ArchivistaBox Dolder: 10,000 pages…
Egg, 5 April 2014: *A few weeks after having first* ***introduced ArchivistaBox Dolder on this site,*** *today's blog is about how many pages ArchivistaBox Dolder can scan in an hour. On its own, the smallest ArchivistaBox Dolder is capable of scanning 2000 pages per day. This is a deliberately conservative estimate. The title of this article suggests that a volume of 2000 pages is attainable, but this blog is about more than ArchivistaBox Dolder; what we want to show is the potential that can be achieved through optimisation.*



Dolder with Scanner iX500: Full house after 300 pages?
The first scans were somewhat sobering. After approx. 300 pages, the ArchivistaBox Dolder signalled "full house". Even the **Fujitsu iX500 scanner** (30 pages/60 images per minute) had outperformed the ArchivistaBox Dolder. A thorough analysis established that the **automatic removal of blank pages takes up around 2 seconds of computing time per page.** Since other ArchivistaBoxes have greater computing power, this is not noticeable on the faster systems. On ArchivistaBox Dolder, however, it is grimly apparent.
So the problem to solve was how to optimise blank page recognition. Extensive testing was carried out with a wide range of libraries. The **solution implemented was faster by a factor of 4, because the pages were reduced in size prior to the test** (though blank pages still remain blank), but each page still took around 0.5 seconds. That meant that 3,500 pages could be scanned in one hour – the equivalent of 7 ring binders.
A faster scanner had to be found…
Now 3,500 pages per hour is really not too bad, but it was established during the test (after optimisation), that both processors (CPUs) were not working to full capacity during scanning. So, for this reason alone, a faster scanner ought to be possible. The **new Fujitsu fi-7160** was chosen. This is **capable of delivering 120 images (60 pages) per minute for well under €1000. In addition, the fi-7160 renders a level of service hitherto unattainable in**

this price range.

During the first scans with the fi-7160, ArchivistaBox could not keep pace; it’s not meant to do more than 4,500 pages per hour. However, further substantive analysis revealed that our image processing library was very slow at setting DPI values (image resolution), requiring much more than 0.5 seconds to record two times two bytes. This is because the entire image is calculated twice over. That seemed pointless, and so it was deactivated. Now 7,200 pages could be handled in one hour.

The fastest scanner and further optimisation

When attempting to get a better performance out of the **fi-6670 (80 pages and 160 images),** it was discovered that pages loaded in landscape format had to be turned by the A3 scanner before they could be filed in the database. Had they been loaded in portrait format, handling speed would have been reduced by around 20%, making the fi-6670 only a little faster than the fi-7160. The libjpeg (Linux standard) library is, in fact, pretty nippy. On the **quicker ArchivistaBoxes, rotating a JPEG page takes (more or less) 0.3 seconds. With ArchivistaBox Dolder, however, rotation takes more than 1 second.** This doesn’t mean that ArchivistaBox Dolder is generally slower, but simply that CPU-intensive things take more time, due to the unit’s lower power consumption (6 Watts per CPU under load). If that were not the case, the CPU would have to be booted “power-hungry” and, at the end of the day, a whirring fan also eats up the watts…

And now, I’m happy to present the jpegtran results. This program allows images in JPEG format to be turned loss-free:

```
time jpegtran -rotate 90 job0085.img > job0085.jpg
real 0m1.035s
user 0m0.900s
sys 0m0.132s
```

After a lot of searching, we finally hit upon the **libjpeg-turbo** library. Here is the result of turning a page of A4 with 300dpi through 90 degrees:

```
time /opt/libjpeg-turbo/bin/jpegtran -rotate 90 job0085.img
> job0085.jpg
real 0m0.time507s
```

user 0m0.372s
sys 0m0.128s
1.035/0.507 gives 2.04 but, in order to get **more than a factor of 2, the "Turbo" is quicker** and, because of that, has rightly gained acceptance in all ArchivistaBoxes. However, 160 x 0.5 seconds on ArchivistaBox Dolder comes to 80 seconds, which the scanning process (even under libjpeg-turbo) cannot match even at full speed. In other words, the combination of ArchivistaBox Dolder and fi-6670 makes no sense whatsoever; the fi-7160 delivers almost the same performance for a lot less money.

ArchivistaBox Dolder scans 10,000 pages per hour

The new **Fujitsu fi-7180 is particularly well suited,** since pages are loaded in portrait format, the unit scans 80 pages and 160 images per minute and it represents much better value than the A3 scanner fi-6670. All the ArchivistaBox systems come with the full range of scanner drivers. Simply attach the scanner to the box and trigger the scanning procedure using the keypad. ArchivistaBox Dolder **scans and processes at over 160 pages per minute. This gives (more or less) 10,000 pages per hour.**

To mark the occasion, we filmed a short video, premiered here on this blog, on scanning with the ArchivistaBox Dolder and the fi-7180:

The video lasts 1 minute 30 seconds – 30 seconds for the introduction of the components and then 60 seconds of scanning. Less than optimal light conditions and a very ordinary mobile telephone camera mean the quality is not good, but the video clearly demonstrates the speed at which ArchivistaBox Dolder can process double-sided scanned colour images (300dpi) with no time lag.

Optimisation makes sense on all ArchivistaBoxes

The high scanning performance of the ArchivistaBox Dolder mean that in 10 hours (i.e. in one day – whether the fi-7180 can get through this in one go remains open to question) a total volume of 100,000 pages would be executed. But does this actually make any sense? To be honest, **anyone needing to process 100,000 pages per day should not necessarily be looking at the ArchivistaBox Dolder.** The same question can, however, be put the other way around. If between several tens and several hundreds of pages need to be handled each and every day, then it would make a lot of sense.

Optimisation, of course, means additional costs. When implemented correctly, however, optimisation can bring **enormous savings,** and this, of course, is not limited to ArchivistaBox Dolder, but is the case with **all our ArchivistaBox systems.** Thanks to optimisation, ArchivistaBox Dolder now works as quickly as the other ArchivistaBox systems prior to optimisation. So, with this in mind, happy working with our ArchivistaBox systems!

P.S.: this blog may give the impression that we are in raptures about Fujitsu scanners. That impression would not be false, because our experience with these scanners has been very good. Alongside the Fujitsu scanners (in the 25 page range), there are many other reasonably-priced duplex scanners that work well with ArchivistaBox (keyword SANE). But we are only able to unconditionally recommend those devices which we have thoroughly tested. Against that, ***without reservation or testing, we can recommend equipment that delivers the data per network. Such scanners (mostly multi-functional devices) are available in the €100 plus price bracket.*** *Make sure that the data can be delivered in PDF format by SMB (Windows folder) or FTP.*

P.P.S.: ArchivistaBox systems are able to ***create full text search indices and searchable PDF files from any documents automatically.*** *Per day, ArchivistaBox* ***Dolder is capable of processing 10,000 colour pages or 20,000 black/white pages.*** *The text recognition function of ArchivistaBox Matterhorn is able to process ten times more documents and, with additional scan stations, this* ***can be extended to a daily volume of more than 1 million documents.***