Network Working Group
Internet-Draft
Intended status: Informational
Expires: 17 July 2026

B.A. Fisher
DPA R&D

13 January 2026

UZPIF Outbound Indexing for Search Engines and AI
draft-dpa-uzpif-outbound-indexing-00

Abstract

This document proposes an outbound, opt-in mechanism for web content discovery and indexing, complementing or replacing traditional inbound crawling models such as those governed by the Robots Exclusion Protocol (REP; [RFC9309]). In the proposed approach, servers proactively initiate authenticated outbound connections to trusted indexers (search engines or AI systems) using identity-bound grants, enabling explicit consent for indexing, freshness signalling, and content usage policy communication.

The mechanism integrates with identity-centric frameworks such as the Universal Zero-Port Interconnect Framework (UZPIF; [UZPIF]) and supports both traditional search engines and AI-driven indexing and retrieval systems. It aims to reduce unsolicited crawling abuse, improve signal quality for indexers, and provide site owners with positive control over discoverability in an era of increasing AI content consumption.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 17 July 2026.

Copyright Notice

Copyright (c) 2026 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (https://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

Table of Contents

1. Scope and Status

This document is an Internet-Draft and represents work in progress. It is published to enable structured technical review, interoperability discussion, and disciplined specification development around outbound, consent-first indexing mechanisms for UZPIF-style transports.

The material is a research artefact. It does not claim technical completeness, production readiness, or endorsement by the IETF or any other standards body, and it is not presented as a standards-track specification.

It is designed for experimentation, operator feedback, and profile-driven deployments. It does not require changes to the HTTP protocol, but it can carry or reference HTTP-origin content.

During conversion from internal research documents into IETF XML, care has been taken to:

* preserve a clear distinction between normative and informative content;

* use requirement language (e.g., "MUST", "SHOULD", "MAY") only where behaviour is intentionally specified;

* avoid any implication of registry finalisation, mandatory implementation, or standards-track status; and

* maintain intellectual-property neutrality, with no implied patent grants or licensing commitments beyond the IETF Trust copyright licence applicable to Internet-Draft text.

Ongoing research, implementation, performance validation, and real-world pilot work remain outside the scope of this Internet-Draft text and may be pursued separately.

2. Executive Summary

This document defines an outbound indexing model in which content publishers (servers) initiate outbound, authenticated sessions to trusted indexers to advertise content availability, request refresh, and communicate explicit usage constraints. The model is intended as a complement to inbound crawling and robots.txt-based opt-out signalling.

The core components are:

* *Outbound Discovery Session:* A publisher-initiated secure session to an indexer service, established over UZPIF and secured using identity-bound handshakes (e.g., TLS-DPA).

* *Discovery Grant:* An identity-bound, purpose-scoped authorisation object that grants an indexer permission to index, cache, summarise, or otherwise process specific content for specific purposes.

* *Policy Communication:* Machine-readable statements about permitted uses (e.g., search indexing, snippet generation, retrieval augmentation, AI training), retention, attribution, and derivative generation.

* *Freshness Signalling:* A method for publishers to explicitly request refresh or indicate change without exposing unauthenticated inbound endpoints.

The proposal is compatible with legacy web publishing and can be adopted incrementally. It is especially suited to "zero-port" deployments where inbound crawling is undesirable or impossible.

3. Terminology

Requirements Language: The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119] and [RFC8174] when, and only when, they appear in all capitals.

This Internet-Draft is primarily exploratory; requirement language is used sparingly and only where behaviour is intentionally specified.

Publisher A server or origin that wishes to make content available for indexing.

Indexer A service (e.g., search engine, AI retrieval system, dataset builder) that consumes content for indexing, ranking, retrieval, summarisation, training, or related processing.

Trusted Indexer An Indexer whose identity is known to, and explicitly authorised by, a Publisher via a Discovery Grant.

Outbound Discovery Session A Publisher-initiated secure session to an Indexer service endpoint; within this session, discovery, policy, and content transfer messages may be exchanged.

Discovery Grant A cryptographically bound authorisation object that

conveys consent and scope (what may be indexed, by whom, and for what purposes).

Content Scope A set of resources to which a Discovery Grant or policy applies (e.g., URL set, content-hash set, semantic collection, feed).

Usage Purpose A declared intent for automated processing, such as traditional search indexing, snippet generation, retrieval augmentation, or AI training.

Freshness Signal A notification indicating that indexed content may have changed and that refresh is desired.

UZPIF Session A secure, identity-bound connectivity substrate defined by [UZPIF], typically established via outbound connections to one or more Rendezvous Nodes.

4. Introduction

Traditional web indexing relies on inbound crawling, where automated clients (crawlers) initiate connections to servers and respect opt-out signals such as robots.txt (as standardised in the Robots Exclusion Protocol; [RFC9309]). While effective for many years, this model exposes servers to unsolicited traffic, abuse from malicious crawlers, and challenges in enforcing preferences - particularly as AI systems increasingly use crawled content for training or real-time retrieval.

Recent developments, including crawler best practices ([draft-illyes-aipref-cbcp]) and discussions on AI-specific controls, highlight limitations of opt-out regimes. Inbound crawling assumes servers are reachable and willing to respond, which conflicts with emerging zero-port and zero-trust architectures.

This document describes an outbound opt-in alternative: servers explicitly initiate authenticated connections to authorised indexers (traditional search engines or AI agents) when they wish to be discovered or refreshed. Discovery requests are bound to cryptographic identities and policy grants, enabling fine-grained control over who may index content and for what purpose (e.g., traditional search, AI training, or summarisation).

The approach builds on identity-first transports such as the Universal Zero-Port Interconnect Framework (UZPIF; [UZPIF]) and Universal Zero-Port Transport Protocol (UZP; [UZP]), where endpoints establish outbound-only sessions. It provides a proactive, consent-first model suited to both human-readable web content and AI-driven search consumption.

This is a research proposal intended for experimentation and discussion, particularly in contexts where reducing inbound exposure and strengthening consent are priorities.

5. Problem Statement

The inbound crawling model is built on an assumption of open reachability: crawlers discover a server, initiate inbound connections, and learn policies after connecting. For modern deployments - especially those seeking to minimise exposed attack surface - this is an inversion of the desired trust model.

Specific limitations include:

* *Unsolicited load and abuse:* REP is advisory, and many automated clients do not comply. Even compliant crawlers can produce significant load when scaled across multiple indexers and AI agents.

* *Weak identity and accountability:* A User-Agent string is not a strong identity; it is easy to spoof and difficult to bind to policy obligations.

* *Purpose ambiguity:* The same content acquisition can be used for search indexing, summarisation, retrieval augmentation, or AI training. Without explicit purpose signalling and enforcement, site operators cannot make informed consent decisions.

* *Incompatibility with zero-port architectures:* If a Publisher exposes no public inbound listening ports, inbound crawling becomes impossible by design.

* *Policy enforcement gaps:* Opt-out mechanisms require a crawler to first connect and then choose to comply, rather than enforcing access authorisation at session establishment.

The outbound indexing model aims to preserve the benefits of web discoverability while shifting control to the Publisher, making consent explicit, identity-bound, and enforceable within authenticated channels.

6. Design Goals

The mechanism defined in this document has the following goals:

* *Opt-in discoverability:* indexing and refresh occur only when the Publisher chooses to contact an Indexer.

* *Identity binding:* all sessions are authenticated and bound to cryptographic identities, enabling durable accountability.

* *Purpose limitation:* Publishers can grant indexing permission for specific purposes (e.g., search indexing) while denying others (e.g., AI training).

* *Policy expressiveness:* Publishers can express usage constraints, retention expectations, attribution requirements, and derivative permissions.

* *Freshness signalling:* Publishers can efficiently request refresh without being continuously crawled.

* *Transport independence:* the mechanism should operate over UZPIF sessions and may be profiled for other outbound-friendly transports.

* *Incremental adoption:* the mechanism complements existing protocols such as REP and does not require immediate ecosystem-wide migration.

7. Architectural Overview

At a high level, outbound indexing replaces "Indexer-driven fetch" with a Publisher-initiated session that can carry announcements, policy, and (optionally) content. Under UZPIF, both Publisher and Indexer may maintain outbound connectivity to one or more Rendezvous Nodes (RNs), which stitch permitted sessions.

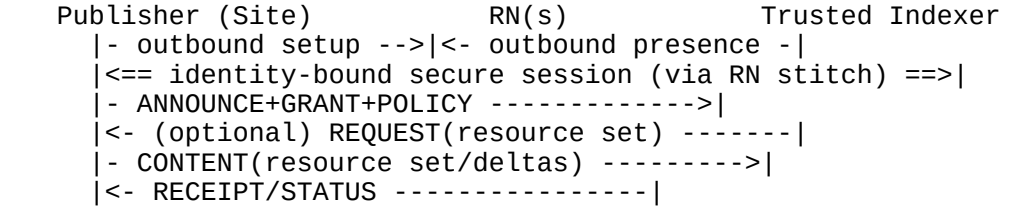

Figure 1: Outbound indexing model: Publisher initiates the session to a trusted Indexer

The session is initiated by the Publisher, but once established it is a bidirectional secure channel in which the Indexer may request specific resources and the Publisher may provide them. The key property is that the Publisher does not expose an unauthenticated public inbound service for discovery.

7.1. Roles and Relationships

The model distinguishes three relationships:

* *Publisher <-> Indexer:* A trust and consent relationship expressed via Discovery Grants and enforced via authenticated sessions.

* *Publisher <-> RN:* A connectivity relationship in which the Publisher maintains outbound sessions to one or more RNs, as defined in [UZPIF].

* *Indexer <-> RN:* An analogous connectivity relationship enabling stitching to Publishers that authorise the Indexer.

This document focuses on Publisher-to-Indexer semantics and does not redefine UZPIF stitching or transport behaviour.

8. Trust and Identity Model

Outbound indexing relies on cryptographic identities for both Publishers and Indexers. In UZPIF deployments, these identities are typically represented by certificates or equivalent credentials issued within an identity plane (e.g., Pantheon as described in [UZPIF]), and sessions are established over secure channels such as TLS-DPA ([TLS-DPA]).

A Publisher MUST authenticate the Indexer identity before sending any content beyond minimal discovery metadata. An Indexer MUST authenticate the Publisher identity before accepting Discovery Grants, policy, or content.

Identity binding serves two purposes:

* *Consent enforcement:* Discovery Grants are bound to specific identities and cannot be meaningfully replayed by unauthorised parties.

* *Operational accountability:* Publishers can select and audit trusted Indexers, and Indexers can maintain verifiable provenance of content acquisition.

8.1. Discovering Indexer Service Endpoints

This document does not mandate a single Indexer discovery mechanism. A Publisher may discover Indexer identities and endpoints through out-of-band agreements, operator-curated trust lists, or an identity plane such as Pantheon ([UZPIF]).

A Publisher SHOULD treat Indexer discovery as a trust decision comparable to granting API access. Blind acceptance of unsolicited Indexer identities reintroduces abuse vectors that outbound indexing is intended to reduce.

9. Discovery Grants

A Discovery Grant conveys explicit permission from a Publisher to an Indexer, scoped by content and purpose. It is an authorisation artefact, not merely a preference hint.

Discovery Grants are conceptually aligned with grant structures described in [UZPIF]. This document defines additional, indexing-specific fields and semantics.

9.1. Grant Properties

A Discovery Grant SHOULD include at least:

* *Issuer:* the Publisher identity that issues the grant.
* *Audience:* the Indexer identity authorised to use the grant.
* *Scope:* a Content Scope describing what may be indexed or retrieved.
* *Purposes:* one or more Usage Purposes for which the Indexer is authorised.
* *Constraints:* optional limits such as maximum fetch rate, retention period, or required attribution.
* *Expiry:* a time limit after which the grant is no longer valid.
* *Signature:* a cryptographic signature or equivalent proof binding the grant to the Issuer.

Discovery Grants MUST be bound to the authenticated identities observed in the Outbound Discovery Session. An Indexer MUST reject a grant if the authenticated Publisher identity is not the Issuer, or if the authenticated Indexer identity is not the intended Audience.

9.2. Scope Examples

Content Scope may be expressed in multiple ways, depending on deployment:

* *URL prefix scope:* allow indexing for all resources under a given origin and path prefix.

* *Feed scope:* allow indexing for resources enumerated in a signed feed.

* *Hash set scope:* allow indexing for content objects identified by cryptographic hashes.

* *Semantic collection:* allow indexing for a named collection (e.g., "docs", "blog", "product-catalogue") maintained by the Publisher.

10. Policy Communication

Outbound indexing provides a channel for Publishers to communicate content usage policy to Indexers in a form that is:

* bound to authenticated identities;

* associated with explicit Content Scope; and

* auditable and revocable.

The policy data model is intentionally generic. Deployments MAY use the vocabulary defined by the IETF AIPREF working group (e.g., [draft-ietf-aipref-vocab]), or they MAY define private purpose tokens under bilateral agreement.

10.1. Policy Elements

A policy statement SHOULD be able to express:

* *Allowed purposes:* which Usage Purposes are permitted (or denied) for a given scope.

* *Derivation:* whether summaries, snippets, embeddings, or other derived artefacts are permitted.

* *Training:* whether AI training or model fine-tuning is permitted.

* *Attribution:* requirements for attribution or source linking in downstream displays.

* *Retention:* permitted retention duration for cached copies or extracted features.

* *Redistribution:* whether indexed content may be redistributed or provided to third parties.

Policy is not an access-control mechanism by itself; it is enforced by the combination of Discovery Grants, authenticated sessions, and Indexer compliance. However, unlike REP, policy is exchanged in an authenticated context where non-compliance can be attributed to a specific identity.

10.2. Illustrative Policy Example

The following is a non-normative example of a policy object that permits traditional search indexing and snippet generation, but denies training and long-term retention:

```
{
  "scope": "https://example.com/docs/*",
  "allowed_purposes": [
    "search.index",
    "search.snippet",
    "rag.retrieve"
  ],
  "denied_purposes": ["ai.train", "ai.finetune"],
  "derivatives": {
    "summary": "allowed",
    "embeddings": "allowed",
    "snippets": "allowed"
  },
  "retention": {
    "cached_copy_days": 14,
    "embedding_days": 30
  },
  "attribution": {
    "required": true,
    "link_back": true
  }
}
```

Figure 2: Example policy object (illustrative)

The encoding, signing, and canonicalisation of policy objects are open for profiling and experimentation. A deployment MAY sign policy objects as part of a Discovery Grant, or MAY carry them as a separate signed artefact within the session.

11. Protocol Operation

This section describes a baseline operational sequence. The message formats used inside the Outbound Discovery Session are intentionally abstract in this version of the document; the focus is on semantics and security properties.

11.1. Session Establishment

A Publisher initiates an Outbound Discovery Session to an Indexer endpoint using UZPIF connectivity ([UZPIF]) and an identity-bound secure channel (e.g., TLS-DPA; [TLS-DPA]).

The Publisher MUST verify the authenticated Indexer identity before sending any sensitive content. The Indexer MUST verify the authenticated Publisher identity before accepting grants or indexing data.

11.2. Announcement and Grant Presentation

Once the secure session is established, the Publisher sends:

* a discovery announcement identifying the Publisher, site scope, and desired indexing actions (new discovery, refresh, or revocation);

* one or more Discovery Grants; and

* associated policy statements for the relevant Content Scopes.

The Indexer MUST validate Discovery Grants and reject any announcement that lacks sufficient authorisation for the requested purposes.

11.3. Content Transfer Modes

This document defines two conceptual modes for content transfer:

* *Publisher Push:* the Publisher proactively sends content objects or deltas to the Indexer.

* *Indexer Request within Session:* the Indexer requests specific resources inside the established session, and the Publisher responds over the same channel.

Both modes preserve the "no unauthenticated inbound ports" property because the Indexer does not initiate a new network connection to the Publisher.

A Publisher MAY choose to offer only one mode. For example, a Publisher with strict egress policy may prefer request/response within a session, while a Publisher with pre-generated feeds may prefer push.

11.4. Freshness Signalling

A Publisher MAY send Freshness Signals to request that an Indexer refresh previously indexed content. Freshness Signals SHOULD be lightweight and SHOULD include enough metadata for the Indexer to prioritise work (e.g., change timestamps, object identifiers, or hashes).

Freshness Signals do not, by themselves, grant permission; they operate under the permissions already established by Discovery Grants and policy.

11.5. Receipts and Auditability

An Indexer MAY provide receipts indicating that it has accepted content for indexing and the purposes under which it will be processed. Receipts can improve auditability and facilitate contractual enforcement in commercial relationships.

This document does not specify a receipt format; profiles may define signed receipts that include:

* the grant identifier used;
* the content scope covered;
* the declared processing purposes;
* timestamps; and
* a signed Indexer identity statement.

11.6. Revocation

A Publisher MUST be able to revoke consent. Revocation may apply to a specific grant, a policy scope, or an entire Publisher-to-Indexer relationship.

Upon receiving a valid revocation instruction from an authenticated Publisher, an Indexer SHOULD cease further acquisition under the revoked scope and SHOULD follow the Publisher’s stated retention and deletion policy.

A Publisher SHOULD treat revocation as an operational and legal relationship issue. Technical signalling can communicate intent and scope, but enforcement ultimately depends on Indexer compliance.

12. Relationship to Existing Mechanisms

Outbound indexing is designed to be complementary to existing web controls and does not attempt to obsolete them.

12.1. Relationship to the Robots Exclusion Protocol

REP ([RFC9309]) is an opt-out signalling mechanism interpreted by automated clients that initiate inbound connections. It is widely deployed and remains relevant for legacy crawling.

Outbound indexing differs in that it:

* is opt-in by default;

* operates over authenticated, identity-bound channels; and

* supports explicit purpose limitation and richer policy statements.

A Publisher MAY use REP for the general web while using outbound indexing for high-value relationships with specific trusted indexers. An Indexer MAY choose to prioritise outbound indexing signals when present, as they can be higher quality and fresher than crawl-derived heuristics.

12.2. Relationship to AI Preference Signalling

The IETF AIPREF working group is developing vocabulary and attachment mechanisms for expressing usage preferences (e.g., [draft-ietf-aipref-vocab] and [draft-ietf-aipref-attach]).

Outbound indexing does not compete with these efforts. Instead, it provides an authenticated delivery channel for the same or compatible preference statements, including in environments where HTTP acquisition is not the primary mechanism or where inbound HTTP access is intentionally restricted.

13. Security Considerations

Outbound indexing reduces exposure to unsolicited inbound traffic by eliminating the need for publicly reachable discovery endpoints. However, it introduces new considerations around trust, grant handling, and policy enforcement.

Implementations SHOULD consider:

* *Indexer impersonation:* Publishers must authenticate Indexer identities; otherwise, an attacker could harvest content by posing as an Indexer.

* *Grant replay:* Discovery Grants should be bound to identities and sessions; replay in a different context should be rejected.

* *Scope escalation:* Indexers must enforce the declared Content Scope and Purposes; ambiguous scopes should be avoided.

* *Confused deputy:* Publishers should avoid issuing broadly scoped grants that a downstream Indexer could use to justify unexpected processing.

* *Compromised Indexers:* A trusted Indexer compromise can lead to large-scale misuse; publishers should prefer short-lived grants and revocation readiness.

* *Metadata leakage:* Even announcing content existence can reveal sensitive information; publishers should consider minimal announcements and staged disclosure.

When used with UZPIF ([UZPIF]) and TLS-DPA ([TLS-DPA]), outbound indexing benefits from identity-bound handshake properties and reduced scanning surface. This document does not define cryptographic primitives; it relies on the referenced transports for channel security.

14. Privacy Considerations

Outbound indexing provides publishers with positive control over who may access content for automated processing. This can reduce privacy harms associated with indiscriminate crawling.

Publishers SHOULD consider:

* minimising announcement metadata to what is necessary for indexing;

* scoping grants narrowly to avoid unintended disclosure;

* using short-lived grants and explicit retention policy; and

* auditing relationships with trusted indexers.

Indexers SHOULD consider:

* providing transparency about processing purposes and retention;

* supporting publisher revocation and deletion requests; and

* limiting onward disclosure of content to third parties unless explicitly permitted.

15. IANA Considerations

This document has no IANA actions.

16. References

16.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

16.2. Informative References

[draft-ietf-aipref-attach] Thomson, M. and M. Nottingham, "Associating AI Usage Preferences with Content in HTTP", Work in Progress, Internet-Draft, draft-ietf-aipref-attach, <https://datatracker.ietf.org/doc/html/draft-ietf-aipref-attach>.

[draft-ietf-aipref-vocab] Keller, P. and M. Thomson, "A Vocabulary For Expressing AI Usage Preferences", Work in Progress, Internet-Draft, draft-ietf-aipref-vocab, <https://datatracker.ietf.org/doc/html/draft-ietf-aipref-vocab>.

[draft-illyes-aipref-cbcp] Illyes, G., Kuehlewind, M., and A. Kohn, "Crawler best practices", Work in Progress, Internet-Draft, draft-illyes-aipref-cbcp, <https://datatracker.ietf.org/doc/html/draft-illyes-aipref-cbcp>.

[RFC9309] Koster, M., Illyes, G., Zeller, H., and L. Sassman, "Robots Exclusion Protocol", RFC 9309, DOI 10.17487/RFC9309, September 2022, <https://www.rfc-editor.org/info/rfc9309>.

[TLS-DPA] Fisher, B. A., "TLS-DPA: An Identity-Bound Security Protocol for Traditional, Overlay, and Zero-Port Transports", Work in Progress, Internet-Draft, draft-dpa-tls-dpa, <https://datatracker.ietf.org/doc/html/draft-dpa-tls-dpa>.

[UZP] Fisher, B. A., "UZP: Universal Zero-Port Transport Protocol", Work in Progress, Internet-Draft, draft-dpa-uzp-transport, <https://datatracker.ietf.org/doc/html/draft-dpa-uzp-transport>.

[UZPIF] Fisher, B. A., "The Universal Zero-Port Interconnect Framework (UZPIF): An Identity-Centric Architecture for Post-Port Networking", Work in Progress, Internet-Draft, draft-dpa-uzpif-framework, <https://datatracker.ietf.org/doc/html/draft-dpa-uzpif-framework>.

Author’s Address

Benjamin Anthony Fisher
DPA R&D Ltd (https://www.dpa-cloud.co.uk)
Email: b.fisher@dpa-cloud.co.uk
URI: https://orcid.org/0009-0004-4412-2269