Ridiculously simply SIP creator

By: OneClick eArchiving project

Ph.D Anssi Jääskeläinen

Research Manager

Xamk / Digitalia

digitalia-aj@new-python-dev: ~/extraspace/eclipse-workspace/GIT/c
20+03:00', 'File:FileInodeChangeDate': '2022:09:04 12:57:20+03:00
ile:FileTypeExtension': 'DOCX', 'File:MIMEType': 'application/vn
'ZIP:ZipRequiredVersion': 20, 'ZIP:ZipBitFlag': 6, 'ZIP:ZipComp
IP:ZipCRC': 1466929458, 'ZIP:ZipCompressedSize': 358, 'ZIP:ZipUn
xml', 'XMP:Title': '', 'XMP:Subject': '', 'XMP:Creator': 'Jääske
XML:LastModifiedBy': 'Jääskeläinen Anssi', 'XML:RevisionNumber'
te': '2022:08:31 16:30:00Z', 'XML:Template': 'Normal.dotm', 'XML
Characters': 452, 'XML:Application': 'Microsoft Office Word', 'X
ML:ScaleCrop': False, 'XML:HeadingPairs': ['Otsikko', 1], 'XML:T
False, 'XML:CharactersWithSpaces': 530, 'XML:SharedDoc': False,

2022:09:04 12:57:20+03:00 and after 2022-09-04T12:57:20+03:00
2022:09:04 12:57:20+03:00 and after 2022-09-04T12:57:20+03:00
2022:09:04 12:57:20+03:00 and after 2022-09-04T12:57:20+03:00
1980:01:01 00:00:00 and after 1980-01-01T00:00:00
2022:08:15 10:05:00Z and after 2022-08-15T10:05:00Z
2022:08:31 16:30:00Z and after 2022-08-31T16:30:00Z
False and after False
ata for 5 files
7de1-058c-4183-bc2c-1ca159c6bf76 metadata to a dictionary
mets.xml filesection and structmap for id, f10f7de1-058c-4183-bc2
tadata for SIP id f10f7de1-058c-4183-bc2c-1ca159c6bf76
METS.xml file
enerated /tmp/runtimeProcessing/f10f7de1-058c-4183-bc2c-1ca159c6b
TS.xml file
nd in mets.xml -->[]
a for 1 files

mp/runtimeProcessing/f10f7de1-058c-4183-bc2c-1ca159c6bf76/repres
{'SourceFile': '/tmp/runtimeProcessing/f10f7de1-058c-4183-bc2c-1
890d7/METS.xml', 'ExifTool:ExifToolVersion': 12.16, 'File:FileNa
f10f7de1-058c-4183-bc2c-1ca159c6bf76/representations/ef297737-89
leModifyDate': '2022:09:04 12:57:22+03:00', 'File:FileAccessDate
'2022:09:04 12:57:22+03:00', 'File:FilePermissions': 644, 'File
MIMEType': 'application/xml', 'XMP:MetsXmlns': 'http://www.loc.go
890d7', 'XMP:MetsLabel': '', 'XMP:MetsType': 'Mixed', 'XMP:MetsC
//earkcsip.dilcis.eu/profile/E-ARK-CSIP.xml', 'XMP:MetsSchemaLoc
rds/mets/mets.xsd\n   http://www.w3.org/1999/xlink http://ww
eu/XML/METS/CSIPExtensionMETS https://earkcsip.dilcis.eu/schema/D
IPExtensionMETS', 'XMP:MetsMetsHdrCreatedate': '2022:09:04 12:5
MP:MetsMetsHdrOaispackagetype': 'SIP', 'XMP:MetsMetsHdrAgentRole
sMetsHdrAgentOthertype': 'SOFTWARE', 'XMP:MetsMetsHdrAgentName':
umentID': 'repMETS.xml', 'XMP:MetsAmdSecId': 'amdSec', 'XMP:Mets
repmetsfileSecGrp', 'XMP:MetsFileSecFileGrpUse': 'Data', 'XMP:Me
1f9c30', 'XMP:MetsFileSecFileGrpFileMimetype': 'application/vnd.
XMP:MetsFileSecFileGrpFileChecksumtype': 'SHA-256', 'XMP:MetsFil
MetsFileSecFileGrpFileChecksum': '84c8f40fd3f2771bd3a2597e946606
pFileUse': 'Datafile', 'XMP:MetsFileSecFileGrpFileSize': 14319,
b528247157b92b5a821/tietoja-kysymyksia.docx', 'XMP:MetsFileSecFi
FLocatType': 'simple', 'XMP:MetsStructMapLabel': 'CSIP', 'XMP:Me
tsstructMap', 'XMP:MetsStructMapDivLabel': 'f10f7de1-058c-4183-b
MapDiv', 'XMP:MetsStructMapDivDivLabel': 'data', 'XMP:MetsStruct
DivDivFptrFileId': 'ID-818c16c5-701c-4552-92af-3535711f9c30'}
2022:09:04 12:57:22+03:00 and after 2022-09-04T12:57:22+03:00
2022:09:04 12:57:22+03:00 and after 2022-09-04T12:57:22+03:00
2022:09:04 12:57:22+03:00 and after 2022-09-04T12:57:22+03:00
ta for 1 files
DC.xml file
a for 1 files

mp/runtimeProcessing/f10f7de1-058c-4183-bc2c-1ca159c6bf76/metada
sing/f10f7de1-058c-4183-bc2c-1ca159c6bf76/metadata/descriptive/D
'DC.xml', 'File:Directory': '/tmp/runtimeProcessing/f10f7de1-05
Size': 557, 'File:FileModifyDate': '2022:09:04 12:57:22+03:00',

OneClick eArchiving

Funded from CEF-TC-2020-2

- Total budget 187 500€
- 21/08/01 – 23/01/31

Originally an idea from AIT, thank you Sven

AIT dropped out

Xamk, EKA, Disec and PIQL carried on with a modified idea

Targets

- Simple automated SIP creation
- Archivematica → EARK IP conformance
- Disec YKSA archive → EARK IP conformance
- Hands on tutorial production

EKA collection migrations

Automated SIP creation

Multiple SIP creation tools exists, but are those simple to use?

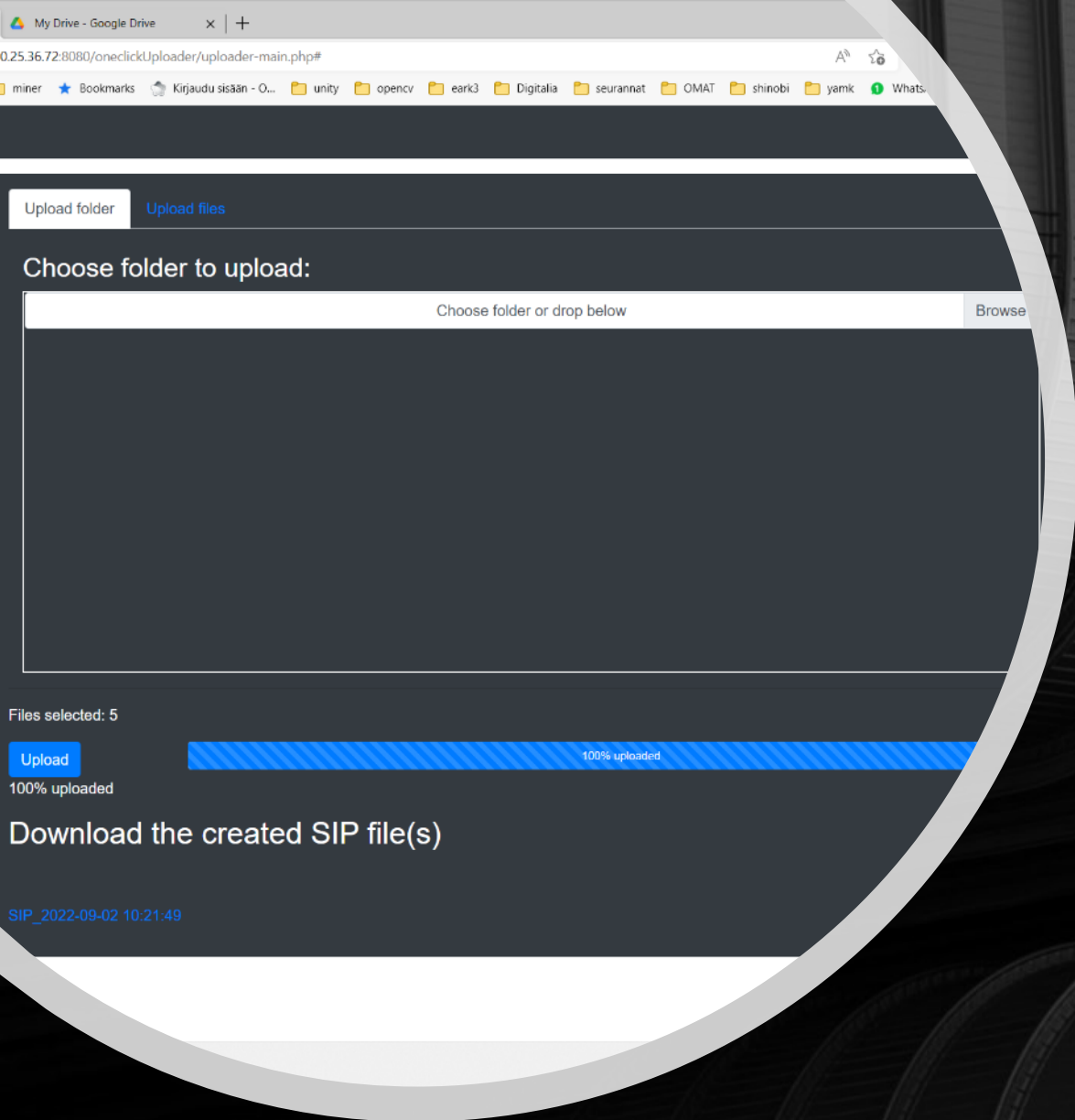Idea → Just add some files/folders and automate everything else

It should be   easy to take into use

easy to use

system independent

→ Dockerized online solution

Basic workflow

End user view	• Upload file(s) & folder(s) • Wait for the SIP to appear

Workflow – system view

Upload	• PHP 8.1
Initiate	• watchdog (watched folder)
Wait	• for upload(s) to be completed
Check	• for viruses
Create	• the dummy SIP structure
Do	• magical things on the background
Construct	• the SIP according to above step
Validate	• the SIP
Produce	• a downloadable package
Return	• the package to the correct webUI

Magical things on the background

- Generate uuid4 identifiers
- Find provided metadata (or generate it on the fly)
- Detect language
- Calculate file sha256 hashes
- Get some metadata for the files
- Fill DC & METS templates
- Fix some date/time formats (explained on problems)
- Replace some identified mime types (explained on problems)

Live demo

```
If (own-computer==true && vpn-
operational == true){
    livedemo
}
Else{
    Play video
}
```

Encountered problems

- Folder uploads and maintaining its structure
- Delayed file copying (within the filesystem)
- Different data/time formats
- Fuzzy specifications & even fuzzier validation results
- Too accurate file type detections
  - application/vnd.ms-pki.seccat & image/x-canon-cr2
    - Validation is based on IANA mime types list
- Files without endings
- Files which exiftool cannot read
- Modifying php upload sizes / timeouts within docker
- Yet to be solved:
  - Creation time of the file(s) is changed on upload
  - Subfolders via webUI / docker

Taking into use

- Install docker (Does not need docker-compose)
- Clone the repo
  - Not yet publicly available. Will be during this autumn, but don't know when ☺
- Change the values in
  - ports.conf (port to be used)
  - config.ini (ip and port settings and metadata filenames)
  - uploads.ini (if larger files || longer execution time is needed)
- Run docker commands
  - docker build -t oneclick .
  - docker run -e TZ=Europe/Helsinki -ti -d -p 8081:8081 oneclick (use the port you defined in config files)
- Should be up and running
  - Check with docker ps

Contact

- Anssi.Jaaskelainen@xamk.fi
- Digitalia.fi
- Digitalia.xamk.fi
- LinkedIn