Archiver

Auto-generated book for Archiver

- ArchiverServiceAgreement
- ArchiveOverHTTP
- ArchiveOverFilesender
- ArchiveFileMod
- Archiver
- ArchiveOverCLI

ArchiverServiceAgreement

Az NIIFI Archiver szolgáltatás célja, hogy intézmények számára hosszú távú, szalagos adattárolás lehetőségét biztositsa.

Föderatív azonosításnál megkövetelt adatok

A szolgáltatás használata céljából az alábbi személyes adatok megadása szükséges:

- *eduPersonPrincipalName* a felhasználó egyedi azonosításához
  - ezzel azonosítja a felhasználót
  - azért van szükség globális azonosítóra, mert a https://webform.niif.hu oldalon feladott szalag-igénylésben használt identitással össze kell tudni kapcsolni a munkamenetet
- *e-mail címre* a portálmotornak van szüksége. Használatával az üzemeltetők szükség esetén értesíteni tudják a felhasználókat a szolgáltatással kapcsolatos problémákról és eseményekről. Az e-mail címet kizárólag a szalagos archiválás szolgáltatáshoz szorosan kapcsolódó módon használjuk fel, harmadik félnek nem adjuk ki.

A szolgáltatás használata során a felhasználó profilja további adatokkal bővülhet, azonban ezek kizárólag a szolgáltatással összefüggő adatok lehetnek. A föderáción keresztül megkapott adatokat, ill. a felhasználó profiljához tartozó adatokat az NIIF Intézet védett módon tárolja, harmadik félnek nem adja ki.

A szolgáltatás használata kizárólag jelen Adatvédelmi Nyilatkozat elfogadása után lehetséges.

Adatok módosíthatósága

A felhasználó a föderatív azonosítás során átadott adatokat nem módosíthatja.

Adattárolás vége

Az NIIF Intézet a profiladatokat a szolgáltatás életciklusának végéig tárolja. A felhasználó rendelkezhet a személyes adatainak és profiljának törléséről. Ezzel egyidőben a hozzá rendelt

szalagokhoz való jogosultságait is elveszti.

Kapcsolat

Az adatok kezelője az **NIIF Intézet**:

- Cím: 1132 Budapest, Victor Hugo utca 18-22, Magyarország
- Tel: +36 1 4503060
- E-mail: archiver *AT_* niif.hu

Az adatkezeléssel kapcsolatos kérdéseivel a fenti e-mail címen elérhető szakemberekhez lehet fordulni.

Szalagokon tárolt adatokhoz való hozzáférés

Meghatározások

Szalag Tulajdonos

- A szalagegységek és meghajtók tulajdonosa az NIIF Intézet. A szalagokhoz fizikai hozzáférés a felhasználók számára nincs lehetőség. **Szalag Adminisztrátor, ill. Felhasználó**
- Minden szalaghoz tartozik egy vagy több adminisztrátor, aki a szalagokon levő adatokat írhatja és olvashatja. Az NIIF Archiver szolgáltatás szempontjából a Felhasználó a Szalag Adminisztrátor.

Adat Tulajdonos

- Az, aki az adatok adminisztrátorának személyéről rendelkezhet. Az NIIF Archiver szolgáltatásában az Adatok Tulajdonosa az Adat Adminisztrátort azonosító intézmény.

Személyes adatok kezelése

Amennyiben a szalagokon tárolt adatok személyes adatokat tartalmaznak, az archiválás előtt a Szalag Adminisztrátorának gondoskodnia kell arról, hogy a személyes adatok jogszerűen tárolhatók legyenek az NIIF Intézet infrastruktúráján.

A szalagokon esetlegesen tárolt személyes adatok kezelője a Szalag Adminisztrátora. A Szalag Tulajdonosa adatfeldolgozónak számít. Az adatok esetleges titkosításért a Szalag Adminisztrátora felelős.

Hozzáférés biztosítása

A szalagokon levő adatokról kizárólag a hozzá rendelt Szalag Adminisztrátora rendelkezhet. Egy szalaghoz tartozó Adminisztrátort az NIIF Intézet kizárólag az alábbi esetekben változtathatja meg (ezáltal hozzáférést biztosítva a szalagon tárolható adatokhoz):

1. amennyiben ezt a Szalag Adminisztrátor azonosított módon kéri elektronikus azonosítás után vagy teljes bizonyító erejű magánokiratban;
2. az Adat Tulajdonos intézmény képviseletre jogosult vezetője cégszerűen aláírt írásos igénylésben jelzi;
3. bíróság jogerős határozata alapján.

Felelősség elhárítása

A felhasználó tudomásul veszi, hogy az NIIF Intézet nem vállal felelősséget a szolgáltatás használatából fakadó károkért, így különösen:

- a szalagok fizikai meghibásodásából fakadó adatvesztésért;
- a metaadatok sérüléséből fakadó adatvesztésért;
- a szalagokon található adatokhoz való illetéktelen hozzáférésből eredő károkért;
- a szolgáltatás szüneteltetéséből vagy megszüntetésből fakadó károkért.

Az adatok tartalmát, ill. a személyes adatkezelés jogosultságát az NIIF Intézet nem vizsgálja. Az adatok tárolásának jogszerűségéért kizárólag a Szalag Adminisztrátora a felelős. A Szalag Tulajdonosa nem tehető felelőssé a Szalag Adminisztrátora által elkövetett esetleges visszaélésekért.

ArchiveOverHTTP

Az archiválás célja az, hogy olyan adatokat mentsünk le, amelyre valószínűleg sohasem lesz szükség, de ha mégis, akkor nagy biztonsággal helyreállítható legyen.

Ez a leírás azt feltételezi, hogy kevés file-t kell szalagra írni. Amennyiben sok állományról van szó, akkor az alábbi lehetőségek közül választhatunk:

- `tar` (zip, rar stb.) segítségével egy file-t csinálunk belőle
  - Ha kevés a helyünk lokálisan, akkor használjuk a `tar --remove-files` opcióját arra, hogy az archiválandó adatokat "belemozgassuk" az archívumba, pl így:
  - `tar -c --remove-files -f cgp.tar cgp`
- kihasználjuk a rekurzív archiválás lehetőségét (DAV vagy SFTP használatával)

Kiszolgálás Apache webszerverrel

Az Apache nem hatékony nagy statikus állományok kiszolgálására, mivel ehhez nagyon sok memóriát foglal(hat). Erre megoldás a mod_xsendfile használata, amellyel egy egyszerű script segítségével hatékonyan szolgálhatjuk ki a file-okat, és tetszés szerint autorizálhatunk. Az autorizáció különösen fontos akkor, ha érzékeny adatokat akarunk menteni, hiszen nem szeretnénk, ha egy publikus webszerverről bárki le tudná azokat tölteni.

Webszerver konfiguráció

Tetszőleges működő http(s) webszerver konfig megteszi. Az alábbi két sort szúrjuk be:

```
XSendFilePath /path/to/archive/
XSendFile on
```

Fontos, hogy az `XSendFilePath` a webszerver *DocumentRoot*-ján kívül legyen, ellenkező esetben xsendfile-on kívüli módon is elérhetőek lehetnek az anyagok.

Kiszolgáló PHP script

```
<?php
```

```
# Place the IP addresses in this array that are authorized to download the
# archives
$trusted_networks = array(
  "193.225.50.212",
  #TODO IPv6!
);

# You should not allow directory listing, and you probably want to be explicit
# about what you serve through this script
$served_files = array(
  "proba.txt",
  "cgp-metadata.txt",
  "cgp.tar",
  "szfe-orphaned_20130318.tar",
  "maildirs2-orphaned_20130318.tar",
);

# The directory where the archives reside
$dir = "/srv/maildirs/archive";

$ip = $_SERVER['REMOTE_ADDR'];
$file = $_GET['file'];

if (!$file) {
  echo "Missing 'file' parameter";
  exit(1);
}
if (!preg_match('/^[a-z0-9A-Z\-_]+\.?[a-z0-9A-Z\-]*$/',$file)) {
  echo "Invalid 'file' parameter";
  exit(1);
}

if (in_array ($ip, $trusted_networks)) {
  if (in_array ($file, $served_files)) {
    header("Content-Type: application/octet-stream");
    header("Content-Disposition: attachment; filename=\"$file\"");
    header("X-Sendfile: $dir/$file");
    exit(0);
  } else {
```

```
      echo "Sorry, this file can not be served!";
      exit(1);
    }
  } else {
    echo "Your IP ($ip) is not authorized!\n";
    exit(1);
  }

exit(0);
```

A scriptet nevezzük el mondjuk `archiver-get.php`-nek. A programhoz való hozzáférést az Apache további eszközeivel is szűkíthetjük (Basic Auth, IP korlátozás stb), de ezek már redundáns védelmet nyújtanak.

Tiszta URL-ek

Ugyan az archiver támogatja a query stringek (`?file=foobar.txt`) használatát, azonban ez kényelmetlen, mivel át kell nevezni a végeredményt ahhoz, hogy értelmes nevet kapjunk. Ezért mindenképpen érdemes egy egyszerű `mod_rewrite` trükköt használni:

```
<IfModule mod_rewrite.c>
  RewriteEngine on
  RewriteRule ^(.*)$ /archiver-get.php?file=$1 [L]
</IfModule>
```

A .htaccess file-t én az archiver könyvtárban helyeztem el, ezután az URL így nézett ki:

- https://szerverneve.niif.hu/archiver/proba.txt

Debian megjegyzés

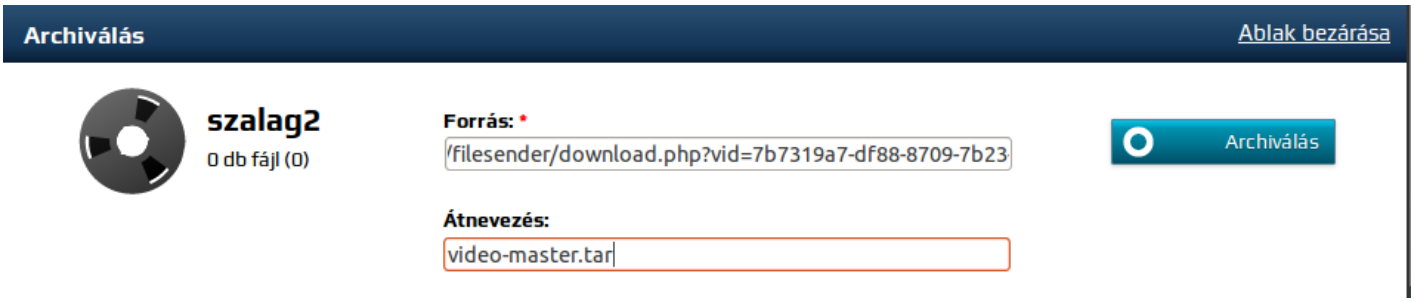
Nagy file-ok kiszolgálására a `mod_xsendfile`-nak legalábba 0.11-es verzióját kell használnunk. Ez (a szöveg írása idején) nincs a Debian Squeeze-ben, azonban minden további nélkül használhatjuk itt is a wheezy-s bináris deb csomagot.

ArchiveOverFilesender

Az archiver aszinkron módon működik, így közvetlenül a böngészőből nem tölthetünk fel archiválandó file-okat, azonban van arra lehetőség, hogy a **Filesender** szolgáltatását használva a desktop gépünkről archiváljunk adatokat.

Az ötlet nagyon egyszerű:

1. Feltöltjük az állományt a filesenderre
2. Saját magunknak elküldjük a file-t
3. Levélben megkapjuk a file letöltési URL-jét. Ezt kimásoljuk.
4. Az archiváláshoz a '?' karakter elé be kell szúrni: download.php.
5. A mentéshez mindenképpen file nevet kell megadni.



ArchiveFileMod

Az archiválandó fájl- és könyvtárnevek nem tartalmazhatnak szóköz karaktert, ezért azokat szükséges megfelő formátumra hozni a következő két parancs futtatásával:

```
find -name "* *" -type d | rename 's/ /_/g' # do the directories first
find -name "* *" -type f | rename 's/ /_/g'
```

Az első parancsot addig kell újra és újra futtatni, amíg hiba nélkül le nem fut!

Archiver

Archiver

Üdvözöljük A NIIF Archiver WIKI oldalán! Az Archiver az Intézet tagintézményeinek létrehozott szolgáltatása, és archiválási lehetőséget biztosít adattároló szalagokra. További információt a GYIK menüben kaphat a szolgáltatásról. Figyelem! A mentett adatokért az Intézet semmilyen felelősséget nem vállal. Az Archiver a https://archiver.niif.hu/ oldalon érhető el. Hozzáférést igényelni az Intézet szolgáltatásigénylő oldalán lehet.

Tippek, trükkök

- Fájl- és könyvtárnevek megfelelő formára hozása
- Archiválás HTTP-n keresztül
- Archiválás Filesenderen keresztül (pl. desktop gépről)
- Archiválás használata API-n keresztül

- Archiver adatvédelmi nyilatkozat
- Archiver felhasználói útmutató (külső link)

ArchiveOverCLI

ArchiveOverCLI

Lehetőség van az archiválást nem csak a webes felületen keresztül használni. Erre a célra ki lett alakítva egy RESTAPI interfész, amin keresztül elérhetőek a szükséges funkciók.

Az API használatát segédletekkel megtalálhatod https://api.archiver.niif.hu/?help oldalon.

A használathoz szükséges *token* a webes felületen a *Token* fül alatt érhető el.

Funkcionalitások kivonatolt segédlete:

Szalag műveletek

- útvonal: **/tape**
  - *GET* - lekérdezhető az elérhető szalagok azonosítói
    - paraméterek: -
    - OPCIONÁLIS paraméterek:
      - **tape_id** - amennyiben csak egy szalag adatai szeretnénk lekérdezni
    - válasz:
      - Egy JSON, mely tartalmazza az elérhető szalagok azonosítóit és neveit.

```
[
    { "id" : <szalag id>,
        "name" : "szalag neve"
    },
    ...
]
```

    - példa:

```
curl --header "token:<TOKEN>" -X GET 'https://api.archiver.niif.hu/tape/'
```

- *POST* - szalag nevét lehet beállítani
  - paraméterek:
    - **tape_id** - szalag azonosítója
    - **name** - beállítandó név
  - válasz: -
  - példa:

```
curl --header "token:<TOKEN>" https://api.archiver.niif.hu/tape/ --data
"tape_id=<ID>&name=<NAME>"
```

- *DELETE* - szalag tartalmának törlése
  - paraméterek:
    - **tape_id** - melyik szalag tartalmát szeretnénk törölni
  - válasz: -
  - példa:

```
curl --header "token:<TOKEN>" -X DELETE
'https://api.archiver.niif.hu/tape/?tape_id=<ID>'
```

Feladatokkal kapcsolatos műveletek

- útvonal: **/archive**
  - *POST* - új archiválási feladata beküldése
    - paraméterek:
      - **tape_id** - szalag azonosító
      - **url** - archiválandó forrás elérhetősége Amennyiben az url tartalmaz '&'-t, azt '%26'-val kell helyettesíteni.
    - OPCIONÁLIS paraméterek:
      - **name** - átnevezés
      - **email** - értesítendő email címe
      - **level** - értesítési szintszintek:
        - 1 - csak a sikeres archiválásról kérek értesítés
        - 2 - sikeres és hibás archiválásról is kérek értesítést
    - válasz:
      - A sikeresen beküldött feladat ID-ja.
    - példa:

```
curl --header "token:<TOKEN>" https://api.archiver.niif.hu/archive/ --data
"tape_id=<ID>&url=<URL>"
curl --header "token:<TOKEN>" https://api.archiver.niif.hu/archive/ --data
```

```
"tape_id=<ID>&url=<URL>&name=<NAME>"
```

- *GET* - lekérdezhető a beküldött feladat állapota
  - paraméterek:
    - **job_id** - lekérdezendő feladat ID-je
  - válasz:
    - Egy JSON, mely tartalmazza a feladat és az archivált fájl(ok) státuszát. A "directories" elem csak hibás könyvtár létrehozáskor jelenik meg a JSON-ban.

```
{
    "state":"feladat állapota",
    "directories" : [
        {
            "dir":"könyvtár neve",
            "state":"error",
            "error_message":"HTTP error: 500 \"Internal Server Error\""
        },
        ...
    ],
    "files" : [
        {
            "file":"fájl neve elérési úttal",
            "state":"complete",
            "checksum":"c9b96a05482a91970309f8087b3ac9c6",
            "ts":"2015-04-14 15:53:01"
        },
    ...
    ]
}
```

  - példa:

```
curl --header "token:<TOKEN>" -X GET
'https://api.archiver.niif.hu/archive/?job_id=<ID>'
```

- útvonal: **/restore**
  - *POST* - új visszaállítási feladata beküldése
    - paraméterek:
      - **tape_id** - szalag azonosító
      - **url** - visszaállítás helye Amennyiben az url tartalmaz '&'-t, azt '%26'-val kell helyettesíteni.

- **file** - visszaállítandó fájl, több verzió esetén a legutolsó változat kerül visszaállításra
- OPCIONÁLIS paraméterek:
  - **name** - átnevezés
  - **email** - értesítendő email címe
  - **level** - értesítési szint szintek:
    - 1 - csak a sikeres visszaállításról kérek értesítés
    - 2 - sikeres és hibás visszaállításról is kérek értesítést
- válasz:
  - A sikeresen beküldött feladat ID-ja.
- példa:

```
curl --header "token:<TOKEN>" https://api.archiver.niif.hu/restore/ --data
"tape_id=<ID>&url=<URL>&file=<FILE>"
curl --header "token:<TOKEN>" https://api.archiver.niif.hu/restore/ --data
"tape_id=<ID>&url=<URL>&file=<FILE>&name=<NAME>"
```

- *GET* - lekérdezhető a beküldött feladat állapota
  - paraméterek:
    - **job_id** - lekérdezendő feladat ID-je
  - válasz:
    - Egy JSON, mely tartalmazza a feladat és a visszaállított fájl(ok) státuszát. A "directories" elem csak hibás könyvtár létrehozáskor jelenik meg a JSON-ban.

```
{
    "state":"feladat állapota",
    "directories" : [
        {
            "dir":"könyvtár neve",
            "state":"error",
            "error_message":"HTTP error: 500 \"Internal Server Error\""
        },
        ...
    ],
    "files" : [
        {
            "file":"fájl neve elérési úttal",
            "state":"complete",
            "checksum":"c9b96a05482a91970309f8087b3ac9c6",
            "ts":"2015-04-14 15:53:01"
        },
```

```
    ...
    ]
}
```

- példa:

```
curl --header "token:<TOKEN>" -X GET
'https://api.archiver.niif.hu/restore/?job_id=<ID>'
```

Fájl műveletek

- útvonal: **/file**
  - *GET* - lekérdezhető a szalagon található fájlok listája
    - paraméterek:
      - **tape_id** - szalag azonosító
    - OPCIONÁLIS paraméterek:
      - **path** - megadható, hogy a szalagon milyen path alatt található fájlokat listázzon csak. Természetesen a lista tartalmazza a path alatt található alkönyvtárakban lévő fájlokat is.
    - válasz:
      - Egy JSON, mely tartalmazza a szalagon elérhető fájlokat és azok méretét.

```
[
    { "file":"fájl neve elérési úttal",
        "size":"méret bájtban"
    },
    ...
]
```

  - példa:

```
curl --header "token:<TOKEN>" -X GET
'https://api.archiver.niif.hu/file/?tape_id=<ID>'
curl --header "token:<TOKEN>" -X GET
'https://api.archiver.niif.hu/file/?tape_id=<ID>&path=<PATH>'
```