



**DLF SPRING FORUM**

**San Diego**

**April 2005**

# **THE UNIVERSITY OF MICHIG**GOOGLE**: FACT AND FANCY**

**Mark Sandler**

**Collection Development Officer**

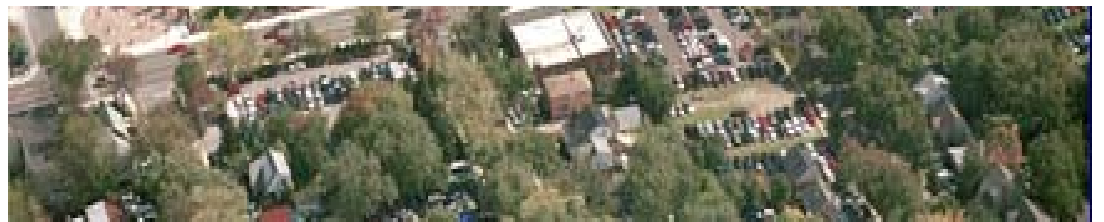
**University of Michigan**

# THE UNIVERSITY OF MICHIGAN

“leaders and best”



Google™





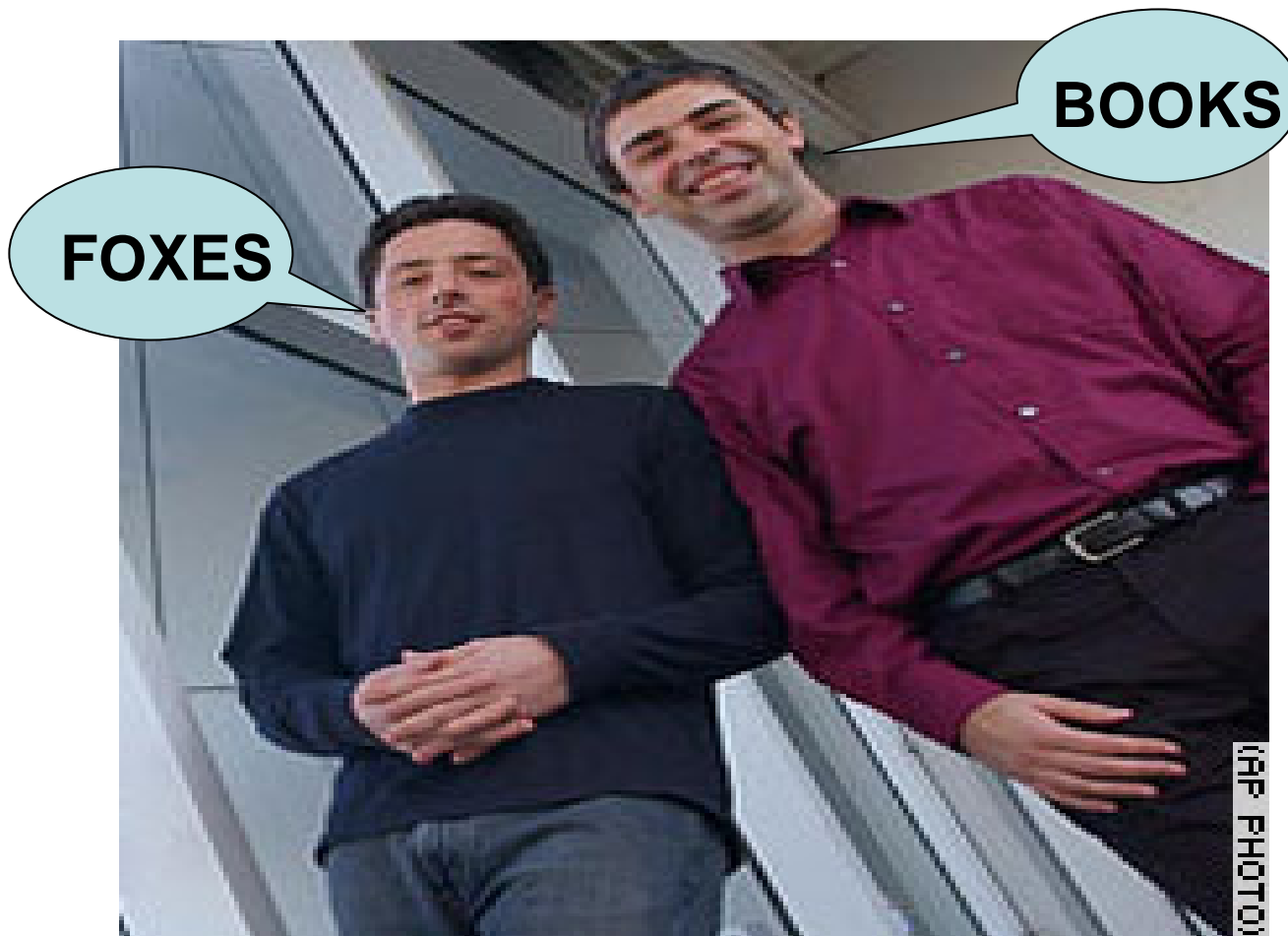
## TODAY'S COMMENTS

- ❖ **What's going on?**
- ❖ **Why is it going on?**
- ❖ **What's changed as a result?**
- ❖ **What needs to change as a result?**





# TWO WILD AND CRAZY GUYS





# DO WHAT?

*Google's mission is to organize the world's information and make it universally useful and accessible*

**“working to digitize the book collections of several major research libraries and make this content searchable through Google”**





# DO WHAT AGAIN?

## ❖ Digitize Michigan's entire stacks collection— 7 million volumes

- Excludes special collections, Bus. Ad. & Law
- Images are high resolution color jpgs
- OCR
- Metadata

## ❖ Return files to Michigan

- 600 dpi bitonal TIFFs, with 300dpi jpeg2000 illustrations
- Files to include OCR and headers





## HOW ARE THEY DOING IT?

- ❖ **Scanning work site has been established on campus**
- ❖ **Staff hired by Google**
- ❖ **Files transferred for a range of optimizing touches**
- ❖ **Finished files to be added to Google Print and derivative returned to Michigan**



# THE GOOGLE WORKSTATION (CONFIDENTIAL)





# THE ANN ARBOR WORK GROUP



# WHAT WILL IT LOOK LIKE?

<http://print.google.com/googleprint/library.html>

**Quality and quantity**

**In copyright and public domain volumes**

**Browsing and searching**

**Downloading and printing**

**Links to booksellers and libraries**





# DO WHY?

“project's aim is simple: help **maintain the preeminence of books and libraries** in our increasingly Internet-centric culture by making these information resources an integral part of the online experience. We hope to **guide more users to their local libraries; to digital archives of some of the world's greatest research institutions; and to out-of-print books** they might not be able to find anywhere else – all while carefully **respecting authors' and publishers' copyrights.**”





# NO, REALLY?

**Why is Google REALLY doing this?**

**To make more money?**

**To enhance the value of their search engine?**

**So Microsoft doesn't?**

**To make the world a better place?**

**All of the above?**

**None of the above?**





# WHY WOULD MICHIGAN DO IT?

- ❖ **We believe Google will successfully reformat a substantial portion of our collection**
- ❖ **We believe the files will be produced to our standards**
- ❖ **The files will be returned to us**
  - **An archive**
  - **An access tool**
  - **Can use in collaboration with others**
- ❖ **We believe this will democratize library access**
- ❖ **These basic files will free up time and money for selective enhancement**





## **“SOOK IT AND SEE”**

**Michigan’s primary interest is to receive high quality digital files for as much of our legacy print collection as possible.**

**And, we wish Google good luck with their implementation, as well.**





# WHAT IFs?

## ❖ What if they can't do what they hope they can?

- Inadequate quality
- Too few titles
- Destroy print copies in the digitization process
- Capture a lot but can't organize it
- Can't/won't migrate files forward
- Lose a lot of money

## ❖ What if they can do what they hope they can?

- Will it undermine print culture?
- Will advertising corrupt the reader experience?
- Will readers drown in too many volumes?
- Will users print too much?
- Will Google make a lot of money, and is that okay?



# WHAT ARE WE DOING IN THE MEANWHILE?

- ❖ Making of America style projects
- ❖ Text Creation Partnership
- ❖ Enhancements to DLXS
- ❖ Institutional Repository
- ❖ Scholarly Publishing Office
- ❖ Commercial acquisitions



# NEW LIFE FOR OLD BOOKS



The screenshot shows a Netscape browser window titled "Making of America". The address bar displays "http://moa.umdl.umich.edu/". The browser's toolbar includes buttons for Back, Forward, Stop, Refresh, Home, AutoFill, Print, and Mail. Below the address bar, there are links to Net.juke, Early English Books Online, header, Google, and University of Michigan Online Directory. The main content area features a large, ornate banner with the text "Making of America" in a cursive font, set against a background of a steam locomotive. Below the banner, there are three main sections: "Go to MoA Books", "Go to MoA Journals", and "About MOA". Each section has a small MOA logo and a brief description. The "About MOA" section includes a link to "About MOA" and a link to "contact". The "contact" link is accompanied by an illustration of a typewriter. The browser's sidebar on the left contains links to Favorites, History, Search, Scrapbook, and Page Holder.

Back Forward Stop Refresh Home AutoFill Print Mail

Address: <http://moa.umdl.umich.edu/>

Net.juke Early English Books Online header Google University of Michigan Online Directory

## Making of America

 [Go to MoA Books](#)

 [Go to MoA Journals](#)

**M**aking of America (MOA) is a digital library of primary sources in American social history from the antebellum period through reconstruction. The collection is particularly strong in the subject areas of education, psychology, American history, sociology, religion, and science and technology. The collection currently contains approximately 8,500 books and 50,000 journal articles with 19th century imprints. The project represents a major collaborative endeavor in preservation and electronic access to historical texts.

The Making of America collection is made up of images of the pages in the books and journals. When you find something you want to look at, you will see a scanned image of the actual pages of the 19th century volume. Optical Character Recognition (OCR) has been performed on the images to enhance searching and accessing the texts - for more on the OCR process see [About MOA](#).

Making of America is growing and changing. Over five thousand new titles have been added, bringing Making of America to over two and a half million pages online. For a short time, the journals and books will be divided into two separate collections; you will not be able to search both at once.

 [MOA about](#)

 [MOA help](#)

 [MOA contact](#)



# THE MAKING OF AMERICA

- ❖ **10,000 volumes**
- ❖ **American Imprints**
- ❖ **1850-1877**
- ❖ **Housed in storage**
- ❖ **Predominantly brittle**
- ❖ **600 dpi TIFFS with dirty OCR searchability**
- ❖ **750,000 page views per month**





# EXTENDING MOA

- ❖ **Internally funded reformatting of 5,000+ volumes per year**
  - **Homeopathy**
  - **Great Lakes**
  - **Public Papers of the Presidents**
  - **New York City History**
  - **Dentistry**
  - **Mathematics**
- ❖ **30,000 books freely accessible online and available for hard copy purchase**



# THE TEXT CREATION PARTNERSHIP(S)



## Early English Books *Online* Text Creation Partnership

Basic Search

Boolean Search

Proximity Search

Bibliographic Search

Word Index

Browse

[About EEBO-TCP](#)

comments to:  
[dtps-info@umich.edu](mailto:dtps-info@umich.edu)



# TEXTS ARE DIFFICULT TO READ (AND KEYBOARD)

There begynneth the lyfe of the blessed  
martyr saynte Thomas.



**T**he martir  
saynte Tho-  
mas was son  
to Gylberte Bequet a  
burgess of the Cite of  
London. And was bo-  
rne in y<sup>e</sup> place where as  
now standeth the chur-  
che called saynte Tho-  
mas of Aheris. And this Gylberte was a good deu-  
te man / and toke the crosse upon hym / and wente on  
pylgrimage to the holy lande / and had a scruaunt w<sup>th</sup>  
hym. And whan he had acomplished his pylgrima-  
ge / he was taken homwarde by the hethenmen / and  
brought in prisson of a pryncer named Amersat where  
longe tyme he / and his self shyp suffred moche payne  
and sorowe. And the pryncer hadde great affeccon to-  
warde this Gylberte / and had oft comunycon with  
hym of the cristen feyth / & of y<sup>e</sup> roialme of Englande  
by whiche conuersacion: it fortuneth / that y<sup>e</sup> daughter  
of this pryncer had especial loue vnto this gylberte / &  
was famplier with hym: and on a tyme she disclosed  
his loue to hym / sayenge if he wolde promyse to wed  
his / she shulde forsake frendes / heritage / and countre  
for his loue / and become cristen / and after longe co-  
munycon betwene them / he promysed to wed hyr

Record as:	Textura	Italic	Bastarda	Rotunda
a	<b>a</b>	<i>a</i>	<b>a</b>	<b>a</b>
b	<b>b</b>	<i>b</i>	<b>BB</b>	<b>b</b>
c	<b>c</b>	<i>c</i>	<b>c</b>	<b>c</b>
d	<b>dd</b>	<i>d</i>	<b>dd</b>	<b>d</b>
e	<b>e</b>	<i>e</i>	<b>e</b>	<b>e</b>
f	<b>f</b>	<i>f</i>	<b>f</b>	<b>f</b>
g	<b>g</b>	<i>g</i>	<b>g</b>	<b>g</b>
h	<b>hh</b>	<i>h</i>	<b>h</b>	<b>hh</b>
i	<b>i</b>	<i>i</i>	<b>i</b>	<b>i</b>
j		<i>j</i>		



# ADDING AMERICAN TEXTS



Evans Early American Imprint Collection  
Text Creation Partnership

Simple Search  
Boolean Search  
Proximity Search  
Citation Search

[About evans-TCP](#)

comments to:  
[dtps-info@umich.edu](mailto:dtps-info@umich.edu)





# AND GALE'S 18TH CENTURY COLLECTION



ECCO  
TCP

EIGHTEENTH CENTURY  
COLLECTIONS ONLINE

- ◆ GOALS & STRATEGIES
- ◆ PRODUCTION
- ◆ BECOME A PARTNER
- ◆ NEWS



## Text Creation Partnership

The University of Michigan and Thomson-Gale are cooperating in a Text Creation Partnership to create 10,000 accurately keyed and fully searchable SGML/XML text editions from among the 150,000 titles available in the Eighteenth Century Collections Online (ECCO) corpus. ECCO is a significant collection of titles relating to the history of eighteenth century Britain, and the Text Creation Partnership seeks to create enduring digital text editions of the most frequently studied works. Gale is producing digital page images and searchable OCR (Optical Character Recognition) for the overall collection. The standards-based editions created and owned by Text Creation Partnership will link directly to the corresponding Gale page images and will provide a research and instructional tool of enduring scholarly value.



# TEXT CREATION PARTNERSHIP

- ❖ 10,200 accurately keyboarded and SGML/TEI encoded works completed (about 3,500 per year)
- ❖ Funding committed to reach 25,000 across EEBO, Evans, and ECCO (support from over 150 libraries)
- ❖ ARL cost for EEBO will be \$2.00 per volume for files that we **OWN**
- ❖ Local implementation
- ❖ Work group of 10 FTE at Michigan and Oxford
- ❖ Excellent vendor support







# DLXS ENHANCEMENTS

- ❖ **25 to 30 partner institutions**
- ❖ **Enhanced newspaper capability**
- ❖ **Unicode indexing was implemented and display is in the works**
- ❖ **XSLT**



# DEEP BLUE

## Institutional Repository

- Built on D-Space
- Coming on-line this Spring
- 8,000 items opening day collection





# SCHOLARLY PUBLISHING OFFICE

## ❖ **Outlet for campus publications**

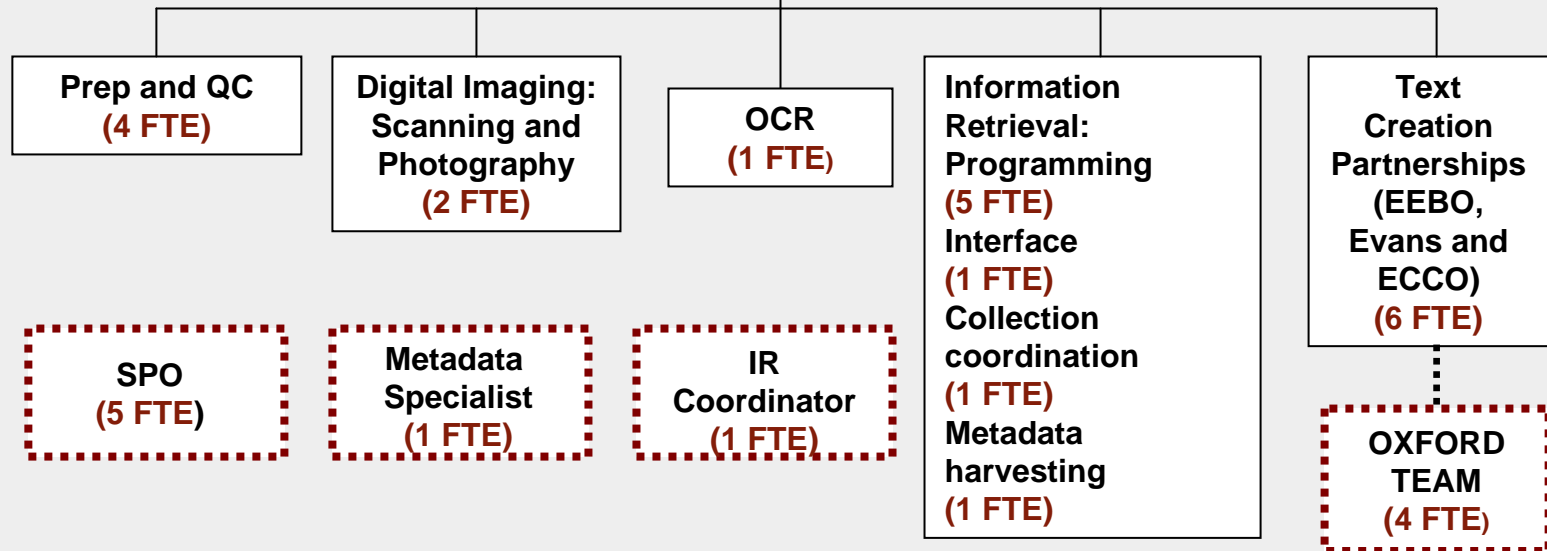
- Michigan Quarterly Review
- Philosopher's Imprint
- Japanese Studies Monographs
- International Institute
- Original Monographs; JEP
- SPARC

## ❖ **Revenue generating projects**

- LLMC
- ACLS History E-Book project



# MICHIGAN'S DIGITAL LIBRARY CAPACITY





# ARE WE STILL BUYING ANYTHING?

- ❖ **Shaw-Shoemaker**
- ❖ **The Making of Modern Economy**
- ❖ **The Making of Modern Law**
- ❖ **STM Backfiles**
- ❖ **Shakespeare**
- ❖ **Empire Online**
- ❖ **Microfilm**
  - **Prange**
  - **Women's magazines**
  - **Spanish Newspapers**



Government	Percentage
Current government	85%
Opposition	10%
No answer	5%



# EARTH TO LIBRARIES

❖ **Not a time for passivity-- get in the game!**

- **Local history collections**
- **Local literature-- TEI encoded editions?**
- **Images**
- **Rare and fragile books**
- **Special collections**
- **Newspapers**
- **Microfilm**
- **Campus resources**



# REALLY, I HADN'T HEARD THAT



- **Budgets are bad**
- **There's too much work**
- **Cooperation is hard**
- **Everything is changing**
- **Publishers are greedy**
- **Faculty are demanding**
- **Staff skills are scarce**
- **Nobody has time to think**
- **I'm scared to send my books offsite**
- **The Provost doesn't understand**





# THE WORK TO BE DONE

- ❖ **Compile communities of content for communities of users**
- ❖ **Develop specialized tools for specialized users**
- ❖ **Selectively upgrade important texts, heavily used texts, and texts of enduring research and instructional value**
- ❖ **Integrate with other important digital library corpora (even the French!)**
- ❖ **Find cooperative approaches for digital and print archiving**



# QUESTIONS AND COMMENTS

