Low-Cost, On-Demand Film Digitisation and Online Delivery

Matt Garner (matt.garner@findmypast.com)

findmypast.com
family history in the making

Abstract

- Hundreds of millions of pages of microfilmed material are not being digitised at this time due to insufficient collective demand to garner commercial investment.

- Using a novel application of recent technological advancements in imaging, data processing and storage, the cost of digitising and online delivery of microfilm can be reduced dramatically, even beyond the threshold that individual family historians are to directly and fully sponsor the digitisation and delivery of individual reels.

- An on-demand business model can be established which is driven and funded completely by end users. This essentially allows a library with a film collection to make the breadth of their collection available online with no direct costs.

findmypast.com
family history in the making

The Worldwide Family History Industry

- Boom of online record availability is fuelling a global, rapidly expanding public interest in family history.

- Primary reasons for this boom is the decrease in access barriers (increase in convenience) to historic and contextual documentation, facilitated primarily by the advent and penetration of broadband Internet.

findmypast.com
family history in the making

The Industrial Digitisation Paradigm

- Many libraries and archives are beginning to digitise their collections. Although the volume of online collections is growing rapidly, only a small percentage of what has been filmed is currently available online.

- The logical order of digitisation employed by each organisation is principally concentrated around those document collections that have the largest general interest (such as census and vital records) and present the most rapid return on investment.

- Due to the high cost of film digitisation, there is effectively a fixed break-even point where many previously filmed collections won’t qualify for commercial digitisation until present market conditions change.

- The on-demand model presented in this presentation is a complete reversal of this traditional paradigm.

findmypast.com
family history in the making

Family History and the Mainstream Market

- In the last few years, family history has gained a foothold in the mainstream consumer market, especially in the United Kingdom.
- Success in the mass market is facilitated by the comprehensiveness and ease-of-use of online record collections.
- Novice family historians quickly give up when required documentation is not available online, especially in places where offline family history libraries or centres are not easily accessible.
- Making the entire breadth of large libraries' filmed collections available online on an on-demand model will have a dramatic positive effect not only for seasoned historians but the mainstream market as well.

findmypast.com
family history in the making

The Cost Points of Digitisation

- Digital Imaging
- Post-Processing
- Image Compression
- Offline Storage
- Labour
- Administrative Overhead

findmypast.com
family history in the making

Technological Advancements Lower Cost

- Numerous technological advancements have lowered the cost of digitisation and online delivery across the board.
  - Advanced Optics
  - High-Performance CPUs
  - Bandwidth
  - Low-Cost Spinning Disks
  - Commodity Server Hardware

findmypast.com
family history in the making

Commercial Film Digitisation

- Very small, specialised market for industrial digitisation equipment.
- Mostly enterprise, academic and governmental players.
- Top-of-the-line NextScan Eclipse range can run from $60k to more than $120k per machine.
- Not necessarily complex optics nor control software.
- 1000% profit margin simply because big players can afford it.

findmypast.com
family history in the making

Low-Cost, High-Quality Optics

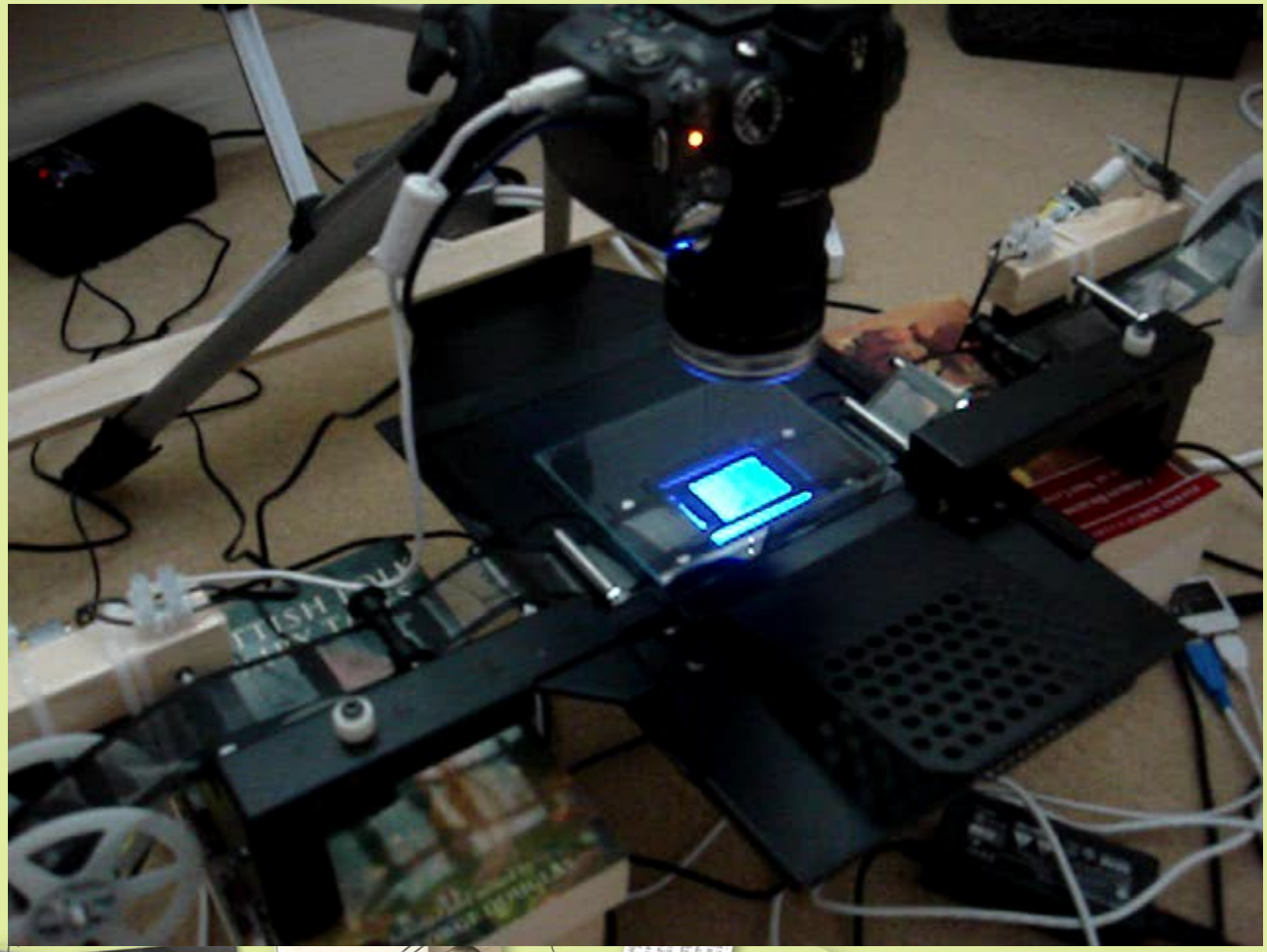
- Moore’s Law applies to Digital Imaging.
  - Digital cameras with double-digit megapixel sensors are now readily available.
- High-end Commericial Digital Cameras include:
  - Self-focusing and powerful zoom optics
  - Ability to mate with additional lens arrays
  - Include professional Software Development Kits (SDKs) for real-time PC-driven imaging.

Proof-of-Concept Automated Film Scanner

- High-end Canon 12 megapixel camera and multi-element lenses
- USB-driven motors
- Machine vision algorithms control film advancement based on live viewfinder
- Programmatically controlled focusing, aperture and exposure settings
- Effective Resolution:
  - Film Surface: >9600 DPI
  - Letter-size Document: 320 DPI
  - Output: 4000x3000 pixels

findmypast.com
family history in the making

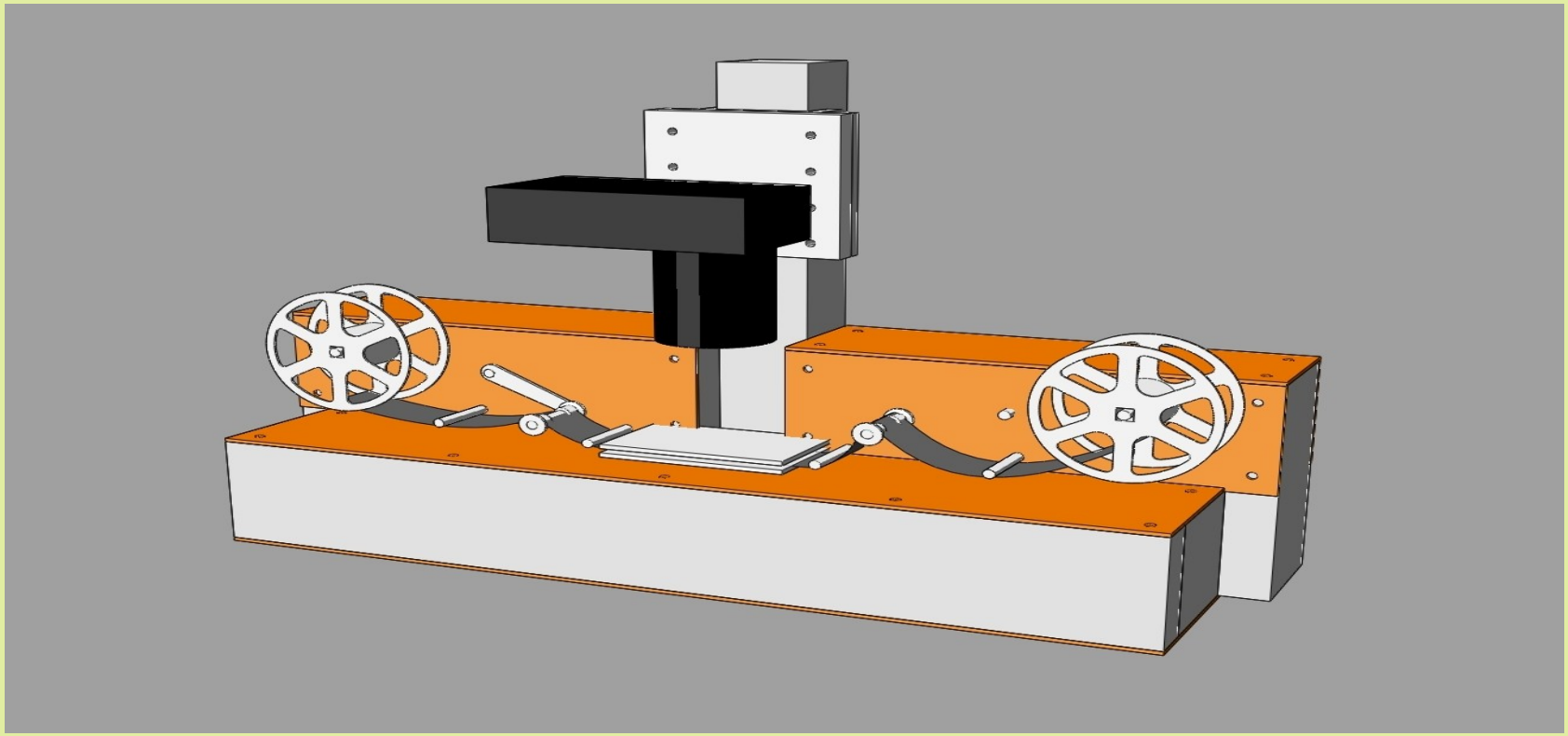
Proof-of-Concept in Motion (Slow Mode)



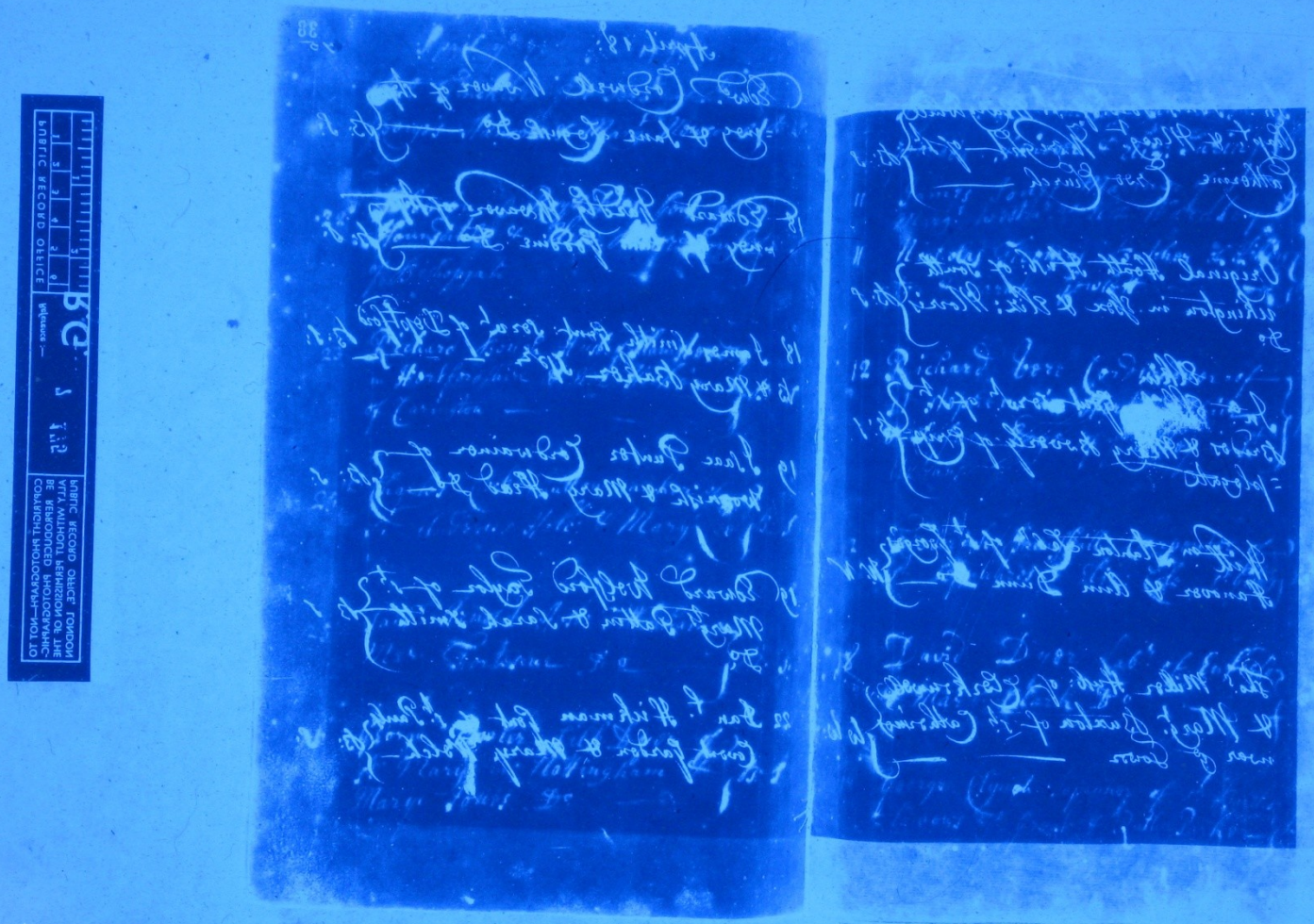
findmypast.com
family history in the making

Industrial Rendering

- Currently being manufactured



Raw Output (London Fleet Marriages)



Sample Post-Processed Output (London Fleet Marriages)

Scanning Automation

- Complete end-to-end, start-to-stop automation required
- Complex machine vision algorithms track documents and control the reel positioning and high-resolution captures
- A low resolution image ribbon representing the reel as a whole is recorded and programmatically checked to verify capture of every image
- This is not ribbon scanning in earnest where the film is primarily imaged in a single, high-resolution continuous stream. This results in substantially faster frame capture rates; however, per machine costs are dramatically increased as appropriate camera sensors cost on the order of $16000 each

findmypast.com
family history in the making

A Massively Parallel Scanner

- Film scanner currently averages 12 pages per minute (on 35mm film) compared to 200ppm on industrial ribbon scanners
- The complete scanner system, including the accompanying PC workstation, can be manufactured for well under $2000
- Several dozen, if not hundreds, of identical scanner units may be manufactured for the same price as purchasing a single, high-speed unit
- Although a single reel may take an hour or two to read, many reels can be read simultaneously.

findmypast.com
family history in the making

Advantages of Scanner Banks

- Many advantages to having a bank of scanners instead of investing in single, high capacity units:
  - Scanners typically only need supervision during film loading/unloading
  - A single operator can manage many machines simultaneously, moving from one to the next, avoiding idleness
  - Plenty of spare components and even spare machines may be available in case of any failures. Single machine failures minimally impact overall performance.

Post-Processing on an Inexpensive Cluster

- Techniques required to prepare images for delivery:
  - Chanel levelling
  - Conversion to grayscale
  - Image inversion
  - Image rotation
  - Frame border cropping
  - Digital correction of lens effects such as barrel distortion and vignetting.
  - Compression
- All techniques can be performed programmatically.

findmypast.com
family history in the making

Distributed, Parallel Processing

- Linux, network boot
- Full-time image processing cores
- Commodity hardware
- Any machines, even older ones, can be added to increase throughput
- Cross-platform client can run on Windows machines only activating when the machine is idle for even additional throughput.
- Each reel is 5GB to 10GB prior to processing; 1GB to 2GB post.

findmypast.com
family history in the making

Online Delivery

- SATA drives available at less than $0.01/GB
- Multi-Terabyte RAID systems available for under $2000
- Fully-managed virtual storage such as Amazon Simple Storage Service is a comparable alternative

- Adobe Flex provides building blocks for a virtual, web-based microfilm reader. Of course, many additional features are possible

- External hard drive shuttles used to transport data from scanning facility to datacentre to minimise bandwidth costs

findmypast.com
family history in the making

The On-Demand Model from a User's Perspective

1. An end user searches an online catalogue of the archive or library.
2. The end user identifies film of interest and pays an affordable, up-front fee.
3. The user is told that they will be notified by email in a reasonable period of time (perhaps one to two weeks) when the film becomes available online.
4. The film is checked out from the repository and transported to the scanning facility.
5. The film is imaged, enhanced, compressed and copied to online storage.
6. An email is sent to the user notifying them of film availability.
7. The user logs into the website and accesses a Flex-based application that displays the images.
8. The original film is returned to the repository.

findmypast.com
family history in the making

Digital Rights Management

- Many possible extensions to the on-demand model to make sure that the rights of the original resource owner are respected.
- For example:
  - Limit on the time the film is available online
  - How many images may be saved locally or printed
- Alternatively, assuming appropriate rights, the images may be continuously available online to the requesting user, or perhaps made available to other users or moved to offline digital storage.

findmypast.com
family history in the making

On-Demand Model in Practice

- Using combination of technology and as much automation as possible, it is commercially feasible to lower the cost of digitising a reel and delivering it online to $10 or less, with sufficient volume.
- This fee could be directly passed onto the end user and would be comfortably within the budget of most family historians.
- On-demand digitisation would eliminate their potentially lengthy wait for online access to more obscure materials of interest.
- Also opens up digitisation options to small libraries that otherwise could not sponsor a digitisation project.
- Purely economy of scale. Higher volumes and build-out lower cost and, in turn, increase demand

findmypast.com
family history in the making

On-Demand in Practice

- Perhaps a bigger risk is being overwhelmed by excessive demand. This would increase fulfillment wait times for end users. This may be mitigated while capacity is scaled out by adjusting the price upward.

- The on-demand model is complementary to large-scale digitisation efforts. Due to the comparatively higher expense per record and lack of transcription/searchability, large-scale projects will generally yield the most usable product and the best value to the end user.

- However, since only a portion of filmed materials are commercially viable for large-scale treatment, the on-demand method is most suitable for providing access to the remainder of the records.

findmypast.com
family history in the making

On-Demand in the Future

- Only a fixed amount of filmed material available
- New media being captured directly to digital
- The on-demand model has a limited commercial viability
- Full capital investment can be recouped well within the period of viability permitting a graceful wind-down and wrap-up in the black

findmypast.com
family history in the making

Summary

- Hundreds of millions of pages of microfilmed material are not being digitised at this time due to insufficient collective demand to garner commercial investment.

- Using a novel application of recent technological advancements in imaging, data processing and storage, the cost of digitising and online delivery of microfilm can be reduced dramatically, even beyond the threshold that individual family historians are to directly and fully sponsor the digitisation and delivery of individual reels.

- An on-demand business model can be established which is driven and funded completely by end users. This essentially allows a library with a film collection to make the breadth of their collection available online with no direct costs.

findmypast.com
family history in the making