

# OD: Project Presentation

[github.com/vogelgesang/upc-od/](https://github.com/vogelgesang/upc-od/)

Franz Bermeo  
Maryam Pashmi  
Adrian Vogelsgesang

# Our goals

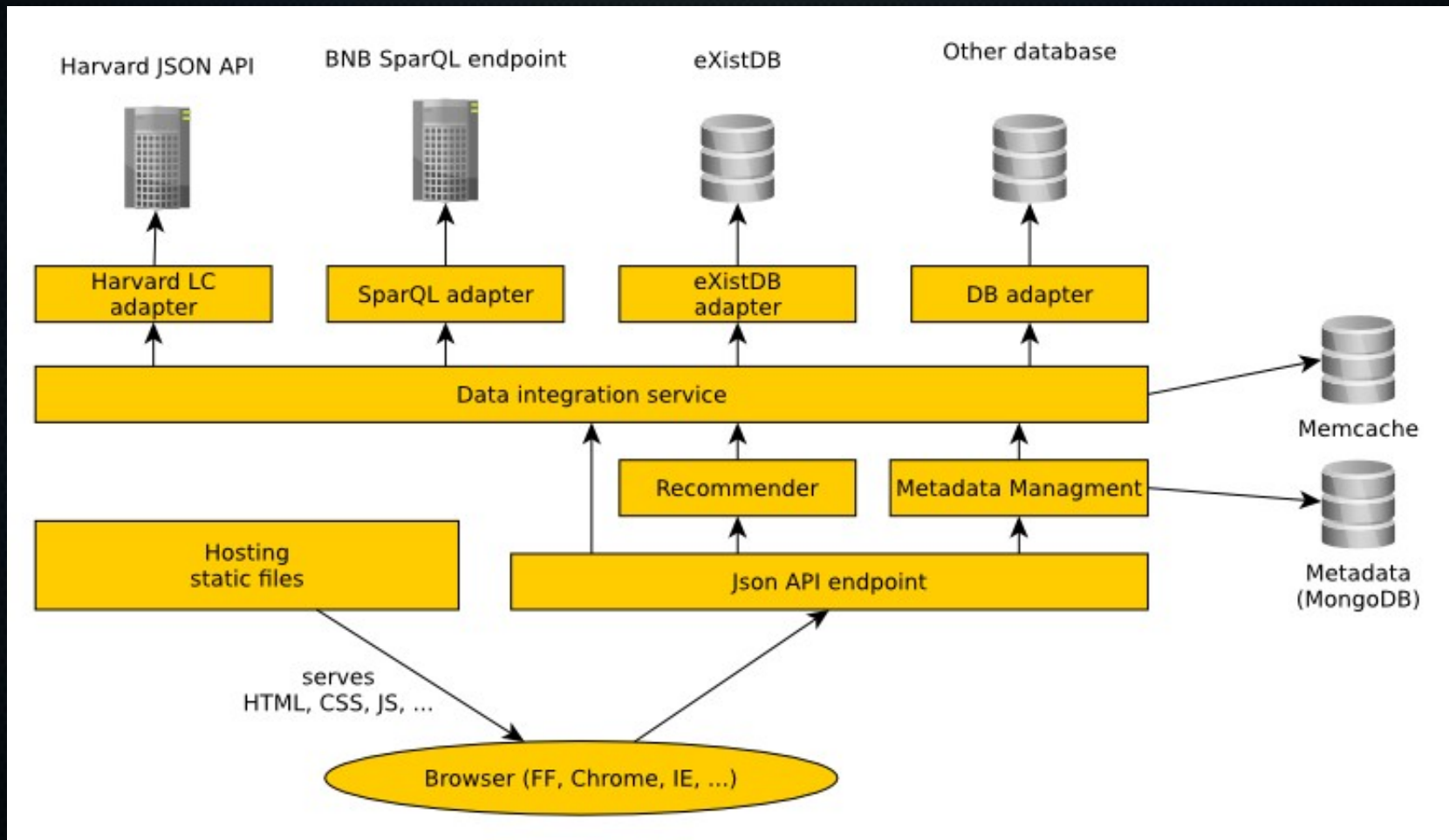
- Book recommendations
- Flexibility
- Realism (→ no invented/synthetic data)
- Speed! (real time)

# Data Sources

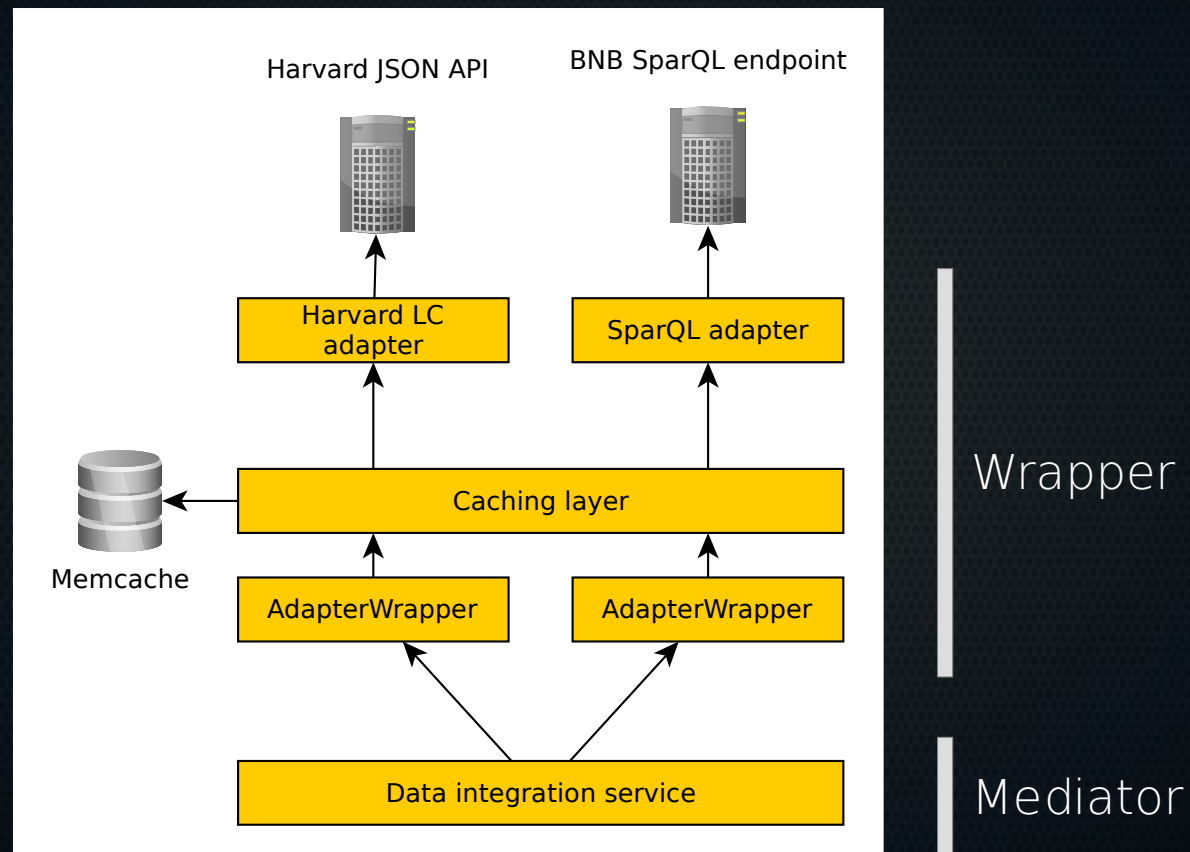
- Cern Bibliography (CC0; XML/Marc21)  
→ eXistDb, Xquery
- Harvard Library Cloud (CC0; JSON Api)  
→ no database, access in real time
- British National Library (CC0; RDF)  
→ no database; access using SparQL



# System architecture



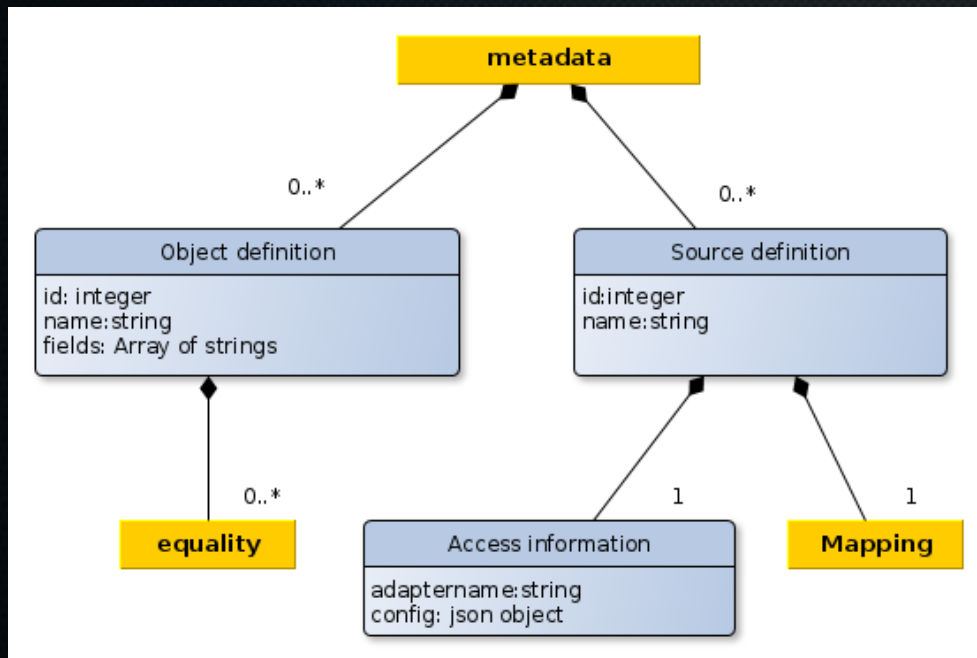
# Data Integration



# The mediator algorithm

- Initialize information base = empty
- Broadcast initial query to all sources
- Until convergence:
  - Merge incoming data with our information base
  - Create new queries from the information base
  - Broadcast these queries to all sources
- And all of this: asynchronous :)

# Meta data repository



```

{
  "_id": "538359929c2946772d6b27c2",
  "name": "Sparql BNB British Library",
  "adapter": {
    "name": "sparql",
    "config": {
      "limit": 20,
      "sparqlEndpoint": "http://bnb.data.bl.uk/sparql"
    }
  },
  "mapping": [
    {
      "sourceType": "http://purl.org/ontology/bibo/Book",
      "mappedType": "book",
      "fieldMapping": {
        "title": [ "http://purl.org/dc/terms/title" ],
        "isbn": [ "http://purl.org/ontology/bibo/isbn13" ]
      }
    }
  ],
}
  
```



# Live Demo

[Home](#) [Schema](#) [Sources](#) [Raw Query](#) **Consolidated Query** [Recommendations](#)

## Consolidated Query

Object type:

Conditions:

Send query

Results:

```
{
  "id": [
    {
      "s": "536cb0f1f0bc857b348d20b3",
      "i": "B28D9029-5803-F228-C525-57E289E88A1B"
    },
    {
      "s": "538359929c2946772d6b27c2",
      "i": "http://bnb.data.bl.uk/id/resource/008796327"
    },
    {
      "s": "5358f92317bbf13036efbb71",
      "i": "98365"
    },
    {
      "s": "5358f92317bbf13036efbb71",
      "i": "98365"
    }
  ],
  "type": "book",
  "fields": {
    "language": [
      "English",
      "eng"
    ],
    "title": [
      "The National Physical Laboratory",
      "The National Physical Laboratory : a history"
    ],
    "publication_place": [
      "Bristol"
    ],
    "publisher": [
```



# Conclusion: the dark side

- Recommendations missing
- Only basic remapping
- Problem: different backgrounds of team members
- eXistDb: index configuration :'(
- Escaping of special characters in SparQL
- VMs are not that easy...

~~Failure~~

Success



# Conclusion: the bright side

- Asynchronous querying
- Provision of partial results
- Splitting the Wrapper into AdapterWrapper and Adapter
- Metadata: LiveEdit
- Secure architecture
- Very flexible SparQL adapter

