Archive-in-a-box

(Internal Demonstrator)

Service-Concept

M 1.4.1.1

Version 1.15 (31/08/2012)
Work Package 1.4
Responsible Partner Technische Universität Darmstadt

DARIAH-DE
Aufbau von Forschungsinfrastrukturen
für die e-Humanities

This research and development project is / was funded by the German Federal Ministry of Education and Research (BMBF), fund number 01UG1110A to M, and managed by the Project Management Agency of the German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt, PT-DLR).
**Project:** DARIAH-DE: Aufbau von Forschungsinfrastrukturen für e-Humanities  
**BMBF Fund number:** 01UG1110A to M  
**Duration:** March 2011 until February 2014  
**Document status:** Draft  
**Dissemination level:** DARIAH-DE-internal  
**Authors:**  
Rainer Becker, Technische Universität Darmstadt  
Richard Eckart de Castilho, Technische Universität Darmstadt

### Revisions:

<table>
<thead>
<tr>
<th>Date</th>
<th>Author</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>01.12.2011</td>
<td>RB</td>
<td>First Draft</td>
</tr>
<tr>
<td>03.03.2012</td>
<td>RB, REC</td>
<td>complete Restructuring (requirements, new candidates)</td>
</tr>
<tr>
<td>03.05.2012</td>
<td>RB</td>
<td>complete Restructuring (dependencies, interoperability-criteria)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>according to DARIAH-DE PM-Reallocation</td>
</tr>
<tr>
<td>10.05.2012</td>
<td>RB</td>
<td>Result of work in Wiki as ‘table’ in appendix 1</td>
</tr>
<tr>
<td>11.06.2012</td>
<td>RB</td>
<td>Rewrite of document-structure; integration 1. evaluation-steps</td>
</tr>
<tr>
<td>13.05.2012</td>
<td>RB</td>
<td>Specification of evaluation-criteria and ranking-calculation</td>
</tr>
<tr>
<td>15.06.2012</td>
<td>RB</td>
<td>Integration of results of first evaluation (according to proceeding work in Wiki; result in appendix 1)</td>
</tr>
<tr>
<td>10.07.2012</td>
<td>REC</td>
<td>Revising; simplified evaluation-calculation; dSpace-Description</td>
</tr>
<tr>
<td>16.07.2012</td>
<td>RB</td>
<td>Candidate descriptions; result, recommendation, future, update of appendix 1</td>
</tr>
<tr>
<td>23.07.2012</td>
<td>RB</td>
<td>Revision following comments from other working package members; revision of service-, candidate descriptions and corresponding evaluation.</td>
</tr>
<tr>
<td>31.07.2012;</td>
<td>REC</td>
<td>Formatting, spelling, grammar, revision of service-, candidate descriptions and corresponding evaluation.</td>
</tr>
<tr>
<td>06.08.2012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>08.08.2012</td>
<td>RB</td>
<td>Service-, candidate descriptions and corresponding evaluation</td>
</tr>
<tr>
<td>13.08.2012</td>
<td>REC / RB</td>
<td>Last Revision for first circulation</td>
</tr>
<tr>
<td>31.08.2012</td>
<td>RB / REC</td>
<td>Revisions according to feedback (Consortium)</td>
</tr>
</tbody>
</table>
1. 'Archive-in-a-box': Introduction ................................................................. 4
   1.1. Aim of the document ............................................................................. 4
   1.2. Impact / Background of this document ............................................... 4
   1.3. Future of this document, next steps (until 2014) .................................. 5
2. DARIAH-specific requirements for an 'Archive-in-a-Box' ......................... 5
   2.1. Textual sources for extraction of requirements ....................................... 5
   2.2. Requirements ....................................................................................... 7
   2.3. First result ............................................................................................ 9
   2.4. DARIAH-DE Dependencies (Interoperability) ....................................... 9
3. Selection of Candidates for an 'Archive-in-a-Box' .................................... 10
   3.1. Collection of a first list of candidates ..................................................... 10
   3.2. Invalid candidates ............................................................................... 12
       3.2.1. EASY 2 .......................................................................................... 12
       3.2.2. Faces ............................................................................................. 12
       3.2.3. Mopseus ......................................................................................... 12
       3.2.4. Phaidra .......................................................................................... 12
       3.2.5. VIRR ............................................................................................. 13
       3.2.6. VuDL .............................................................................................. 13
       3.2.7. ZOpenArchives ............................................................................. 13
   3.3. Candidates for further evaluation .......................................................... 13
4. Criteria ......................................................................................................... 14
   4.1. DARIAH-DE interoperability (max: 5 Points) ....................................... 14
   4.2. Customization- / Extension-Efforts (max: 5 Points) ............................... 14
   4.3. Sustainability of solution (max: 5 Points) ............................................. 15
   4.4. Calculation of results .......................................................................... 15
5. Results ......................................................................................................... 16
   5.1.1. DocPortal ......................................................................................... 16
   5.1.2. dSpace ............................................................................................ 17
   5.1.3. ePrints ............................................................................................. 18
   5.1.4. Invenio ............................................................................................. 19
   5.1.5. OPUS 3 ........................................................................................... 20
   5.1.6. OPUS 4 ........................................................................................... 22
   5.1.7. PubMan .......................................................................................... 23
   5.1.8. TextGridRep ................................................................................... 24
   Overview: Table of Results (simplified via point-values) ............................. 26
6. Expectably interoperable and sustainable candidates / work for the next milestone ................................................................. 26
7. Appendix 1: Evaluation-Matrix (complex) ................................................. 28
8. Appendix 2: possible logo ......................................................................... 31
1. 'Archive-in-a-box': Introduction

Society nowadays not only depends on paper anymore, it more and more gets computational. This has implications. Like natural scientists, scientists in the humanities have to deal more and more with large amounts of research-data. This trend accelerates as not only the social sciences, but research in general is embracing computational resources and instruments.

This document aims to provide a concept for researchers in the humanities on how to employ digital archives, with a particular focus on philological data. The document gives an overview of software for digital archives already available at the time of writing as well as an evaluation scheme and first result for these products according to DARIAH-DE requirements. From that scheme a recommendation is derived which products could be further integrated into the DARIAH infrastructure to gain a stable and trustworthy in-a-box-experience (preferably in a form of a preconfigured and updateable virtual machine) for all kinds of archival use scenarios in the future research.

1.1. Aim of the document

This document should be understood as the first of two documents. This first document aims to

(1) provide an overview of the already available repository-packages at the time of writing (for a quick overview see table 1 in section 3.1),

(2) take into account interoperability with the DARIAH-DE infrastructure services (AAI, generic search etc.) and the standards they use as key criteria for a recommendation for further integration (see section 2.4),

(3) provide an orientation through recommendations for further testing (according to (1)+(2), see section 5).

1.2. Impact / Background of this document

The document at hand should as well be mainly understood as a set of recommendations. These recommendations could apply if you

• plan to introduce a new archival system at your institute / centre or for your own project or network
  -- or --

• plan to migrate existing archival systems and their data to a new platform
  -- under the condition that --
• you want to do either of this in a way that is sustainable as well as interoperable with main parts of the coming DARIAH-DE-infrastructure (specifications that currently could be taken into account).

1.3. Future of this document, next steps (until 2014)

The document at hand will be followed by a second document. This second document will build on the recommendations provided here and will deliver more in-depth information on the recommended products through tests performed as well as experiences gained on their integration into the DARIAH-DE infrastructure. This information can be such as:

• Installation instructions and pitfalls of each procedure, as well as solutions and workarounds. These instructions may serve for the preparation of a virtual machine (VM) image which can be used to test-drive a particular product and it may serve as input for data centers that wish to offer the product as an application service (AaaS).

• Identification of forks of each package software-development to follow for sustaining open and secure DARIAH-DE-compatibility (including support-addresses etc.).

• Introduction and discussion of DARIAH-DE-specific patches, extensions and/or plug-ins which should be included into VM images to ensure out-of-the-box interoperability with the DARIAH-DE infrastructure.

The second document is scheduled for February 2014.

2. DARIAH-specific requirements for an 'Archive-in-a-Box'

This section deals primarily with requirements originating from DARIAH itself towards archive-software. To start with a look at the software-candidates for further consideration, you can skip this section and continue to section 3. You may still want to take a look at section 2.4, interoperability, for an introduction of standards and formats used in the DARIAH infrastructure, which are key criteria for the evaluation (section 3/4) and its results according to each software package (section 5) as well as our early recommendations (section 6).

2.1. Textual sources for extraction of requirements

In order to get an overview of the central objectives of DARIAH-DE with respect to the 'Archive-in-a-Box' intentions and to legitimate the present form of this document, we collected relevant passages from the initial DARIAH-DE proposal (as well as two other documents) and slightly refined on them. We marked central source-passages, subsumed them under an identifier (**AIB-SRC-number**, followed by reference to the
page) and subsequently extracted some central and/or some relational key requirements for further use within this document.

To cite and subsume the eight main source-passages we found:


**(AIB-SRC-2)** p.20 – "Obwohl DARIAH nicht zentral Langzeitarchivierung (LZA) durchführt, wird DARIAH in diesem Task lokale LZA-Aktivitäten unterstützen. Hierzu zählt vor allem die Integration von Technologien und Konzepten aus der LZA-Forschungslandschaft (z.B. aus PLANETS, CASPAR, nestor, kopal) in DARIAH, z.B. in das Softwarepaket Archive-in-the-box. Ebenso werden die in VCC 3 entwickelten Standards und Konzepte in die technische Infrastruktur integriert, was durch das VCC 1 vorbereitet werden kann."


**(AIB-SRC-6)** p.74 – "Wo sich AP 3.2 auf Wissenschaftler als Zielgruppe konzentriert, adressiert AP 3.4 speziell Forschungsdatenarchive und große Datensammlungen. Neben der Etablierung von interoperablem Datenstandards (siehe AP 3.3) berät dieses AP dabei auch in der Einbettung von Datenverwaltungsmechanismen in Forschungsdatensysteme. Darunter können Werkzeuge z.B. zur Datenvalidierung sein (z.B. JHOVE ) oder auch Prozeduren z.B. zur Beschreibung und Qualitätssicherung von Forschungsdaten. Technische Vorschläge werden hierbei eng mit AP 1.4 abge-
stimmt und wo möglich auch direkt z.B. im Archive-in-the-box Dienstkonzept umgesetzt."

(AIB-SRC-7) (Source: VCC1 Document of Work) "This service consists of an annotated reference list to existing software solutions for establishing a research archive. All solutions are evaluated for their pros and cons with respect to specific institutional/technical contexts. A how-to for ensuring interoperability with other DARIAH components will be included. This service does not, in itself, include support, training or advice on installation, data ingest, preservation, or maintenance. Furthermore this service does not include Persistent Identifier (PID) minting or institutional AAI facilities for the organization. If required PID services can be provided by DARIAH’s Persistent Identifier Service.

Contributions for DARIAH construction

- DARIAH-DE - ESciDoc (in which MPG has an important role)
- DARIAH-NL - EASY II (created by DANS)
- DARIAH-AT - Phaidra"

(AIB-SRC-8) (Source: DARIAH-Report D8.1.2 (preparation phase), pp.46/47, not cited because of great text quantity): this passage was not put into further consideration, because in contrast to DARIAH-project-proposal and VCC DOW it was outdated and in part too specific for affiliation

2.2. Requirements

The following high-level requirements (in bold) were extracted from these sources. Mentioned under each requirement is list of the sources (see section 2.1) as well as a set of comments that specify the extracted requirements onward.

AIB-REQ-1: AIB candidate is a software package

- AIB-SRC-4 - software to be installed locally ("bei sich ... installieren können")

AIB-REQ-2: AIB candidate must be locally installable

- AIB-SRC-1 - local technical systems ("lokale technische Systeme")
- AIB-SRC-4 - software to be installed locally ("bei sich ... installieren können")
- Comment 1: AIB candidate must not be a cloud service
- Comment 2: it may be offered by DARIAH partners as a hosted service
- Comment 3: it should be popular amongst archive hosters

AIB-REQ-3: AIB candidates are recruited from existing archive software

- AIB-SRC-1 - assistance for employing existing software ("Orientierungshilfen für bestehende Software")
- AIB-SRC-3 - designed and implemented ("entworfen und umgesetzt")
- Comment 1: a AIB candidate does not have be developed from scratch
AIB-REQ-4: AIB candidates must interoperate with DARIAH infrastructure

- AIB-SRC-1 - extensions for integration with DARIAH infrastructure
- AIB-SRC-4 - compatible with the DARIAH infrastructure ("mit der DARIAH-Infrastruktur kompatibel")
- Comment 1: a AIB candidate should support common standards (e.g. OAI-PMH, Dublin Core,...)
- Comment 2: DARIAH infrastructure services should support common standards

AIB-REQ-5: AIB candidates may be selectively extended to interoperate with the DARIAH infrastructure

- AIB-SRC-1 - extension for an integration ("bei einigen ausgewählten ... Anpassungen für eine Integration")
- AIB-SRC-2 - integration of technologies and concepts from LTA research ("Integration von Technologien und Konzepten aus der LZA-Forschungslandschaft; DARIAH&LZA")
- AIB-SRC-3 - designed and implemented ("entworfen und umgesetzt")
- Comment 1: not all AIB candidates are supposed to be extended
- Comment 2: a AIB candidate should support a plug-in system

AIB-REQ-6: AIB candidate should be easy to install and use

- AIB-SRC-3 - easily deployable repository environment ("einfach aufzusetzende Repositoriums- umgebung")
- AIB-SRC-3 - quick and simple ("schnell und unkompliziert")
- Comment 1: a AIB candidate should offer a VM, and installer or a installable package (e.g. RPM, Deb,...)
- Comment 2: as most AIB candidates do not offer that, part of the AIB implementation should be to document and facilitate the installation process
- Comment 3: to determine ease of use, online demonstration installations may be helpful
- Comment 4: to determine if installation is easy, online demo installations are not sufficient

AIB-REQ-7: AIB demonstrator is primarily conceptual work

- AIB-SRC-3 - AIB concept (Archive-in-the-Box-Konzept)
- AIB-SRC-4, AIB-SRC-6 - AIB service concept (Archive-in-the-Box Dienstkonzept)
- Comment 1: this does not contradict AIB-REQ-5/AIB-SRC-3 because the implementation is limited to extending selected candidates to interoperate with the DARIAH infrastructure
- Comment 2: for a proper evaluation of AIB candidates, test installations and experiments regarding the integration of DARIAH infrastructure may be necessary
2.3. First result

Summarizing the requirements extracted from the DARIAH-DE proposal, successful AiB-candidates would be

(a) already existing software packages (Req 1, 3),
(b) easily install-/useable on a local scale (Req 6, 2),
(c) already interoperating with according parts of DARIAH-DE basically (Req 4),
(d) extensible to improve interoperability with the DARIAH-DE infrastructure (Req 5).

This implies that inside DARIAH-DE the work on 'archive in a box' could only be

(e) conceptual work mainly (Req 7).

2.4. DARIAH-DE Dependencies (Interoperability)

While criteria (a) and (b) are self-explanatory, criteria (c) and (d) need to be further elaborated to make the relevant information explicit (regarding DARIAH-DE interoperability) in order to recommend a particular software-package.

The following services currently planned by DARIAH-DE (August 2012) are of interest for our recommendation of archive products. Any evaluated archive product should be compatible with as many as possible of the respective standards used by these services. Central DARIAH-DE-Services (bold) and their standards (italics) read as follows.

DARIAH Hosting – The hosting service can be used to run archive software directly on DARIAH hardware resources. This can be done either for production purposes or for some short-time testing. Offered service range from the provision of a clean VM with preferred operating system up to assisted installation of the requested archive software solution and help by integration with DARIAH infrastructure.

DARIAH-AAI – The authentication and authorization infrastructure (AAI) service is currently based on Shibboleth and OpenLDAP. Its preferred solution is Shibboleth and the underlying SAML 2.0 standard, especially the two profiles WebSSO and ECP.

DARIAH-BP –The bitstream preservation (BP) offers a file-based service, which provides longterm and redundant binary data storage using a storage virtualization system or a storage resource of choice. It can be accessed by a self-defined, HTTP-based RESTful API and is protected by the DARIAH-AAI. An archive-in-a-box system can use this service to store data objects as well as backups of archive's meta data. A solution may already define a storage abstraction layer for which a plug-in for the DARIAH-BP could be implemented. While metadata of each tested archive-application is usually stored in a database, it may be possible to store a backup of this metadata in the BP service as well.
DARIAH-PID – The persistent identifier (PID) service provides location-independent access to digital objects. Under the hood, the CNRI handle system is used.

DARIAH-CR – The collection registry (CR) provides a central service for administering and finding different digital archives. Metadata from this registry is harvested using OAI-PMH through DARIAH Collection Level Description Application Profile (DCLAP). This service does not directly interact with an archive-in-a-box solution. An archive owner could, however, set up an archive that hosts a collection and then manually register that collection in the DARIAH-CS.

DARIAH-SR – The schema registry (SR) represents crosswalks between metadata schemes that can be used to relate metadata from different collections using different metadata schemes. This function is primarily used by the DARIAH-GS. An archive-in-a-box solution does not directly interact with this service. An archive owner should use any of the metadata schemes registered in this service for the collection to be properly searchable using the DARIAH Generic Search service.

DARIAH-GS – The generic search (GS) service deals with finding of objects and evaluates relevance of collections in DARIAH-CR using crosswalks of the DARIAH-SR. The service is based on Lucene / Solr.

3. Selection of Candidates for an 'Archive-in-a-Box'

3.1. Collection of a first list of candidates

After carefully extracting the requirements for an 'Archive in a Box' as well as specifying the main interoperability features with primary DARIAH-DE services successful candidates should offer, we took a look at the already available products that could be employed to implement an 'Archive in a Box'.

For the evaluation, we leave out each underlying infrastructure project such as Fedora Commons, MyCoRe, eSciDoc, and TextGridRep, because most of them do not provide any end-user oriented user interface directly. Regarding applications, sometimes even additional effort is required to implement a generic or project specific user interface (see VIRR or FACES).

The following table should provide an overview of all initially evaluated candidates, each with a link to a support site. Because some of the packages are quite complex in their structure and involve different levels / layers (while each of this layers has to be tested separately because of possible alternatives on a particular layer) this table includes hints to the respective underlying component of each of the candidates as well.

---

1 See also another previous (2004) attempt on this matter:
<table>
<thead>
<tr>
<th>Candidate</th>
<th>Based on</th>
<th>Created by / supported through</th>
<th>Included</th>
</tr>
</thead>
<tbody>
<tr>
<td>DocPortal</td>
<td>-&gt; MyCoRe</td>
<td><a href="http://www.mycore.de:8291/content/below/index.xml">http://www.mycore.de:8291/content/below/index.xml</a></td>
<td>Yes</td>
</tr>
<tr>
<td>dSpace</td>
<td>-&gt; Fedora Commons</td>
<td><a href="https://wiki.duraspace.org/display/DSpace/Home">https://wiki.duraspace.org/display/DSpace/Home</a></td>
<td>Yes</td>
</tr>
<tr>
<td>Easy II</td>
<td>-&gt; Fedora Commons</td>
<td>DANS: <a href="http://www.dans.knaw.nl/en/content/categorieen/projecten/easy-ii-electronic-archiving-system-fedora">http://www.dans.knaw.nl/en/content/categorieen/projecten/easy-ii-electronic-archiving-system-fedora</a></td>
<td>No</td>
</tr>
<tr>
<td>ePrints</td>
<td>standalone</td>
<td><a href="http://www.eprints.org">http://www.eprints.org</a></td>
<td>yes</td>
</tr>
<tr>
<td>eSciDoc</td>
<td>-&gt; Fedora Commons</td>
<td>MPDL: <a href="https://www.escidoc.org">https://www.escidoc.org</a></td>
<td>indirectly</td>
</tr>
<tr>
<td>Fedora Commons</td>
<td></td>
<td>DuraSpace: <a href="http://fedora-commons.org/">http://fedora-commons.org/</a></td>
<td>indirectly</td>
</tr>
<tr>
<td>Invenio</td>
<td>standalone</td>
<td>CERN: <a href="http://cdsware.cern.ch/invenio/documentation.html">http://cdsware.cern.ch/invenio/documentation.html</a></td>
<td>yes</td>
</tr>
<tr>
<td>MyCoRe</td>
<td></td>
<td><a href="http://www.mycore.de/index.html">http://www.mycore.de/index.html</a></td>
<td>indirectly</td>
</tr>
<tr>
<td>Mopseus</td>
<td>-&gt; Fedora Commons</td>
<td>Athena: <a href="http://194.177.192.14/mopseus/quick_guide_en.html">http://194.177.192.14/mopseus/quick_guide_en.html</a></td>
<td>no</td>
</tr>
<tr>
<td>OPUS 3</td>
<td>standalone</td>
<td><a href="http://www.opus-repository.org/opus3/index.html">http://www.opus-repository.org/opus3/index.html</a></td>
<td>yes</td>
</tr>
<tr>
<td>OPUS 4</td>
<td>standalone</td>
<td><a href="http://opus4web.zib.de">http://opus4web.zib.de</a></td>
<td>yes</td>
</tr>
<tr>
<td>Phaidra</td>
<td>standalone</td>
<td>Universität Wien: <a href="http://phaidraservice.univie.ac.at/das-system-phaidra/">http://phaidraservice.univie.ac.at/das-system-phaidra/</a></td>
<td>no</td>
</tr>
<tr>
<td>TextGridLab</td>
<td></td>
<td><a href="http://www.textgrid.de">http://www.textgrid.de</a></td>
<td>no</td>
</tr>
<tr>
<td>TextGridLab</td>
<td>-&gt; TextGridRep</td>
<td><a href="http://www.textgridrep.de">http://www.textgridrep.de</a></td>
<td>yes</td>
</tr>
<tr>
<td>VIRR</td>
<td>-&gt; eSciDoc</td>
<td>MPDL: <a href="http://virr.mpdl.mpg.de">http://virr.mpdl.mpg.de</a></td>
<td>no</td>
</tr>
<tr>
<td>VuDL</td>
<td>standalone</td>
<td><a href="http://vudl.org">http://vudl.org</a></td>
<td>no</td>
</tr>
<tr>
<td>ZOpenArchives</td>
<td>standalone</td>
<td><a href="http://www.pentila.com/solutions/education/zoai-1/zopenarchives-1/zopenarchives">http://www.pentila.com/solutions/education/zoai-1/zopenarchives-1/zopenarchives</a></td>
<td>no</td>
</tr>
</tbody>
</table>
3.2. Invalid candidates

Some candidates on this initial list lacked central key-features that a successful candidate for an ‘Archive-in-a-Box’ should have, thus they were removed from the list:

- EASY 2
- Faces (MPIB)
- Mopseus
- Phaidra
- VIRR
- VuDL
- ZOpenArchives

The following section will give a short comment why each of the candidates was removed from the list in the first place. It will be followed by the list of the remaining candidates that came into closer examination.

3.2.1. EASY 2

The main reason for removing EASY 2 is the currently relatively closed nature of the solution. The package lacks open documentation and options for download of the package or its sourcecode. So through accessibility- and time-issues we could not investigate it further for the concept at hand. We may take another look at the package and its advancements by contacting the developers for the following milestone, if this situation changes.

3.2.2. Faces

Faces is an application built on eSciDoc. It was removed from the candidate list because it seems to be custom-built for a particular project: a lifespan database of adult emotional facial stimuli, mainly pictures. It does currently not seem to fit into the mainly philological / textual orientation of DARIAH-DE.

3.2.3. Mopseus

Mopseus – though it is produced and supported by a DARIAH-partner – was removed from the list of candidates because it seems to no longer be developed or maintained and therefore would not be a sustainable solution.

3.2.4. Phaidra

The main reason for the exclusion of Phaidra from the list was the current closed nature of the solution software-side and a relatively major lack of key features of our criteria-list. We may take another look at the package and its advancements by contacting the developers for the next milestone if the situation changes.
3.2.5. VIRR

VIRR is an application built on eSciDoc. The main reason for its exclusion was that this package seems to be custom-built for a particular project only – on various legal artifacts of the period of the Holy Roman Empire.

3.2.6. VuDL

Although VuDL meets many of our requirements and supports some standards (METS, OAI, scanning options, eXist-XML-database, plans incorporation of MODS, ePub and TEI, has user-groups as well as an issue tracker), its focus is strongly on image data and thus does not fit into the mainly philological / textual orientation of DARIAH-DE. According to your intentions and usage you may take a closer look but you will not get DARIAH-DE support currently.

3.2.7. ZOpenArchives

The main reason for the removal of ZOpenArchives from the list was that software-development for this package seemed to be discontinued thus this seems not to be a sustainable solution.

3.3. Candidates for further evaluation

These remaining candidates were evaluated further (see table 1).

Table 2: list of evaluation-candidates for an 'Archive-in-a-Box'

<table>
<thead>
<tr>
<th>Candidate</th>
<th>Based on</th>
<th>Created by / supported through</th>
</tr>
</thead>
<tbody>
<tr>
<td>DocPortal</td>
<td>-&gt; MyCoRe</td>
<td><a href="http://www.mycore.de:8291/content/below/index.xml">http://www.mycore.de:8291/content/below/index.xml</a></td>
</tr>
<tr>
<td>dSpace</td>
<td>-&gt; Fedora Commons</td>
<td><a href="https://wiki.duraspace.org/display/DSPACE/Home">https://wiki.duraspace.org/display/DSPACE/Home</a></td>
</tr>
<tr>
<td>ePrints</td>
<td>standalone</td>
<td><a href="http://www.eprints.org">http://www.eprints.org</a></td>
</tr>
<tr>
<td>eSciDoc</td>
<td>-&gt; Fedora Commons</td>
<td>MPDL: <a href="https://www.escidoc.org">https://www.escidoc.org</a></td>
</tr>
<tr>
<td>Invenio</td>
<td>standalone</td>
<td>CERN: <a href="http://cdsware.cern.ch/invenio/documentation.html">http://cdsware.cern.ch/invenio/documentation.html</a></td>
</tr>
<tr>
<td>OPUS 3</td>
<td>standalone</td>
<td><a href="http://www.opus-repository.org/opus3/index.html">http://www.opus-repository.org/opus3/index.html</a></td>
</tr>
<tr>
<td>OPUS 4</td>
<td>standalone</td>
<td><a href="http://opus4web.zib.de">http://opus4web.zib.de</a></td>
</tr>
<tr>
<td>TextGridLab</td>
<td>-&gt; Text-GridRep</td>
<td><a href="http://www.textgrid.de">http://www.textgrid.de</a></td>
</tr>
<tr>
<td>TextGridRep-Search</td>
<td>-&gt; Text-GridRep</td>
<td><a href="http://www.textgrid.de/">http://www.textgrid.de/</a></td>
</tr>
</tbody>
</table>
4. Criteria

As base of our further evaluation we differentiated initial criteria and added them into a matrix while meanwhile differentiating the latter. V3 of this matrix formed the basis to build the following evaluation and its main criteria on. (see result in appendix 1)

4.1. DARIAH-DE interoperability (max: 5 Points)

A successful candidate for an 'Archive-in-a-box' should support at least three out of the following five central DARIAH-DE Infrastructure criteria (interfaces; one point of one of the criteria is given when all specs apply to the candidate).

1. **DARIAH-AAI**: we consider this service supported if a system supports Shibboleth and/or LDAP (preferred: Shibboleth)

2. **DARIAH-BP**: we consider that a system potentially supports the DARIAH-BP if it provides a storage abstraction layer that supports alternative or (at best) multiple different storage backends.

3. **DARIAH-PID**: we consider a system compatible with the DARIAH-PID service if it provides integration with the CNRI handle system.

4. **DARIAH-GS**: we consider a system to be compatible with this service if it offers metadata to be harvested via OAI-PMH.

5. **DARIAH Hosting**: Though we do not plan to use the according services itself, we expect a system to be suitable for hosting if cloud-based and/or commercially supported hosting-services exist. Their existence is taken as a hint that stable 'boxed'-scenarios can be build.

4.2. Customization- / Extension-Efforts (max: 5 Points)

If an otherwise successful candidate does not initially support all DARIAH-DE Interoperability-criteria, it at least should have relatively minimal extension efforts in the remaining fields. The according point-value is oriented on each result regarding 4.1. (marginal differences result from fine-granularity and move within a continuum ranging from 0 to 5 points with 0,5 point-steps). The candidate receives

- 5 points if only none or minor extension-efforts are conceivable in remaining fields;
- 3 points if achievable extension-efforts need to be made
- 0 point if the extension-efforts are not achievable within DARIAH-DE; this leads to an exclusion of the candidate out of considerations.
4.3. Sustainability of solution (max: 5 Points)

An adequate candidate should support as much sustainability criteria as possible (one point for one criterion counts when at least two of each respective properties apply to the candidate).

1. Tested solution, scalability, stable **use-cases** (quantity, language support), DINI-support.
3. Currently **conceivable**, community-driven future development; **User-Feedback**: online as well as offline (nearness).
4. Options for **migration** (esp. common standards for importing / exporting).
5. Financial support / stable **institutional support** (e.g. governmental / educational agency).

4.4. Calculation of results

The points for each solution are then summed up to a total score.

For the next milestone we will examine the highest-ranking products.

We recommend that products not reaching at least 9 points should not be considered for common use by DARIAH-DE at the present time (except for special scenarios as mentioned in the corresponding sections).
5. Results

According to a DARIAH-DE-internal, wiki-based evaluation-table we used (a screen-shot of the matrix in its latest state is found in appendix 1), we came to the following conclusions regarding the remaining candidates ('1p' means '1 point given in the corresponding context'). For more references, also take a look at appendix 1.

5.1.1. DocPortal

<table>
<thead>
<tr>
<th>DARIAH-DE interoperability</th>
<th>2,5 / 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customization- / Extension-Efforts</td>
<td>2,5 / 5</td>
</tr>
<tr>
<td>Sustainability of solution</td>
<td>3 / 5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>8 / 15</strong></td>
</tr>
</tbody>
</table>

DocPortal is an example application for publication management of MyCoRe, a German semi-commercial framework for digital asset management, both under GPL. Both products work on major modern operating systems.

MyCoRe supports LDAP (and IP-Ranges) for authentication, but does not offer a stable support for Shibboleth currently (0.5p). MyCoRe offers a VFS layer\(^2\) with S/FTP and Samba backends, so a DARIAH-BP backend should be realizable (1p). It does not support CNRI (just URN with DNB resolver). There is, through workarounds, support for OAI-PMH (1p) while it delivers metadata through Z39.50/ID3 and data trough SOAP for searching. It does currently support main media types except TEI. Thus, DocPortal has some DARIAH-DE-interoperability in the current state (2,5 of 5 points).

Though the solution is well documented, there are no semi-commercial hosting options currently, some DARIAH-DE-extension-efforts (2,5 of 5 points) could be expected, esp. around CNRI-Handles.

Regarding sustainability (3 of 5 points), the solution has DINI-certified instances (1p), has good and sustained channels for user-/developer-feedback (1p) and supports extensions via modules and plug-ins (even through a 'Geschäftsstelle', 1p). But it only offers Pica3-migration possibilities, hast just a few use-cases currently and no English language support anymore (though I18N is supported).

With an overall result of 8 out of 15 points in our evaluation, the solution could have some promising sides, but must undergo some changes in the next time to be recon-

\(^2\) [http://dev.thulb.uni-jena.de/viewvc/archive/branches/MYCORE_DIST/mycore/documentation/ProgGui-de/ProgrammerGuide.odt?revision=367&view=co&pathrev=485](http://dev.thulb.uni-jena.de/viewvc/archive/branches/MYCORE_DIST/mycore/documentation/ProgGuide-de/ProgrammerGuide.odt?revision=367&view=co&pathrev=485)
sidered for future evaluation. In its current state DARIAH-DE would have to invest considerable time to extend this solution to meet our requirements (excluded at this point for an official DARIAH-DE ‘Archive in a Box’).

5.1.2. dSpace

<table>
<thead>
<tr>
<th>DARIAH-DE interoperability</th>
<th>3,5 / 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customization-/ Extension-Efforts</td>
<td>3,5 / 5</td>
</tr>
<tr>
<td>Sustainability of solution</td>
<td>4 / 5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>11 / 15</strong></td>
</tr>
</tbody>
</table>

dSpace is an archival system for electronic prints and other kinds of digital academic content that builds up on Fedora Commons though a BSD-like license. In fact, both were developed by the same group of people in the US, dSpace in the first place for MIT, later on it got supported by HP. Both products are implemented in Java and work on major modern operating systems and network platforms.

dSpace supports Shibboleth when running behind a Shibboleth-enabled Apache HTTPD server. Thus, it could be interoperable with the DARIAH-AAI service (1p). dSpace supports multiple storage backends including local files and SRB. There is a bit-preservation service available called 'DuraCloud' which can be (commercially) integrated within dSpace. Thus, integration with the DARIAH-BP could be feasible, but must be tested (0,5p). dSpace seems to directly support the CNRI Handle service so it should be possible to integrate the DARIAH-PID service (1p). While supporting major textual media-types, excluding TEI, OAI-PMH metadata can be harvested from dSpace. Thus, it should interoperate with the DARIAH Generic Search in major parts (1p). Overall, this solution should offer some DARIAH-DE-interoperability (3,5/5 points).

Mild DARIAH-DE-extension-efforts need to be made (3,5/5 points), in particular testing, integrating and possibly extending to use DARIAH-BP and implementing a TEI workaround. Some Language-files for German are available.

There seem to be channels for user-/ developer-feedback, in Europe via 'ambassadors'. Development is in part community driven, sometimes 'nearness' could be a problem (1p). dSpace offers support and hosting-options mainly in commercial frameworks (1p). It is extensible via a plug-in system while trusted (0,5p) humanities-solutions must be found around the hundreds of working, mostly middle-scale in-

---

3 http://www.dspace.org/1_7_0Documentation/Storage%20Layer.html#StorageLayer-BitstreamStore
4 https://wiki.duraspace.org/display/DSPACE/I18nSupport
stances (none DINI-certified currently, 1p). dSpace offers at least basic migration functionalities (0,5p). It should be a sustainable, although not a mainly european solution (no experiences of DARIAH-partners can be integrated and only external, partly commercial channels could be used) – (3,5 of 5 points).

Overall 11 out of 15 points: dSpace could be a good choice for an Application-as-a-Service with relatively reduced time-investment in existing complex frameworks (recommendation). For the next milestone it will be tested in wider contexts with a look if the at first glance relatively good interoperability-factors meet expectations and also to existing hosting-options, forks of code-evolution and appropriation-needs for text-based European humanists.

5.1.3. ePrints

<table>
<thead>
<tr>
<th>DARIAH-DE interoperability</th>
<th>3,25 / 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customization- / Extension-Efforts</td>
<td>2,5 / 5</td>
</tr>
<tr>
<td>Sustainability of solution</td>
<td>5 / 5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>10,75 / 15</strong></td>
</tr>
</tbody>
</table>

ePrints is an UK-based document-management solution, mainly for educational purposes. Its is written in Perl, licensed under GPL and runs on major modern operating systems via RPM or MSI (MSI is explicitly not recommended).

ePrints does support LDAP and Shibboleth (the latter seemingly with minor difficulties, 0,75p). Since version 3.2, ePrints supports a hybrid storage architecture which allows to distribute data across different storage locations including local storage and cloud storage. Therefore, extending this with a plug-in to use the DARIAH-BP as another cloud storage seems possible (1p). At least two instances does support CNRI through a workaround (must be tested; 0,5p), OAI-PMH is supported (1p). Its search can be extended with plug-ins, e.g for Xapian. Search results can be exported to several formats and services (including OpenSearch, BibTeX, twitter, Google-Earth, and RSS). ePrints has some DARIAH-DE-interoperability in the current state (3,5 of 5 points), in some cases not straight forward.

This does result in some DARIAH-DE-extension-efforts (2,5 of 5 points), esp. regarding the situation around slightly differing formats and testing a CNRI-workaround. Some language-files should need to be translated. It should be considered that

---

5 http://www.eprints.org/software/training/3.2/admin/storage_tutorial.php, also see: http://eprints.soton.ac.uk/265818/
6 http://www.arl.org/sparc/bm~doc/IR_Guide_&_Checklist_v1.pdf; also eMails with Tim Brody, Lead Developer
ePrints is implemented in Perl while most efforts currently underway in DARIAH-DE are realized in Java.

Regarding sustainability (5 of 5 points), the solution not only has DINI-certified installations and a cybermetrics rating of 31 (1p) but as only candidate supports 'Repository Policy options' through OpenDOAR. ePrints supports metadata extraction from major textual input-formats including different TeX-flavors but not TEI (1p). Its migration-options include a very wide variety of major formats (dSpace, OAI-PMH, Dublin Core, JSON ArXiv and others, 1p). Themes / branding in ePrints seems easy (lib/themes) and language-customization is supported. It has some sustained channels for user- / developer-feedback including training, courses, events, hosting and institutionalized consultancy via a trustworthy european university-background (University of Southampton; 1p), has a plug-in / app-concept (67 plug-ins currently, 1p) and seems to have a sustained future development.

With an overall result of 11 out of 15 points in our evaluation, the solution is recommended according to sustainability and open formats but for the next milestone it needs to be tested interface- and format-side to see if the found workarounds regarding interoperability meet expectations.

5.1.4. Invenio

DARIAH-DE interoperability 2 / 5
Customization- / Extension-Efforts 1,5 / 5
Sustainability of solution 4 / 5
Total 7,5 / 15

Invenio is a Swiss digital library and document repository. The code is written in Python and released under the GPL. It runs on major modern Linux platforms through a tarball and RPM.

Invenio does support LDAP and Shibboleth (1p). There seems to be no storage abstraction layer by which the DARIAH-BP could be integrated and no information on a dedicated bit-preservation-interface or destined hosting-options. Invenio does not seem to support CNRI handles straight forward. OAI-PMH is supported (1p). Its search-functionality is done through a Google-like multi-level API with citation-metrics, referencing and search-engine-support. Invenio has a reduced DARIAH-DE-interoperability in the current state (2 of 5 points).
This situation does result in considerable DARIAH-DE-extension-efforts (BP, PID; 1,5 of 5 points). As with some other solutions, main parts of the source-code are rather uncommented currently, but support should work. It should be considered that Invenio is implemented in Python while most efforts currently underway in DARIAH-DE are realized in Java.

Regarding sustainability (4,0 of 5 points), the solution has 30 installations worldwide. It seems to be really robust, esp. regarding a huge quantity of documents (over 1 Mio. tested in one instance (1p)). Invenio supports languages though I18N (currently English) but seems to support just a few major textual input-formats (excluding TEI). Its migration-options include major formats (MarcXML, Google Scholar, arXiv, oaidec2 and others; 1p). Invenio, while themes / branding seems easy, has some channels for user- / developer-feedback including meetings via trustworthy university-backgrounds (1p), currently only around high-energy physics (CERN, DESY, EPFL, FNAL, SLAC). It supports extensions via 'modules' (1p) and seems to have a sustained future development (through participating on some at first rather closed CERN-forks of development later on).

With an overall result in our evaluation of 7,5 out of 15 points, the solution seem to be recommendable only for specific projects, e.g. projects with large scale in a rather closed form, possibly DARIAH-DE-searchable, but without any handle-system – with considerations on an own BP-solution as well as a forked code-evolution besides physics-disciplines (excluded at this point for an official DARIAH-DE 'Archive in a Box').

**5.1.5. OPUS 3**

<table>
<thead>
<tr>
<th>DARIAH-DE interoperability</th>
<th>1,5 / 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customization- / Extension-Efforts</td>
<td>1,5 / 5</td>
</tr>
<tr>
<td>Sustainability of solution</td>
<td>4 / 5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>7 / 15</strong></td>
</tr>
</tbody>
</table>

OPUS 3 is a German document-server (institutional repository), mainly used for publications of Universities ('Hochschulschriften') in most cases. Looking at the access-side of the software itself, it seems to be a rather closed solution currently, although the sourcecode is under the GPLv2. OPUS 3 runs on major modern operating systems and network platforms via a tarball, but without any installer or script.

OPUS 3 does not support LDAP or Shibboleth regarding AAI (just IP-Range). Currently, there seems to be no storage abstraction layer, although, this issue seems to
be on the roadmap for the near future (0.5p). It does not support CNRI (just URN-PIDs for DNB), OAI-PMH is supported (even enriched with XMetaDissPlus2.0 for its – in our evaluation rather singular – DNB-interface, 1p). OPUS 3 does not come with an integrated search engine, instead it relies on external search engines like Google for its full text. To sum it up, OPUS 3 has a weak DARIAH-DE-interoperability in its current state (1,5 of 5 points).

Although the solution is well documented and the rather closed situation regarding sourcecode could have beneficial sides for closed solutions, that does result in some major DARIAH-DE-extension-efforts (1,5 of 5 points), which includes needs to customize some language-support-files.

Regarding sustainability (4 of 5 points), the solution has a lot of DINI-certified installations and a middle global cybermetrics\(^7\) ranking (1p). While its textual input-formats only support PostScript and PDF, its migration-options include major formats (Pica, BibTex, Dublin Core, XMetaDissPlus, epubli.de; 1p), some (esp. the DNB-interface) surely may come in handy in use-cases we did not think of at start. OPUS 3 has channels for user- / developer-feedback via trustworthy former DFG- and BMBF-financed university-backgrounds (1p), seems to support extensions (1p) and seems to have a sustained future development.

With an overall result of 7 out of 15 points in our evaluation, the solution could have some promising solutions user-side but while its use is rather limited, it must undergo heavy changes to be reconsidered for future evaluations. In its current state DARIAH-DE would have to invest considerable effort to extend this solution to meet requirements (only recommended in limited form within intended and closed use-cases – then also see OPUS 4; excluded at this point of recommendation as an official DARIAH-DE 'Archive in a Box').

\(^7\) http://repositories.webometrics.info/index.html
5.1.6. OPUS 4

<table>
<thead>
<tr>
<th>Evaluation Area</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>DARIAH-DE interoperability</td>
<td>2,25 / 5</td>
</tr>
<tr>
<td>Customization- / Extension-Efforts</td>
<td>2 / 5</td>
</tr>
<tr>
<td>Sustainability of solution</td>
<td>4,5 / 5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>8,75 / 15</strong></td>
</tr>
</tbody>
</table>

OPUS 4 is a German document-server (institutional repository), its intended use is for publications of Universities ('Hochschulschriften'). Other than OPUS 3, it has open access to the software itself and its sourcecode. It is written in PHP and Java and is licensed under the GPL 3. It runs on major Linux platforms via a tarball with unix-installation-scripts.

OPUS v4.22 (German) does only support LDAP (and IP-Range) regarding AAI (0,25p). There is currently no storage abstraction layer that could be used to plug in the DARIAH-BP service (currently only in planning state), but KOBV offers hosting\(^8\) (1p). OPUS 4 does not support CNRI, but at least URN-UUID-PIDs with xepicur-resolver for DNB, OAI-PMH is supported (enriched with XMetaDissPlus2.0 for its DNB-interface; 1p). Its search-functionality is based on Solr. To sum it up, OPUS 4 has a relatively mixed DARIAH-DE-interoperability in the current state (2 of 5 points).

This does result in some DARIAH-DE-extension-efforts (2 of 5 points), esp. regarding more info on the situation around BP. As with other solutions, some parts of the source-code seem to be rather undocumented / uncommented currently, but support seems to work.

Regarding Sustainability (4,5 of 5 points), the solution has DINI-certified (2010) installations besides a lot others (1p). It seems it be robust regarding a medium quantity of documents (10000 tested elsewhere) and secure, as it supports, while being undocumented, GPG (0.5p). It also supports major languages other than German (currently only in v4.02 of Opus) and major textual input-formats (excluding TEI). Its migration-options include major formats (incl. OPUS 3, citations and bibliographies), some (esp. the enhanced DNB-interface) may come in handy in special cases (1p). Themes / branding seems easy. OPUS 4 has some channels for user- / developer-feedback including workshops via trustworthy former DFG- and BMBF-financed university-backgrounds (1p), seems to support extensions (1p) and seems to have a sustained future development.

\(^8\) [http://www.kobv.de/bib_opus_archvierung.html](http://www.kobv.de/bib_opus_archvierung.html)
With an overall result of 8,5 out of 15 points in our evaluation, the solution is not recommended beyond intended use-cases (with middle document quantity) though it has extension-potentials beyond this in a variety of ways.

5.1.7. PubMan

DARIAH-DE interoperability 3,5 / 5  
Customization- / Extension-Efforts 3,5 / 5  
Sustainability of solution 5 / 5  
**Total** 12 / 15

PubMan is a document management application built on eSciDoc, which is a German system for eScience-aware knowledge and information management, build on Fedora by the DARIAH-DE partner MPDL. The code of the infrastructure-solution as well at its tested application PubMan ('Publication Management') run under a CDDL 1.0 (OSI)-licence. Both work on major modern operating systems via Java and an lzPack installer.

eSciDoc does support Shibboleth and LDAP (1p) regarding AAI. As storage backed, eSciDocs repository service currently builds on Fedora Commons, REST is supported. There seems to be no direct plug-in mechanism at the moment to support alternative storage backends, iRODS- and SRB-support seems to be only planned.\(^9\) An alternative integration scenario where metadata is stored in eSciDoc and actual documents are stored in an alternative storage has been considered within the TextGrid project (0,5p).\(^10\) For metadata MODS and METS are supported. The PID Manager supports CNRI through the Handle System of CNRIs Digital Object Architecture (1p), OAI-PMH is supported (in certain versions; 1p). eSciDoc's and PubMan's search-functionality run through REST, SRW/SRU (eSciDoc) and SOAP as well as CoNe and OpenSearch (PubMan). Together both have some DARIAH-DE-interoperability in the current state (3,5 of 5 points).

This should result in some relatively minor DARIAH-DE-extension-efforts (3,5 of 5 points), esp. testing.

Regarding Sustainability (5 of 5 points), the solution has no DINI-certified installations currently but semi-commercial, university-driven services that help getting DINI certification (1p) e.g. through MPI or 'KnowEsis' (consultancy, support, development, im-

plementation, even hosting). eSciDoc/PubMan supports major languages, internationalization is done currently (1p); it has a working service / app-concept; major textual input-formats should be supported, TEI-support and other key features for textual work are currently been developed in MPs project ‘digitization lifecycle’ (1p). PubMan’s migration-options include a variety of major formats (ArXiv, eDoc, eSciDoc, BibTeX) and include even Zotero- and Wordpress-interfaces (1p). Themes / branding on the double-solution seems to be relatively easy. Both have channels for user- and developer-feedback, including user meetings, mailing lists and different support teams through trustworthy MPI-backgrounds that as well guarantee eSciDoc’s sustainable development through elaborate plans to distribute future tasks (1p).

With an overall result of 12 out of 15 points in our evaluation, the solution is recommended for further testing for the next milestone. If results meet expectations, resulting time could be invested to test other candidates or custom interfaces that seem to be build currently (e.g. to TextGrid).

5.1.8. TextGridRep

<table>
<thead>
<tr>
<th>DARIAH-DE interoperability</th>
<th>4 / 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customization- / Extension-Efforts</td>
<td>3,5 / 5</td>
</tr>
<tr>
<td>Sustainability of solution</td>
<td>5 / 5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>12,5 / 15</strong></td>
</tr>
</tbody>
</table>

TextGrid consists of several components including TextGridLab, an elaborate desktop application for scientists dealing with textual sources through digital state-of-the-art tools, TextGridRep, a repository-solution and TextGridRepositorySearch, a search-solution for such repositories. All parts are 'mainly' under LGPL 3. TextGridLab is a Java-based stand-alone application that runs on major current operating systems, while TextGridRep and its respective search are web-applications / web-services. Currently a rather independent TextGridRep-VM is being worked on in addition (access to beta is given, no documentation available yet). We build our evaluation on information on an early candidate of this latter solution.

TextGridRep thus does support LDAP and Shibboleth (1p), while OAI-PMH and MARC 21 are not supported currently. TG-CRUD is an included service largely similar to the DARIAH-BP, both are being worked on by a DARIAH-DE Partner. The solution offers possibilities of accessing multiple storage points, it is conceivable to implement a TextGrid-based solution on top of the DARAH-BP (1p). Textgrid uses the GWDG Handle System (1p) and offers sophisticated search-functionality through Lu-
cene (supports Corpus Search, a semantic as well as a management and analysis system with linguistic / lexicographical capabilities). Hosting is beening worked on currently through the development of specific, stable VM-images (1p). TextGridRep has good DARIAH-DE-interoperability currently (4 of 5 points). This does result in minor DARIAH-DE-extension-efforts, which would be accomplished mainly though Partners other than Darmstadt (3,5 of 5 points)\textsuperscript{11}, e.g. OAI-PMH support, DARIAH-BP integration and some new GUI esp. for an extended TextGridRep-VM with major DARIAH-DE-Interoperability and support at least for AiB-use-cases and scenarios being tested in the near future (see section 6, p.27).

Regarding Sustainability (5 points), this leading german repository solution has support on DINI-certified installations (1p). It supports German and English languages as well as major scientifically required textual input-formats – currently it is the only solution with a working TEI solution out-of-the-box (1.5p). Its migration-options include TextGrid itself as well as METS and will support other formats, extensions for some eSciDoc-Interoperability are being worked on,\textsuperscript{12} Themes / branding of the TextGridRep-VMs seem to be easy (1p). TextGrid has sustained and near channels for user- / developer-feedback including issue trackers, workshops, user summits through a trustworthy, former BMBF-financed university-background and wide networks including a robust sustainability-strategy for future development (1.5p).

With an overall result of 12,5 / 15 points in our evaluation, this solution of a DARIAH-DE-partner (in form of a special TG-Rep-VM currently under development) is recommended for future tests.

\textsuperscript{11} according to a telephone-conference on 21.08.012
\textsuperscript{12} http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_R1-3-2_Wege_zur_Verknuepfung_von_eSciDoc_und_TextGrid__erstes_Konzept_.pdf
Overview: Table of Results (simplified via point-values)

As a result of our evaluation we constructed a basic overview of all candidates evaluated (see table 3) with respective results as point-values (for in-depth evaluation results see appendix 1).

Table 3: evaluated candidates for an 'Archive-in-a-Box' with respective point-values

<table>
<thead>
<tr>
<th>Candidate</th>
<th>Based on</th>
<th>Created by / support through</th>
<th>Score (I= Interoperability; E=Extension Efforts; S=Sustainability)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DocPortal</td>
<td>--&gt;MyCoRe</td>
<td><a href="http://www.mycore.de:8291/content/below/index.xml">http://www.mycore.de:8291/content/below/index.xml</a></td>
<td>I 2,5; E 2,5; S 3 = 8 (excl.)</td>
</tr>
<tr>
<td>dSpace</td>
<td>standalone</td>
<td><a href="https://wiki.duraspace.org/display/DSPACE/Home">https://wiki.duraspace.org/display/DSPACE/Home</a></td>
<td>I 3,5; E 3,5; S 4 = 11</td>
</tr>
<tr>
<td>ePrints</td>
<td>standalone</td>
<td><a href="http://www.eprints.org">http://www.eprints.org</a></td>
<td>I 3,25; E 2,5; S 5 = 10,75 (nice-to-have)</td>
</tr>
<tr>
<td>Invenio</td>
<td>standalone</td>
<td>CERN: <a href="http://cdsware.cern.ch/invenio/documentation.html">http://cdsware.cern.ch/invenio/documentation.html</a></td>
<td>I 2; E 1,5; S 4 = 7,5 (excl.)</td>
</tr>
<tr>
<td>OPUS 3</td>
<td>standalone</td>
<td><a href="http://www.opus-repository.org/opus3/index.html">http://www.opus-repository.org/opus3/index.html</a></td>
<td>I 1,5; E 1,5; S 4 = 7 (excl.)</td>
</tr>
<tr>
<td>OPUS 4</td>
<td>standalone</td>
<td><a href="http://opus4web.zib.de">http://opus4web.zib.de</a></td>
<td>I 2,25; E 2; S 4,5 = 8,75 (excl.)</td>
</tr>
<tr>
<td>PubMan</td>
<td>--&gt; eSciDoc</td>
<td>MPDL: <a href="http://test-pubman.mpdl.mpg.de:8080/pubman/">http://test-pubman.mpdl.mpg.de:8080/pubman/</a></td>
<td>I 3,5; E 3,5; S 5 = 12</td>
</tr>
<tr>
<td>TextGridLab</td>
<td>--&gt; Text-GridRep</td>
<td><a href="http://www.textgrid.de">http://www.textgrid.de</a></td>
<td>I 4; E 3,5; S 5 = 12,5</td>
</tr>
</tbody>
</table>

6. Expectably interoperable and sustainable candidates / work for the next milestone

According to our findings and the results of the evaluation, for the next milestone M1.4.1.2 we will do test-installations, in-depth-examinations and (if needed) customizations of the following products in this order of priorities:

1. TextGridRep (VM; 12,5 points),
2. PubMan (eSciDoc; 12 points)
3. dSpace (11 points),
4. ePrints (10,75 points)
The first three candidates will be put through full integration tests with DARIAH-DE. Priority (4.) is a merely 'nice-to-have'-candidate we will test if we got enough time left.

The planned future tests of the candidates for M1.4.1.2 will include experimental scenarios to see possibilities, open questions of DARIAH-DE-interoperability as well as usability of each solution in action: to extend the candidates and gain at least one stable candidate.13

Some of these planned scenarios for each candidate would be:

- **Scenario 1:** migrate. An institution or individual delivers already existing digital data: migrate it into the DARIAH-DE Infrastructure via 'archive-in-a-box'.

- **Scenario 2:** introduce. In an already existing humanities center no digital data does exist yet: actions are taken to step inside the DARIAH-DE Infrastructure through 'archive-in-a-box'.

- **Scenario 3:** be compatible. Two or more humanities data centers converge technologically: through usage of an 'archive-in-a-box'.

- **Scenario 4:** be useable. Each solution will be tested by at least two textual humanists according to their specific scientific needs.

- **Scenario 5:** show advantages. Each solution will be put through tests to evaluate advantages to regular forms of dealing with concerned matters in the humanities (research-question, workflow, concentration on main research-goals)

According to the results of the document at hand, to ensure and sustain current as well as future DARIAH-DE interoperability, the next milestone M1.4.1.2 will also include and document patches as well as extensions for its recommended solution(s) (at least according to each finding of expectable extension-needs above).

The result will be ready-to-use software-solutions in form of pre-configured virtual machines that work 'out-of-the-box'.

---

13 While in some parts of development the document at hand lacked to follow designated internal procedures (according to M2.2.4, p.7) after a internal PM reallocation, M1.4.1.2 will follow them.
7. Appendix 1: Evaluation-Matrix (complex)

references: https://dev2.dariah.eu/wiki/display/DARIAHDE/AP1.4+Archive-in-a-Box

<table>
<thead>
<tr>
<th>Project</th>
<th>Documentation</th>
<th>Helpdesk</th>
<th>Support</th>
<th>Test PID integration</th>
<th>Test BP integration</th>
<th>Test Platform</th>
<th>Documentation PDF</th>
<th>Documentation Help-Files: CC</th>
<th>Testability Required</th>
<th>Testability Provided</th>
</tr>
</thead>
<tbody>
<tr>
<td>jOAI</td>
<td>[2.0] (Devel)</td>
<td>[2.0]</td>
<td>[2.0]</td>
<td>[2.0]</td>
<td>[2.0]</td>
<td>[2.0]</td>
<td>[2.0]</td>
<td>[2.0]</td>
<td>[2.0]</td>
<td>[2.0]</td>
</tr>
<tr>
<td>Hibernate</td>
<td>[3.4] (Devel)</td>
<td>[3.4]</td>
<td>[3.4]</td>
<td>[3.4]</td>
<td>[3.4]</td>
<td>[3.4]</td>
<td>[3.4]</td>
<td>[3.4]</td>
<td>[3.4]</td>
<td>[3.4]</td>
</tr>
<tr>
<td>MyCoRe</td>
<td>[2.2] (Devel)</td>
<td>[2.2]</td>
<td>[2.2]</td>
<td>[2.2]</td>
<td>[2.2]</td>
<td>[2.2]</td>
<td>[2.2]</td>
<td>[2.2]</td>
<td>[2.2]</td>
<td>[2.2]</td>
</tr>
</tbody>
</table>

Note: Each matrix entry contains the following information:
- "[x]" indicates that the feature is included.
- "<" indicates that the feature is not included.
- "[de]" indicates that the feature is available in German.
- "<de>" indicates that the feature is not available in German.

**Software packages that can fit into the role of 'archive-in-a-box':**

- Apache 2.2 (recent), Wise 1.7.1
- SVN-Platform
- DocuServer
- Hibernate [3.4 (Devel)]
- MySQL 4.1/5
- PostgreSQL
- MySQL
- GPG
- Apache 2.2, PHP 5.3
- CentOS 6.2
- Fedora
- Mac, Win
- Linux

**Internal/External Libraries/Tools:**

- [x] GSA-LDP (Berlin) [2.0] (Devel)
- [x] Wise (system-wide)
- [x] Hibernate [3.4 (Devel)]
- [x] MySQL
- [x] GPG
- [x] Apache 2.2, PHP 5.3

**Development Environments:**

- Linux, Mac, Win

**Architecture:**

- [x] SOA (GSA-LDP, Berlin) [2.0] (Devel)
- [x] Wise (system-wide)
- [x] Hibernate [3.4 (Devel)]
- [x] MySQL
- [x] GPG
- [x] Apache 2.2, PHP 5.3

**Platforms:**

- [x] Wise 1.7.1 (Testing, development)
- [x] Hibernate [3.4 (Devel)]
- [x] MySQL
- [x] GPG
- [x] Apache 2.2, PHP 5.3
- [x] CentOS 6.2
- [x] Fedora
- [x] Mac, Win
- [x] Linux

**Documentation:**

- [x] Wise (system-wide)
- [x] Hibernate [3.4 (Devel)]
- [x] MySQL
- [x] GPG
- [x] Apache 2.2, PHP 5.3

**Support:**

- [x] Wise (system-wide)
- [x] Hibernate [3.4 (Devel)]
- [x] MySQL
- [x] GPG
- [x] Apache 2.2, PHP 5.3

**Testability:**

- [x] Wise (system-wide)
- [x] Hibernate [3.4 (Devel)]
- [x] MySQL
- [x] GPG
- [x] Apache 2.2, PHP 5.3

**Contact Academic work:**

- [x] Wise (system-wide)
- [x] Hibernate [3.4 (Devel)]
- [x] MySQL
- [x] GPG
- [x] Apache 2.2, PHP 5.3

**Issues:**

- [x] Wise (system-wide)
- [x] Hibernate [3.4 (Devel)]
- [x] MySQL
- [x] GPG
- [x] Apache 2.2, PHP 5.3

**Project page:**

- [x] Wise (system-wide)
- [x] Hibernate [3.4 (Devel)]
- [x] MySQL
- [x] GPG
- [x] Apache 2.2, PHP 5.3

**Appendix 1:**

- [x] Wise (system-wide)
- [x] Hibernate [3.4 (Devel)]
- [x] MySQL
- [x] GPG
- [x] Apache 2.2, PHP 5.3

**Contact:**

- [x] Wise (system-wide)
- [x] Hibernate [3.4 (Devel)]
- [x] MySQL
- [x] GPG
- [x] Apache 2.2, PHP 5.3

**Support:**

- [x] Wise (system-wide)
- [x] Hibernate [3.4 (Devel)]
- [x] MySQL
- [x] GPG
- [x] Apache 2.2, PHP 5.3
Institutional and support

Commercial support

2006

yes

==

2007

no

==

2008

no

==

2009

no

==

2010

no

==

Commercial support

DARIAH offers a number of services that can be used by archive producers. These include:

- Filesystem: digital objects are stored in file systems for easy access and backup.
- Search: digital objects are stored in a searchable (partially or fully) replicated file system using CSV, RDF, XMB.
- Export: digital objects are stored in a downloadable (partially or fully) replicated file system using OAI, XMB, CSV, RDF.
- Search engine support: digital objects are stored in a search engine for easy access. (i.e. Google, Yahoo).
- Personal collections: digital objects are stored in personal collections as user-defined entities.
- Installation and support: digital objects are stored in a distributed (and optionally replicated) file storage based on the predecessor of iRODS.
- Cybermetrics Ranking: digital objects are stored as files on the file system.
- User feedback: digital objects are stored as files on the file system.
- Migration of existing data: digital objects are stored as files on the file system.
- Specifications / Definitions: digital objects are stored as files on the file system.
- Professional support: digital objects are stored as files on the file system.
- Search engine support: digital objects are stored as files on the file system.
- Commercial support: digital objects are stored as files on the file system.
- Search engine support: digital objects are stored as files on the file system.
- Commercial support: digital objects are stored as files on the file system.
- Professional support: digital objects are stored as files on the file system.
- Search engine support: digital objects are stored as files on the file system.
- Commercial support: digital objects are stored as files on the file system.
- Professional support: digital objects are stored as files on the file system.
- Search engine support: digital objects are stored as files on the file system.
- Commercial support: digital objects are stored as files on the file system.
- Professional support: digital objects are stored as files on the file system.
- Search engine support: digital objects are stored as files on the file system.
- Commercial support: digital objects are stored as files on the file system.
- Professional support: digital objects are stored as files on the file system.
- Search engine support: digital objects are stored as files on the file system.
- Commercial support: digital objects are stored as files on the file system.
- Professional support: digital objects are stored as files on the file system.
- Search engine support: digital objects are stored as files on the file system.
- Commercial support: digital objects are stored as files on the file system.
- Professional support: digital objects are stored as files on the file system.
- Search engine support: digital objects are stored as files on the file system.
- Commercial support: digital objects are stored as files on the file system.
- Professional support: digital objects are stored as files on the file system.
- Search engine support: digital objects are stored as files on the file system.
- Commercial support: digital objects are stored as files on the file system.
- Professional support: digital objects are stored as files on the file system.
- Search engine support: digital objects are stored as files on the file system.
- Commercial support: digital objects are stored as files on the file system.
- Professional support: digital objects are stored as files on the file system.
- Search engine support: digital objects are stored as files on the file system.
- Commercial support: digital objects are stored as files on the file system.
- Professional support: digital objects are stored as files on the file system.
- Search engine support: digital objects are stored as files on the file system.
- Commercial support: digital objects are stored as files on the file system.
- Professional support: digital objects are stored as files on the file system.
- Search engine support: digital objects are stored as files on the file system.
- Commercial support: digital objects are stored as files on the file system.
- Professional support: digital objects are stored as files on the file system.
repositories schemes registered in this service for the collections. The platform has been successfully deployed at the National Library of the Netherlands.

Additional tools – The Generic Search (GS) service provides a generic search function for the collection.

User feedback – Users can rate digital objects and provide comments.

Workflow support – Supports workflow processes for digital objects.

Subscription support – Supports subscription processes for digital objects.

Command line interface – The command line interface is provided for advanced users.

File format support – Supports various file formats for digital objects.

Indexing – The indexing mechanism is supported for digital objects.

Customizable UI/branding – The platform supports customizable user interfaces and branding.

Metadata schemes – The platform supports various metadata schemes.

Metadata search – The platform supports metadata search.

User feedback – Users can rate digital objects and provide comments.

Workflow support – Supports workflow processes for digital objects.

Subscription support – Supports subscription processes for digital objects.

Command line interface – The command line interface is provided for advanced users.

File format support – Supports various file formats for digital objects.

Indexing – The indexing mechanism is supported for digital objects.

Customizable UI/branding – The platform supports customizable user interfaces and branding.

Metadata schemes – The platform supports various metadata schemes.

Metadata search – The platform supports metadata search.

User feedback – Users can rate digital objects and provide comments.

Workflow support – Supports workflow processes for digital objects.

Subscription support – Supports subscription processes for digital objects.

Command line interface – The command line interface is provided for advanced users.

File format support – Supports various file formats for digital objects.

Indexing – The indexing mechanism is supported for digital objects.

Customizable UI/branding – The platform supports customizable user interfaces and branding.

Metadata schemes – The platform supports various metadata schemes.

Metadata search – The platform supports metadata search.

User feedback – Users can rate digital objects and provide comments.

Workflow support – Supports workflow processes for digital objects.

Subscription support – Supports subscription processes for digital objects.

Command line interface – The command line interface is provided for advanced users.

File format support – Supports various file formats for digital objects.

Indexing – The indexing mechanism is supported for digital objects.

Customizable UI/branding – The platform supports customizable user interfaces and branding.

Metadata schemes – The platform supports various metadata schemes.

Metadata search – The platform supports metadata search.

User feedback – Users can rate digital objects and provide comments.

Workflow support – Supports workflow processes for digital objects.

Subscription support – Supports subscription processes for digital objects.

Command line interface – The command line interface is provided for advanced users.

File format support – Supports various file formats for digital objects.

Indexing – The indexing mechanism is supported for digital objects.

Customizable UI/branding – The platform supports customizable user interfaces and branding.

Metadata schemes – The platform supports various metadata schemes.

Metadata search – The platform supports metadata search.

User feedback – Users can rate digital objects and provide comments.

Workflow support – Supports workflow processes for digital objects.

Subscription support – Supports subscription processes for digital objects.

Command line interface – The command line interface is provided for advanced users.

File format support – Supports various file formats for digital objects.

Indexing – The indexing mechanism is supported for digital objects.

Customizable UI/branding – The platform supports customizable user interfaces and branding.

Metadata schemes – The platform supports various metadata schemes.

Metadata search – The platform supports metadata search.

User feedback – Users can rate digital objects and provide comments.

Workflow support – Supports workflow processes for digital objects.

Subscription support – Supports subscription processes for digital objects.

Command line interface – The command line interface is provided for advanced users.

File format support – Supports various file formats for digital objects.

Indexing – The indexing mechanism is supported for digital objects.

Customizable UI/branding – The platform supports customizable user interfaces and branding.

Metadata schemes – The platform supports various metadata schemes.

Metadata search – The platform supports metadata search.

User feedback – Users can rate digital objects and provide comments.

Workflow support – Supports workflow processes for digital objects.

Subscription support – Supports subscription processes for digital objects.

Command line interface – The command line interface is provided for advanced users.

File format support – Supports various file formats for digital objects.

Indexing – The indexing mechanism is supported for digital objects.

Customizable UI/branding – The platform supports customizable user interfaces and branding.

Metadata schemes – The platform supports various metadata schemes.

Metadata search – The platform supports metadata search.

User feedback – Users can rate digital objects and provide comments.

Workflow support – Supports workflow processes for digital objects.

Subscription support – Supports subscription processes for digital objects.

Command line interface – The command line interface is provided for advanced users.

File format support – Supports various file formats for digital objects.

Indexing – The indexing mechanism is supported for digital objects.

Customizable UI/branding – The platform supports customizable user interfaces and branding.

Metadata schemes – The platform supports various metadata schemes.

Metadata search – The platform supports metadata search.

User feedback – Users can rate digital objects and provide comments.

Workflow support – Supports workflow processes for digital objects.

Subscription support – Supports subscription processes for digital objects.

Command line interface – The command line interface is provided for advanced users.

File format support – Supports various file formats for digital objects.

Indexing – The indexing mechanism is supported for digital objects.

Customizable UI/branding – The platform supports customizable user interfaces and branding.

Metadata schemes – The platform supports various metadata schemes.

Metadata search – The platform supports metadata search.

User feedback – Users can rate digital objects and provide comments.

Workflow support – Supports workflow processes for digital objects.

Subscription support – Supports subscription processes for digital objects.

Command line interface – The command line interface is provided for advanced users.

File format support – Supports various file formats for digital objects.

Indexing – The indexing mechanism is supported for digital objects.

Customizable UI/branding – The platform supports customizable user interfaces and branding.

Metadata schemes – The platform supports various metadata schemes.

Metadata search – The platform supports metadata search.

User feedback – Users can rate digital objects and provide comments.

Workflow support – Supports workflow processes for digital objects.

Subscription support – Supports subscription processes for digital objects.

Command line interface – The command line interface is provided for advanced users.

File format support – Supports various file formats for digital objects.

Indexing – The indexing mechanism is supported for digital objects.

Customizable UI/branding – The platform supports customizable user interfaces and branding.

Metadata schemes – The platform supports various metadata schemes.

Metadata search – The platform supports metadata search.

User feedback – Users can rate digital objects and provide comments.

Workflow support – Supports workflow processes for digital objects.

Subscription support – Supports subscription processes for digital objects.

Command line interface – The command line interface is provided for advanced users.

File format support – Supports various file formats for digital objects.

Indexing – The indexing mechanism is supported for digital objects.

Customizable UI/branding – The platform supports customizable user interfaces and branding.

Metadata schemes – The platform supports various metadata schemes.

Metadata search – The platform supports metadata search.

User feedback – Users can rate digital objects and provide comments.

Workflow support – Supports workflow processes for digital objects.

Subscription support – Supports subscription processes for digital objects.

Command line interface – The command line interface is provided for advanced users.

File format support – Supports various file formats for digital objects.

Indexing – The indexing mechanism is supported for digital objects.

Customizable UI/branding – The platform supports customizable user interfaces and branding.

Metadata schemes – The platform supports various metadata schemes.

Metadata search – The platform supports metadata search.

User feedback – Users can rate digital objects and provide comments.

Workflow support – Supports workflow processes for digital objects.

Subscription support – Supports subscription processes for digital objects.

Command line interface – The command line interface is provided for advanced users.

File format support – Supports various file formats for digital objects.

Indexing – The indexing mechanism is supported for digital objects.

Customizable UI/branding – The platform supports customizable user interfaces and branding.