

Workflow zum Hochladen von Mediendateien

- Upload-Workflow
 - Metadaten eintragen
 - Prüfsumme berechnen
 - Dateien hochladen
 - Feierabend
 - Abgelehnte Dateien korrigieren
- Korrekturlieferungen
- Vorgehen bei falschen Prüfsummen
- Limits

Upload-Workflow

Metadaten eintragen

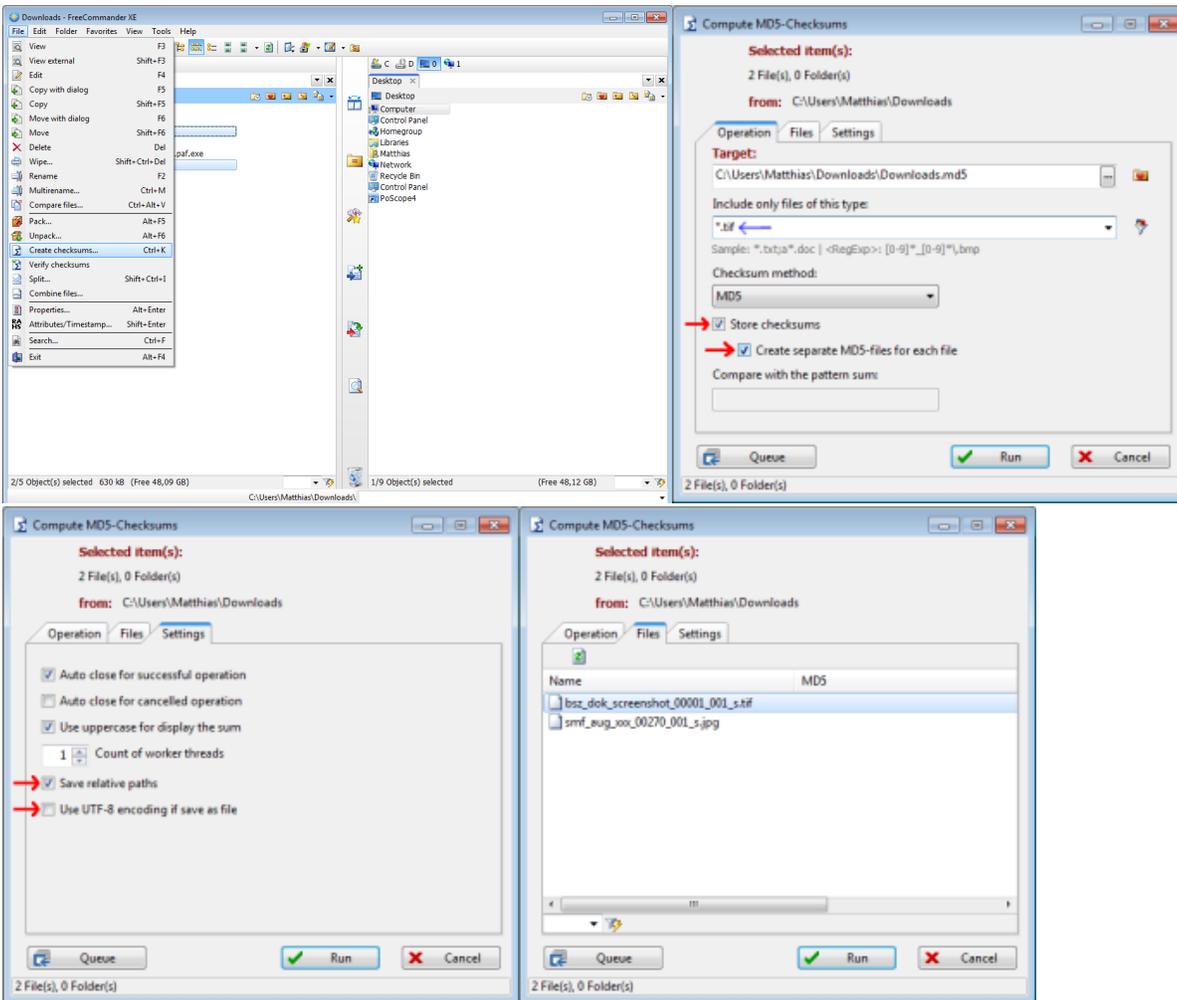
- XMP-Metadaten in den Header der Bilddatei eintragen oder aktualisieren, falls erforderlich. Siehe [Metadatenfelder in Dateiname und Dateiheder \(XMP\)](#).
- Bilddatei lokal entsprechend der [Dateinamenskvention für die Medienbereitstellung](#) benennen. Auf Kleinschreibung achten!
- Darauf achten, dass die Aufnahmekennung nicht schon vorhanden ist. D.h., wenn es Aufnahmekennung 001 schon gibt, für weitere Bilder Aufnahmekennung 002 usw. vergeben. ([Korrekturlieferungen siehe unten](#).)

Prüfsumme berechnen

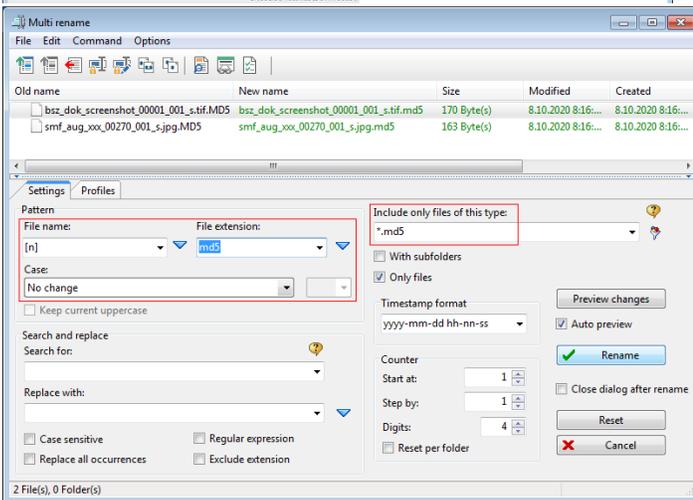
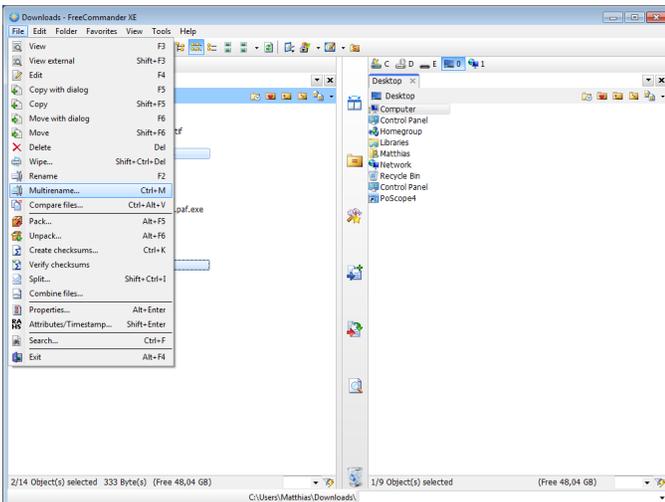
Es wird dringend empfohlen, Prüfsummen für alle Bilddateien zu berechnen, um Übertragungsfehler auszuschließen. Lieferungen ohne Prüfsummen werden angenommen und verarbeitet – in der Hoffnung, dass keine Bitfehler bei der Übertragung aufgetreten sind.

In FreeCommander: Gewünschte Dateien (nicht den beinhaltenden Ordner!) auswählen, dann Datei Prüfsummen berechnen auswählen. Prüfsummen müssen in separaten Dateien gespeichert werden, mit relativen Pfaden und nicht mit UTF-8 Kodierung (siehe rot markierte Häkchen in den Screenshots). Die Angabe der Zieldatei sowie sämtliche anderen Häkchen unter "Settings" sind irrelevant.

Sonderfall: Wenn bereits Prüfsummendateien vorhanden sind und diese aktualisiert werden sollen, ist es nicht praktikabel einzeln hunderte TIFF-Dateien (aber nicht die MD5-Dateien) auszuwählen. Stattdessen alle Dateien auswählen und im Feld zum Dateinamensfilter (blaue Markierung) mit *.tif (bzw. *.jpg) die Verarbeitung auf TIFF- (bzw. JPEG-)Dateien einschränken.



FreeCommander erstellt die Dateien als .tif.MD5 (Großbuchstaben). Deshalb alle Dateien auswählen, Multirename starten, im Feld Dateierdung md5 (Kleinbuchstaben) eingeben und auf *.md5 einschränken. Das Feld Groß-Kleinschreibung sollte hier auf "No change" stehen; FreeCommander ändert die Dateierdung trotzdem auf Kleinschreibung. (Mit Multirename können auch Großbuchstaben im Dateinamen "verkleinert" werden. Das muss dann aber vor der Berechnung der Prüfsummen erfolgen!)



Die Prüfsummendatei für xyz_abc_xxx_12345_001_s.tif muss xyz_abc_xxx_12345_001_s.tif.md5 oder xyz_abc_xxx_12345_001_s.md5 heißen, d.h. mit Endung .tif.md5 oder .md5 und komplett in Kleinbuchstaben (alternativ mit .jpg, wenn das Bild eine JPEG-Datei ist – nicht empfohlen wegen Qualitätsverlust). Die Prüfsummendatei darf keinen Pfad enthalten und der Dateiname in der Prüfsummendatei muss dem Namen der Bilddatei entsprechen. Folgende Datei würde beispielsweise abgelehnt:

```
# MD5 checksums generated by FreeCommander (http://www.freecommander.com) 10.08.20
047642F8A7DF1D4663C8587EF4F49C7C *Dokubilder\bsz_dok_Screenshot_00001_001_a.tif
```

Korrekt ist die Angabe nur mit Dateiname, z.B. eine Prüfsummendatei bsz_dok_screenshot_00001_001_s.tif.md5 (oder bsz_dok_screenshot_00001_001_s.md5) für bsz_dok_screenshot_00001_001_s.tif:

```
# MD5 checksums generated by FreeCommander (http://www.freecommander.com) 10.08.20
047642F8A7DF1D4663C8587EF4F49C7C *bsz_dok_screenshot_00001_001_s.tif
```

Der oben gezeigte Prozess erstellt Dateien, die diesen Anforderungen genügen, sofern die eigentlichen Dateien und nicht der beinhaltende Ordner ausgewählt werden.

Bei bereits vorhandener Prüfsummendatei: Darauf achten, dass

- der Dateiname in der Prüfsummendatei kleingeschrieben ist und dem Namen der Bilddatei entspricht.
- der Dateiname in der Prüfsummendatei keinen Pfad enthält.

Dateien hochladen

Bild- und Prüfsummendatei auf SFTP-Server hochladen (Ordner `upload` für Produktion, `test` für Test). Die Dateien müssen direkt in `upload` bzw. `test` sein und nicht in Unterordnern davon! Dateien in Unterordnern werden ignoriert (insbesondere auch Dateien in `invalid`).

Bei der Bilddatei können Master (`_m.tif`) und Submaster (`_s.tif`) vorhanden sein oder nur Submaster oder nur Master. Wenn ein Master oder Submaster nachgeliefert wird, muss entweder nur die zu ergänzende Datei hochgeladen werden (empfohlen), oder die Aggregation muss vorher in `expo.media` gelöscht werden (nicht empfohlen). Bei Ergänzungslieferungen wird das Medienobjekt in `imdas pro` ersetzt, d.h. eventuelle Texte für Onlinekataloge müssen vom alten Objekt kopiert werden.

Feierabend

Feierabend machen. Der Import läuft über Nacht (ab 19:00).

Abgelehnte Dateien korrigieren

Nach dem nächtlichem Import den Ordner `upload\invalid` (bzw. `test\invalid`) prüfen, sofern vorhanden. Dieser enthält abgelehnte Dateien, die nicht der Dateinamenskonvention entsprechen, deren Prüfsummen inkorrekt waren oder die eine bestehende Datei ersetzen würden. Darauf achten, dass

- der Dateiname keine Großbuchstaben enthält oder Bestandteile im Dateinamen fehlen.
- der Dateiname alle Unterstriche (Trennzeichen für Dateinamenszerlegung bei Verarbeitung) an der den richtigen Stellen enthält.
- die Inventarnummer keine Unterstriche enthält (Unterstrich ist in der Inventarnummer nicht als Trennzeichen erlaubt, siehe [reduzierte Inventarnummern](#).)

Dateien, deren Name nicht der Dateinamenskonvention entspricht oder die schon vorhanden sind, lokal entsprechend umbenennen, Prüfsummen neu berechnen und erneut hochladen. Wenn nur die Prüfsumme inkorrekt war, entfällt das Umbenennen natürlich.

Dateien sollten nicht auf dem Server umbenannt werden, weil dann der Dateiname in der Prüfsummendatei nicht mehr mit dem Dateinamen der Bilddatei übereinstimmt. Außerdem besteht das Risiko, sie versehentlich im Ordner "invalid" zu belassen, was dann zu Verwirrung führt wenn sie nicht importiert werden. (Der Inhalt des Ordners "invalid" wird beim Import ignoriert.)

Der Ordner `upload\invalid` (bzw. `test\invalid`) kann gelöscht werden, wenn die Dateien darin erneut hochgeladen wurden (oder anderweitig nicht mehr benötigt werden). Dann zeigt die Existenz des `invalid`-Ordners, dass Dateien abgelehnt wurden. Alternativ die entsprechende Bilddatei darin löschen, wenn erneut hochgeladen, und den Ordner regelmäßig auf abgelehnte Dateien überprüfen.

Korrekturlieferungen

Wenn eine Datei (Master/Submaster) *ergänzt* werden soll, handelt es sich nicht um eine Korrekturlieferung! Stattdessen nur die zu ergänzende Datei mit korrektem Dateinamen gemäß dem normalen Prozess hochladen. Sie wird dann automatisch zur bestehenden Aggregation hinzugefügt.

- Wenn nur eine der Dateien (Master oder Submaster) ersetzt werden soll, die andere (Submaster bzw. Master) von `expo.medoa` herunterladen, wenn lokal nicht mehr vorhanden.
- Die bestehende Aggregation auf `expo.media` löschen. Wenn sie dort nicht vorhanden ist, weil ihre Verarbeitung gescheitert ist (und nur dann!), kann auf diesen Schritt verzichtet werden.
- Optional aber empfohlen: Vorhandenes Medienobjekt in `imdas pro` vor Upload löschen, auch wenn nur eine der Versionen (Submaster oder Master) ersetzt werden soll.
- Danach die aktuelle(n) Datei(en) für diese Aggregationen hochladen, wie bei einer normalen Lieferung. Wenn nur eine davon (Master oder Submaster) ersetzt wurde, muss trotzdem die andere (Submaster bzw. Master) neu mit hochgeladen werden.

Der Rest des Vorgangs ist identisch zu einer normalen Lieferung. (Korrekturlieferungen sind nichts anderes als eine Löschung gefolgt von einer normalen Lieferung.)

Vorgehen bei falschen Prüfsummen

Wenn die Prüfsumme einer Datei nicht dem entspricht, was in der zugehörigen Prüfsummendatei angegeben ist, geht das System sicherheitshalber davon aus dass Datenkorruption vorliegt, und bricht die Verarbeitung komplett ab. Die entsprechende Fehlermeldung hat das Format:

```
ERROR: Fehler beim Verarbeiten der TrafoPipeline: SftpDownloadPipe: Bei der Checksummen-Prüfung der Datei /mnt/lsdf/imdas/smf/mne/afri/i-0819/001/smf_mne_afri-i-0819_001_s.tif.md5 ist ein Fehler aufgetreten: Für die Datei /mnt/lsdf/imdas/smf/mne/afri/i-0819/001/smf_mne_afri-i-0819_001_s.tif stimmt die generierte Checksumme d3f68440233fd84599a6e4c1470b2b4f stimmt nicht mit der originalen Checksumme 9f6029ad021dd981adf7002f4ac0ac52 überein!
```

Gründe können sein:

- Bilddatei wurde nach Berechnung der Prüfsumme verändert (deshalb sollte die Prüfsumme als letzter Schritt berechnet werden!)
- Bilddatei passt nicht zur Prüfsummendatei (z.B. es wurde eine ältere Version des Submasters hochgeladen, als die zu der die Prüfsumme berechnet wurde)
- Bilddatei wurde nicht korrekt übertragen

Zur Fehlerbehebung ist die Prüfsumme der lokal vorhandenen Datei entsprechend dem obigen Workflow neu zu berechnen. Danach muss die Prüfsummendatei erneut hochgeladen werden.

Datenkorruption ist bei Übertragung mittels SFTP technisch extrem unwahrscheinlich. Der empfohlene Workflow ist so gestaltet, dass die organisatorischen Fehler minimiert werden.

Limits

Pro Lieferung sind folgende Limits einzuhalten. Wenn noch unverarbeitete Daten im upload-Ordner liegen sollten, ist von einem Verarbeitungsfehler auszugehen und von einer erneuten Lieferung abzusehen.

Limit	Maximum	Anmerkungen
Datenmenge gesamt	55GB	Bei Lieferung aus einem Backlog zu liefernder Dateien sollten pro Tag ca. 50GB geliefert werden, also ziemlich dicht am Limit.
Anzahl Bilddateien	TBD	Bei Lieferung von Master + Submaster zählt pro Aggregation nur der Submaster. Aktuell einfach so viele Dateien liefern, bis die Datenmenge erreicht ist.
Auflösung Einzelbild	16-bit: 65Mpx 8-bit: >80Mpx	Breite×Höhe, d.h. 65Mpx = nicht größer als 8000x8000, 6000x11000, 4000x16000, 2000x33000. Genaueres Limit für 8-bit ist noch zu ermitteln.
Dateigröße Einzelbild	(16-bit: 400MB) (8-bit: >240MB)	Ausschlaggebend ist nur die maximale Auflösung. Unkomprimierte Bilder brauchen aber bei 16-bit 6 Byte pro Pixel und bei 8-bit 3 Byte pro Pixel, d.h., aus der Dateigröße ergibt sich meistens die Auflösung.
Ende des Uploads	19:00:00	Der Ingest-Job startet um 19:00 Uhr. Dateien, die zu diesem Zeitpunkt noch nicht vollständig übertragen waren, können beim Ingest korruptiert werden (der noch nicht übertragene Teil kann fehlen).

Extrem große Einzelbilder (z.B. 33098x4758 = 158M Pixel) können nach vorheriger Absprache mit dem BSZ importiert werden. Sie dürfen nicht mit einer normalen Lieferung geliefert werden, weil sie wegen Ressourcenlimits auf dem Server tagsüber separat verarbeitet werden müssen.