IAB Tech Lab OpenRTB
Ads.txt Public Spec – Public Comment Draft

This Draft Document is in Public Comment from May 17 until June 19, 2017. Please send your comments to OpenRTB@iabtechlab.com

This document has been developed by the OpenRTB Working Group.

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

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
Jennifer Derke, Director of Product, Programmatic and Data, IAB Tech Lab
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 Nicholas Product Manager, Criteo
Rachel Nyswander Thomas, Vice President, Operations and Policy, Trustworthy Accountability Group (TAG)
Shree Madhavapeddi, Product Manager, Google
Bill Simmons, CTO, DataXu
Sam Tingleff VP, Buyer Engineering, Rubicon Project
Ian Trider Director, RTB Platform Operations, Centro
Mike Zaneis, CEO, Trustworthy Accountability Group (TAG)

TABLE OF CONTENT

1. ABSTRACT

As part of a broader effort to eliminate the ability to profit from counterfeit inventory in the open digital advertising ecosystem, Ads.txt provides 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.

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

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: "/ads.txt" and HTTP request header containing "Content-Type: text/plain".

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 obey the declarations.

Other HTTP status codes such as redirects should be obeyed until a resource is found.

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 is forbidden (HTTP 403) the advertising system should interpret the directive that no advertising is currently authorized.

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 or records, separated by line breaks. The records consist of a set of lines of the form:

<FIELD> <SEPARATOR> <FIELD> <SEPARATOR> <FIELD> <SEPARATOR> <FIELD>

3.3 THE RECORD

The following defines the contents within each field. We refer to the IAB OpenRTB [7], Google AdX [8] 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. This should be the operational domain of the system, if that is different than the corporate domain, to facilitate WHOIS and reverse IP lookups. To establish clear ownership of the delegate system.
Field #2	Seller 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 real-time bid request callouts within IAB OpenRTB 'Publisher.ID' or the Google AdX 'seller_network_id' fields. For

		IAB OpenDirect it is 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 website (content owner) directly operates the account on the system in field #1. A value of 'RESELLER' indicates that the website has authorized the system in field #1 to operate the account and resell ad space. Other types may be added in the future.
Field #4	Certification Authority ID	(Optional) An ID that uniquely identifies the advertising system within a certification authority. A current certification authority the Trustworthy Accountability Group (aka TAG), and the TAGID would be included here [11].

3.4 SYNTAX DEFINITION

We are avoiding creating a BNF-like formal description here.

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

The consumer systems should liberally interpret any sequence of whitespace, tabs and commas as a single field separator. If the data is obviously corrupted or malformed the contents of the file should be ignored. No field should contain an embedded field separator (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.

Allowed identifiers in field #2 are typically integers, yet may vary and be strings. For reference, OpenRTB's Publisher.ID is a string field [14] and AdX's seller_network_id is an int32 field [15].

Comment lines are denoted by the character "#". Any line containing "#" should inform the consumer to ignore the entire line and data within it.

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.

Additional extensions should be enclosed within comment lines in a syntax defined by the extender and consumers of the extension.

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 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. 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

```
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 INTEROPERABILITY

Implementers 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.

5.3 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.

5.4 SUBDOMAINS

When writing spiders, implementers should request the /ads.txt from the domains that are driving significant requests for advertising.   Typically, it will be sufficient to truncate the domain to having one domain under the public suffix list [12], however implementers may find that some subdomains will have unique advertising systems installed for example http://myblog.someblogplatform.com/ads.txt, and as such implementers may need to check these URLs as well.

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 the first version of the specification. The purpose of the 60 day public comment period is to address them and listen to new ideas.

6.3 IDEA: SUBDOMAINS?

A publisher who owns large number of domains and like to specify the same list of authorized sellers for some or all of those domains - is there a way to make it easy to "inherit" or "reuse" an ads.txt across domains while maintaining the authorization aspect that's inherent in the fact that the ads.txt is hosted on the specific domain? Conversely the domain may have a parent domain to refer to for a canonical list.

One idea is to refer to other related and/or included ads.txt for consumption by robots.

```
include: http://subdomain1.example.com/ads.txt
include: http://subdomain2.example.com/ads.txt
include: http://example.com/ads.txt
related: http://sisterexample.com/ads.txt
```

6.4 IDEA: RESELLER BY REFERENCE?

Will this work if the publisher trusts the exchange to ONLY sell the Pubs inventory from a controlled list of exchanges? If this is part of the spec the Pub must trust the exchange, and the buyer has to keep track of this and decide if they trust the exchange.

```
include_reseller:  http://silverexchange.com/ads.txt
```

6.5 FUTURE DIRECTIONS

Future directions include covering mobile apps and other non-web environments, allowed ad formats, etc.

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, 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://developers.google.com/ad-exchange/rtb/downloads/realtime-bidding-proto.txt