

# **‘Super - Simple - Static - Sustainable’ a low resource repository jam**

Progress Report  
2024-06-06



## Workshop core values and aims

- THIS IS **NOT A COMPETITION**.
- We will encourage participants to **learn new skills and to work with new collaborators** and create teams with a diverse range of background and levels of experience
- Produce **demonstrable code/techniques for creating, managing and disseminating simple static repositories** / archives *with real data* (where possible) that can *immediately* be used in low resource environments

But it might be a *\*bit\** like the old OR “Developer Challenge”

We will recap the inspiration for this session, present & discuss **Suleman’s “principles and design goals”:**

- Simplicity/Minimalism – only as much system as needed
- No imposition on users – no logins unless necessary
- No Internet - should work even when offline
- Simple preservation – as simple as copying files
- Flexibility in objects/services – no arbitrary restrictions
- Superimposed and hierarchical information – because we understand this
- Platform agnosticism – any OS, any hardware
- Pre-processing – fewer cycles needed when live

Inspired by the rousing OR-2023 closing keynote

<https://zenodo.org/records/8111569>

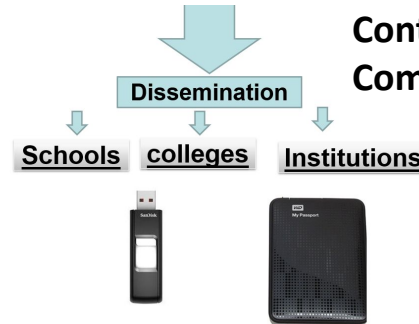
# Internet in a Box

## 1. Context

- Population 3.8
- Schools > 1400
- Students 600,000
- Conventional Library costly
- Internet penetration below 1%

## 2. Problem

Difficult to provide info



## 3. Solution

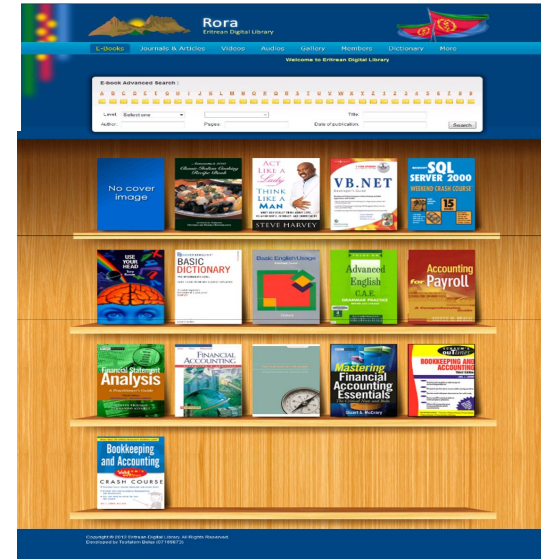
Low End Server

A router or Access point 50 users

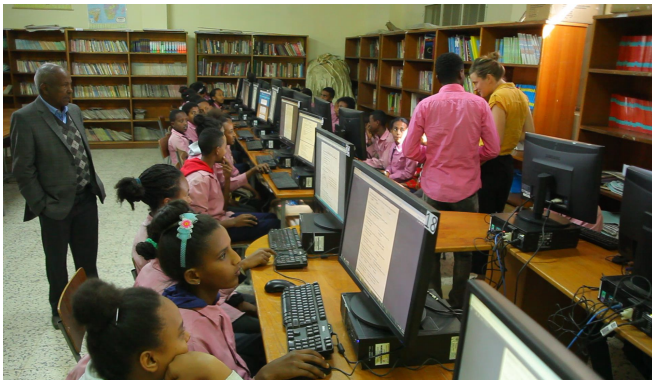
Website

Content

Communities in Diaspora



## 4. Result



# Easy Open

A user-friendly top level organization for self-contained browsable packages

Patrick Stickler, Peter Sefton, ...

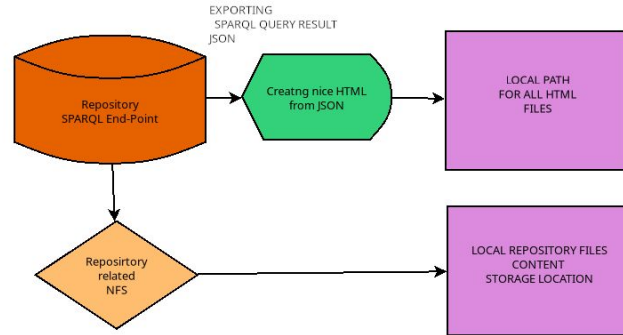
**OPEN.html**    #Required  
**content/**    #Required  
**metadata/**    #Optional  
**resources/**    #Optional



```
OPEN.html        # redirects to content/ro-crate-preview.html
content/
  ro-crate-metadata.json
  ro-crate-preview.html
  ro-crate-preview_files/
  ...
  bag1/
    bagit.txt
    bag-info.txt
    manifest-sha512.txt
    data/
      example.txt
metadata/
  manifest.txt
    # content/ro-crate-metadata.json
    # content/bag1/bag-info.txt
```

# Generating Static repository from the Dynamic one

## (Mahsa and Andjela)



<b>thingid</b>	Thing_2ae71d85-8525-4cd3-8352-c633c992b0d2
<b>title</b>	Slotmanifestatie WK2020 Vernieuwingsimpulsen bij woningrenovatie
<b>sub_title</b>	
<b>description</b>	
<b>object_type</b>	conference
<b>category</b>	
<b>publication_date_published</b>	
<b>publication_status</b>	
<b>type_of_resource</b>	text
<b>status</b>	OPEN
<b>date_created</b>	2016-01-29 18:00:37.719000
<b>date_modified</b>	2016-01-29 18:00:37.719000
<b>emailaddress</b>	
<b>embargo</b>	
<b>access_condition</b>	AccessCondition_6f35b3c0-43c9-4567-baf4-0fa1d53ede57
<b>access_conditions</b>	<b>access_value</b> © 2013 Delft University of Technology
<b>classification</b>	
	Collection_570619ec-9388-455c-ba19-b55caf421def
	Contributor_8f827723-22e2-4dab-9f2d-41300db70101

<b>thingid</b>	Thing_d9509685-ec42-4dba-82d7-f23ff1c1661c
<b>title</b>	Influence of microencapsulated phase change material (PCM) addition on (micro) mechanical properties of cement paste
<b>sub_title</b>	
<b>description</b>	Excessive cracking can be a serious durability problem for reinforced concrete structure drawbacks, mainly related to strength reduction. In this work, a range of experimental microcapsules was shown to depend on the capsule diameter and the temperature, i.e., 'less extent due to changes in porosity caused by their inclusion. Finally, a novel microcapsule will contribute to future developments of cementitious composites incorporating phase change material (PCM) in concrete.
<b>object_type</b>	journal article
<b>category</b>	Scientific
<b>publication_date_published</b>	2017-7-27
<b>publication_status</b>	Published
<b>type_of_resource</b>	
<b>status</b>	OPEN
<b>date_created</b>	2017-10-06T08:40:52.73+02:00
<b>date_modified</b>	2024-05-15T05:46:22.944+02:00
<b>emailaddress</b>	
<b>embargo</b>	
<b>access_condition</b>	AccessCondition_770d8dac-4809-48db-8c69-ac77de4ccd89
<b>access_conditions</b>	<b>access_type</b> use and reproduction © 2017 B. Šavija, H. Zhang, E. Schlangen
<b>classification</b>	
<b>collection</b>	Collection_d241379b-0f48-4a9a-9540-032b044c54b5
<b>contributor</b>	Contributor_02ba14e7-8984-46e3-8614-c4c8bcb220aa
<b>file_container</b>	FileContainer_c4770206-cd13-4c92-9c9e-7769d99dc80e
<b>identifier</b>	Identifier_9fc9136e-e46b-44e0-9503-61da8cc86396

<b>thingid</b>	Thing_a6725b28-7350-4ef3-9c4d-6fa87516cdd3
<b>title</b>	History Urbanism Resilience
<b>sub_title</b>	Planning and Heritage
<b>description</b>	Proceedings of the 17th International Planning History Society Conference, Delft, Netherlands, July 17-21, 2016.@en
<b>object_type</b>	book
<b>category</b>	Scientific
<b>publication_date_published</b>	2016
<b>publication_status</b>	Published
<b>type_of_resource</b>	
<b>status</b>	OPEN
<b>date_created</b>	2024-05-30 05:25:43.874
<b>date_modified</b>	2024-05-30 05:25:43.878
<b>emailaddress</b>	
<b>embargo</b>	
<b>access_condition</b>	
<b>access_conditions</b>	
<b>classification</b>	
<b>collection</b>	Collection_d241379b-0f48-4a9a-9540-032b044c54b5
<b>contributor</b>	Contributor_1f26e274-a0aa-440d-9edc-b534cebac0ee
<b>file_container</b>	FileContainer_cc2b8671-4635-443c-a178-7709ef81cf92
<b>identifier</b>	Identifier_fa74f1dd-7aa4-4fd2-a3d2-4a6f50ea24f3
<b>identifiers</b>	

# Simple export of Zip + HTML in Easy Open Format | Hyeokjin Kwon (TU Delft / 4TU.ResearchData)

## German ASA Questionnaire Translation - Part 1: Translation and Formative Assessment

doi: 10.4121/1975af9a-9b58-4dde-ae58-58e1001ef553.v1

### Cite 1. Problem: Downloaders need to manage metadata separately.

DATASET

by Boleslav Khodakov, Emma Bokel, [Nele Albers](#), Andrea Bönsch, Jonathan Ehret, Willem-Paul Brinkman

This is the data and analysis code for the formative assessment of the German translation of the Artificial Social Agent (ASA) Questionnaire. This questionnaire can be used to characterize interactions between humans and ASAs.

#### Studies

The formative assessment consists of 3 rounds of assessing the correlation between original English questionnaire items and their German translations. English items with too low correlation scores in a round are retranslated and re-assessed in the subsequent round. The assessment consists of people who indicate having German as their first and primary language, being fluent in English, and being bilingual rating the English items and their German translations for the human-ASA interaction depicted in a 30-second video clip. The studies were run on the online crowdsourcing platform Prolific.

#### Data

For each round, we provide the ratings for the English and German questionnaire items as well as data on the characteristics of the participants (age, gender, highest completed education level from their Prolific profiles). A random identifier called "rand\_id" can be used to link the data files.

#### HISTORY

2023-06-26 first online, published, posted

#### PUBLISHER

4TU.ResearchData

#### FORMAT

.docx, .md, .pdf, .txt, .sav, .csv, .tex, .bib, .zip, .py, .rmd

#### FUNDING

- This work is part of the multidisciplinary research project Perfect Fit, which is supported by several funders organized by the Netherlands Organization for Scientific Research (NWO), program Commit2Data - Big Data & Health (project number 628.011.211). Besides NWO, the funders include the Netherlands Organisation for Health Research and Development (ZonMw), Hartstichting, the Ministry of Health, Welfare and Sport (VWS), Health Holland, and the Netherlands eScience Center.

#### ORGANIZATIONS

TU Delft, Faculty of Electrical Engineering, Mathematics and Computer Science, Department of Intelligent Systems, Interactive Intelligence

## DATA

FILES (1)

852,964 bytes MD5: f62e2eb9eb1a08782d4b488b8a477f97 German\_ASAQ\_Translation\_Formative.zip

download all files (zip)  
852,964 bytes unzipped

### 2. Idea: Let's include metadata in ZIP



USAGE STATS

195 63

views downloads

CATEGORIES

Artificial Intelligence and Image Processing  
Information and Computing Sciences

KEYWORDS

Artificial social agents, Chatbot, Robots  
Virtual agents

TIME COVERAGE

18 May 2023 - 4 June 2023

LICENCE

CC BY 4.0

EXPORT AS...

RefWorks, BibTeX, Reference Manager  
NLM, DC, CFF

```
wsgi.py
60 class ApiServer:
+ 3839 def ui_download_static_all (self, request, dataset_id, version):
+ 3840     """Implements /sdownloader/items/<id>/versions/<version>"""
+ 3841
+ 3842     dataset, metadata = self._accessible_files_for_dataset (request, dataset_id, version
+ 3843 =version)
+ 3844
+ 3845     if dataset is None or metadata is None:
+ 3846         return self.error_403 (request)
+ 3847
+ 3848     try:
+ 3849         file_paths = []
+ 3850
+ 3851         # Create a temporary directory to store the files
+ 3852         tempdir = tempfile.TemporaryDirectory(suffix=dataset_id, prefix=version)
+ 3853         temp_dir = tempdir.gettempdir()
+ 3854
+ 3855         # Create a metadata file for the static download
+ 3856         metadata_file = f"{temp_dir}/metadata.json"
+ 3857         metadata_fileinfo_file = f"{temp_dir}/metadata_fileinfo.json"
+ 3858         self.log.info ("Metadata file: %s", metadata_file)
+ 3859         self.log.info ("Metadata fileinfo file: %s", metadata_fileinfo_file)
+ 3860         with open (metadata_file, "w", encoding="UTF-8") as mfile1:
+ 3861             json.dump (dataset, mfile1)
+ 3862         with open (metadata_fileinfo_file, "w", encoding="UTF-8") as mfile2:
+ 3863             json.dump (metadata, mfile2)
+ 3864
+ 3865         file_paths.append ({"fs": metadata_file, "n": "metadata.json"})
+ 3866         file_paths.append ({"fs": metadata_fileinfo_file, "n": "metadata_fileinfo.json"})
+ 3867
+ 3868         zipfly_object = zipfly.ZipFly (paths = file_paths)
+ 3869         writer = zipfly_object.generator ()
+ 3870         response = self.response (writer, mimetype="application/zip")
+ 3871
+ 3872
```

### 3. Patch for Proof of Concept

```
hkwon 08:22 0s ~/Downloads $ ls -al German_ASAQ_Translation_Formative.zip
-rw-r--r-- hkwon staff 884241 Thu Jun 6 08:18:36 2024 -- German_ASAQ_Translation_Formative.zip
Downloads $ unzip German_ASAQ_Translation_Formative.zip > /dev/null
Downloads $ cd German_ASAQ_Translation_Formative/
Downloads/German_ASAQ_Translation_Formative $ ls -al
drwx----- hkwon staff 160 Thu Jun 6 08:14:22 2024 -- .
drwx----- hkwon staff 2560 Thu Jun 6 08:22:46 2024 -- ..
-rw-r--r-- hkwon staff 852964 Thu Jun 6 08:08:37 2024 -- German_ASAQ_Translation_Formative.zip
-rw-rw-r-- hkwon staff 33500 Mon Jun 26 07:49:40 2023 -- README.html
-rw-r--r-- hkwon staff 10123 Thu Jun 6 08:14:15 2024 -- ro-create-metadata.json
```

README.html : The static file including Javascript/CSS

ro-create-metadata.json : All metadata of the dataset



# EPrints to Easy Open static export

Here's last year's slides... (sorry Rory ;) Actually, here's a SVN commit:

**~/tools/static\_site.pl**

Rev 7182 – 4607 d 19 h (**Monday 24th October 2011**)

Author: tdb01r (Tim Brody)



Log message:

\* experimental tool to generate a static version of an EPrints repository

- Started enhancing it to output the site in EasyOPEN as previously described.
- Outputs an EPrintsXML metadata file, thumbnails, documents, browse views
- Browser-based search using Lunr.js <https://lunrjs.com/> *A bit like Solr, but much smaller and not as bright.*

# EasyScrape: 'Remote site generation' (impl. Easy Open)

Kim Shepherd, Kiflom Michael Kahsay, Marco Schirone

 **Scrappy** 

Spiders for feeds:  
OAI-PMH, RSS, Atom

Crawl/harvest repo,  
output static stuff

- Markdown (readable, usable, site generators like it)
- Resources (PDFs, data, images)

## No Internet?!

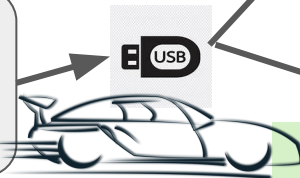
Crawler + generator + storage  
+ friend with Internet =

- Distribute offline repo
- Host 'online' stuff locally

## Generate site

- Jekyll
- 11ty
- Pelican

EasyOpen, OCFL, Bagit..



Publish  
local  
content on  
Internet



Host Internet  
content locally



[\*\*https://github.com/OpenRepositoriesConference/Super-Simple-Static-Sustainable\*\*](https://github.com/OpenRepositoriesConference/Super-Simple-Static-Sustainable)

