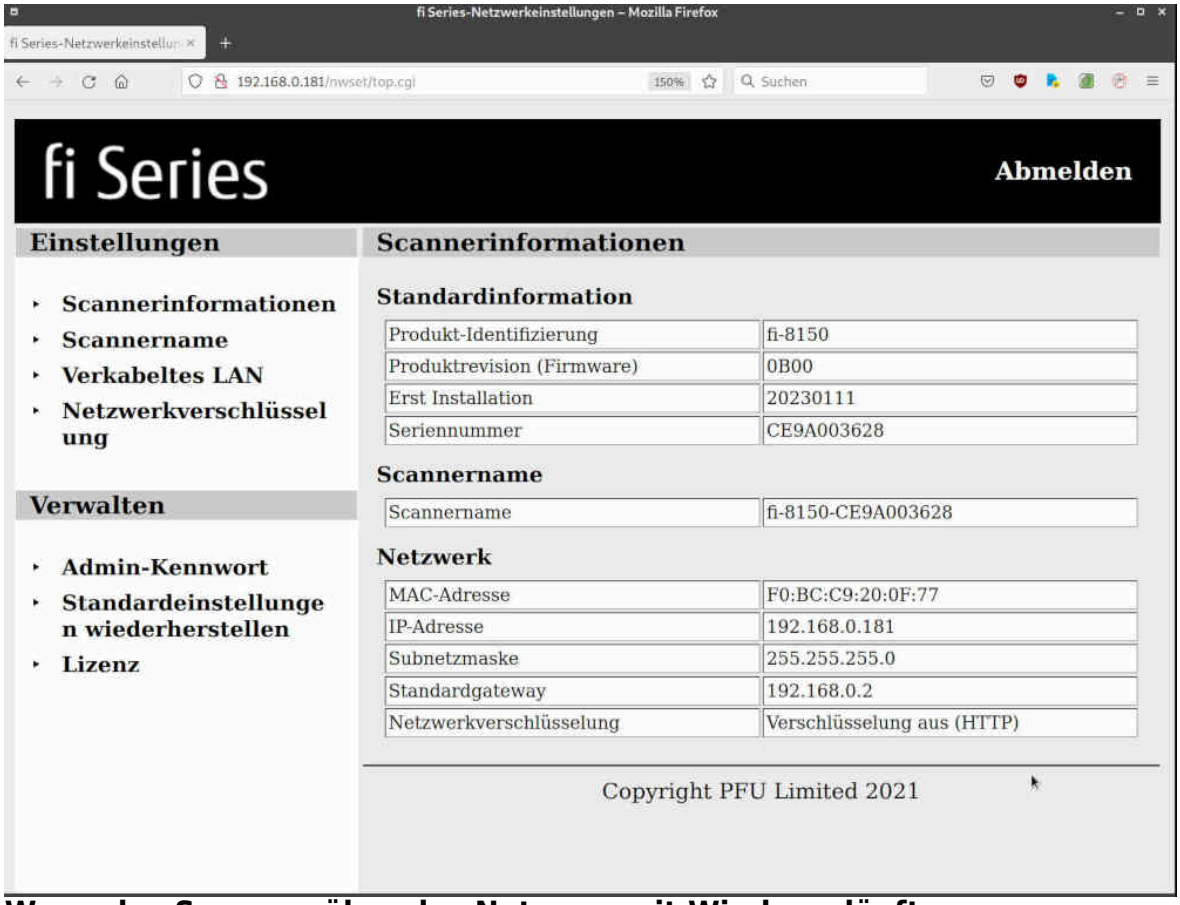
Mit 40 Zeilen zur ArchivistaBox als Netzwerkscanner

Egg, 11. Janaur 2023: *Voller Freude wurde die vergangenen Tage ein netzwerkfähiger Scanner von Fujitsu bestellt. Positiv vermerkt werden kann, dass das Gerät problemlos mit der ArchivistaBox arbeitet. Weniger positiv vermerkt werden muss, dass — obowhl das Gerät über einen Netzwerkanschluss verfügt — das Scannen auf einen Freigabeordner nur mit Windows gelingt. Andere Betriebssystem, z.B. Mac oder Linux bleiben aussen vor. Grund genug, einen solchen Job mit der ArchivistaBox zu realisieren.*

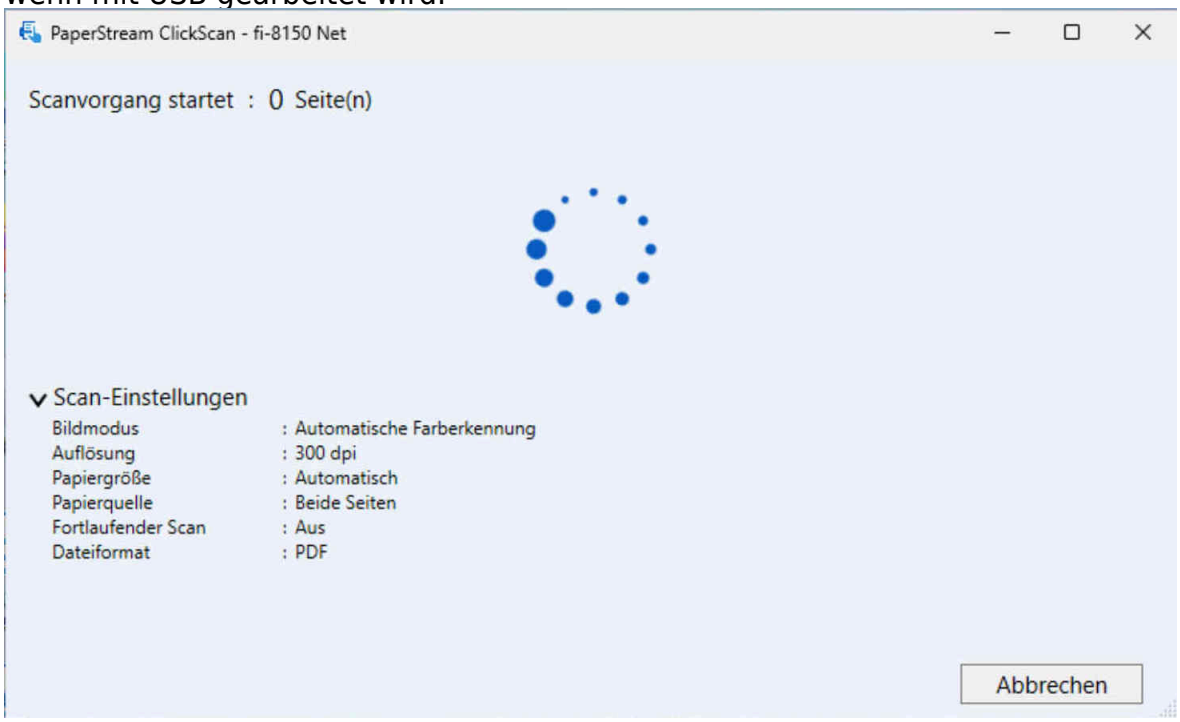
Matchentscheidend beim Scannen ist der Einzug

Vor einiger Zeit wurde an dieser Stelle ein **Brother-Scanner vorgestellt.** Dieses Gerät kann bequem über ein Web-Interface konfiguriert werden und sendet die Scans mühelos auf ein Freigabelaufwert. Leider überzeugen die Brother-Geräte — vom Einzug her betrachtet — im harten Büroalltag nicht. Rein vom Einzug her arbeiten die Fujitsu-Geräte schlicht und einfach in einer anderen Liga. Voller Freude wurde deshalb ein Fujtsu fi-8150 bestellt.

Das Gerät verfügt über USB und Netzwerkanschluss. Über USB kann das Gerät an eine jede ArchivistaBox angehängt werden. Leider entpuppte sich der Netzwerkanschluss — um es noch wohlwollend auszudrücken — als Niete. Zwar können die IP-Kenndaten über das Web-Interface festgelegt werden (Password ist password). Leider ist es nicht möglich, in diesem Interface Scans festzulegen, um z.B. auf einen Netzwerkpfad zu scannen.



Wenn das Scannen über das Netz nur mit Windows läuft
Wer mit dem Fujitsu fi-8150 über das Netzwerk scannen möchte, benötigt zwingend Windows und ist überdies eine ganze Weile damit beschäftigt, bis die Software aufgespielt ist bzw. gescannt werden kann. Für ein netzwerkfähiges Gerät ist dies äusserst enttäuschend. Dies umso mehr, als das Gerät durchaus für Mac und Linux beworben wird. Dies trifft ja auch zu, jedoch eben nur dann, wenn mit USB gearbeitet wird.



Script der ArchivistaBox «hilft» dem Scanner auf die Sprünge
Nun ist die ArchivistaBox in erster Linie ein DMS (Dokumenten Management System). Jedoch, eine jede ArchivistaBox lässt sich per Skript erweitern und so wurde dem fi-8150 das beigebracht, was eingentlich von «Haus» aus hätte funktionieren können, das Scannen in einen Freigabeordner. Zunächst das Skript:

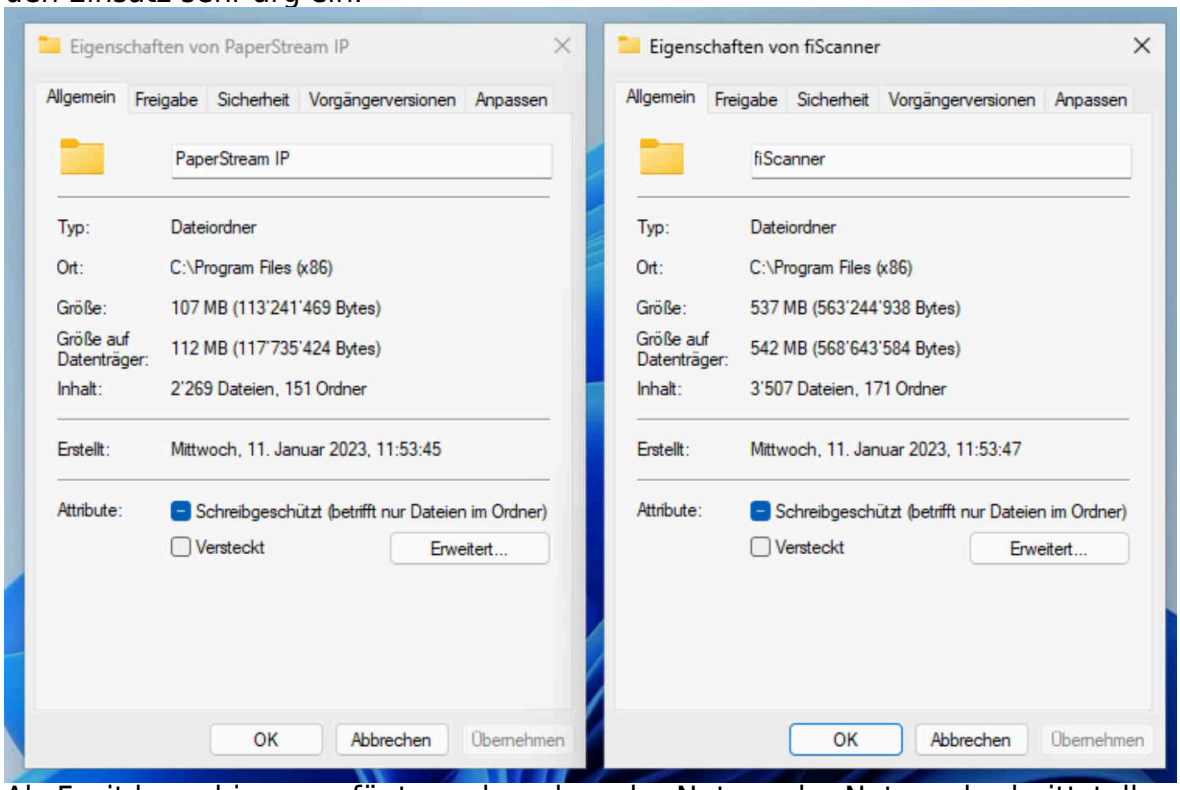
```
#!/usr/bin/perl
use lib qw(/home/cvs/archivista/jobs);
use AVJobs;
my ($host,$db,$user,$pw,$lnr) = @ARGV;
my $net = "/mnt/net";
my $smbhost = "//192.168.0.161/pdfs";
my $smbuser = "avbox";
my $smbpwd = "archivista";
logit("open $db at $host with $user for $lnr");
my $dbh=MySQLOpen($host,$db,$user,$pw);
if ($dbh) {
  if (HostIsSlave($dbh)==0) {
    logit("login in $db is sucessfully");
    my $lnr2 = ($lnr*1000)+1;
    my $sql = "select Quelle from archivbilder where
Seite=$lnr2";
    my @row = $dbh->selectrow_array($sql);
    if (length($row[0])>0) {
      logit("we got a pdf file for $lnr");
      mkdir $net if !-d $net;
      my $err=0;
      my $mounted = `df | grep $net`;
      if ($mounted eq "") {
        my $cmd = "mount -t cifs $smbhost $net ".
          "-o username=$smbuser,password=$smbpwd";
        $err=system($cmd);
        logit("$err=>$cmd");
      } else {
        logit("$smbhost at $net already mounted");
      }
      if ($err==0) {
        my $fname = "$net/$db-$lnr.pdf";
        logit("try to write $fname");
        writeFile($fname,\$row[0]) if !-e "$fname";
      }
    }
  }
  $dbh->disconnect();
}
```

Anzufügen gilt es hier, dass die Angaben bei **smbhost, smbuser und smbpwd** dem Freigabepfad (z.B. des Windows-Rechners oder des Macs), dem dortigen Benutzer und Passwort entsprechen müssen.
Weiter muss das Skript so abgelegt werden, dass es direkt nach der OCR-Erkennung aufgerufen wird. Dabei ist der folgende Pfad auf der ArchivistaBox zwingend zu verwenden:
/home/data/archivista/cust/autofields
Innerhalb dieses Pfads muss das Skript erstens den Namen der Datenbank entsprechen, die zum Scannen verwendet werden soll. Zweitens muss dem Datenbanknamen ‚ocr.pl' hinzugefügt werden. Bei der Datenbank ‚archivista' muss das Skript folglich ‚archivistaocr.pl' lauten.
Wird nun entweder über WebDMS oder das **Keypad gescannt,** so werden die

gescannten Seiten nach der Texterkennung direkt in das gewünschte Verzeichnis kopiert. Dadurch, dass die Texterkennung selber über mehrere CPU-Kerne oder gar Rechner laufen kann, ist hier eine enorme Geschwindigkeit möglich. Auf der **ArchivistaBox Everst** z.B. können pro Sekunde etwa 200 Seiten verarbeitet werden.

Quizfrage: 40 Zeilen oder 600 MByte Software ?

Das obige Skript umfasst ca. 40 Zeilen. Selbstverständlich kann der Prozess an beliebige Bedürfnisse angepasst werden (z.B. Scannen direkt zum PC der Chefetage). Demgegenüber benötigt die von Fujitsu ausgelieferte Software über 600 MByte unter Windows. Die 600 MByte wären wohl verkraftbar, dass aber absolut keine Software für Mac oder Linux zur Verfügung steht, dies schränkt den Einsatz sehr arg ein.



Als Fazit kann hier angefügt werden, dass der Nutzen der Netzwerkschnittstelle der 8000er-Serie durch das auf Windows beschränkte Scannen arg limitiert ist. Zwar kostet das Kombipaket aus einer ArchivistaBox Dolder mit einem Fujitsu fi-7140 etwas mehr als der Fujitsu fi-8150-Scanner. Dafür lässt sich das Gespann auf fi-7140 und ArchivistaBox deutlich flexibler in beliebige Rechnerwelten einbinden. Und zwar selbst dann, wenn es nur darum geht, möglichst effizient PDF-Dateien zu erstellen.

Nachbemerkung: *Die Version 2023/I ist nur notwendig, sofern Geräte der 8000er-Serie von Fujitsu zum Einsatz kommen sollen. Für das Erstellen der PDF-Dateien mit den 7000er-Geräten kann das Skript auch auf älteren ArchivistaBoxen zum Einsatz kommen.*