

# Making Data Exchange More Efficient

References, Affiliations, Citations, APIs, etc.

# Topics

- Harvesting Methods
- Metadata Schema
- Metadata Scope
- Standardization
  - authors, affiliations
- References
  - unique identifiers
- Matching articles with e-prints

# Harvesting Methods

- FTP server
  - Springer
- Download from proprietary systems
  - IOP Stacks
  - Elsevier CONSYN
- OAI-PMH
  - Hindawi, PoS, arXiv
- RESTful API
  - APS
- Screen Scraping
  - small publishers

# Metadata schema

- JATS
  - APS, EDP, Elsevier, Hindawi
  - upcoming: WSP
- NLM PubMed
  - WSP, IOP
- DC
  - PoS
- custom format
  - arXiv

# Metadata Scope

- Common set of basic metadata
  - References
  - subject categories
  - arXiv identifier
  - additional material

# Standardization/Identification

- Authors
  - ORCID
  - Inspire ids
  - email addresses
  - ....
- Affiliations

# References

## improve capture and identification

- extraction from PDF
  - arXiv, publishers if no feeds
- publisher feeds: structured metadata
  - sometimes unstructured
- identifiers
  - DOI
  - ISBN

# Matching Preprints with Articles

- by arXiv id
  - JHEP
  - APS partially
- Error propagation
  - wrong DOI on arXiv
  - wrong pubnote



# Author and Citation Data

upcoming:

RDF/SparQL interface to author and citation  
data