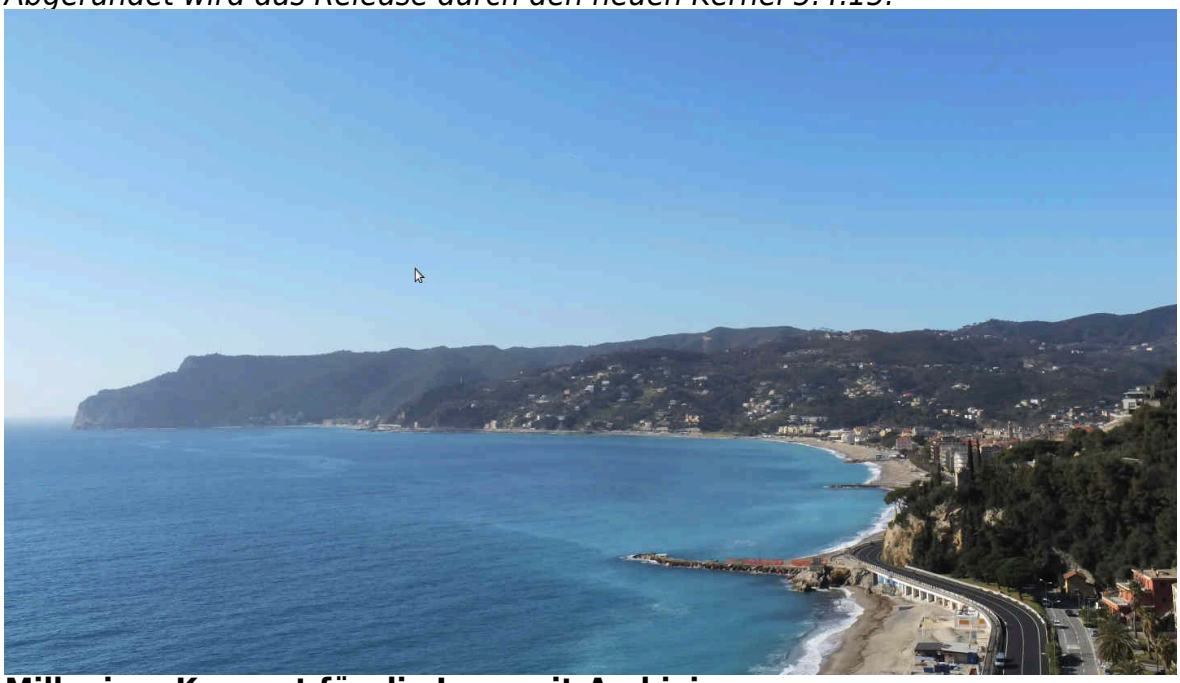
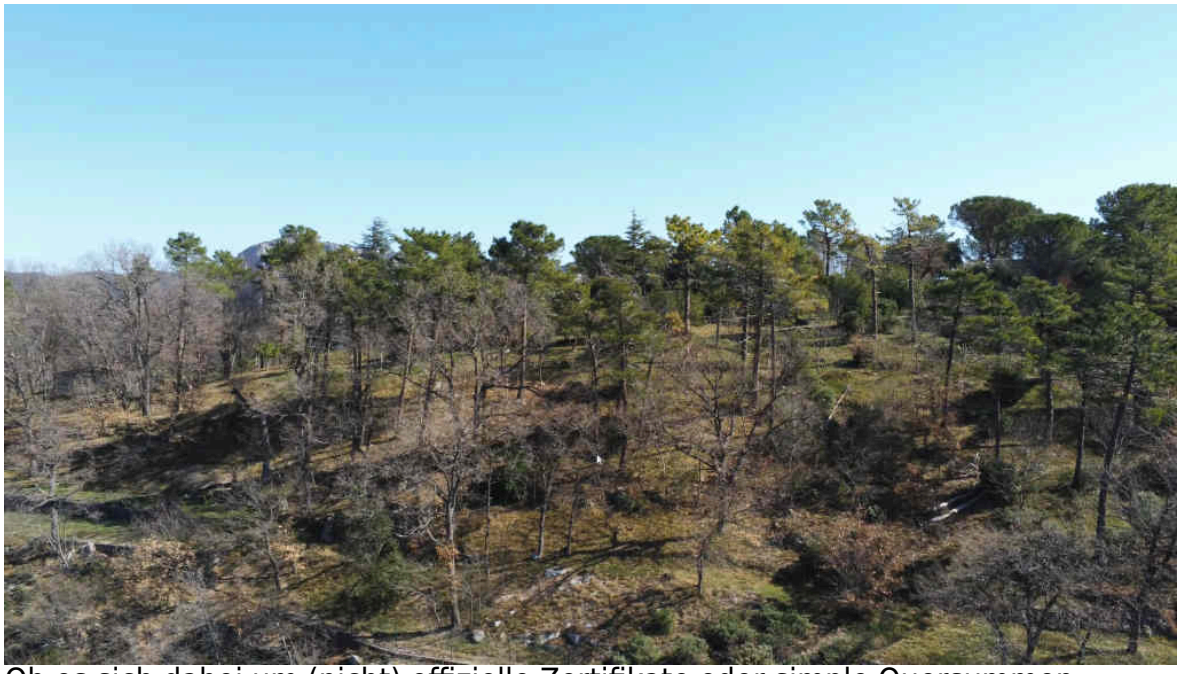
ArchivistaBox & AVMultimedia mit Langzeit-Archivierung
Egg, 31. Januar 2020: *Sowohl AVMultimedia wie auch die ArchivistaBox 2020/II sind neu in der Lage, optische Datenträger mit bis zu 100 GByte (BDR-XL) zu erstellen. Damit können (dies ist gerade bei multimedialen Inhalten wichtig) neu auch umfangreichste Projekte auf optische Medien archiviert werden. Daneben gibt es im neuen Release Support für die neuesten AMD-Prozessoren (z.B. Ryzen 3950X) mit bis zu 32 Threads und auch die Unterstützung für Datennträger mit bis zu 16 TByte wurde implementiert. Abgerundet wird das Release durch den neuen Kernel 5.4.15.*



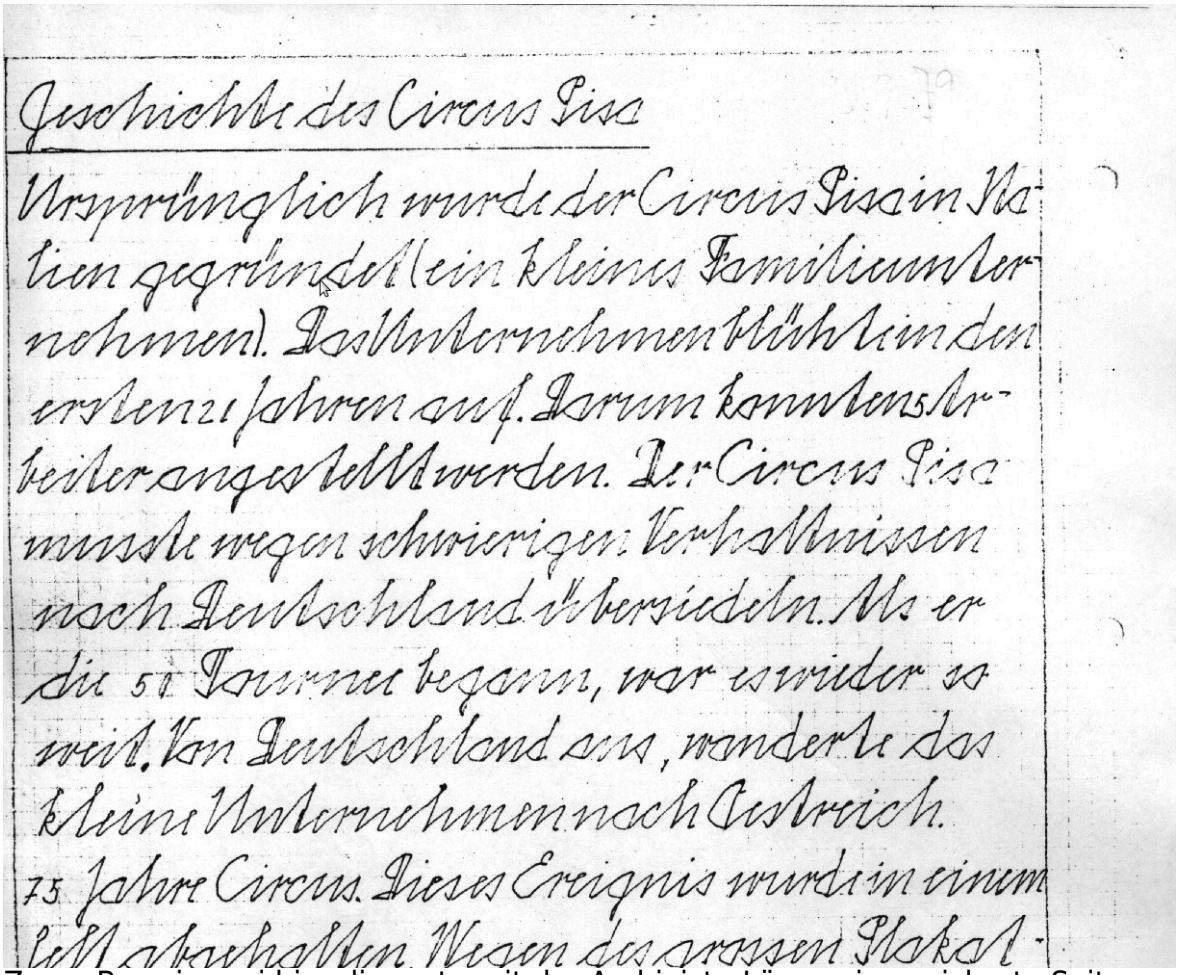
Millenium-Konzept für die Langzeit-Archivierung
Wer heute Daten speichert, der verwendet dabei ausschliesslich Festplatten. Moderne SSD-Disk bieten extrem schnelle Zugriffszeiten. Allerdings geht dabei oft vergessen, dass Daten auf Festplatte «nur flüchtig» geischert werden. Genauso schnell wie diese auf die Datenträger geschrieben werden können, sind sie auch wieder gelöscht, oder noch schlimmer, auch unbemerkt geändert. Gerade bei Systemen im Bereich Document Management (DMS) als auch bei die Archivierung kommen daher seit geraumer Zeit Prüfverfahren zur Anwendung, um sicherzustellen, dass die Daten nicht geändert wurden.



Ob es sich dabei um (nicht) offizielle Zertifikate oder simple Quersummen-Prüfverfahren (Stichwort MD5) handelt, spielt dabei keine Rolle. Mit den entsprechenden Prüfdaten kann zwar einwandfrei gesagt werden, ob Daten verändert wurden oder nicht; es kann aber nicht verhindert werden, dass die Daten unbeabsichtigt und/oder mutwillig geändert werden.
Die Konsequenz dieses Unterschieds sind sich viele Anwender/innen nicht wirklich bewusst. Was nützt das beste Prüfverfahren, um später feststellen zu müssen, die Daten wurden doch geändert, die digitalen «Originale» sind bei Änderungen unwiderruflich verloren.
Letztlich helfen dabei auch keine Festplatten-Dublikate weiter, denn konventionelle Datenträger «entleeren» sich über die Zeitdauer von einigen Jahren. Mit anderen Worten, Inhalte von Festplatten, die z.B. 10 Jahre im Save gesichert werden, bei diesen Datenträgern besteht die grosse Gefahr, dass diese Daten gar nicht mehr gelesen werden können.
Den einzigen wirksamen Schutz bieten optische Medien, bei denen die Daten nur einmal schreibbar gesichert werden. Diese Datenträger bieten (bei richtiger Lagerung freilich) jenen Schutz, den wichtige Informationen verdienen. Die ArchivistaBox bietet deshalb seit mittlerweile mehr als 20 Jahren genau das Konzept der Sicherung der Daten auf optische Datenträger mit.

«Geschichte» der Archivierung mit optischen Datenträgern

Im Jahre 1998 waren CDR-Datenträger, die um die 700 MByte Kapazität umfassen, gut geeignet, um damalige Datenbestände zu sichern. Bei Schwarz/Weiss können bei 700 MByte (bzw. 50 KByte pro Seite) immerhin 14’000 Seiten pro optische Disk archiviert werden.
Auch wenn es bereits damals DVD-Laufwerke bzw. die entsprechenden Medien gab, so hatten die damaligen DVD-Datenträger nicht den besten Ruf, für Langlebigkeit einzustehen, siehe dazu auch den guten Beitrag bei **https://de.wikipedia.org/wiki/DVD.**
Gegenüber DVD-Datenträgern boten CDR-Datenträger eine sehr gute Prognose. 30 bis 50 Jahre Lesbarkeit gaben die Hersteller auf die entsprechenden Datenträger. Die erste Version der Archivista-Software stammt aus dem Jahre 1998, die ersten erstellen Datenträger, die gefunden werden konnten, stammen vom April 1996, was immerhin ca. 25 Jahren entspricht. Kurz und gut, die Daten konnten problemlos gelesen werden (im übrigen selbst silberbeschichtete CDR).



Geschichte des Circus Pisa

Ursprünglich wurde der Circus Pisa in Italien gegründet (ein kleines Familienunternehmen). Das Unternehmen blühte in den ersten 12 Jahren auf. Darum konnten Mitarbeiter angestellt werden. Der Circus Pisa musste wegen schwierigen Verhältnissen nach Deutschland übersiedeln. Als er die 50 Tournee begann, war es wieder so weit. Von Deutschland aus, wanderte das kleine Unternehmen nach Oestreich.
75 Jahre Circus. Dieses Ereignis wurde in einem [illegible] Wesen des grossen Plakat-

Zum «Beweis» sei hier die erste mit der Archivista-Lösung je gesicherte Seite publiziert. Es handelt sich dabei um die erste Seite einer Geschichte, die der heutige Geschäftsführer 1978 bei einem Schreibwettbewerb zum Gastspiel des bekannten Kabarettisten Emil Steinberges beim Zirkus Knie 1977 einreichte. Als Preis gab es im ein Buch zum Gastspiel des Entertainers sowie eine Tonaufnahme in der Form einer Schallplatte des Auftritts. Das Manuskript wurde 1996 digitalisiert; die Schallplatte selber konnte damals nicht digitalisiert werden. Mittlerweile ginge selbst dies, für ca. **100 Franken bzw. Euro gibt es Abspielgeräte für die alten Vinylplatten.**

Wie oben erwähnt, selbst gebrannte DVD-Datenträger wurden intern nur wenige erstellt, doch auch diese ca. 15 Jahre alten DVDs konnten ohne Schwierigkeiten gelesen werden. Entscheidend für die Lesbarkeit ist ferner, dass die Daten in einem allgemein lesbaren Format gesichert wurden. Bei der Archivista-Lösung ist dies seit jeher gegeben, weil von allen Daten immer auch gerasterte Bilddaten erstellt werden.

Wären die Daten mit Festplatten, Bandlaufwerken und/oder WORM-Datenträgern erstellt worden (alle drei Formate galten damals als das Mass aller Dinge in der Branche), so wären diese Daten kaum derart einfach verfügbar, wenn überhaupt, denn die optischen Datenträger im 5.25 Zoll-Format haben einen unbestreitbaren Vorteil, die Grösse der Datenträger (eben 5.2 Zoll) bleibt über Jahrzehnte gleich und die Daten werden einzig einmal beschreibbar gesichert.

ArchivistaBox und AVMultimedia 2020/II mit M-Disk

Der grösste Nachteil der bisherigen Sicherung auf die optischen Datenträger beruht in der sehr begrenzten Speicherkapazität der CDR- (700 MByte) und DVD-Medien (4.2 GByte). Wenn Festplattengrössen mit 16 TByte vefügbar sind, so ergäbe dies selbst bei DVD-Datenträgern 3900 Scheiben. Bei gewogenen 40

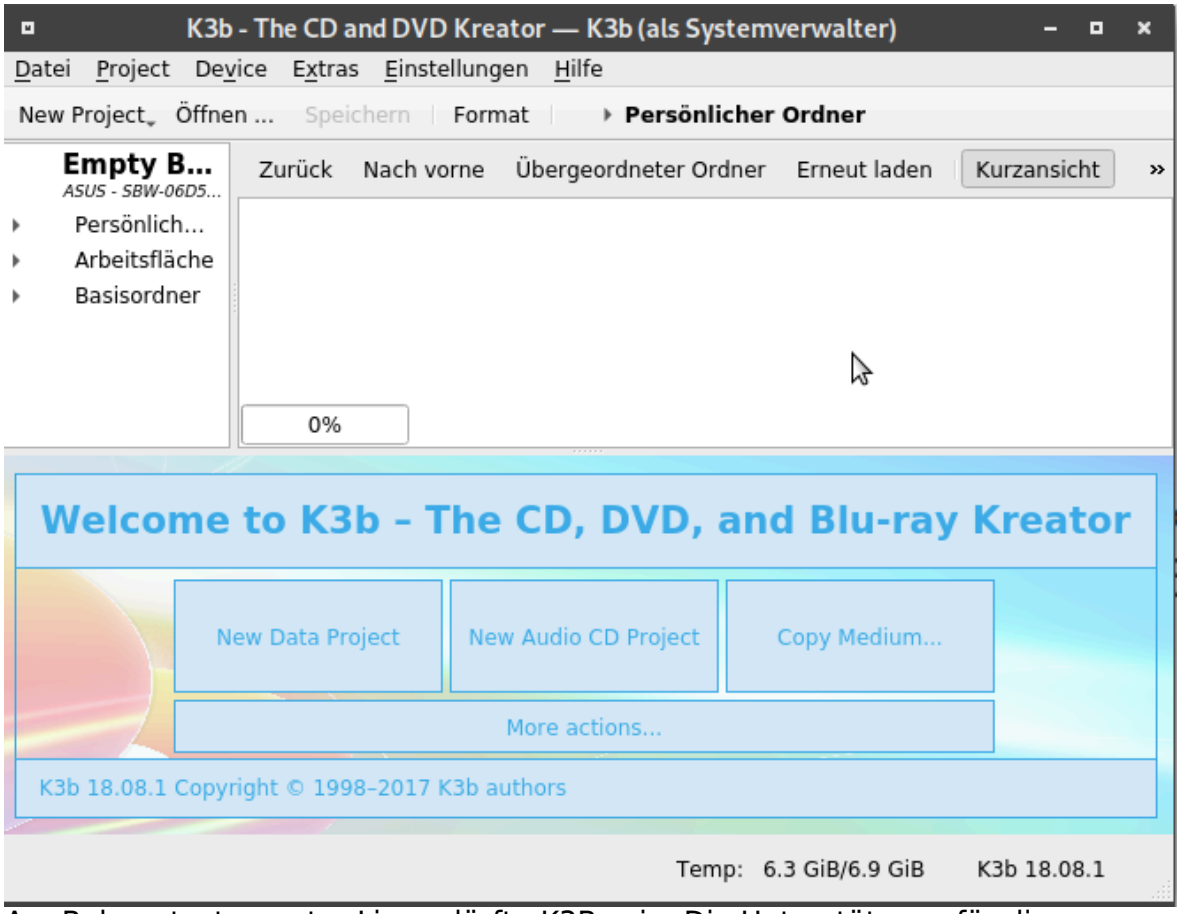
Gramm pro DVD ergäbe dies 156 Kilogramm.
Darauf liessen sich zwar über 335 Milliarden Dokumente archivieren, die wiederum satte 1.3 Mio Tonnen Gewicht verursachen würden. Nur eben, im Jahre 2020 geht es längst nicht nur darum Schriftgut zu archivieren. Vielmehr geht es darum, auch Ton- und Videomaterial langfristig verfügbar zu halten. Eine Stunde 4K Filmmaterial in guter Qualität benötigt um die 10 bis 20 GByte an Daten.
Bei 16 TByte Kapazität können (bei 10 GB 4K-Material pro Stunde) «nur» noch 1600 Stunden Filmmaterial archiviert werden. Und ja, es kann ja nicht sein, dass 1 Stunde Material in 4K bereits mehr als 2 DVDs füllt. Seit einigen Jahren gibt es deshalb BlueRay-Disks, die ca. 25 GByte speichern können. Mittlerweile lassen sich BlueRay auch bereits mehrlagig (BDR-XL) beschreiben, womit pro Scheibe bis zu 100 GByte Daten gesichert werden können.
Bei Verwendung von BDR-XL ergeben sich für 16 TB noch 160 Datenträger, bei 40 Gramm pro Disk ergibt dies noch 6.4 Kilogramm Gewicht. Das ist zwar noch immer ca. sechs- bis achtmal mehr als eine vergleichbare 3.5-Zoll-Festplatte, aber insgesamt doch sehr «verschmerzbar».
Wichtig ist weiter, dass es neben den BDR-XL-Scheiben jene im M-Disk-Format gibt. Erstere sind günstiger, letztere bieten eine deutlich längere Lebensdauer. Gemäss Herstellerangaben wird die Haltbarkeit auf bis zu 1000 Jahre angegeben (M-Disk=Millenium-Disk), dies im Unterschied zu den normalen BDR-XL-Datenträgern, bei denen die Lebensdauer ca. 20 bis 25 Jahre betragen soll. Selbstverständlich gelten diese Angaben nur für den Fall, dass die Medien richtig gelagert werden, d.h. bei tiefer Luftfeuchtigkeit, konstanter niedriger Temperatur und ohne Sonneneinstrahlung, doch gelten diese Anforderungen für alle Datenträger (inkl. Festplatten).

K3B, Brasero und Xfburn im Test

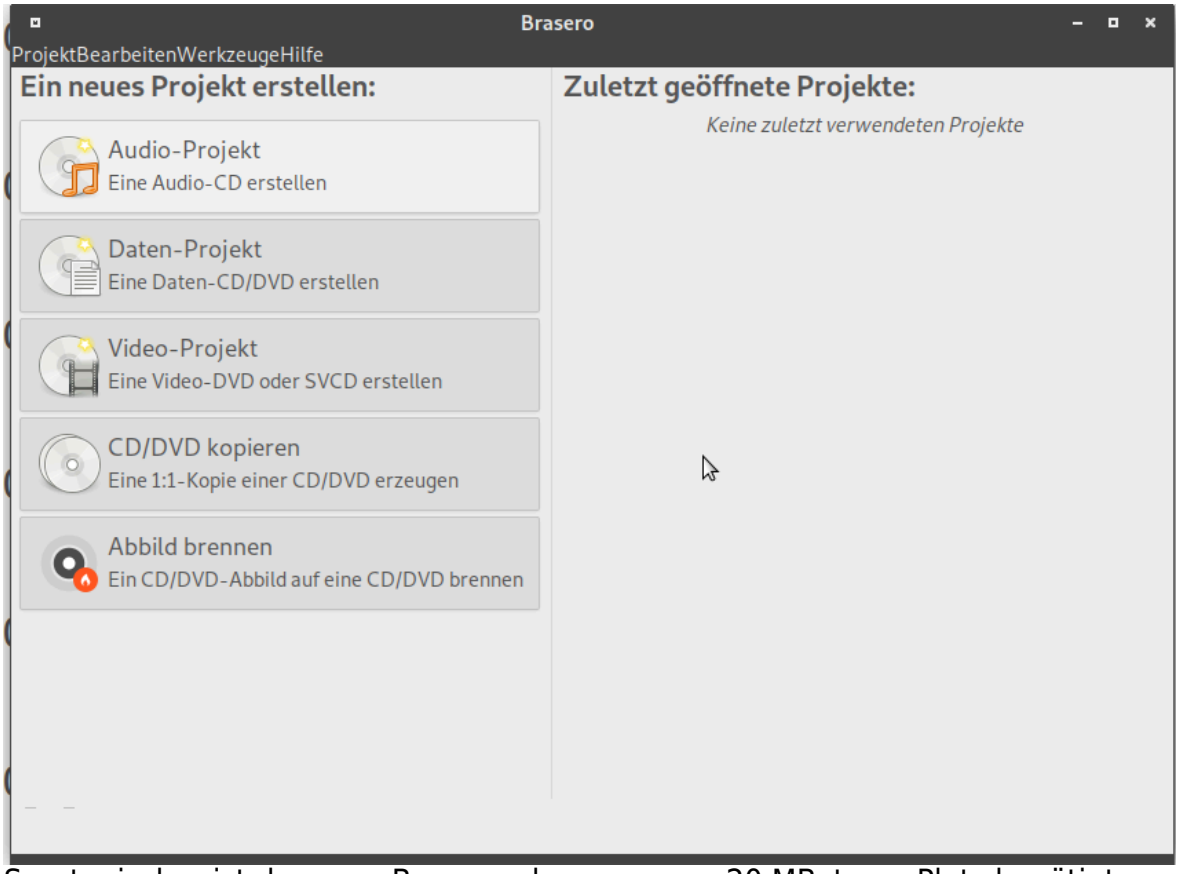
Eine gute Unterstützung von BlueRay (bzw. BDR-XL) war unter Linux lange Zeit nicht oder nur rudimentär vorhanden. Mittlerweile (bzw. mit aktuellen Linux-Kerneln) stehen die entsprechenden Treiber zur Verfügung. Das Schreiben solcher Medien mit Konsolenprogrammen (z.B. xorriso) gelingt ohne Schwierigkeiten. Dazu ein Beispiel für die Konsole:

```
xorriso -as mkisofs -iso-level 3 -o /home/data/data.iso
/home/data/archivista/*
growisofs -Z /dev/sr0=/home/data/data.iso
```

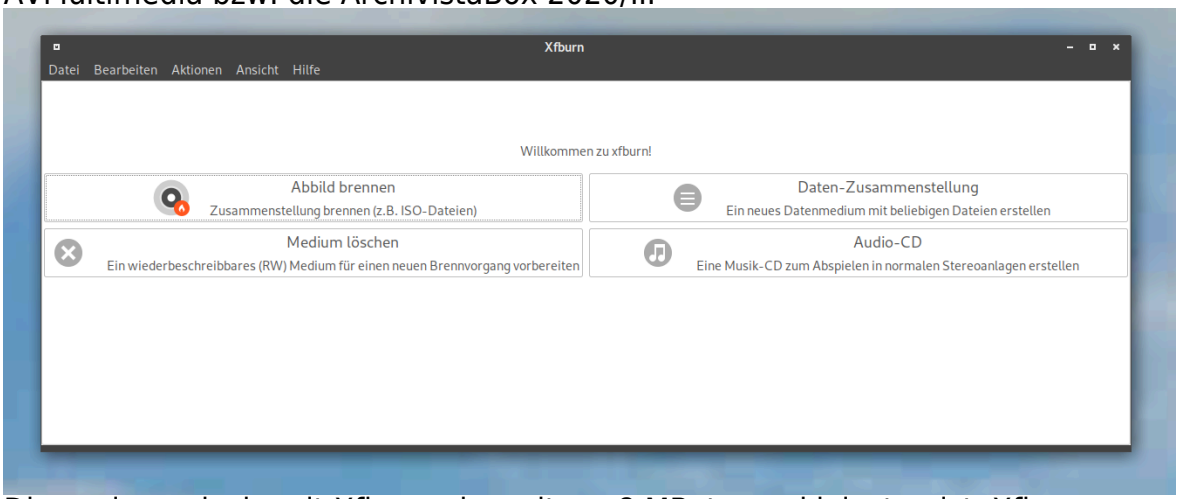
Nun ist das Arbeiten mit der Konsole nicht das, was sich «Normalsterbliche» üblicherweise wünschen. Daher galt es für die ArchivistaBox 2020/II sowie AVMultimedia, ein geeignetes grafisches Tool zu evaluieren und letztlich in die Distribution auch einzubinden.



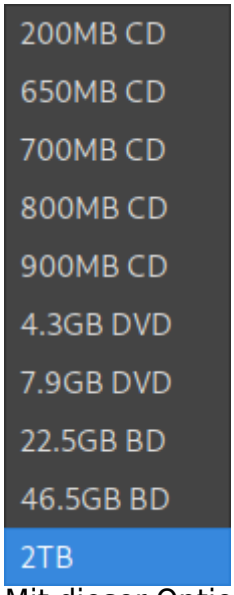
Am Bekanntesten unter Linux dürfte K3B sein. Die Unterstützung für die verschiedenen Formate ist sehr gut, allerdings benötigt K3B um die 150 MByte an Software. Einmal abgesehen von der Grösse bietet K3B zwar eine unglaubliche Fülle von Parametern, doch können diese beim Einstieg mitunter auch sehr verwirrend sein.



Spartanischer ist dagegen Brasero, das «nur» ca. 20 MByte an Platz benötigt. Allerdings war Brasero (zumindest nicht mit der mit AVMultimedia getesteten Version 3.12.2) nicht in der Lage, BDR-XL-Medien mit 100 GByte zu erstellen. Aus diesem Grunde fand letztlich auch Brasero nicht den Weg nach AVMultimedia bzw. die ArchivistaBox 2020/II.



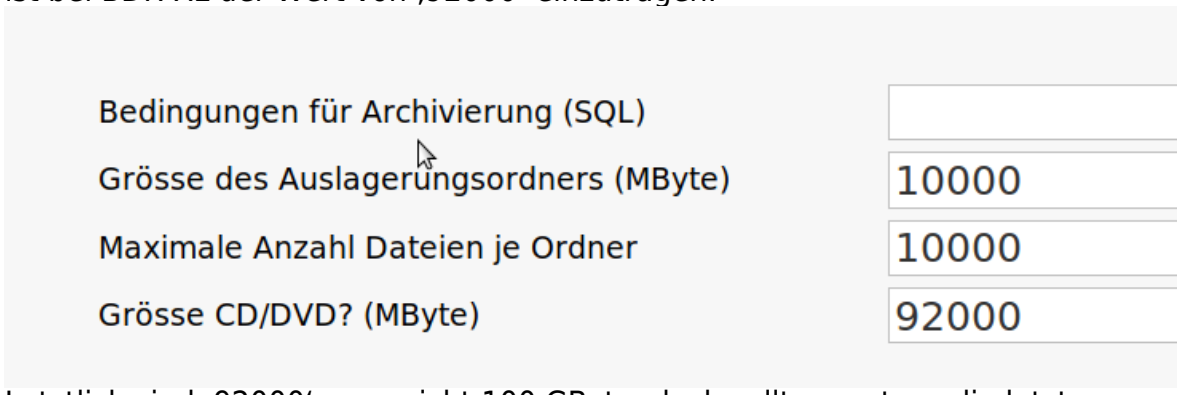
Dies gelang einzig mit Xfburn, das mit ca. 8 MByte am kleinsten ist. Xfburn bietet wenige, dafür aber zielführende Optionen und ist in der getesteten Version 0.5.5 in der Lage, BDR-XL mit 100 GByte Umfang zu erstellen. Etwas gewühnungsbedürft ist einzig, dass beim Erstellen der Image-Datei die Option 2 TByte auszuwählen ist.



Mit dieser Option konnten problemlos BDR-XL-Medien ertellt werden. Dabei empfiehlt es sich sehr, immer zuerst eine ISO-Datei (Abbild) zu erstellen, und erst diese Datei danach auf die optische Scheibe zu brennen.

Optionen für ArchivistaBox 2020/II in WebAdmin

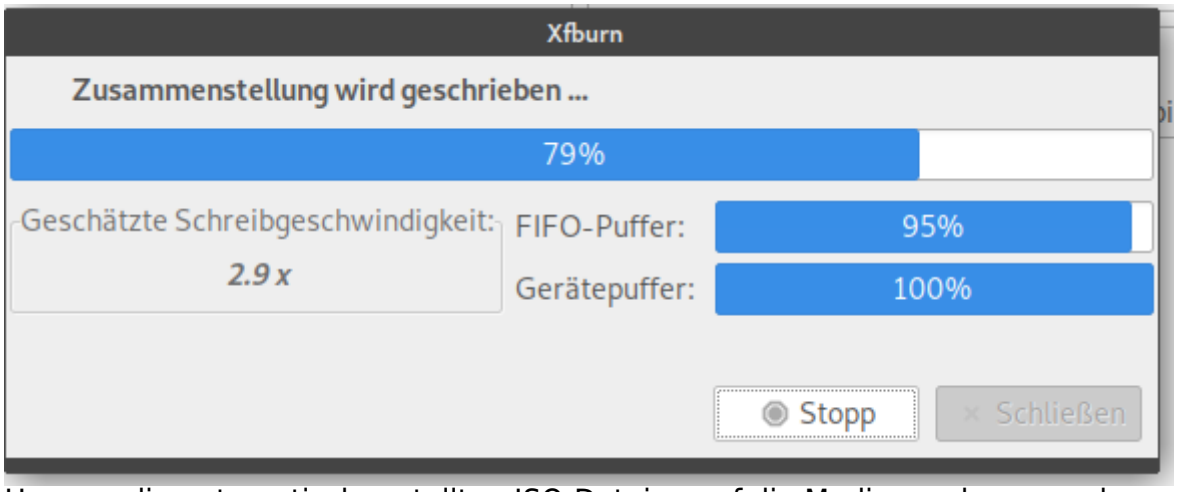
Um nun Daten der ArchivistaBox mit bis zu 100 GByte zu archivieren, sind in WebAdmin drei Parameter anzupassen:. Bei ‚Grösse des Auslagerungsordners' dürften 10000 MByte sinnvoll sein, ebenfalls ist zu empfehlen, die maximale Anzahl der Dateien auf 10000 zu begrenzen und bei ‚Grösse CD/DVD? (MByte)' ist bei BDR-XL der Wert von ‚92000' einzutragen.



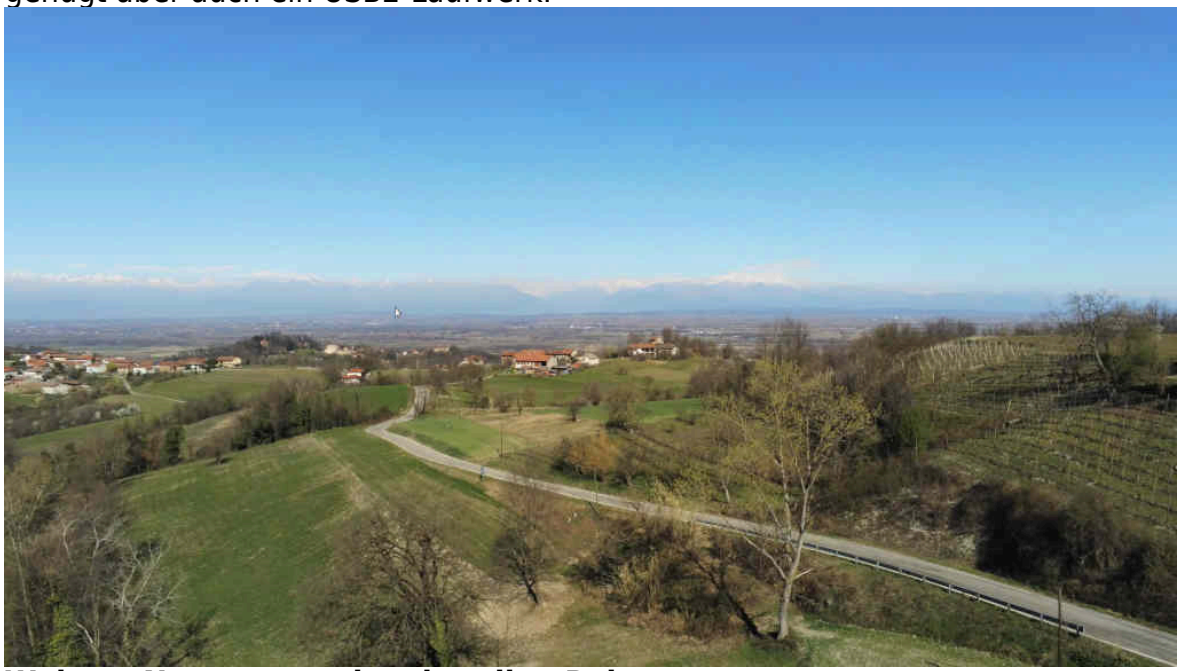
Letztlich sind ‚92000' zwar nicht 100 GByte, doch sollten erstens die letzten Blöcke eines optischen Mediums nie komplett beschrieben werden (statistisch gesehen ist dort die Fehlerrate am grössten) und zweitens bezeichnen sich die 100 GByte auf die Rechnung mit 100 * 1000 MByte * 1000 KByte * 1000 Bytes. Wird jeweils mit 1024 geteilt (mathematisch korrekte Rechnungsweise), so fasst die 100 GByte BDR-XL «nur» gute 93 GByte.

Dies soll aber die neuen Möglichkeiten nicht grundsätzlich schmälern. Dank dem BDR-XL-Format ist es möglich, weit über 150 CDR-Disk auf einen Datenträger zu schreiben.

Sofern die obenstehenden Optionen in WebAdmin korrekt eingetragen sind, werden bei der Archivierung die entsprechenden ISO-Dateien automatisch erstellt. Sie finden diese im Ordner ‚/var/lib/vz/template/iso', wobei jeweils immer der Name der Datenbank und de des ersten Ordners verwendet wird (Beispiel ‚archivista_ARCH0002.iso').



Um nun die automatisch erstellten ISO-Dateien auf die Medien zu brennen, kann unter ‚Multimedia' das Programm ‚Xfburn' gestartet werden. Dazu sollte vorgängig bereits ein Medium eingelegt werden, denn nur so wird der Brenner auch korrekt erkannt. Bei der ersten Maske ist ‚Abbild brennen' sowie die gewünschte Datei aus dem Ordner ‚/var/lib/vz/tempalte/iso' zu wählen. Danach kann der Prozess mit ‚Abbild brennen' ausgelöst werden. Der Brennvorgang dauert mit 90 GByte etwas mehr als eine Stunde.
Grundsätzlich sollten immer jeweils zwei Datenträger erstellt werden. Ob dabei zwei unterschiedliche Fabrikate und/oder Laufwerke verwendet werden, darüber liesse sich streiten. Marktführer bei den Datenträgern ist Verbatim, entsprechende Medien sind aber auch bei Sony erhältlich, bei den Laufwerken gibt es aktuell eine stattliche Anzahl von Herstellern (getestet wurden die Laufwerke von Asus mit USB3.1. Rein von der zu schreibenden Datenmenge genügt aber auch ein USB2-Laufwerk.



Weitere Neuerungen im aktuellen Release
Mit Release 2020/II wird neu der Linux-Kernel 5.4 ausgeliefert. Dieser Kernel-Zweig wird für die nächsten Jahre zur Verfügung stehen. Unterstützt werden dabei die neuesten AMD-Prozessoren. Getestet wurde die neue Version mit dem Ryzen 3950X mit 16 Kernen bzw. 32 Threads. Die Leistung dieses Prozessors ist beachtlich. 4K-Filme lassen sich damit fast in Echtzeit erstellen (ca. 70 Minuten für 60 Minuten Filmmaterial).

Ebenfalls realisiert wurde eine bessere Unterstützung für UEFI- und Legacy-Bios-Derivate. In der aktuellen Version wird getestet, ob entweder UEFI oder mit Legacy gebootet wird. Entsprechend wird dabei auch die Software auf der Festplatte eingerichtet. Wer bisher mit Legacy arbeitete, kann dies weiter machen, wer mit UEFI bootet, bei dem wird für die Installation immer auch UEFI eingerichtet. Neu hingegen ist die komplette Unterstützung für Festplatten über 2 TB, dies sowohl für UEFI als auch für UEFI.
Wer detaillierte Informationen erhalten müchte, findet die entsprechenden Hinweise (in Englisch) im entsprechenden Forum von **Devuan.**

Langfristige Archivierung über Jahrzehnte

Zum Abschluss dieses Blogs sei noch auf einen **Beitrag zum gleichen Thema aus dem Jahre 2008** (dies sind immerhin 12 Jahre) verwiesen. Darin wird all das ausgeführt, was damals für die Archivista-Lösung bereits Bestand hatte und heute noch immer Gültigkeit besitzt. All jene, die lieber in die Zukunft schauen, genügt das Wissen, das mit der Version 2020/II nicht nur multimediale Inhalte archiviert werden können, sondern eben auch über Jarhzehnte sauber auf optische Datenträger ausgelagert werden können.