



Our aim

Assess how scientific resources can be effectively, regularly and reliably exchanged across systems using the ResourceSync protocol¹.

```
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9" xmlns:rs="http://www.openarchives.org/rs/terms/">
  <rs:md capability="resourcelist" at="2017-01-03T09:00:00Z" />
  <url>
    <loc>http://example.com/res1</loc>
    <rs:md hash="md5:1584abdf8ebdc9802ac0c6a7402c03b6" type="application/pdf" />
    <rs:ln rel="describedby" href="http://example.com/res1_dublin_core_md.xml" type="application/xml" />
  </url>
  <url/>
</urlset>
```

Figure 1. ResourceSync Resource List

ResourceSync characteristics

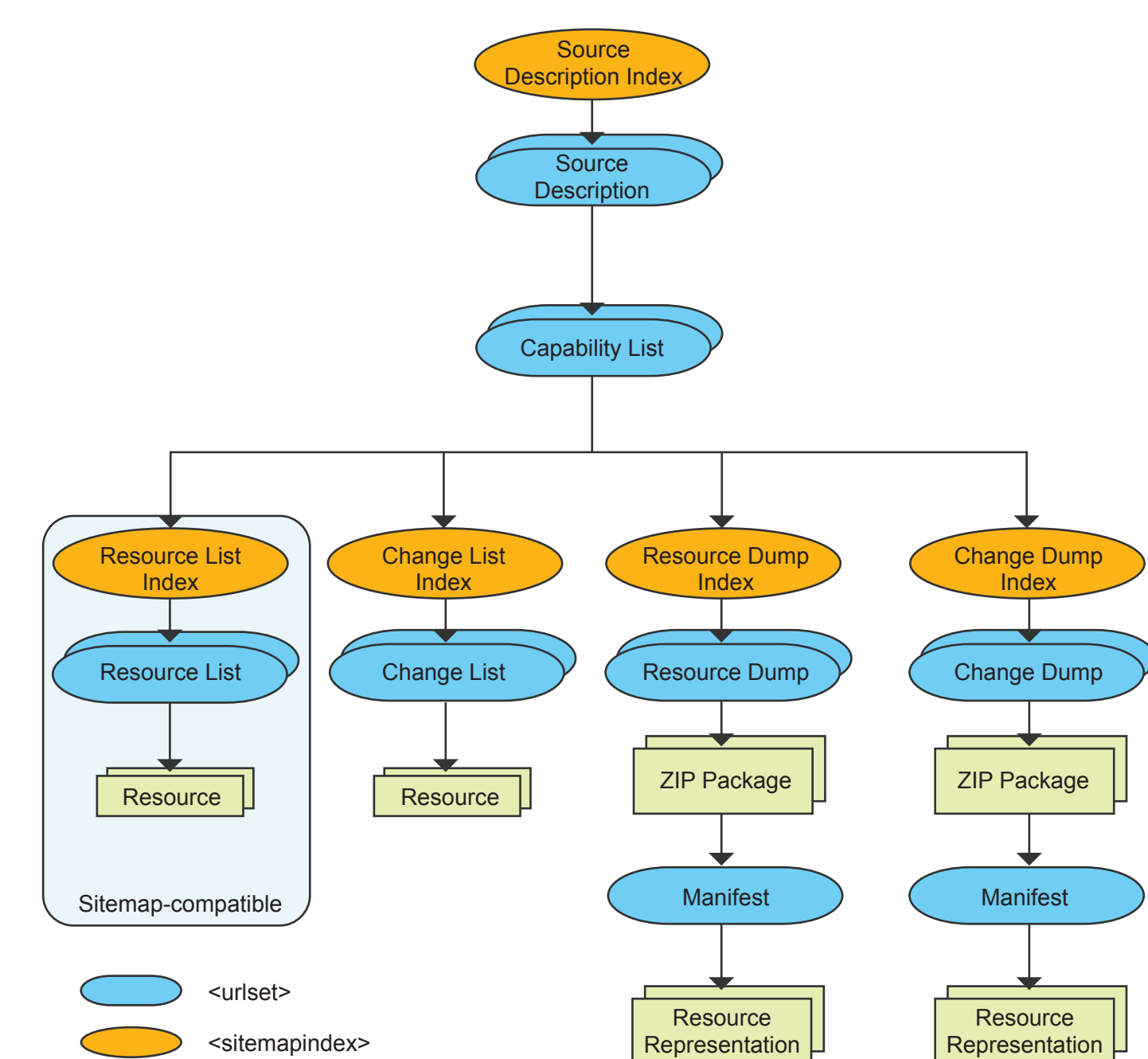


Figure 2. ResourceSync framework structure²

- a) Explicit link between metadata and the described resource
- b) Designed to allow synchronisation of resources, not just metadata
- c) Web-centric

The problem

Current technology for exchanging data across repositories based on a 15 year old technology Open Archives Initiative – Metadata Harvesting Protocol (OAI-PMH)³.

OAI-PMH is:

- Not scalable for large quantities of resources
- Suffers from inconsistent implementations
- Does not deal with resources, only metadata

| IR Software | # | AVG (Record/s) | Variance |
|-----------------|-----|----------------|----------|
| Eprints | 225 | 106.4 | 6.9E+05 |
| Dspace | 76 | 414.2 | 4.2E+06 |
| OJS | 35 | 14.1 | 9.3E+01 |
| Digital Commons | 16 | 19.2 | 4.6E+02 |
| Fedora | 2 | 186.7 | 5.0E+04 |

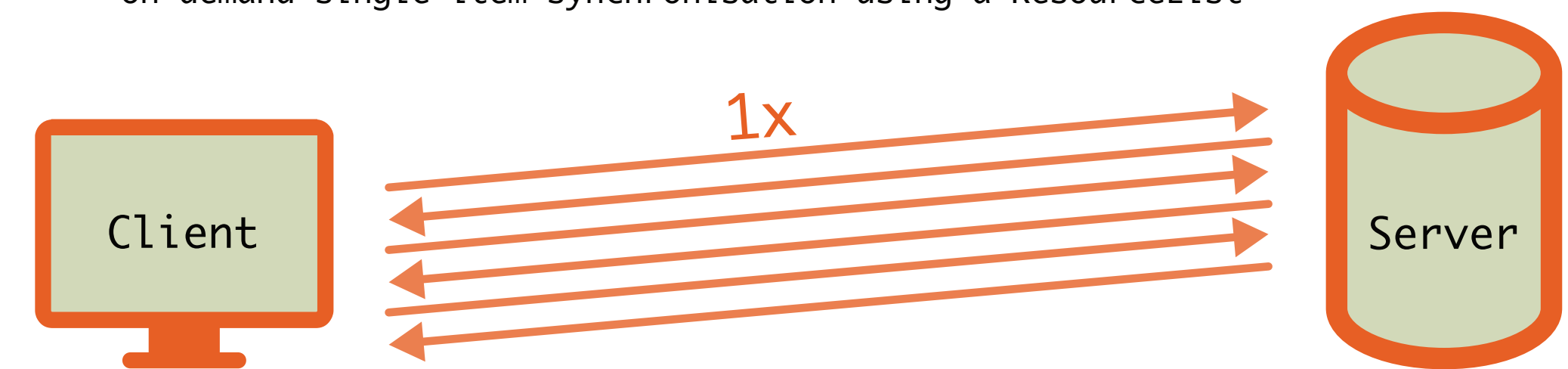
Figure 3. Differences in OAI-OMH performance

References

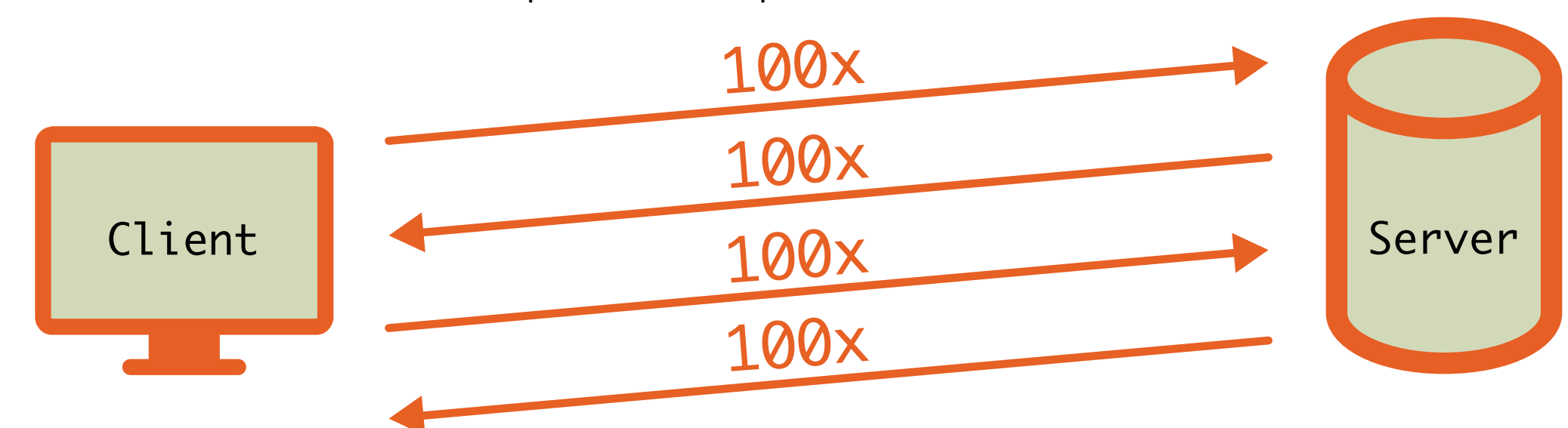
1. Open Archives Initiative – ResourceSync Framework Specification <http://www.openarchives.org/rs/toc>
2. ResourceSync Framework Specification (ANSI/NISO Z39.99-2017) <http://www.openarchives.org/rs/1.1/resourcesync>
3. Open Archives Initiative – Protocol for Metadata Harvesting <https://www.openarchives.org/pmh/>

Synchronisation approaches

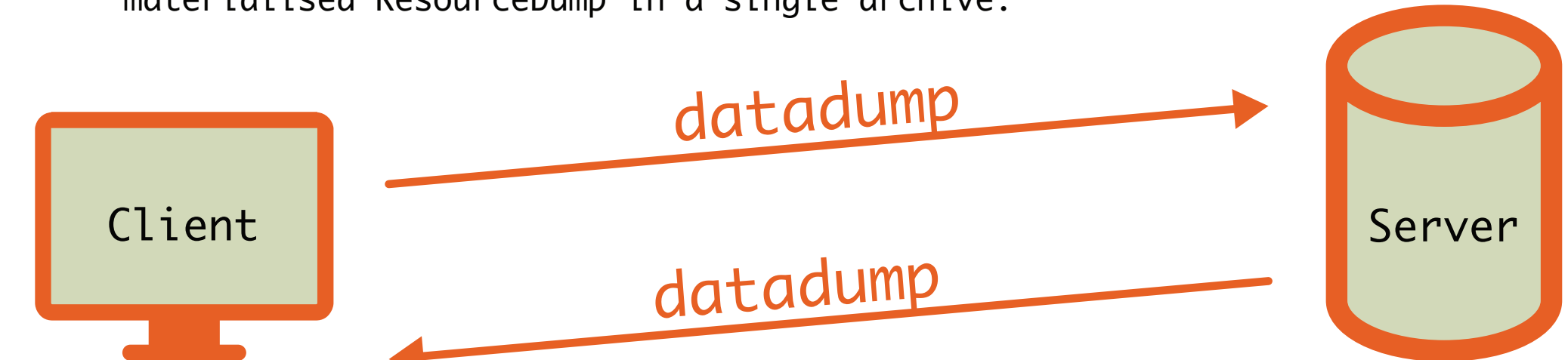
1) Standard ResourceSync – on demand single item synchronisation using a ResourceList



2) Batch ResourceSync – on demand ResourceDump with multiple records in an archive.



3) ResourceSync dump – materialised ResourceDump in a single archive.



Comparing OAI-PMH and ResourceSync: initial results

Conducted a set of experiments comparing OAI-PMH with ResourceSync.

1. Developed a scalable implementation of ResourceSync client and server
2. Ran experiments
3. Evaluating and analysing the results
4. Will disseminate the results
5. Will reach out to external partners to test and productionise this technology

| Repository | OAI-PMH | RS standard | RS batch100 | RS batch500 | RS batch1000 |
|------------------------|---------|-------------|-------------|-------------|--------------|
| Open University | 25.50 | 1.54 | 93.10 | 85.99 | 122.54 |
| White Rose | 3.98 | 2.72 | 64.22 | 104.08 | 160.17 |
| Oxford University | 572.58 | 1.64 | - | 92.22 | 135.80 |
| Cambridge University | 266.12 | 2.84 | 47.13 | 64.41 | 140.68 |
| University of Trieste | 291.88 | 2.61 | 90.25 | 153.00 | 157.50 |
| Manchester | 93.42 | 2.40 | 57.53 | 92.41 | 143.73 |
| Southampton University | 118.36 | 2.74 | - | 141.95 | 116.82 |

Figure 4. Metadata harvesting performance

| Repository | Discovered Requests (OAI-PMH) | Duration (s) |
|----------------------|-------------------------------|--------------|
| Open University | 9,378 | 13,530 |
| White Rose | 25,110 | 30,292 |
| Oxford University | 20 | 2547 |
| Cambridge University | 10,054 | 34,724 |
| Glasgow Enlighten | 21,920 | 7,950 |

Figure 5. Number of requests per downloaded resource

Conclusion

Ongoing project. Outputs:

- Paper benchmarking OAI-PMH against ResourceSync across a range of scenarios
- Scalable implementation of ResourceSync server and client