

# Building Capacities and Communities for Digital Scholarship

The “Digging Deeper, Reaching Further: Libraries  
Empowering Users to Mine HathiTrust Digital  
Library Resources” Project

Harriett Green

Collections As  
Data Symposium,  
Library of  
Congress

September 27,  
2016

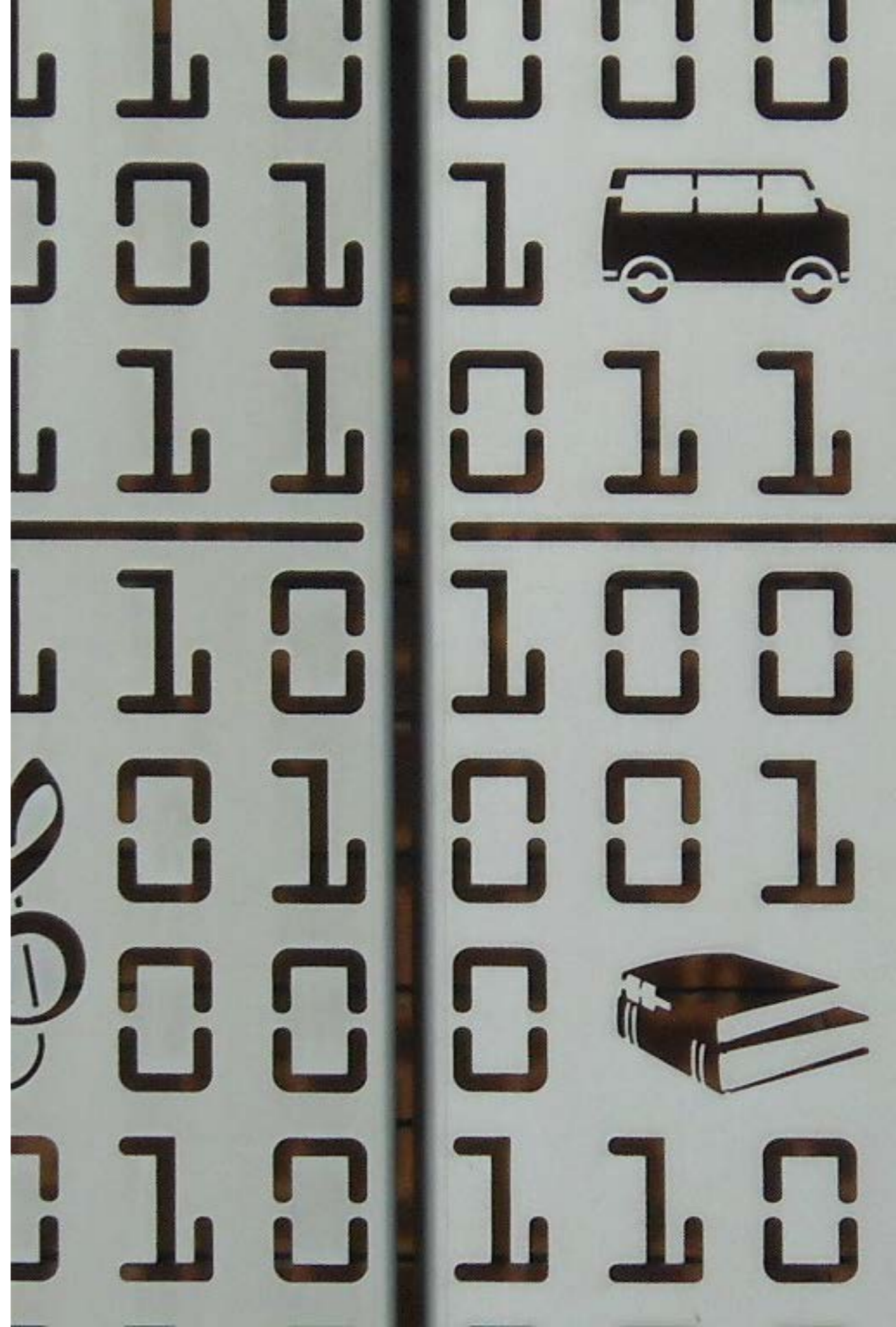
# The Changing Landscape for Research Libraries

“In 2033, the research library will have shifted from its role as a knowledge service provider within the university to become a collaborative partner within a rich and diverse learning and research ecosystem.”

**ARL Strategic Thinking + Design Initiative  
report (2015)**

# Data in the Humanities and Social Sciences

How do we support emergent data-driven research?



Data for  
Scholarly  
Research in  
the  
Humanities  
and Social  
Sciences

“A critical mass of information is often necessary for understanding both the context and the specifics of an artifact or event, and this may include large collections of multimedia content: images, text, moving images, audio.”

**American Council of Learned Societies, *Our Cultural Commonwealth: The report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences* (2006)**

# Building Capacities for Digital Scholarship in Academic Libraries



# Are Academic Libraries Prepared for Digital Scholarship?

“The research library will survive because of the introduction of ever more and newer digital technologies, not in spite of them.”

**Bernard Frischer**

- Davis and Dombrowski, *Divided and Conquered* (2010, NITLE report)
- Zorich, *A Survey of Digital Humanities Centers in the United States* (2008, CLIR)
- *ARL SPEC Kit #326: Digital Humanities* (2012) and *ARL SPEC Kit #350: Supporting Digital Scholarship*

# Intersections of Digital Scholarship + Librarianship

- Studies and reports from Ithaka S+R, Association for Research Libraries (ARL), and Coalition for Networked Information (CNI)
- White and Gilbert (eds.), *Laying the Foundation: Digital Humanities in Academic Libraries* (2016)
- Hartsell-Gundy et al. (eds.), *Digital Humanities in the Library: Challenges and Opportunities for Subject Specialists* (2014)
- ACRL Digital Humanities Interest Group, ADHO Libraries Special Interest Group

# Re-Skilling Librarians: Building Skill Capacity

- University of Maryland Libraries, DH Incubator
- Columbia University Libraries, The Developing Librarian Project: <http://developinglibrarian.org>
- Harvard Library Lab, <http://osc.hul.harvard.edu/liblab/>

**How can we empower all academic librarians to engage with digital scholarship and new forms of research?**



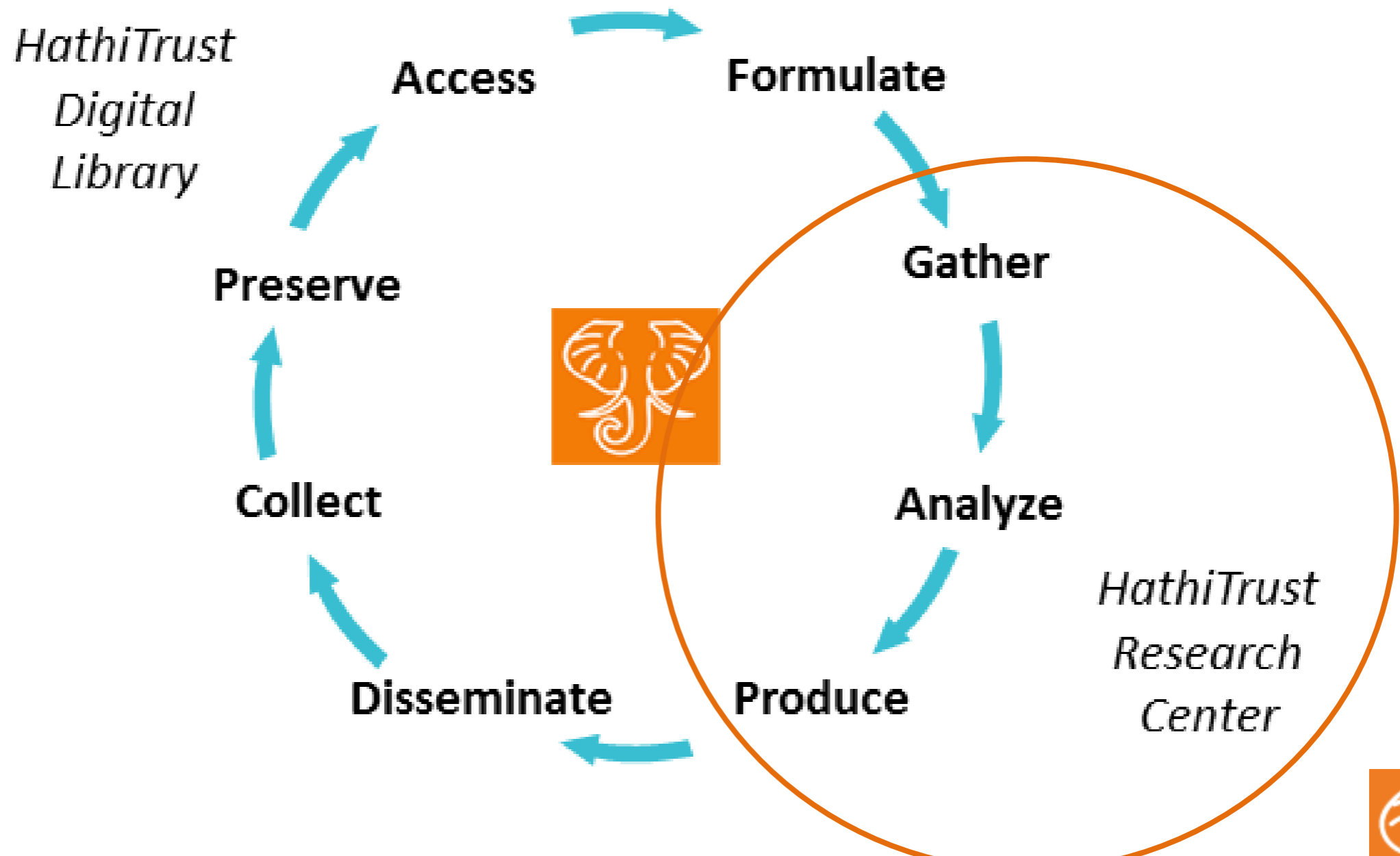
**H T**  
**R C**

# **Digging Deeper, Reaching Further**

Libraries Empowering Users to Mine the  
HathiTrust Digital Library Resources

## The Project

# HathiTrust ecosystem



# HathiTrust Research Center

*How do we enable our digital collections for research use?*

- Jointly led by the University of Illinois at Urbana-Champaign and Indiana University Bloomington
- Facilitates text analysis of HathiTrust Digital Library content
  - Currently accessible: 39% of the collection in public domain
  - Soon: “Non-consumptive research” with in-copyright works
- Focus on large-scale, computational research
- Research & Development
- User Engagement and Assessment

<http://www.hathitrust.org/htrc>



@greenharr

green19@Illinois.edu

# HTRC User Requirements Study: User Personas

## Digital Project Librarian

- Wants flexible, transparent tools
- Role: Research Support staff
- Challenges: Inaccessible data, matching tool to researcher

## Faculty Member

- Wants computational resources
- Role: Experienced Researcher
- Challenges: Collaboration, Finding texts

## Graduate Student

- Wants examples
- Role: New Researcher
- Challenges: Understanding stats, choosing areas of interest

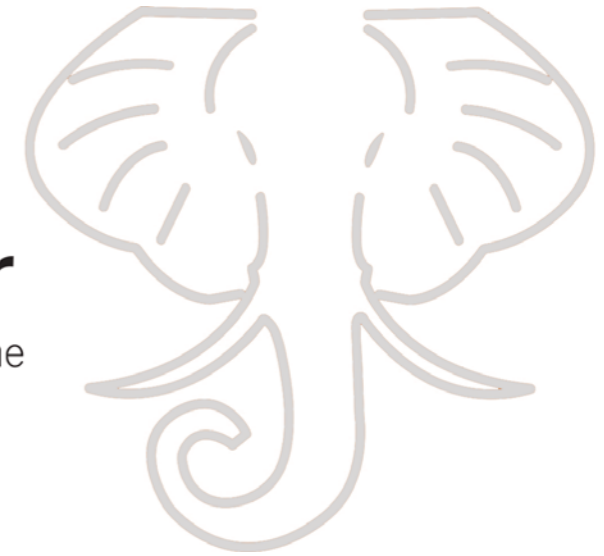
Credits: Alex Kinnaman, Peter Organisciak, Eleanor Dickson

# DDRF: The Goals



**Digging Deeper,  
Reaching Further**

Libraries Empowering Users to Mine the  
HathiTrust Digital Library Resources



- Arm librarians with instructional content and tool skills in digital scholarship and digital humanities;
- Empower librarians to become active research partners on digital projects at their institutions;
- Enable librarians to build foundations for digital scholarship centers and services



## DDRF: the team

- **Principal Investigator:** Harriett Green, University of Illinois at Urbana-Champaign
- **University of Illinois:** Beth Sandore Namachchivaya, Stephen Downie (co-PIs)
- **Indiana University Bloomington:** Angela Courtney (co-PI)
- **Lafayette College:** Neil McElroy, Terese Heidenwolf (co-PI)
- **Northwestern University:** Geoff Morse (co-PI)
- **University of North Carolina:** Amanda Henley (Stewart Varner) (co-PI)

## DDRF: current work

- Developing a suite of curricular resources to train librarians and LIS professionals in approaches and tools frequently used for text analysis research
- 2016-2017: Piloting the workshops at the 5 partner institutions

# DDRF: The Curriculum

Data cleaning  
and  
reformatting

Web Scraping

Build datasets

Command line  
and Python  
scripting

Text analysis  
with HTRC  
tools



# DDRF: To Come...

- Conduct a nationwide “roadshow” of workshops
- Toolkit of open educational resources and curricula for librarians to teach text mining and digital scholarship methods
- Build network of librarians teaching and pursuing text mining methods
- Research study on curriculum development for digital scholarship methods and assessment of strategies for LIS professional development

Looking to the Future:

How Can Librarians  
Prepare for Data-  
Driven Scholarship?



New  
Approaches  
to collection  
curation and  
engagement



@greenharr

green19@Illinois.edu

# Librarians' Engagement in the Research Lifecycle



@greenharr

green19@Illinois.edu

# Building a community of practice

- What infrastructures do libraries need to build or revise for supporting digital scholarship?
- How do we empower librarians to build the skills needed to respond to new research needs?
- How can librarians share skills, materials, and resources for supporting digital scholarship?



# Our Thanks

- This project was made possible by the **Institute for Museum and Library Services**, award #RE-00-15-0112-15
- **HathiTrust Research Center**
- **DDRF Project Partners:** University of Illinois at Urbana-Champaign, Indiana University – Bloomington, Lafayette College, Northwestern University, University of North Carolina at Chapel Hill



@greenharr

green19@Illinois.edu

# Image Credits

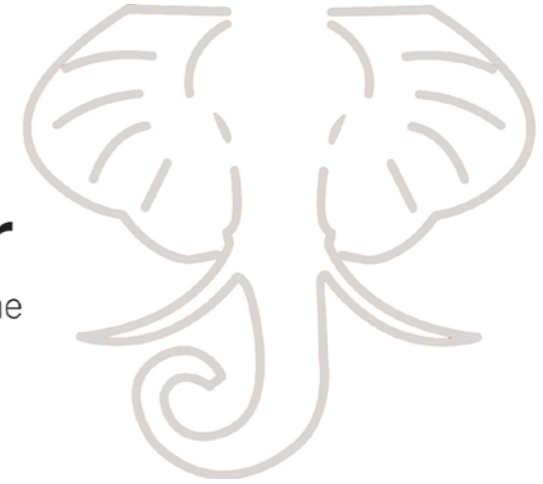
- DDRF project logo, by Leanne Nay
- "Binary," Michael Coghlan, CC-BY-SA 2.0, <https://flic.kr/p/aYEytM>
- Construction, by Matt, CC-BY-NC 2.0, <https://flic.kr/p/7Q5xYn>
- "HK The Arch Overview", Wikimedia Commons, CC-BY 3.0, [https://commons.wikimedia.org/wiki/File:HK\\_The\\_Arch\\_Overview.jpg#/media/File:HK\\_The\\_Arch\\_Overview.jpg](https://commons.wikimedia.org/wiki/File:HK_The_Arch_Overview.jpg#/media/File:HK_The_Arch_Overview.jpg)
- "More Bildsten notebooks" by Jonas Lowgren, on Flickr, CC-BY 2.0, [https://www.flickr.com/photos/jonas\\_lowgren/7406596056](https://www.flickr.com/photos/jonas_lowgren/7406596056)
- Demonicuss, Telescope over Paris, on Flickr, CC-BY-NC-ND 2.0, <https://flic.kr/p/w8URk>
- "38/365 Puzzled," by Mykl Roventine, on Flickr, CC-BY-NC-SA 2.0, <https://www.flickr.com/photos/mykloventine/3261364899>
- "Pipes" by Chris Smart, on Flickr, CC-BY-NC-ND 2.0, <https://flic.kr/p/9Wjju9>

THANK  
YOU!



**Digging Deeper,  
Reaching Further**

Libraries Empowering Users to Mine the  
HathiTrust Digital Library Resources



**Harriett Green**

**green19@illinois.edu | @greenharr**

**The DDRF Project:**

<http://teach.htrc.illinois.edu>

**HathiTrust Research Center:**

<http://www.hathitrust.org/htrc>

**Download the Presentation Handout:**

<https://uofi.box.com/v/DDRFAsData>