Sustaining and modernizing open source repository platforms (in the context of Brazil)

Washington Segundo
Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)

COAR Annual Meeting 2023
17th of May 2023
A Hive

(W. Segundo; Ibict)

COAR Meeting, May 2023
Brazilian Portal of Publications and Scientific Data in Open Access (approx. 4,335,000 items) <https://oasisbr.ibict.br>
Updating versions 4–6 to 7
Requirements
The update procedure sets all the environment automatically, and it can be performed/repeated without affecting previous executions and/or the currently operational DSpace.

- The updater is non-invasive to the DSpace in production.
- It works with copies of the database (dump) and the directories "config," "assetstore," and "solr".
- It only exposes HTTP ports.
- For each installation, it creates random passwords (MD5) for the database.
A ".properties" file provided by the updater, specifying the directory of the "dspace.dir" copy, as well as the addresses for the backend and frontend deployment.

- Migrates all the data registered in the old DSpace to the new one.
- Updates the submission forms to the new standard.
- Migrates the Solr statistics.

Source code (BETA version)
https://github.com/projetos-codic-ibict/migracao-dspace
Research data repository netowork
Research data repository network

(figshare, Apollo, GEO, PLOS, GBIF, Omics DI, NIH, OpenAIRE, Explore, LA Referencia, Oasisbr, Dataverse Project)
3Data

Data from, about and for the research

(Data by W. Segundo; Ibict)
The Oasisbr Portal  The BrCris Project  The Laguna Project  The CoNCiênciA Consortium  The dARK PID  Conclusion

Physical Infrastructures  Funders  Projects  Researchers  Organizations

- Patents
- Datasets
- Articles, Thesis, Dissertations
- Softwares
- Results

(W. Segundo; Ibict)  COAR Meeting, May 2023
Laguna

- Level 6: Observatories; Data Recovery, Editing and Submission Interfaces
- Level 5: Dashboards
- Level 4: APIs RESTful, SPARQL, ...
- Level 3: Semantic Repository
- Level 2: FI
- Level 1: FI, FI, FI, FI, FI
- Level 0: XML, CSV, TSV, JSON, HTML, SQL

Harvesting and update tools

External world

access layers

management layers

(W.Segundo; Ibict)

COAR Meeting, May 2023
ConCiência

DataCite
FIND, ACCESS, AND REUSE DATA

Consortium lead

Organization 1
(Also Consortium Lead role)

Repository

Organization 2

Repository

Organization 3

Repository

(CNPq; ibict)

COAR Meeting, May 2023
Goals

- **Avoid** disconnected multiple PID assignments. Example: a journal article (deposited in two different repositories) with two different DOIs that are not known

- **Facilitate** PID assignments to multiple types of digital objects (articles, thesis, data sets, DMP, instruments, ...) in many diffusion channels (repositories, journals, preprint servers, ...)

- **Generate** services open to publication, diffusion and evaluation systems

- **Preserve** metadata through *blockchain* mechanisms
The blockchain consortium model

1. Organization
2. Group Access
3. On Permission
4. Shared Ledgers
5. Transparency

(W.Segundo; Ibict)
The dARK PoC

Users → html page → Blockchain → web service

Blocks: institucion A, institucion B, institucion C

(W.Segundo; Ibict) COAR Meeting, May 2023
Response example

```json
{
    "noid": "8003/fkwff300001v",
    "external_pids": [
        {
            "id": "0x6cd3205875840306d7d5999315739722ce57f31d3e41a33690a8afdd",
            "schema": "DOI",
            "value": "10.1016/J.IS.2021.101826",
            "owner": "0xf17f52141ebef6c7341fad080c5704d77216b732"
        }
    ],
    "payload": "{'title': 'Blockchain-based Privacy-Preserving Record Linkage: enhancing data privacy in an untrusted environment', 'author': 'Thiago Nobrega', 'type': 'Article'}",
    "external_links": [
        "https://doi.org/10.1016/J.IS.2021.101826"
    ],
    "owner": "0xf17f52141ebef6c7341fad080c5704d77216b732"
}
```

(W. Segundo; Ibict)
Conclusion

- **Cooperation** (in Brazil, Latin America and outside) between institutions and formation of collaborative networks

- Development of models, standards and aggregation tools / treatment / retrieval / data visualization and information

- Solutions, with distributed management, that are simple to use and have low computational cost

(W. Segundo; Ibict)
# Acknowledgements

<table>
<thead>
<tr>
<th>Oasisbr DSpace 7 updating</th>
<th>Dataverse Network</th>
<th>BrCris</th>
<th>Laguna</th>
<th>dARK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lautaro Matas</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Márcio Gurgel</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Renê Júnior</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Thiago Nóbrega</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Edilson Júnior</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Jesús M. Chalco</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Thiago Dias</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Vivian Silva</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Jesiel Viana</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Adilson Pinto</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

(W. Segundo; Ibict)
Thank You