


Zenodo: Preservation Meter

João Gonçalves - [jsvgoncalves@\[gmail.com, github, linkedin, ...\]](mailto:jsvgoncalves@gmail.com)
Supervisor: Lars Holm Nielsen
28 August 2014

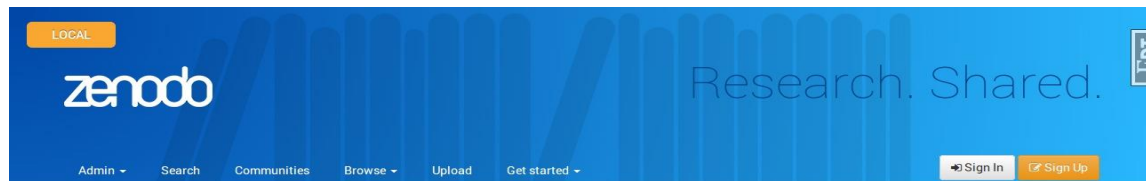
Context

- Digital Preservation



-  An open digital repository for everyone and everything [<http://zenodo.org>]

Problem



08 May 2013

Presentation Open access

Test title

Doe, John ; Smith, Jane

(show affiliations)

Test Description

Preview

1 / ? < > 🔍 📷

Sorry, we were unable to find the document at the original source. Verify that the document still exists.

You can also try to download the original document by clicking [here](#).

Files

Name	Date	Size	
test0.pdf	29 Jul 2014	1.9 kB	Preview Download
test1.pdf	29 Jul 2014	1.9 kB	Preview Download
test2.pdf	29 Jul 2014	1.9 kB	Preview Download

Comments

Publication date:

08 May 2013

DOI

[10.5372/zenodo.8](#)

Collections:

[Presentations](#)

[Open access](#)

License (for files):

[Creative Commons CCZero](#)

Uploaded by:

[tester](#) (on 29 July 2014)

New to ZENODO?

Read more about [features and benefits](#).

Sign Up

Share



Cite as

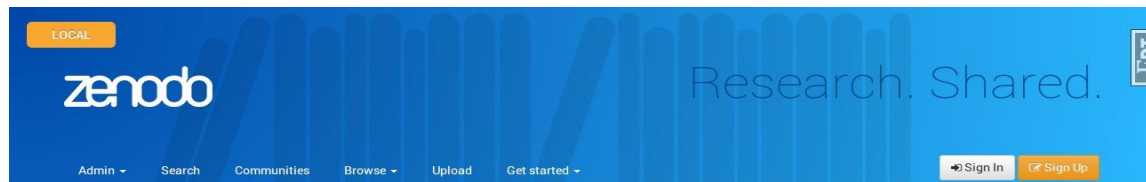
Doe, John *et al* (2013). Test title. ZENODO.

Further citation formats: [DOI Citation Formatter](#).

Export

[BibTeX](#), [DataCite](#), [DC](#), [EndNote](#), [NLM](#), [RefWorks](#), [MARC](#), [MARCXML](#)

Problem



08 May 2013

Test title

Doe, John ; Smith, J.

(show affiliations)

Test Description

Preview

Sorry, we w

Name	Date	Size
------	------	------

test0.pdf	29 Jul 2014	1.9 kB
-----------	-------------	--------

test1.pdf	29 Jul 2014	1.9 kB
-----------	-------------	--------

test2.pdf	29 Jul 2014	1.9 kB
-----------	-------------	--------

Preview

Download

Preview

Download

Preview

Download

Files

Name	Date	Size
------	------	------

test0.pdf	29 Jul 2014	1.9 kB
-----------	-------------	--------

test1.pdf	29 Jul 2014	1.9 kB
-----------	-------------	--------

test2.pdf	29 Jul 2014	1.9 kB
-----------	-------------	--------

Preview

Download

Preview

Download

Preview

Download



Cite as

Doe, John ; Smith, J. (2013). Test title. ZENODO.

Full text citation formats: DOI Citation Formatter.

Export

BibTeX, DataCite, DC, EndNote, NLM, RefWorks
MARC, MARCXML

Problem

- Suitable for preservation?

08 May 2013

Test title

Doe, John ; Smith, Jane

(show affiliations)

Test Description

Preview

1 / ? < > 🔍 🖨

Sorry, we were unable to find the document at the original source. Verify that the document still exists.

You can also try to download the original document by clicking [here](#).

Name	Date	Size	Preview	Download
test0.pdf	29 Jul 2014	1.9 kB		
test1.pdf	29 Jul 2014	1.9 kB		
test2.pdf	29 Jul 2014	1.9 kB		

Comments

Publication date: 08 May 2013

DOI: [10.5272/zzenodo.8](https://doi.org/10.5272/zzenodo.8)

Collections: Presentations, Open access

License (for files): Creative Commons CCZero

Uploaded by: tester (on 29 July 2014)

New to ZENODO? [Sign Up](#)

Read more about [features and benefits](#).

Share

Cite as

Doe, John *et al* (2013). Test title. ZENODO.

Further citation formats: [DOI Citation Formatter](#).

Export

BibTeX, DataCite, DC, EndNote, NLM, RefWorks, MARC, MARCXML

Problem

- Suitable for preservation?

- Raise awareness?

08 May 2013

Test title

Doe, John ; Smith, Jane

(show affiliations)

Test Description

Preview

1 / 2

Sorry, we were not able to find the document at the original source. Verify that the document still exists.

You can also try to download the original document by clicking [here](#).

Name	Date	Size	Preview	Download
test0.pdf	29 Jul 2014	1.9 kB	Preview	Download
test1.pdf	29 Jul 2014	1.9 kB	Preview	Download
test2.pdf	29 Jul 2014	1.9 kB	Preview	Download

Comments

Publication date: 08 May 2013
DOI: [10.5281/zenodo.8](#)
Collections: Presentations, Open access
License (for files): Creative Commons CCZero
Uploaded by: tester (on 29 July 2014)

New to ZENODO?
Read more about features and benefits. [Sign Up](#)

Share
[Twitter](#) [Facebook](#) [Email](#) [+](#)

Cite as
Doe, John *et al* (2013). Test title. ZENODO.
Further citation formats: DOI Citation Formatter.

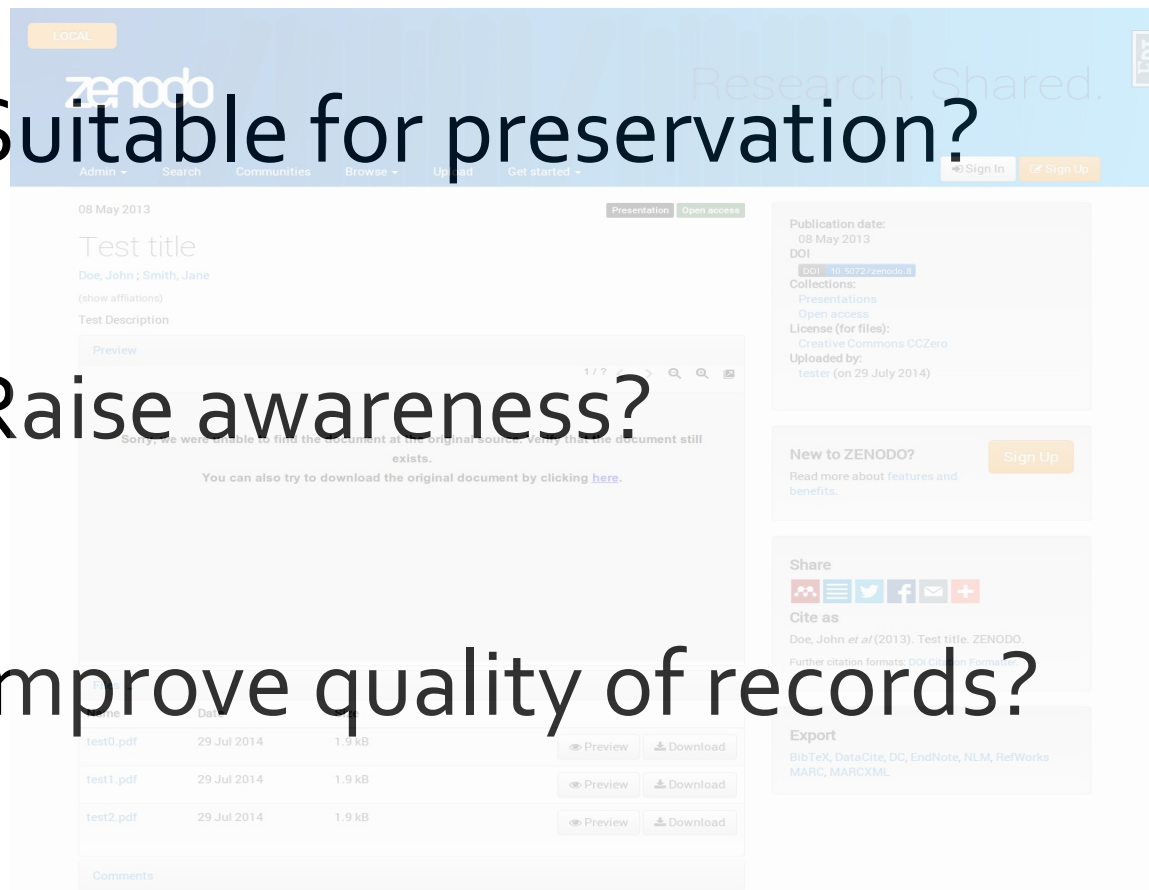
Export
BibTeX, DataCite, DC, EndNote, NLM, RefWorks, MARC, MARCXML

Problem

- Suitable for preservation?

- Raise awareness?

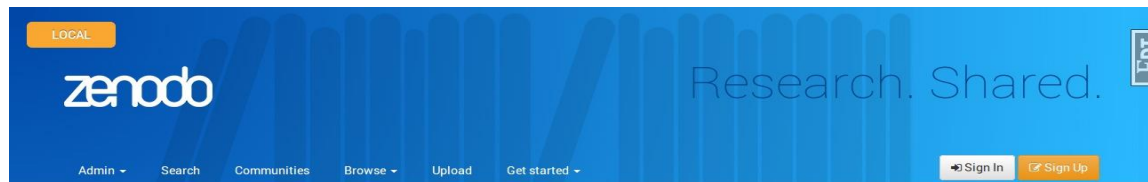
- Improve quality of records?



Objectives

- Add **meter and badge** to records to indicate the **suitability** of the documents for long-term **preservation**
 - Define preservation guidelines and best practices
 - Implement as an independent/plugable module to Zenodo
 - Develop testing and validation

Preservation Meter



08 May 2013

Presentation Open access

Test title

Doe, John ; Smith, Jane

(show affiliations)

Test Description

Preview

1 / ? < > 🔍 📄

Sorry, we were unable to find the document at the original source. Verify that the document still exists.

You can also try to download the original document by clicking [here](#).

Files

Name	Date	Size	
test0.pdf	29 Jul 2014	1.9 kB	Preview Download
test1.pdf	29 Jul 2014	1.9 kB	Preview Download
test2.pdf	29 Jul 2014	1.9 kB	Preview Download

Comments

Publication date:

08 May 2013

DOI

[10.5372/zenodo.8](#)

Collections:

[Presentations](#)

[Open access](#)

License (for files):

[Creative Commons CCZero](#)

Uploaded by:

[tester](#) (on 29 July 2014)

Preservation Meter

100%

Further information: [Digital Preservation Best Practices](#).

New to ZENODO?

Read more about features and benefits.

Sign Up

Share



Cite as

Doe, John *et al* (2013). Test title. ZENODO.

Further citation formats: [DOI Citation Formatter](#).

Export

[BibTeX](#), [DataCite](#), [DC](#), [EndNote](#), [NLN](#), [RefWorks](#)

Preservation Meter

The screenshot shows the Zenodo website interface. At the top, there's a blue header with the Zenodo logo and navigation links. Below the header, the main content area displays a document page for 'Test title' by Doe, John; Smith, Jane. The document is dated 08 May 2013. A 'Preservation Meter' callout box is overlaid on the right side of the page, showing a green bar at 100% and the status 'Well preserved'. The callout box also lists 'Suggested actions: None' and provides a link to 'Further information: Digital Preservation Best Practices.'.

Preservation Meter

100%

Status: Well preserved.

Suggested actions: None.

Further information: [Digital Preservation Best Practices.](#)

Test title
Doe, John ; Smith, Jane
(show affiliations)
Test Description

08 May 2013

Publication date: 08 May 2013
DOI: [10.5281/zenodo.8](#)
Collections: [Presentations](#), [Open access](#)
License (for files): [Creative Commons CC0](#)
Uploaded by: [tester](#) (on 29 July 2014)

Sorry, we were unable to find the document at the original source. Verify that the document still exists.
You can also try to download the original document by clicking [here](#).

Files

Name	Date	Size	Preview	Download
test0.pdf	29 Jul 2014	1.9 kB	Preview	Download
test1.pdf	29 Jul 2014	1.9 kB	Preview	Download
test2.pdf	29 Jul 2014	1.9 kB	Preview	Download

Comments

New to ZENODO?
Read more about features and benefits. [Sign Up](#)

Share
[Reddit](#) [LinkedIn](#) [Twitter](#) [Facebook](#) [Email](#) [+](#)

Cite as
Doe, John *et al* (2013). Test title. ZENODO.
Further citation formats: [DOI Citation Formatter](#).

Export
[BibTeX](#), [DataCite](#), [DC](#), [EndNote](#), [NLN](#), [RefWorks](#)

Preservation Meter

- Score based approach
 - mapping file
type => quality score
 - proprietary vs open

Publication date:

08 May 2013

DOI

DOI [10.5072/zenodo.8](https://doi.org/10.5072/zenodo.8)

Collections:

[Presentations](#)

[Open access](#)

License (for files):

[Creative Commons CCZero](#)

Uploaded by:

[tester](#) (on 29 July 2014)

Preservation Meter

100%

Status:

Well preserved.

Suggested actions:

None.

Further information: [Digital Preservation Best Practices](#).

Preservation Meter

- Score based approach
 - mapping file
type => quality score
 - proprietary vs open
- Single file evaluation

Publication date:

08 May 2013

DOI

DOI [10.5072/zenodo.8](https://doi.org/10.5072/zenodo.8)

Collections:

[Presentations](#)

[Open access](#)

License (for files):

[Creative Commons CCZero](#)

Uploaded by:

[tester](#) (on 29 July 2014)

Preservation Meter

100%

Status:

Well preserved.

Suggested actions:

None.

Further information: [Digital Preservation Best Practices](#).

Preservation Meter

- Score based approach
 - mapping file
type => quality score
 - proprietary vs open
- Single file evaluation
- Multiple file evaluation
 - avg([])
 - compressed formats (zip, tar..)

Publication date:
08 May 2013
DOI
DOI [10.5072/zenodo.8](https://doi.org/10.5072/zenodo.8)
Collections:
[Presentations](#)
[Open access](#)
License (for files):
[Creative Commons CCZero](#)
Uploaded by:
[tester](#) (on 29 July 2014)

Preservation Meter

100%

Status:
Well preserved.
Suggested actions:
None.

Further information: [Digital Preservation Best Practices](#).

Implementation



jsvgoncalves added some commits 23 days ago



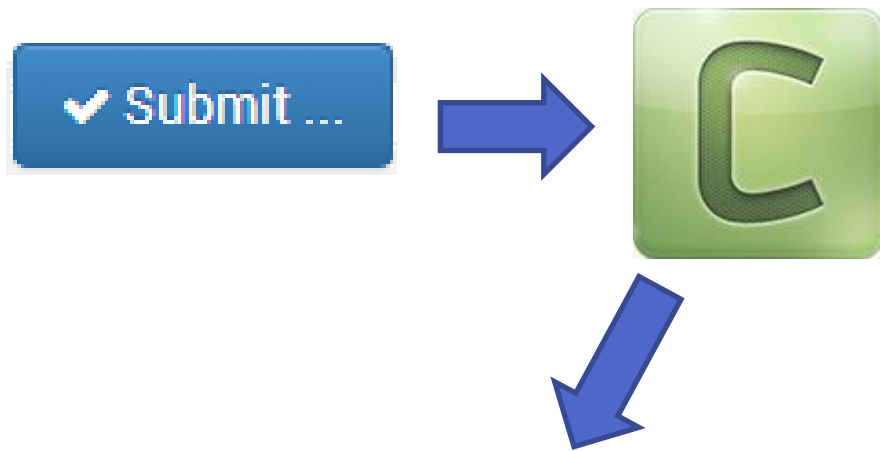
preservationmeter: Add module ...



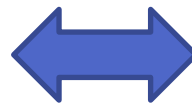
preservationmeter: Starting tests ...

```
▼ modules
  ► communities
  ► deposit
  ► github
  ► inspire
  ▼ preservationmeter
    ▼ templates
      preservation_meter.html
    ▼ testsuite
      __init__.py
      test_api.py
      __init__.py
      api.py
      config.py
      tasks.py
      views.py
```

Implementation



```
def calculate_score(file_path_list):  
    """  
    Receives a list of file paths and  
    calculates their preservation score.  
    """
```



```
PRESERVATIONMETER_FILES_QUALITY = {  
    # Publications  
    '.pdf': 100,  
    '.htm': 100,  
    '.html': 100,  
    '.txt': 100,  
    '.odt': 95,  
    '.rtf': 90,  
    '.docx': 60,  
    '.doc': 40,  
  
    # Datasets  
    '.xml': 100,  
    '.csv': 100,  
    '.xlsx': 60,  
    '.xls': 40,  
    '.por': 100,  
    '.tab': 100,  
    '.tfw': 100,  
    '.dwg': 100,  
    '.shp': 100,  
    '.shx': 100,  
    '.dbf': 100,  
    '.sav': 50,  
    '.dta': 50,  
    '.mdb': 50,  
    '.accdb': 50,  
    '.dbf': 50,  
    '.ods': 50,  
    '.mif': 50,  
    '.kml': 50,  
    '.ai': 50,  
    '.dxf': 50,  
    '.svg': 50,  
  
    # Images & Posters  
    '.tif': 100,  
    '.jpeg': 90,  
    '.jpg': 90,  
    '.tiff': 80,  
    '.raw': 70,  
    '.psd': 50,  
  
    # Presentations  
  
    # Videos  
    '.flac': 100,  
    '.mp4': 100,  
    '.mj2': 100,
```

Implementation

```
import tarfile
```

```
f = 'some-file'
```

```
tarfile.is_tarfile(f)
```

```
> ...
```

```
> IOError: [Errno2]
```


```
import zipfile
```

```
f = 'some-file'
```

```
zipfile.is_zipfile(f)
```

```
> False
```


Development & Testing

tmuxinator +  **.js**

Development & Testing

→ ~ which dev-zenodo
dev-zenodo: aliased to mux s zenodo



```
→ ~  
→ ~ workon zenodo && inveniomange runserver  
fatal: Not a git repository (or any parent up to mount point  
/home)  
Stopping at filesystem boundary (GIT_DISCOVERY_ACROSS_FILESY  
STEM not set).  
/home/joao/.virtualenvs/zenodo/local/lib/python2.7/site-pack  
ages/pyparsing.py:3546: DeprecationWarning: Operator '<<' is  
deprecated, use '<=>' instead  
    ret << Group( Suppress(opener) + ZeroOrMore( ignoreExpr |  
ret | content ) + Suppress(closer) )  
  
* Running on http://0.0.0.0:4000/  
* Restarting with reloader  
fatal: Not a git repository (or any parent up to mount point  
/home)  
Stopping at filesystem boundary (GIT_DISCOVERY_ACROSS_FILESY  
STEM not set).  
/home/joao/.virtualenvs/zenodo/local/lib/python2.7/site-pack  
ages/pyparsing.py:3546: DeprecationWarning: Operator '<<' is  
deprecated, use '<=>' instead  
    ret << Group( Suppress(opener) + ZeroOrMore( ignoreExpr |  
ret | content ) + Suppress(closer) )  
  
workon zenodo && cdvirtualenv src/zenodo && gst && gulp  
→ ~  
→ ~ workon zenodo && cdvirtualenv src/zenodo && gst && gulp  
On branch preservationmeter  
Your branch is up-to-date with 'origin/preservationmeter'.  
  
nothing to commit, working directory clean  
[09:12:06] Working directory changed to ~/.virtualenvs/zenodo  
[09:12:29] Using gulpfile ~/.virtualenvs/zenodo/gulpfile.js  
[09:12:29] Starting 'watch'...  
[09:12:29] Finished 'watch' after 8.85 ms  
[09:12:29] Starting 'default'...  
[09:12:29] Finished 'default' after 7.46 μs
```

Conclusion

- Main objectives accomplished

Conclusion

- Main objectives accomplished
 - And tested :-)

```
..- .Name                               Stmts  Miss  Cover   Missing
-----
zenodo.modules.preservationmeter         0      0   100%
zenodo.modules.preservationmeter.api    118      0   100%
zenodo.modules.preservationmeter.config  0      0   100%
zenodo.modules.preservationmeter.views   3      0   100%
-----
TOTAL                                   121      0   100%
-----

13 tests run in 6.7 seconds.
1 skipped (12 tests passed)
```

Conclusion

- Main objectives accomplished
- Future work:
 - Provide more details about the score
 - Provide feedback even before the record is uploaded
 - Improve evaluation methods
 - Add gamification techniques (?)

CERN Experience

ONE DOES NOT SIMPLY

INSTALL INVENIO

A close-up photograph of a woman with dark hair, crying. Her eyes are closed, and her hand is pressed against her forehead in a gesture of distress or exhaustion. The lighting is soft, highlighting her facial features and the texture of her hair.

I WANT TO WORK

BUT THEY TAKE ME ON TRIPS

CERN Experience

- Learnt lots of new stuff
 - web (python, flask, ...)

CERN Experience

- Learnt lots of new stuff
 - web (python, flask, ...)
 - git

CERN Experience

- Learnt lots of new stuff
 - web (python, flask, ...)
 - git
 - bash

CERN Experience

- Learnt lots of new stuff
 - web (python, flask, ...)
 - git
 - bash
 - greek.

CERN Experience

- Learnt lots of new stuff
 - web (python, flask, ...)
 - git
 - bash
 - greek.
- Working in a big team (processes and workflow)

CERN Experience

- Learnt lots of new stuff
 - web (python, flask, ...)
 - git
 - bash
 - greek.
- Working in a big team (processes and workflow)
- Lots of visits and trips (Google, ETHZ, Grenoble, ...)

CERN Experience

- Learnt lots of new stuff
 - web (python, flask, ...)
 - git
 - bash
 - greek.
- Working in a big team (processes and workflow)
- Lots of visits and trips (Google, ETHZ, Grenoble, ...)
- Awesome time :-)

Zenodo: Preservation Meter

João Gonçalves - [jsvgoncalves@\[gmail.com, github, linkedin, ...\]](mailto:jsvgoncalves@gmail.com)
Supervisor: Lars Holm Nielsen
28 August 2014