"Capital Markets of the World"

The Microfilming and Digitization of the Credit Lyonnais Stock Exchange Price Quotes Collection, 1850-2000

Supervisor: Prof. Dr. **Marc Flandreau**

Project Manager: Prof. Dr. **Michael Schiltz**

Day-to-day support and curation: Graduate Institute Library Staff

Business Partners:

1) Microfilming: Nichimy Corporation (Tokyo, Japan). Tel: 03-3815-1231. E-mail: info@nichimy.co.jp. Liaison: Kudou Hajime (e-mail: kudou@nichimy.co.jp).
2) Publishing (online disclosure):

Sponsors:

1. Executive Summary

The *Capital Markets of the World* project is an attempt to commercialize the scholarly potential of a comprehensive and unique collection of historical financial (multi-market stock exchange quotes gathered by the Credit Lyonnais Bank after its inception in 1863) information and, simultaneously, to work towards the long-term sustainable preservation of the latter. Recognizing the fact that accessibility and preservation of the information contained are different, sometimes conflicting, but nevertheless equally important ambitions, the project takes a hybrid approach. It builds on an analog technology (i.e. microfilm) and a digital one (online disclosure in the form of a searchable database). For the production of both, it employs the expertise of two commercial companies that have been recognized as world leaders in their field.

The project's business model is, after the very early stages, self-sustaining. Proceeds from the sale of access to the digital materials will be fed-back into the microfilming/digitization workflow; commercialization is thereby dependent on the identification of 'priority targets', for which scholarly demand should be considered guaranteed. Their disclosure is made dependent on injections of private and public money.

2. Problem Definition

2.1. Market Overview

"Capital Markets of the World" is an attempt to both a) preserve and b) disclose online the multi-market stock exchange quotes collection gathered by Credit Lyonnais throughout its more than 150 year old history. The collection, hosted at the Graduate Institute (Geneva, Switzerland) finds itself at an exciting yet challenging moment of its history. Its prospect looks bright, as the 21st century technological environment puts it is in a unique position to fill an important gap in available primary sources for the study of economic and financial history, and arguably related fields (economics, world history, the history of print and readership and so on). At the same time, however, many hurdles remain. One of these concerns the *enormity* of the project (currently several hundred meters of archives, amounting to millions of pages to be shot). The scanning and consecutive publication of this invaluable source will have to take place within the span of a decade or more; it will demand exceptional precaution with respect to project workflow and continuity, and this in financial (funding), technological (standardization, open formats), and documentary (making it possible to 'go back in time') terms. The latter is compounded by questions related to the collection's preservation, as preservation puts in place yet other requirements, which often conflict the demand for access and user-friendliness.

2.2. Market Background

The project has been developed out of a growing interest in capitalizing on available historical information (historical newspapers and magazines, archives …) that was previously difficult, cumbersome, and/or very costly to distribute or make accessible.[1] Nowadays, large-scale possibilities for digitization, falling costs of online storage, sharply enhanced technologies of optical character recognition (OCR) etc. make it possible to provide an easily searchable 'backlist' of a journal/newspaper, or an organization's documentary past. The managers of the project under discussion take particular hints from the commercial success and impact of back-listed journals that have traditionally been associated with a) reliability of information, and, by extension, b) their role of 'vetting' or certifying trustworthiness. In this context, commonly known examples with a high scholarly impact include: *New York Times*, *The Economist*, *The Financial Times*, *The New Yorker*,…

[1] Bond, Shannon. 2015. "Magazines and Newspapers Repackage Their Archives." *Financial Times*, April 57.http://www.ft.com/intl/cms/s/0/d3adfc18-d274-11e4-ae91-00144feab7de.html#axzz3jebxh7pu; Davis, Nicola. 2015. "Not Fade Away… How Robots Are Preserving Our Old Newspapers." *The Guardian*, July 5, sec. Books. http://www.theguardian.com/books/2015/jul/05/british-library-digitising-newspapers-boston-spa.

In academia, where the inherently meritocratic orientation dictates the need for state-of-the-art technologies, including technologies for retrieving yet unexplored sources, demand for documentary corpora (datasets, repositories, …) is massive and continuous. *This is certainly so for unique collections of an important international scope, as the project under discussion here.*

In view of its specialized nature (financial information and intelligence), the project discussed here caters foremostly to the scholarly community, in particular economic and financial historians, economic historians, and economists. In this context we note the renewed interest in long-horizon time-series of price information.[2] The project supervisor and manager are furthermore convinced that it will also appeal to researchers in related fields, as there are the history of print and readership, […]. Other interested partners may include international organizations, research services of commercial banks, and so on.

2.3. Market Challenges

In sharp contrast with the overwhelming demand for digital access to informational corpora, there has been a conspicuous and remarkable lack of attention for the *sustainability* of such access (and, by extension, the preservation of the underlying collection). As a matter of fact, even several high profile attempts to make available large troves of information (as in the case of Microsoft's digitization program) were aborted for commercial reasons, and nowadays struggle to survive in the "Internet Archive".

Even more fundamental are issues referred to as 'bit rot', 'legacy software' and the daunting spectre of a 'digital black hole' (after Vint Cerf, Google Corporation). Put simply, digital media turn out to be particularly short-lived. Hard-drives, still the most common form of digital storage in view of their storage capacity, have an estimated lifetime of a mere 6 years.[3] Data migration is possible, but becomes more risky and unreliable over time; and as volume increases, complexity multiplies. Faulty software and unreadable file formats are not necessarily a physical phenomenon: they do not actually decay, but rather suffer from a lack of being responsive and updated with respect to the changing environment in which they reside. Again, the Internet Archive Software Collection[4] and its host, Jason Scott,[5] are host to the sad remains of impermanent of digital data.

[2] Compare, in this respect, the "Long-Term Asset Return Study" released annually by Deutsche Bank's Market Research Division.

[3] "How Long Do Disk Drives Last?" 2015. *Backblaze Blog | The Life of a Cloud Backup Company*. Accessed September 6 2015. https://www.backblaze.com/blog/how-long-do-disk-drives-last/. For a generic discussion of reasons behind disk drive failure, see: "Hard Disk Drive Failure." 2015. *Wikipedia, the Free Encyclopedia*. https://en.wikipedia.org/w/index.php?title=Hard_disk_drive_failure&oldid=679541538.

[4] https://archive.org/details/software

[5] http://www.textfiles.com/

Yet other concerns are hardwired into the digital paradigm itself, more especially into the ease with which digital information can be handled. Malicious code can destroy swaths of information in a matter of seconds; a simple human error may be responsible for altering or deleting any digital 'original'. In short, the digital paradigm does not (yet?) comprise a data secure and future-proof access strategy.

2.4. Market Opportunities

A hybrid, multi-partner approach to preservation and information disclosure may make it possible to guarantee access to digitized materials, while at the same time enabling a mid-term to long-term (500 years) preservation of archival information. Interestingly, there has been a renewed and relatively recent academic interest in the matter.[6] Authors within this research strand recognize that requirements with regard to preservation and (ease of) access have often been conflicting, yet reject the choice for one technology over the other (as is often done in the name of 'opportunity costs'). Instead, they argue for the necessity and viability of both through the adoption of hybrid technologies.

3. Solution

The hybrid approach is at the heart of the *Capital Markets of the World* project too. In the early stages of the project, Marc Flandreau therefore solicited collaboration with Michael Schiltz. Schiltz has set up several projects related to data-infrastructure and -management. Directly related to the project under discussion, he has been responsible for the development of a 'smart catalog' for the archives of the Yokohama Specie Bank at the University of Tokyo (in Japanese; available upon request). With colleagues at the Library of Congress and the library staff of the archive section of the Library of the Economics Faculty of the University of Tokyo, he has as well developed a trusted hybrid approach towards conservation (microfilming and digitization) as a means towards sustainable curation and disclosure of bank archives. As indicated above, the latter has increasingly come to be regarded as a viable solution to the -technologically sometimes conflicting- aims of a) conservation on the one hand, and b) accessibility on the other.

Concretely, the project proposes 1) the use of (analog) *preservation microfilm* and *digital* disclosure, *and* 2) *this through partnering with different companies that have been internationally recognized for their respective competences*. In this context, we confirm being acquainted with the very recent and

[6] Bereijo, Antonio. 2004. "The Conservation and Preservation of Film and Magnetic Materials (1): Film Materials." *Library Review* 53 (6): 323–31. doi:10.1108/00242530410544411; Bereijo, Antonio. 2004. "The Conservation and Preservation of Film and Magnetic Materials (2): Magnetic Materials." *Library Review* 53 (7): 372–78. doi:10.1108/00242530410552313; Voges, Christoph, Volker Märgner, and Tim Fingscheidt. 2008. "Digital Data Storage on Microfilm – Error Correction and Storage Capacity Issues."*Archiving Conference* 2008 (1): 212–15; Voges, Christoph, and Jan Fröhlich. 2011. "Long-Term Storage of Digital Data on Cinematographic Film." *Archiving Conference* 2011 (1): 158–61.

very promising idea to write both computer-readable digital data and human readable pictures/text onto one photosensitive film,[7] but, for reasons of long-term cost and others (highlighted below), do not follow this path.

4. Business Model

Whereas the technological base layer of the project consists of a combination of both analog (microfilm) and digital means, *its true potential is to be found in its partnering with world-class commercial enterprises*.

Importantly, the markets of both microfilming services and scientific publishing are characterized by their oligopolistic, if not semi-monopolistic nature. Nowadays, only one corporation (Fujifilm, Japan) still produces analog film. One is not surprised that one of the very few research-intensive preservation microfilm offering companies is also Japanese. An unchallenged leader of scanning services within Japan proper (it has scanned the historical archives of all of Japan's foremost newspapers), **Nichimy Co.** has been hired by the United States' National Archives to conduct historical research and scanning services. The *Capital Markets of the World - project* has been fast to recognize the potential to embrace the potential of an analog technology that does not only offer a long-term preservation solution, but *also offers an image resolution (around 2,000 dpi) commonly unmatched by digital imaging*. Importantly, it should also be considered the cheapest form of long-term preservation, as there are zero costs with respect to data migration, software maintenance, and so on. Note furthermore that the digitization of CMoW analog microfilm implies only minimal quality loss. In contrast with other microfilm to digitization efforts, in which digitization of an (inherently lower resolution) positive film leads to an 'acceptable' outcome, the CMoW-project foresees *digitization on the basis of high resolution (micro-fine grain, silver halide, direct-image films; 1200 lp/mm) 'direct duplicating' microfilms*.[8]

Referring to table 1 below, we recognize the great potential of digital, 'self contained'[9] microfilm as proposed by Piql,[10] yet regard the solution as unfit for the project under discussion. Piql takes a very promising 'holistic' approach towards preservation, yet its solution for relatively fast retrieval of data presupposes an (admittedly limited) hardware infrastructure (Piql Writer/Piql Reader, and the Cinevator for cinematic film). Piql is thus most ideally geared towards a preservation demand that does

[7] Brudeli, Bjørn H. 2014. "A Holistic Approach to Digital Preservation." *SMPTE Conferences* 2014 (10): 1–11. doi:10.5594/M001579.

[8] See, for a technical specification: http://www.fujifilmusa.com/products/microfilm/duplicating_films/direct_duplicating_film/index.html

[9] Piql digital microfilm is self-contained in that it contains human-readable text explaining how to reproduce and interpret your files. Data retrieval is independent of access to specific technologies and, more important, specific vendors of hardware.

[10] See, for a description of the technology and the company's view of preservation: http://www.piql.com/

not simultaneously plans records disclosure (e.g. government records, administrative registers, and yet other documents that are not intended for (commercial) publication, yet should be preserved for a long-term future in view of, for instance, legal stipulations). For the *Capital Markets of the World* project, B/W analog microfilm[11] is both a cost-cutting factor and a guarantee for sustainability.

Our choice to collaborate with a major publishing house rather than maintaining the digital repository by ourselves has been inspired by a similar concern with minimizing project costs. Not only is it possible to avoid costly investments with regard to personnel, hard- and software environments and maintenance (including the hard-to-guarantee commitment to long-term accessibility, e.g. through digital object identifiers (DOI)), the project manages to make full use of existing networks of dissemination and merchandising that are difficult, if not impossible to replicate from scratch.

The handling of the physical book volumes, (un)bound papers, … is outsourced to yet another trusted partner, closely cooperating with Nichimy Corporation. Volumes are disassembled cautiously, and treated with a wax/spray (?) that slows down the process of decay with a factor of (?). Before microfilming, Nichimy operators furthermore apply a vacuum cleaner especially designed for rare and brittle books, and manuscript boxes. The latter removes dust particles that may hamper the readability of the text contained, and may eventually threaten the condition of the physical original.

The very oligopolistic nature of both the preservation and publishing businesses, coupled with the academic expertise embodied in the Capital Markets of the World - project, creates vast opportunities for what will remain a limited group of project members. *The combination of linguistic skills (Japanese as a guarantee for a liaison with Nichimy) and an existing tradition of cooperation, together with the built-up expertise with respect to archiving is not only unique; it can also easily be adapted to the needs of other archival projects, the existence of which is guaranteed in Geneva*. This project's authors stress furthermore that they are dedicated to mastering the technologies needed for the digitization of analog microfilm as well.

5. Product and Technology

Below, we briefly list the characteristics of different storage technologies:

[11] Color microfilm is a relatively recent alternative, yet it represents a very high cost in comparison to B/W microfilm. Furthermore, color photographic dyes tend to degrade over the long term. As the *Capital Markets of the World* collection mostly consists of textual information (stock exchange quotes and financial journalism), the choice for B/W microfilm is obvious. For a discussion of color microfilm, see: Voges, Christoph, Volker Märgner, and Tim Fingscheidt. n.d. *Investigations on Color Microfilm as a Medium for Long-Term Storage of Digital Data*.

Table 1.

	magnetic tape	optical disks	hard disk drive	cloud	microfilm	Piql preservation services
projected lifetime (years)	30	100(?)*	6	N/A	500	500
analog data					*	*
digital data	*	*	*	*		*
true WORM**		*			*	*
offline format		*			*	*
instant retrieval			*	*		
vendor neutral reading device		***	*		*	*
migration free storage medium					*	*
true archival format						*

*depending on the manufacturer, claims vary considerably. In a rather extreme case, the manufacturer of MDisc has claimed an expected lifetime expectancy of 1,000 (!) years. However, apart from the fact that the retrieval of data from optical disks is vendor dependent (and thus vulnerable to the vendor going out of business), dust, heat and UV light can do severe damage. Besides physical damage, failure of the reflective layer, typically followed closely by degradation of the data layer, are the primary failure modes of all optically recordable disks. In general, because of their fragility and their low data capacity, this project's authors stress that optical discs should be considered unfit for large-scale organizational archiving.

**WORM: "Write Once, Read Many": refers to the unalterable nature of the data once it has been written on the storage medium.

***this is subject of debate: certain types of disks (e.g. Blu-ray disks and MDiscs require a Blu-ray device for data retrieval, yet it is unclear whether these devices will be accessible for the long-term future).

What has been argued in the above paragraphs can easily be distilled from Table 1. In view of the concrete demand for the *preservation* of both the physical collection and the information contained, (silver halide) microfilm is the obvious choice.

With respect to the choice of a digital format for information disclosure, this should be obvious. In order to enhance user friendliness, 1) downloadability (in .pdf or others formats), 2) the use of optical character recognition (OCR), and 3) the possibility of exporting metadata into leading bibliographic software applications (Zotero, EndNote, Refworks, Mendeley, BibTeX), and the use of a 'digital object identifier' (DOI) are options to be considered.

[Indication of the strengths of the partnering scientific publisher]

6. Marketing and Sales

(to be discussed with the partnering publishing house)

7. External Environment / Competition

A comprehensive and almost complete multi-market instrument of primary materials regarding historical stock exchange quotes is not yet available to the academic community. Admittedly, a number of efforts have been made at digitizing stock exchange lists. However, these efforts were typically made at the national level, for the simple reason that there does not exist any repository or archive where sets of multiple national collections are kept. In effect a competing project would see itself confronted with the highly labor-intensive prospect of having to conclude separate licensing agreements with as many national repositories as there are markets to cover. This would furthermore assume that all collections are still in existence, which is pertinently not the case —quite a few collections in the Crédit Lyonnais collection are simply unique.

Below we explore comparisons with a few existing digital projects:

- London's Investors' Monthly Manual: Yale Management School:

see:
http://som.yale.edu/faculty-research/our-centers-initiatives/international-center-finance/data/historical-london

Importantly, the Crédit Lyonnais collection contains the *complete* IMM series starting from 1865, rather than 1869... Note as well that the YSM project has data for St. Petersburg and Shanghai. Russian data is from monthly records; given the use of the Julian calendar so-called 'monthly' data do not correspond to a monthly frequency for other markets. The Crédit Lyonnais collection, by contrast, contains Saint Petersburg *daily* Stock Exchange and Moscow when it opens. The data for Shanghai are annual (the authors have used end-of-month data for December each year), and were collected by means of a secondary source, the *North China Herald* (at the time the most representative journalistic outlet for foreigners in China).

- Paris, official stock exchange: BNF Gallica:

see: http://gallica.bnf.fr/ark:/12148/cb32715573v/date

This collection does not include data on the interbank market which was as large as the official market.

- Vienna:

see: http://www.anno.onb.ac.at/

ANNO, i.e. AustriaN Newspapers Online contains several titles of contemporary (19th century and early twentieth century) economic journalism. Although certainly meaningful when acquainting oneself with trading practices etc., the collection does not have the scope to be comprehensive; also, the financial data contained herein are not formatted in the systematic way the Crédit Lyonnais collection is.

- Others:

There are other projects that have not resulted in any form of online disclosure. Examples include an old project from the Oslo stock exchange, the Antwerp stock exchange, etc. We have heard some years ago of a NY project but could not get confirmation. Usually, existing projects have sought to put extracted "data" online rather than primary materials along with a sorting system.

The Collection's appeal is further corroborated by the existence of online accessible repositories and databases of financial journalism. We mention in this context:

- *The Economist* Historical Archive:

http://gdc.gale.com/products/the-economist-historical-archive-1843-2007/

- *Financial Times* Historical Archive:

http://gale.cengage.co.uk/financial-times-historical-archive.aspx

- *North China Herald*:

http://nch.primarysourcesonline.nl/nch/

Importantly, the Capital Markets of the World collection includes yet other, more specialized sources of contemporary financial journalism (*Burdett's Official Intelligence, The Bankers' Magazine, l'Économiste européen*, *a complete run of the Economist*'s *Investors' Monthly Manual (IMM)*…), that are not yet available online, but that will eventually be included in the microfilming/digitization workflow.

8. Management Team

As indicated above, Michael Schiltz has a proven track record with the preservation of bank archives through his participation in the microfilming of the vast Yokohama Specie Bank archives. Research-wise too, he has demonstrated strong vision and leadership. All of his research projects, culminating in a 5-year role as principal investigator (PI) of a European Research Council Starting Grant,[12] have been brought to fruition within the agreed project time frame. Due to a) his acquaintance with research undertakings that contain a strong element of archival preservation and disclosure (see,

[12] "'Dead End': An Economic and Cultural History of Japan in the Age of the Great Depression, 1927-1937' | ERC: European Research Council." 2015. Accessed September 6 2015. http://erc.europa.eu/dead-end-economic-and-cultural-history-japan-age-great-depression-1927-1937.

for example, his tenure of two Kluge Fellowships at the Library of Congress, Washington DC) and b) his vast network of contacts in (digital) librarianship and archiving, he is in a unique position to bring to the *Capital Markets of the World* project a continuity that is prerequisite to its successful completion.

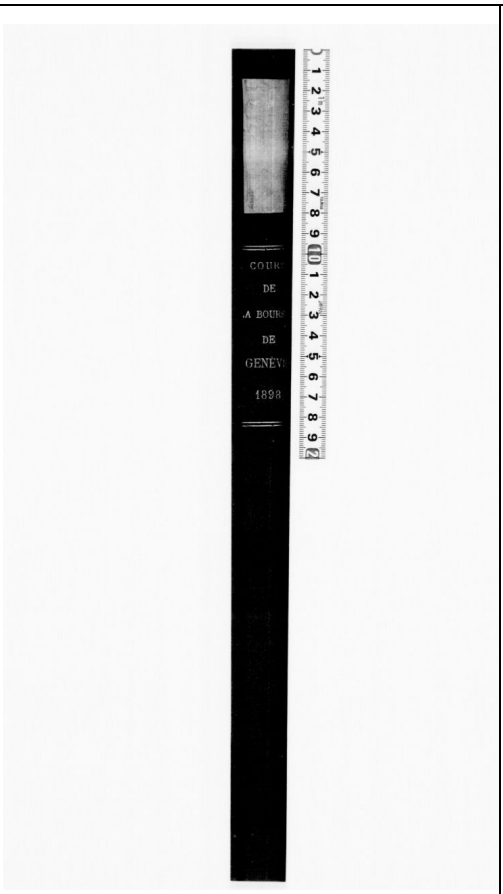
Next, the project profits from the excellent library/archival infrastructure of the Graduate Institute. See, for an overview of collections and expertise in conservation efforts: http://graduateinstitute.ch/home/research/library/archives.html
1 archivist has already been working 100% on the collection's preservation.
The project's Supervisor, Marc Flandreau, spends a considerable amount of research time to the Collection, through the supervision of several Ph.D. candidates. He also intends to follow up closely the conservation/disclosure effort.

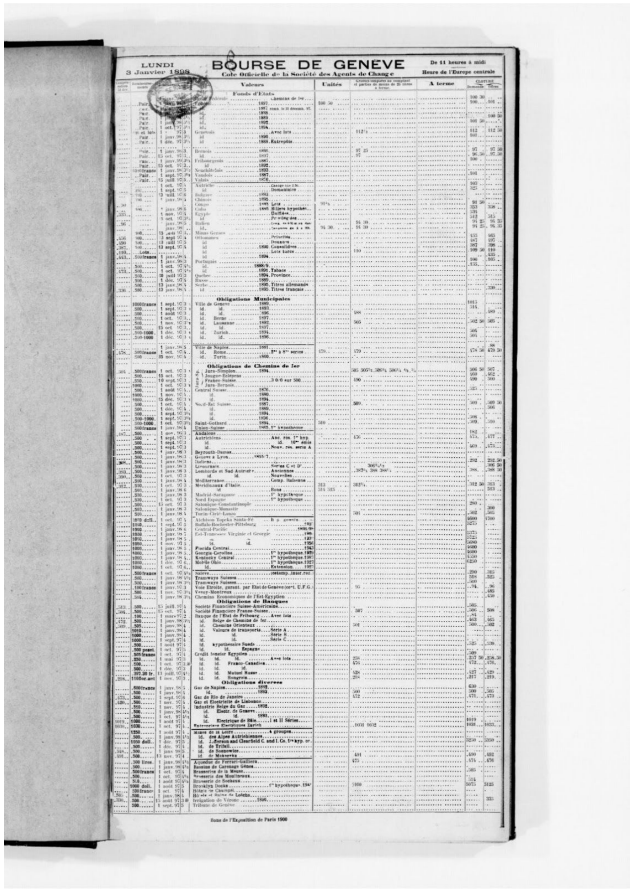
9. Financial Projections and Key Metrics

10. Current Status, Use of Proceeds and Milestones

In November 2014, project leaders Marc Flandreau and Michael Schiltz agreed to a **pilot project** with the Japanese research and scanning services company Nichimy. Aided by a generous donation by Mr. Yves Mirabaud, the complete collection of Geneva stock exchange quotations (contained in the Credit Lyonnais collection) effectively amounted to the first complete 'batch' of documents to be microfilmed. Consisting of 147 bounded volumes, or approximately 113,000 pages, the collection was judged an ideal test-case. This collection is in good or even excellent state, yet, due to its sheer size, the uncommon oblong format of the volumes etc., it was guaranteed to produce several of the challenges that are typical of large-scale scanning projects. The pilot project has proven indispensable in informing the project leaders of a) quality controls (preparations for scanning, cleaning) to be put in place and agreements with respect to workflow, b) important precautions with respect to unbinding of the volumes, c) issues related to the transportation of a collection, d) photographic-technical matters to be taken into account, e) software and hardware to be used for digitizing microfilm and applying OCR, and so on.

Image 1. An illustration of the size and uncommon format of the Geneva Stock Exchange Price Quotes Collection





Although other collections will turn out to contain yet other challenges (e.g. deteriorated or brittle paper quality / torn pages, questions concerning the future implications of unbinding, problems of 'vignetting'[13] when shooting large-format materials,...), the pilot project arguably contained all elements to simulate the workflow that can be used for the processing of all future batches, relatively independent of their size.

Below, such **workflow** is represented schematically. Roughly, each phase consists of 4 elements to be executed simultaneously (A-D), each related to different batches at different points of the workflow process. Note the 'cascade'-like structure of the workflow; all elements are to be considered of similar 'weight' of importance, including with respect to the time frame in which they are to be executed.

Table 2. Approximate Project Workflow

	A. preparation of	B. microfilming and	C. preparation of a	D. online disclosure

[13] "Vignetting." 2015. *Wikipedia, the Free Encyclopedia*. https://en.wikipedia.org/w/index.php?title=Vignetting&oldid=669684787.

	material (precise counting of pages, unbinding, cleaning, …)	production of custom-made acid-free boxes for physical items	'checklist' for OCR-processing; soliciting the creation of market-specific metadata	
I	batch 1			
II	batch 2	batch 1		
III	batch 3	batch 2	batch 1	
IV	batch 4	batch 3	batch 2	batch 1
V	batch 5	batch 4	batch 3	batch 2

etc.

The **order** in which batches are to be microfilmed/digitized is based on an assessment of a) their respective scholarly importance and b) arguably related to the latter, their appeal for academic libraries, research institutions affiliated with international banks and others, international organizations…

The project will rely on injections from public and private bodies only in the early stages of its development, and this to be decided on the nature of the batch to be microfilmed/digitized, the character of its publication ('Green Open Access' versus commercial publication, either as 'pay-per-view' or in the form of a yearly (private or institutional) subscription), or, dependent on the conditions of the donor/investor, yet other specifications. As soon as these 'priority batches' (see table 3 below) are digitally available to the public, a share of the proceeds from subscription fees will be fed back into the microfilming and digitization workflow so as to guarantee a steadily growing corpus of multi-market financial information.

The first 'priority group' is identified as follows:

1. London
2. New York
3. Hamburg
4. Amsterdam
5. Alexandria
6. Buenos Aires
7. Tokyo
8. St. Petersburg

One may object to the exclusion of the Paris stock exchange from the above list, but that has to do with logistical concerns. Comprising several lists, both official and unofficial, the Paris collection is by far the largest series in the collection (approx. 500 boxes). Its microfilming/digitization is therefore taken as the focal point of the second priority group. After the second group, the project authors plan to proceed geographically.

11. Risk Analysis