
15. https://github.com/InteractiveAdvertisingBureau/adstxtcrawler