NAME

make_bagit -- Create a BagIt archive of a folder, according to RFC 8493

SYNOPSIS

make_bagit -b *input_folder*
make_bagit -i *input_folder* **-o** *output_folder*
make_bagit -h | -x

DESCRIPTION

Bash_AVpres is a collection of Bash scripts for audio-visual preservation. One of these small programs is **make_bagit**. It creates an archive of a folder, according to the "BagIt File Packaging Format", as designed by the Library of Congress in the U.S.A. and standardised under RFC 8493. It also supports hash algorithms which are not recommended by RFC 8493.

This script should work correctly with Bash from version 3.2 (released back on 2006-10-11), however we strongly advise using the current version 5.3.9 (released on 2025-12-10).

OPTIONS

BASIC OPTIONS

-b *input_folder*, **--bagit=***input_folder*
input folder

This replaces the *input_folder* by its BagIt archive, containing the original folder, and is faster.

-i *input_folder*, **--input=***input_folder*
input folder

-o *output_folder*, **--output=***output_folder*
output BagIt folder

This copies the content of the *input_folder* into the *output_folder* while creating a BagIt archive, and requires more time.

ADVANCED OPTIONS

The arguments of the advanced options can be overwritten by the user, either by passing a different value when calling the script, or by setting it in the configuration file. Please remember that any string containing spaces must be quoted or its spaces must be escaped.

--bagit_version=*bagit_version*
RFC 8493 defines BagIt version 1.0 and **make_bagit** uses this by default:

--bagit_version=1.0

For compatibility reasons, however, the values from 0.93 to 0.97 are also considered valid and supported.

--bag_info=*text_file*
include additional informations into "bag-info.txt", without any verification

--fetch=*text_file*
include into the BagIt archive a "fetch.txt" file, without any verification

--algorithm=(*md5*|*sha1*|*sha256*|*sha512*|*xxh128*|*xxh64*|*xxh32*|*crc32*)
RFC 8493 considers the four unkeyed cryptographic hash functions MD5, SHA-1, SHA-256 and SHA-512, which is the default:
--algorithm=sha512

make_bagit and **verify_bagit** support all four hash algorithms, but the use of different and/or multiple algorithms in the same BagIt archive is currently not supported.

In addition, **make_bagit** and **verify_bagit** also support the four non-cryptographic algorithm *xxh128*, *xxh64*, *xxh32* and *crc32*. Please note that these are not recommended in the RFC 8493 specification.

The algorithm name can be passed in upper or lower case letters.

--with_scripts=(*yes*|*no*)
include or not the **make_bagit**, **verify_bagit** and **undo_bagit** scripts into the BagIt archive (default is *no*)

The script needs an external command to compute recursively the hash checksums of all elements inside the folder.

--md5='/bin/md5sum'
MD5 command

--sha1='/bin/sha1sum'
SHA-1 command

--sha256='/bin/sha256sum'
SHA-256 command

--sha512='/bin/sha512sum'
SHA-512 command

--xxh128='/bin/xxsum -H128'
xxHash 128 command

--xxh64='/bin/xxsum -H64'
xxHash 64 command

--xxh32='/bin/xxsum -H32'
xxHash 32 command

--crc32='/bin/crc32'
CRC-32 command

--cp=*copy_tool*
The script needs a copy command when the **-i** or **--input** and **-o** or **--output** options are used.

By default, **gcp** is used rather than the classic **cp**, because the GNU flavour offers more possibilities, such as preserving the files' original date and time metadata. The default configuration is:
--cp='/bin/gcp --preserve=mode,timestamps -r'

--exclusion=(*file_list***)**
The exclusion *file_list* is defined as an array and its default value is:
exclusion=('.DS_Store' 'desktop.ini')

To avoid any exclusion, use an empty array:
exclusion=()

INFORMATIVE OPTIONS

-h, --help
display a help message

-x, --options
display the advanced options with their default arguments and indicate if a local configuration has been defined or not

NOTES

BAGIT ARCHIVE STRUCTURE

The structure of a BagIt archive is:

bagit.txt
data/<ARCHIVE_CONTENT>
manifest-<*algorithm*>.txt
[bag-info.txt]
[fetch.txt]
[tagmanifest-<*algorithm*>.txt]

Only the "bagit.txt" and "manifest-<*algorithm*>.txt" files are mandatory, in addition to the "data" folder, which contains the actual content of the archive.

According to RFC 8493, the "bag-info.txt" and "tagmanifest-<*algorithm*>.txt" files are optional and **make_bagit** creates these always. The "bag-info.txt" file contains metadata elements describing the bag and its payload and is intended primarily for human use. Yet the "Payload-Oxum" metadata is used by **verify_bagit** to quickly detect incomplete bags, before performing any checksum validation. The "tagmanifest-<*algorithm*>.txt" file contains the checksum manifest of all the other files generated.

The optional "fetch.txt" file without any verification can be included in a BagIt archive with **make_bagit**. Only its presence can be checked with **verify_bagit**.

In addition, for sustainability and transparency reasons, **make_bagit** can embed its own source code as well as **verify_bagit** and **undo_bagit** ones into the BagIt folder: "make_bagit.txt" is the script used to actually generate the BagIt archive and "verify_bagit.txt" can be used to check its integrity. The action of **make_bagit** can be reverted by the **undo_bagit** command, embedded as "undo_bagit.txt". The embedding can be activated by passing the **--with_scripts='yes'** option.

Note that by default the ".DS_Store" files generated by macOS and the "desktop.ini" files generated by Windows are deleted before the BagIt archive is created. You can modify the default exclusion file list in the configuration file or by passing the **--exclusion=()** option.

Any space in folder and file names is replaced by an underscore sign before the BagIt archive is created.

CONFIGURATION FILE

An external configuration file

${HOME}/.config/AVpres/Bash_AVpres/make_bagit.txt

can be defined, allowing the script to import alternate default values for the following options:

bagit_version
algorithm

md5
sha1
sha256
sha512
crc32
xxh32
xxh64
xxh128
cp
with_scripts
exclusion

Example:
algorithm='xxh128'
with_scripts='yes'

LOG FILES

Temporary log files are stored at
/tmp/AVpres/make_bagit.XXXXXXXXXX

The log files can be used for debugging, for example by running **cat** on the address prompted with fatal error messages:

cat /tmp/AVpres/make_bagit.*XXXXXXXXXX*

SEE ALSO

RFC 8493 -- John A. Kunze, Justin Littman, Liz Madden and Chris Adams: "The BagIt File Packaging Format (V1.0)", October 2018
https://www.rfc-editor.org/info/rfc8493

RFC 1321 -- Ronald L. Rivest: "The MD5 Message-Digest Algorithm", April 1992
https://www.rfc-editor.org/info/rfc1321

RFC 6234 -- Donald Eastlake and Tony Hansen: "US Secure Hash Algorithms (SHA and SHA-based HMAC and HKDF)", May 2011
https://www.rfc-editor.org/info/rfc6234

Yann Collet: "xxHash fast digest algorithm", version 0.2.0, 2023-06-29
https://github.com/Cyan4973/xxHash/blob/dev/doc/xxhash_spec.md

undo_bagit(1) and **verify_bagit**(1); **xxhsum**(1), **md5sum**(1), **sha1sum**(1), **sha256sum**(1) and **sha512sum**(1).

COPYRIGHT

Copyright (c) 2014-2026 by Reto Kromer
Copyright (c) 2026 by Michal Cohen

LICENSE

The **make_bagit** Bash script is released under a 3-Clause BSD License.

DISCLAIMER

The **make_bagit** Bash script is provided "as is" without warranty or support of any kind.