



Long Term Preservation Policy template & The DICE Digital Preservation Service

Thursday, September 15th 2022

Wilko Steinhoff (DANS-KNAW)
Mark van de Sanden (SURF)
Eko Indarto (DANS-KNAW)
Peter Doorn (DANS-KNAW)

Data Archiving and Networked Services

DANS

SURF



EUDAT CONFERENCE
2022

13-15 Sept. 2022 | Athens, Greece



The DICE project has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 under Grant Agreement no. 101017207

T4.3: Long Term Preservation

- Generic Long Term Preservation (LTP) policy template which can be used as a model by providers of both EUDAT and other LTP-services.

Clarify to users:

- What is or can be guaranteed
 - For how long
 - And by whom
-
- Technical specifications and implementation between EUDAT B2SHARE with LTP outsourced in a trusted archive.

DICE aims regarding long-term preservation (LTP):

- To provide relevant EUDAT services with adequate long-term preservation (LTP) policies.
- To implement these policies for at least the B2SHARE and B2SAFE services.
- The policies for long-term preservation will provide data services made available through EOSC with a basic and fundamental solution and associated business model to sustain research data in the long run.

Without LTP policies the accessibility to, and assessment of FAIR research data, cannot be guaranteed.

I. REPORT

1. Long-Term Preservation Policy Report for EUDAT Services (in particular B2SHARE and B2SAFE)

- 1.1 Introduction
- 1.2 Method: Approaches to LTP Policies
- 1.3 Applying the Long-Term Preservation Policy Template
- 1.4 Agreement structure within EUDAT
- 1.5 Cost modelling for LTP
- 1.6 Technical implications related to the LTP Policy Template
- 1.7 Summary of DICE Task 4.3

II. LTP Policy Template

Long-Term Preservation Policy Template

- 1. Objectives, scope and delimitation of this policy
- 2. Ingest
- 3. Archival Storage
- 4. Data Management
- 5. Administration
- 6. Preservation Planning
- 7. Access

Section and Article	Brief description	Data service			
		B2SHARE		B2SAFE	
		Curation	Status	Curation	Status
		Enhanced		Bit preservation	
2. Ingest					
2.1	Submission Information Package for Ingest	X	Yes	X	Users are able to store AIP packages, but this is not mandatory.
2.2	Metadata index and Cataloguing	X	2 publication workflows are supported, including with a reviewed process		No
2.3	Licenses	X	A license can optionally specified		No
2.4	Persistent Identifiers	X	Yes		Yes
	Additional PIDs		Yes		No

Part of D4.2; DICE: Pilots for the integration with other services & platforms:
<http://doi.org/10.23728/b2share.b27ba476b37740d598a87092e1a527f0>

- ▶ LTP Policy Template covers articles on:
 1. Objectives, scope and delimitation of this policy
 2. Ingest
 3. Archival Storage
 4. Data Management
 5. Administration
 6. Preservation Planning
 7. Access
- ▶ Including articles when LTP service has been outsourced (v.s. “in-house”)
- ▶ Follows the Open Archival Information System (OAIS) reference model
- ▶ Defined for 4 curation levels (CTS):
Bit preservation, Basic, Enhanced and Data-level

1. **Objectives, scope and delimitation of this policy**
 - 1.1. The general aim of this preservation policy is to ensure the authentic and trustworthy preservation and accessibility of digital research data and related outputs, henceforth referred to as “datasets”, for reuse in the long term.
 - 1.2. The specific aims of the preservation policy are to:
 - Provide authentic and reliable instances of datasets to researchers;
 - Maintain the integrity and quality of the datasets;
 - Ensure that datasets are managed throughout their lifecycle (e.g. when migrations or changes in metadata are carried out) in the medium that is most appropriate for the task they perform;
 - Ensure that the relevant level of information security is applied to each dataset;
 - 1.3. The scope of this policy is limited to [LTP Institution]’s service for long-term preservation, [LTP Service]. It applies exclusively to datasets held in [LTP Service]. This policy does not consider preservation of other materials, such as [LTP Institution]’s web pages, internal and external documents, and digital objects in any other services the [LTP Institution] provides.
 - 1.4. The [LTP Institution] assumes responsibility for the long-term preservation and accessibility of datasets ingested (by individual deposits or bulk data transfers) into its [LTP Service].
 - 1.5. Long-term preservation and providing sustained access to datasets fits within the remit and mission of [LTP Institution].

Article		Brief description	Curation Level			
			A: Bit preservation	B: Basic Curation	C: Enhanced Curation	D: Data-level curation
2. Ingest						
2A. LTP in-house	2B. LTP outsourced					
2.1	2.6	Submission Information Package for Ingest	x	x	x	x
2.2	2.7	Metadata index and Cataloguing		x	x	x
2.3	2.8	Licenses		x	x	x

Implementation of the LTP policy for B2SHARE in a CTS certified archive (outsourced).

“When an authorized B2SHARE user creates a dataset, an LTP archiving possibility will be offered on the dataset page to that user. The dataset owner can now start the process of transferring the dataset from B2Share to the LTP-archive.”

Two objectives:

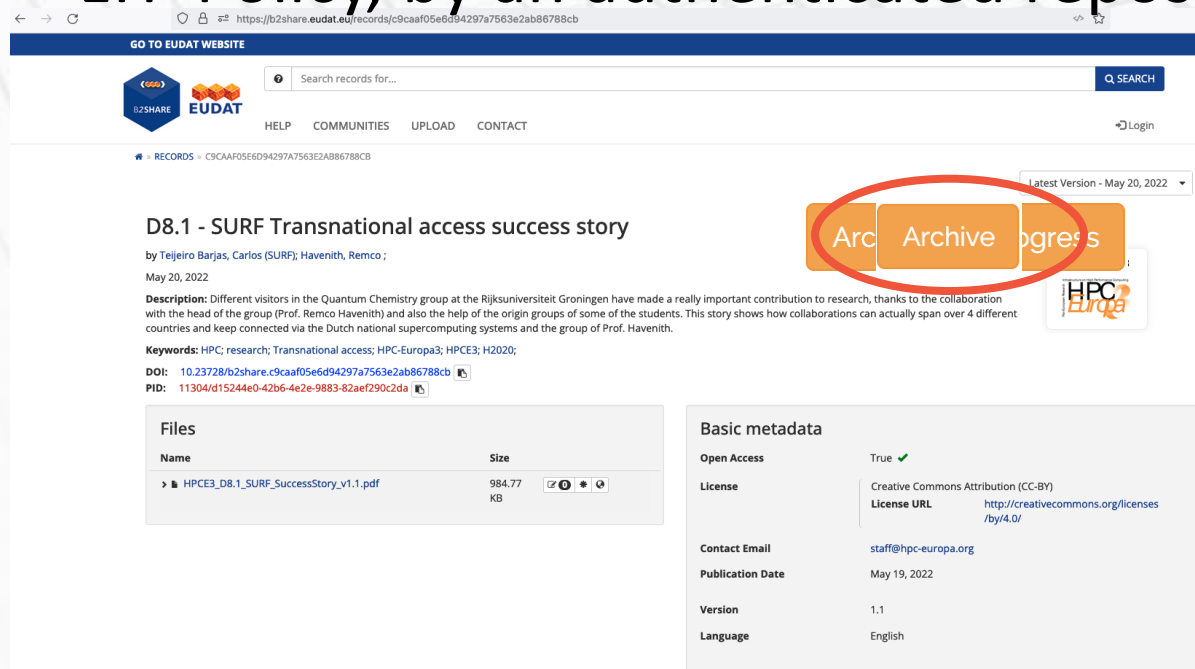
- ▶ Technical specification document (Read the Docs)
- ▶ Reference LTP implementation: DICE Digital Preservation Service

Implementation:

- ▶ “Notification Driven Archiving” using Open Web Standards

Single B2SHARE record archiving

- Single dataset will be archived on demand, according to the repository LTP Policy, by an authenticated repository user.



GO TO EUDAT WEBSITE

Search records for... Q SEARCH

HELP COMMUNITIES UPLOAD CONTACT Login

RECORDS - C9CAAF05E6D94297A7563E2AB86788CB

Latest Version - May 20, 2022

D8.1 - SURF Transnational access success story

by Teijeiro Barjas, Carlos (SURF); Havenith, Remco ;
May 20, 2022

Description: Different visitors in the Quantum Chemistry group at the Rijksuniversiteit Groningen have made a really important contribution to research, thanks to the collaboration with the head of the group (Prof. Remco Havenith) and also the help of the origin groups of some of the students. This story shows how collaborations can actually span over 4 different countries and keep connected via the Dutch national supercomputing systems and the group of Prof. Havenith.

Keywords: HPC; research; Transnational access; HPC-Europa3; HPCE3; H2020;

DOI: [10.23728/b2share.c9caaf05e6d94297a7563e2ab86788cb](https://doi.org/10.23728/b2share.c9caaf05e6d94297a7563e2ab86788cb)

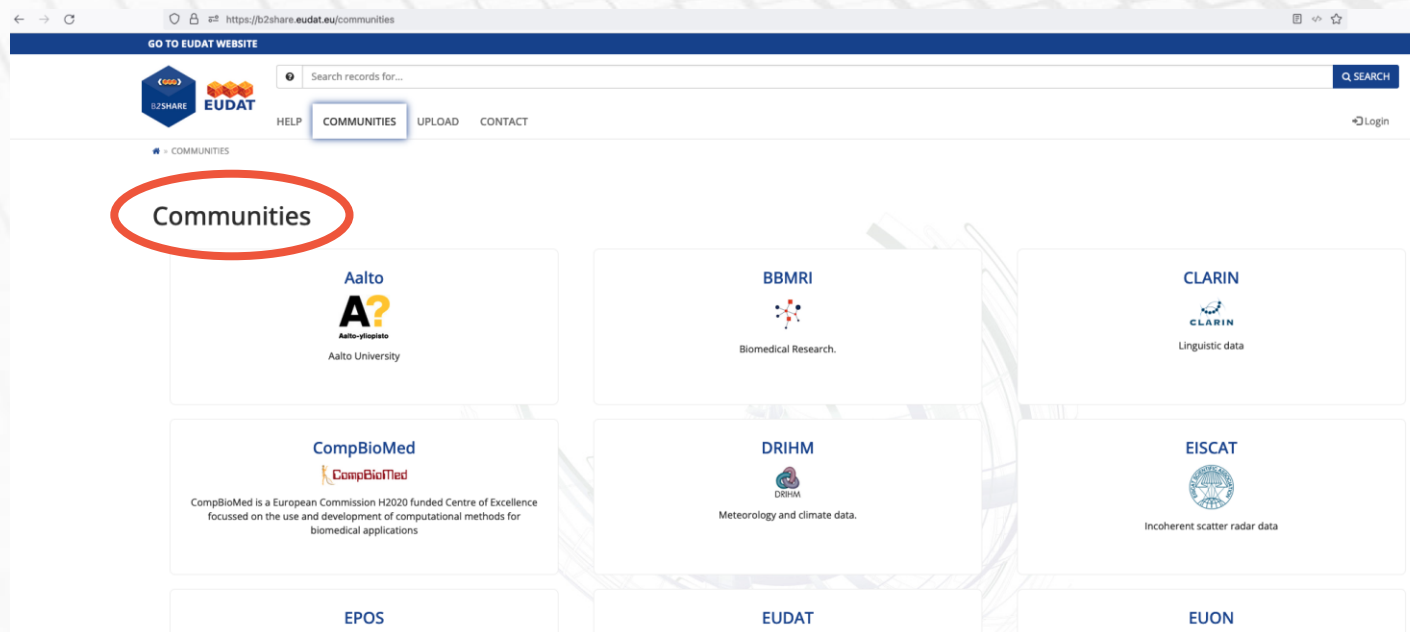
PID: [11304/d15244e0-42b6-4e2e-9883-82ae7290c2da](https://www.eudat.eu/record/11304/d15244e0-42b6-4e2e-9883-82ae7290c2da)

Files	
Name	Size
HPCE3_D8.1_SURF_SuccessStory_v1.1.pdf	984.77 KB

Basic metadata	
Open Access	True ✓
License	Creative Commons Attribution (CC-BY) License URL http://creativecommons.org/licenses/by/4.0/
Contact Email	staff@hpc-europa.org
Publication Date	May 19, 2022
Version	1.1
Language	English

B2SHARE Community's records archiving

- ▶ All records with a B2SHARE community could be auto-archived, according to a community archival agreement.



Use Open Web Standards:

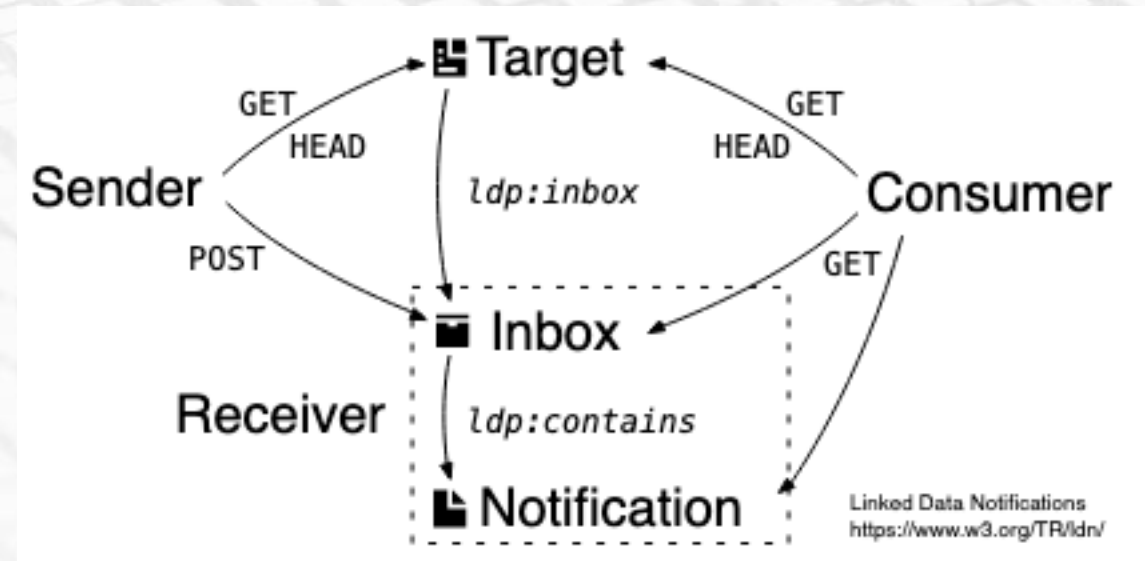
- ▶ Linked Data Notifications (*W3C: Notification system*)
- ▶ ActivityStreams 2.0 (*W3C: Vocabulary*)
- ▶ SignPosting (*IETF RFCs based: Navigational*)

Linked Data Notifications (LDN)

Notification

- W3C Recommendation (Web standard)
- <https://www.w3.org/TR/ldn/>

Protected	
GET	/settings Get Settings
POST	/inbox Post Inbox
HEAD	/inbox Post Inbox
default	
GET	/inbox Get Inbox
GET	/ About
HEAD	/ About



ActivityStreams 2.0 (AS2)

Vocabulary

A model for representing potential and completed activities using the JSON-LD format.

These make up the messaging/notification payloads.

► W3C Recommendation

<https://www.w3.org/TR/activitystreams-vocabulary/>

Activity Types

Activity Type	Purpose	Interoperability-Scope
as:Offer	Indicates that the actor is offering the object. If specified, the target indicates the entity to which the object is being offered.	Web-wide
coar-notify:ReviewAction	An action related to a review, or request for a review, of a resource	Scholarly-communication
coar-notify:ArchiveAction	An action related to LTP archiving, or request for LTP archiving, of a resource	Scholarly-communication

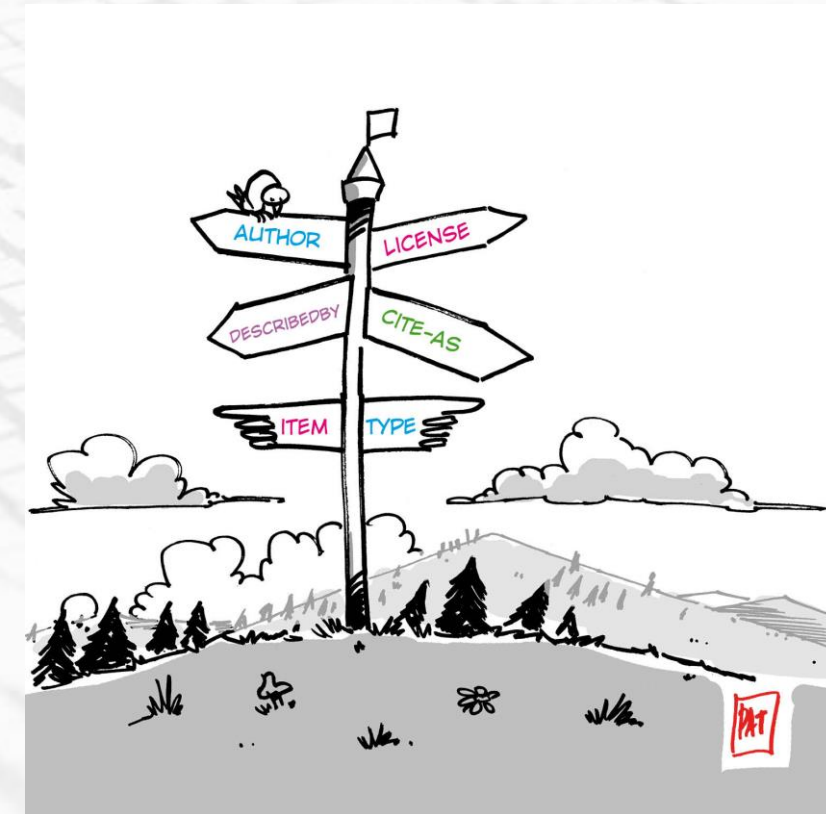
Namespaces:

as : <https://www.w3.org/ns/activitystreams#>

coar-notify : http://purl.org/coar/notify_vocabulary/

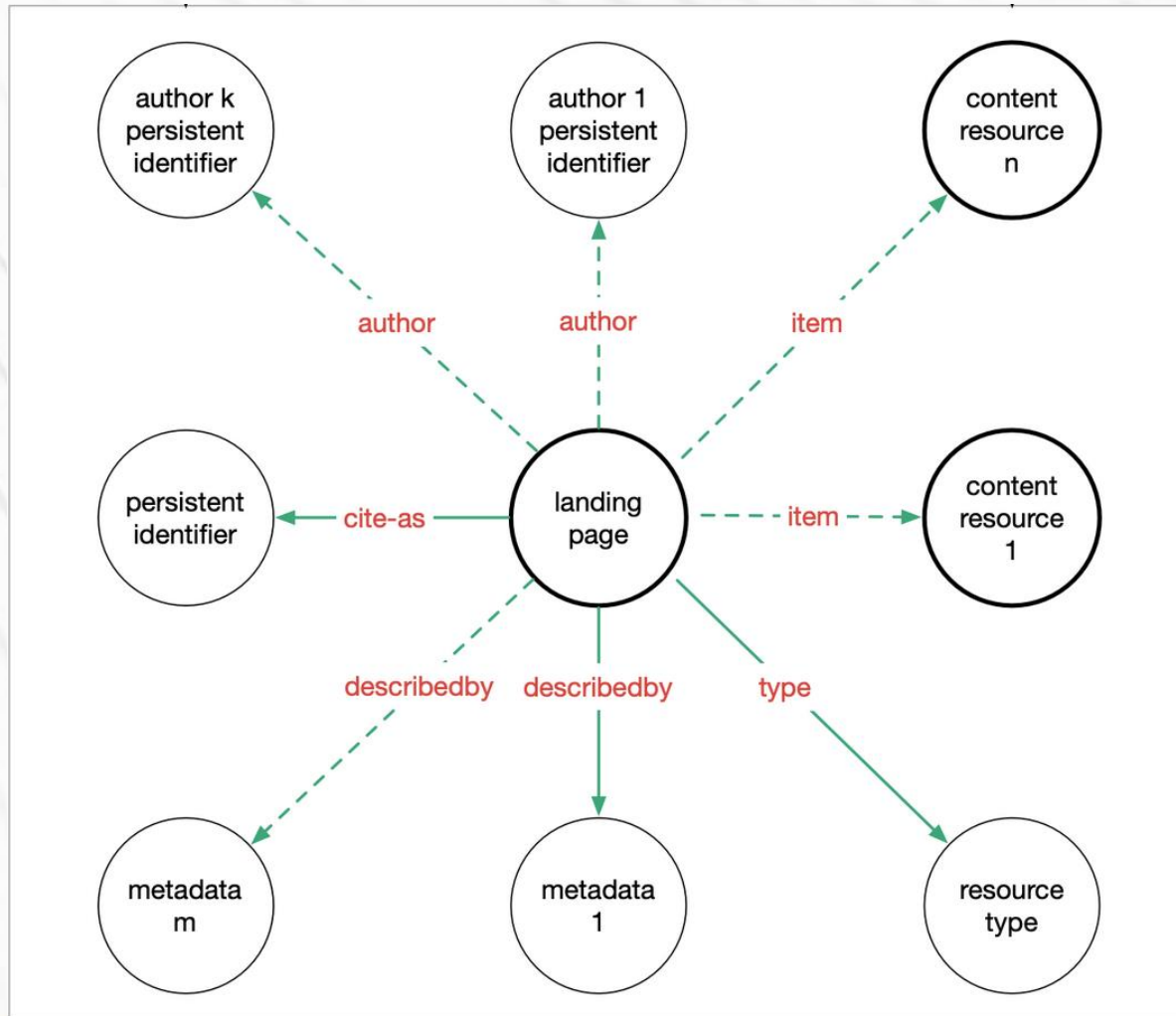
Navigational

- ▶ REST/HATEOAS technology
- ▶ Based on IETF RFCs and IANA registered link relation types
- ▶ Link Set will very soon be an RFC
- ▶ FAIR Signposting Profile:
<https://signposting.org/FAIR>



Cartoon by Patrick Hochstenbach

Signposting



B2SHARE Signposting Module: Link Set example

```
1 {
2   "linkset": [
3     {
4       "anchor": "http://localhost:5000/records/7646043f7a784b5597e257e9f5d7450f",
5       "cite-as": [
6         {
7           "href": "https://doi.org/10.15497/A675341C-F705-4136-B7C3-B9C148556186"
8         }
9       ],
10      "describedby": [
11        {
12          "href": "https://citation.crosscite.org/format?style=bibtex&doi=10.15497/A675341C-F705-4136-B7C3-B9C148556186",
13          "type": "application/x-bibtex"
14        }
15      ],
16      "item": [
17        {
18          "href": "http://localhost:5000/api/files/9d5cb5d6-9a5b-4819-bd45-48e9f7984ed0/myfile",
19          "type": "application/octet-stream"
20        },
21        {
22          "href": "http://localhost:5000/api/files/9d5cb5d6-9a5b-4819-bd45-48e9f7984ed0/myfile-ver0",
23          "type": "application/octet-stream"
24        }
25      ],
26      "license": {
27        "href": "http://creativecommons.org/licenses/by/4.0/"
28      },
29      "type": [
30        {
31          "href": "https://schema.org/AboutPage"
32        }
33      ]
34    },
35    {
36      "anchor": "http://localhost:5000/api/files/9d5cb5d6-9a5b-4819-bd45-48e9f7984ed0/myfile",
37      "collection": [
38        {
39          "href": "http://localhost:5000/records/7646043f7a784b5597e257e9f5d7450f",
40          "type": "text/html"
41        }
42      ]
43    },
44    {
45      "anchor": "http://localhost:5000/api/files/9d5cb5d6-9a5b-4819-bd45-48e9f7984ed0/myfile-ver0",
46      "collection": [
47        {
48          "href": "http://localhost:5000/records/7646043f7a784b5597e257e9f5d7450f",
49          "type": "text/html"
50        }
51      ]
52    }
53  ]
54 }
55 }
```

<https://github.com/EUDAT-B2SHARE/b2share/pull/1813>

Putting it all together: DDPS Components diagram

1) An LTP request will be sent from a B2SHARE landing page to the Archival Bot's inbox, conveying the URL of the landing page of the dataset.

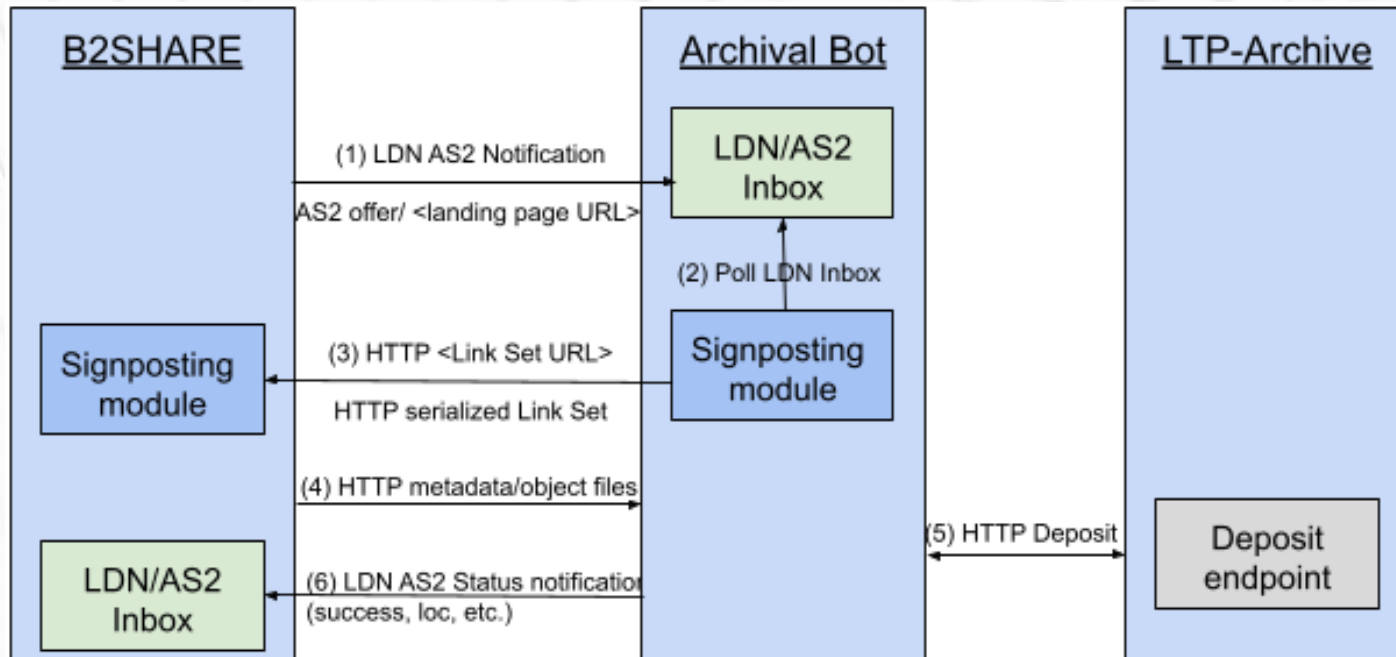
2) The Archival Bot will poll the LDN Inbox for such requests.

3) The Archival bot visits the landing page URL, discovers a Link Set provided via Signposting, and retrieves it.

4) The Archival Bot parses the Link Set to obtain URLs for object files and metadata, and then retrieves those.

5) The Archival Bot deposits an archival package containing metadata and object files to the LTP-archive.

6) The Archival Bot reports back on the status of the deposit.

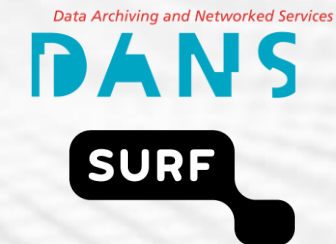


➤ **COAR** *Notify* Project

- Repository and stand-alone review services interoperability Project (Peer review system)
- <https://notify.coar-repositories.org>

➤ Researcher Pod: Mellon-funded Project (UGent)

- Personal research data-pods on the decentralized web
- <https://mellonscholarlycommunication.github.io/spec-notifications>



Thanks for your attention

JOIN OUR COMMUNITY



DICEosc



/company/diceosc

Contact & further information:

- Wilko Steinhoff (DANS-KNAW)
- Mark van de Sanden (SURF)
- Eko Indarto (DANS-KNAW)

Acknowledgements:

Peter Doorn, Hans van Piggelen; Herbert van de Sompel; Patrick Hochstenbach; Martin Klein;
Paul Walk