

ΟΑΙ-ΡΜΗ πρωτόκολλο

Το **ΟΑΙ-ΡΜΗ** (Open Archives Initiative Protocol for Metadata Harvesting) είναι ένα πρωτόκολλο που έχει σχεδιαστεί για τη συλλογή (harvesting) μεταδεδομένων από αποθετήρια ψηφιακού περιεχομένου.

Βασικά σημεία:

1. Σκοπός

- Δημιουργήθηκε για να επιτρέπει σε διαφορετικά αποθετήρια (π.χ. βιβλιοθήκες, αρχεία, ψηφιακές συλλογές) να εκθέτουν τα μεταδεδομένα τους με ενιαίο τρόπο.
- Οι "harvesters" (συλλέκτες) μπορούν να τα αντλήσουν κεντρικά και να δημιουργούν υπηρεσίες αναζήτησης, ευρετηρίασης ή aggregation.

2. Πρωτόκολλο

- Βασίζεται σε **HTTP + XML**.
- Οι κλήσεις γίνονται μέσω HTTP requests σε "endpoints" του αποθετηρίου, με παραμέτρους όπως `?verb=Identify` ή `?verb>ListRecords`.
- Οι απαντήσεις δίνονται σε XML, συνήθως με **Dublin Core** schema, αλλά μπορεί να υποστηρίζονται και άλλα metadata formats.

3. Κύρια Verbs (εντολές ΟΑΙ-ΡΜΗ)

- `Identify` : Επιστρέφει πληροφορίες για το repository (όνομα, admin email, πρωτόκολλο).
- `ListMetadataFormats` : Τι metadata formats υποστηρίζει (π.χ. oai_dc).
- `ListSets` : Λίστες συλλογών (subsets) αν υπάρχουν.
- `ListIdentifiers` : Επιστρέφει μόνο identifiers για records.
- `ListRecords` : Επιστρέφει ολόκληρα τα records με metadata.
- `GetRecord` : Επιστρέφει ένα συγκεκριμένο record με βάση το identifier.

4. Τυπική χρήση

- Ένα αποθετήριο (π.χ. EKT, arXiv, DSpace, EPrints) εκθέτει OAI-PMH endpoint.
- Ένας harvester (π.χ. Europeana, OpenAIRE, Google Scholar) συλλέγει metadata περιοδικά.
- Το metadata ευρετηριάζεται κεντρικά, ενώ το πλήρες περιεχόμενο παραμένει στο source repository.

5. Παράδειγμα κλήσης

http

<https://export.arxiv.org/oai2?verb=Identify>
https://export.arxiv.org/oai2?verb>ListRecords&metadataPrefix=oai_dc
https://export.arxiv.org/oai2?verb=GetRecord&identifier=oai:arXiv.org:2101.00001&metadataPrefix=oai_dc

6. Εφαρμογές

- Βιβλιοθήκες και αποθετήρια ακαδημαϊκών εργασιών.
- Ενοποίηση καταλόγων διαφορετικών οργανισμών.
- Δημιουργία θεματικών μηχανών αναζήτησης (π.χ. μόνο για ιατρικά άρθρα).

Θέλεις να σου δείξω **πρακτικό παράδειγμα** σε C# ή Python για το πώς τραβάς δεδομένα από ένα OAI-PMH endpoint;