Digging into Quilt Data

Dean Rehberger

Director, Matrix

rehberge@msu.edu

Twitter: deanreh

Metaphorical Definition

#DID is quilt making

http://www.pages.drexel.edu/~gkw22/history.html

Metaphorical definition

#DID is quilt making

- Many hands
- Making
- Valued/Devalued
- Reuse/Remix
- Art/Science
- Transformative

DIGGING INTO DATA

to Answer Authorship Related Questions

Digging into Data to Answer Authorship Related Questions seeks to explore authorship studies of visual arts through computational image analyses. Utilizing three datasets of visual works—15th-century manuscripts, 17th and 18th-century maps, and 19th and 21st-century quilts—to investigate what might be revealed about the authors and their artistic lineages by comparing manuscripts, maps, and quilts across four centuries, DID investigates the accuracy and computational scalability of adaptive image analyses when they are applied to diverse collections of image data.

Collaborating Sites

University of Sheffield, UK

University of Illinois at Urbana-Champaign, USA

Michigan State University, USA

Alliance of American Quilts, North Carolina, USA

JISC

Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada

All Divisions and Offices

Office of Digital Humanities

For more information about the Office of Digital Humanities:

odh@neh.gov

The Office of Digital Humanities (ODH) offers grant programs that fund project teams experimenting with digital technologies to develop new methodologies for humanities research, teaching and learning, public engagement, and scholarly communications. ODH funds those studying digital culture from a humanistic perspective and humanists seeking to create digital publications. Another major goal of ODH is to increase capacity of the humanities in applying digital methods

D-Lib Magazine
March 2006

Volume 12 Number 3

ISSN 1082-9873

What Do You Do with a Million Books?

Gregory Crane
Tufts University
<gregory.crane@tufts.edu>

Introduction

The Greek historian Herodotus has the Athenian sage Solon estimate the lifetime of a human being at c. 26,250 days (Herodotus, *The Histories*, 1.32). If we could read a book on each of those days, it would take almost forty lifetimes to work through every volume in a single million book library. The continuous tradition of written European literature that began with the *Iliad* and *Odyssey* in the eighth century BCE is itself little more than a million days old. While libraries that contain more than one million items are not unusual, print libraries never possessed a million books of use to any one reader. The great libraries that took shape in the nineteenth and twentieth centuries were meta-structures, whose catalogues and finding aids allowed readers to create their own customized collections, building on the fixed classification schemes and disciplinary structures that took shape in the nineteenth century.

The digital libraries of the early twenty-first century can be searched and their contents transmitted around the world. They can contain time-based media, images, quantitative data, and a far richer array of content than print, with visualization technologies blurring the boundaries between library and museum. But our digital libraries remain filled with digital incunabula – digital objects whose form remains firmly rooted in traditions of print, with HTML and PDF largely mimicking the limitations of their print predecessors.

How Not to Read a Million Books

by Tanya Clement, Sara Steger, John Unsworth, Kirsten Uszkalo

October, 2008

[Figure 1] First of all, where does the trope of "a million books" come from? It originates, as far as I know, with the Universal Library and its Million Books Project, which began in 2001. The Universal Library is directed by Raj Reddy, professor and former Dean of Computer Science at Carnegie Mellon University; the million books project (funded by NSF and others) was a kind of very large pilot, aimed at digitizing a million books ("less than 1% of all books in all languages ever published"[1]), beginning with partners in India and later expanding to China and Egypt. The "million book" goal was accomplished in 2007, by which time it had been eclipsed by some large commercial projects, including most notably Google Print (now known as Google Book Search), which had begun in secret in 2002 and was unveiled at the Frankfurt Book Fair in October 2004, and which had Harvard's library as one of its initial partners. Google Books aims to scan as many as 30 million books, a number equal to all the titles in WorldCat, and for all we know, they are already about halfway there.[2] Libraries and others have been digitizing books for years, but these massive digitization projects really changed the landscape, and they raised the question "What do you do with a million books?"—a question first asked, I think, by Greg Crane, in D-Lib Magazine, in March of 2006.[3] My answer to that question is that whatever you do, you don't read them, because you can't.

[Figure 2] As Franco Moretti points out, in Graphs, Maps, Trees, we focus on a "minimal fraction of the literary field":

> . . . a canon of two hundred novels, for instance, sounds very large for nineteenth-century Britain (and is much larger than the current one), but is still less than one per cent of the novels that were actually published: twenty thousand, thirty, more, no one really knows —and close reading won't help here, a novel a day every day of the year would take a century or so... And it's not even a matter of time, but of method: a field this large cannot be understood by stitching together separate bits of knowledge about individual cases, because it isn't a sum of individual cases: it's a collective system, that should be grasped as such, as a whole."[4]

I think that what Moretti calls "the quantitative approach to literature" acquires a special importance when millions of books are equally at your fingertips, all eagerly responding to your Google Book Search: you can no longer as easily ignore the books you don't know, nor can you grasp the collective systems they make up without some new strategy—a strategy for not reading.

[Figure 3] Martin Mueller is my collaborator and co-PI on the MONK project, and professor of classics and English at Northwestern University. Martin is fond of citing this poem about not reading, called "The Spectacles":

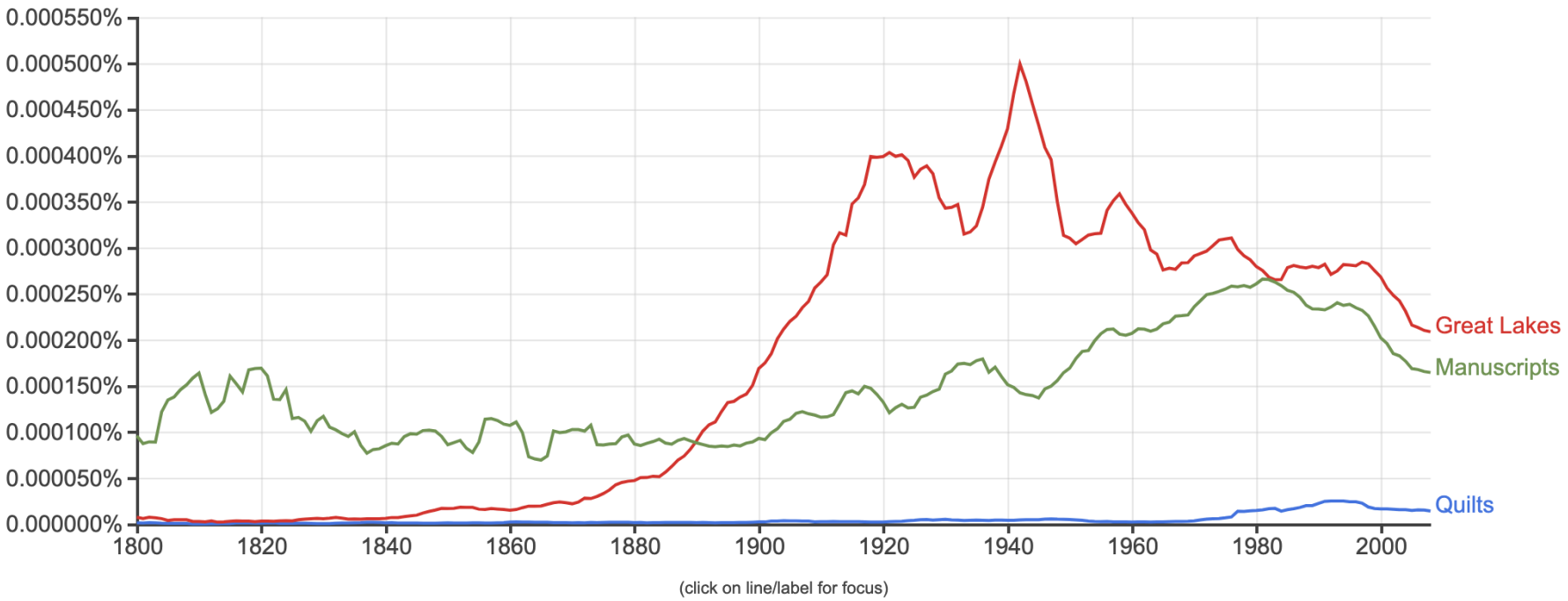
Google Books Ngram Viewer

Graph these comma-separated phrases: Quilts,Great Lakes,Manuscripts case-insensitive

between 1800 and 2008 from the corpus English (2012) with smoothing of 3. Search lots of books

Tweet

Embed Chart



(click on line/label for focus)

Search in Google Books:

1800 - 1886	1887 - 1993	1994 - 1997	1998 - 2003	2004 - 2008	quilts	English
1800 - 1904	1905 - 1937	1938 - 1945	1946 - 1994	1995 - 2008	great lakes	English
1800 - 1821	1822 - 1968	1969 - 1980	1981 - 1992	1993 - 2008	manuscripts	English

- DIGGING INTO THE ENLIGHTENMENT: MAPPING THE REPUBLIC OF LETTERS
- HARVESTING SPEECH DATASETS FOR LINGUISTIC RESEARCH ON THE WEB
- MINING A YEAR OF SPEECH
- RAILROADS AND THE MAKING OF MODERN AMERICA—TOOLS FOR SPATIO-TEMPORAL CORRELATION, ANALYSIS, AND VISUALIZATION
- STRUCTURAL ANALYSIS OF LARGE AMOUNTS OF MUSIC INFORMATION
- TOWARDS DYNAMIC VARIORUM EDITIONS
- USING ZOTERO AND TAPOR ON THE OLD BAILEY PROCEEDINGS: DATA MINING WITH CRIMINAL INTENT

THE QUILT INDEX

Search Quilt Records
keyword/full-record go
Advanced Search

Select Language

Home | About | Collections | Search | Browse | Essays | Galleries | Lesson Plans | FAQ | Contact | Wiki | Blog | Donate

The Sun Sets on Sunbonnet Sue
The Seamster's Union (Local #500)
c. 1979
Michigan State University Museum

Welcome to the Quilt Index

Did you know that know that the Quilt Index now has over 90,000 freely-accessible and searchable records of quilts (from public and private collections, including over 250 museums) and hundreds of related stories, photos, publications, and ephemera?

We hope that you, as a contributor to/and or user of the Index, or perhaps as someone who champions women artists, textile history, regional quilt documentation projects, or other facets of the Index, will take a moment and consider how you might support the Index in your charitable giving. All you need to do is go to Donate Now to express your support through a tax-deductible eligible donation to the Index.

I and the rest of the QI leadership team appreciate your support,

Marsha MacDowell, Ph.D.
Director, The Quilt Index (www.quiltindex.org) Curator and Professor, Michigan State University

More Quilt Index News
Celebrating App Give!

What's New

QI Facebook:

Features

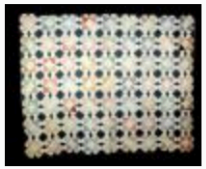
iPhone/iPad app
Android App
Take QI to Go

Quilt Index Wiki
Museums, listings, add your own

Collections
Apply with your collection

Donate
Help QI sustain and grow

Wild Card



The Kensington Block

Reviews

DIGGING INTO DATA

to Answer Authorship Related Questions

Digging into Data to Answer Authorship Related Questions seeks to explore authorship studies of visual arts through computational image analyses. Utilizing three datasets of visual works—15th-century manuscripts, 17th and 18th-century maps, and 19th and 21st-century quilts—to investigate what might be revealed about the authors and their artistic lineages by comparing manuscripts, maps, and quilts across four centuries, DID investigates the accuracy and computational scalability of adaptive image analyses when they are applied to diverse collections of image data.

Collaborating Sites

University of Sheffield, UK

University of Illinois at Urbana-Champaign, USA

Michigan State University, USA

Alliance of American Quilts, North Carolina, USA

JISC

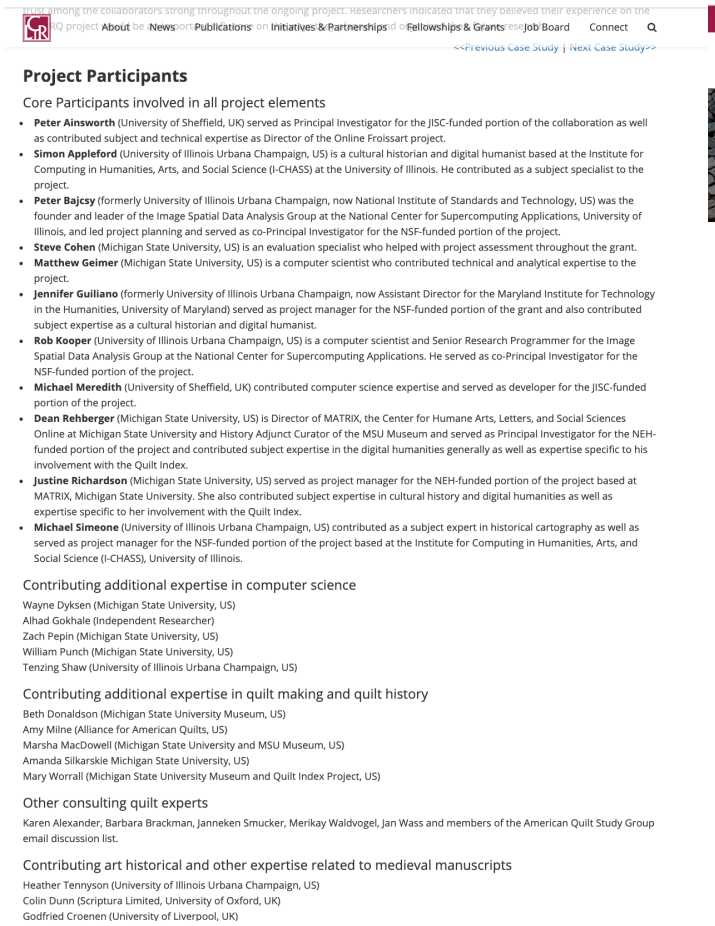
Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada

CLIR Report

- https://www.clir.org/pubs/reports/pub151/case-studies/did-arq/



Project Participants

Core Participants involved in all project elements

- **Peter Ainsworth** (University of Sheffield, UK) served as Principal Investigator for the JISC-funded portion of the collaboration as well as contributed subject and technical expertise as Director of the Online Froissart project.
- **Simon Appleford** (University of Illinois Urbana Champaign, US) is a cultural historian and digital humanist based at the Institute for Computing in Humanities, Arts, and Social Science (I-CHASS) at the University of Illinois. He contributed as a subject specialist to the project.
- **Peter Bajcsy** (formerly University of Illinois Urbana Champaign, now National Institute of Standards and Technology, US) was the founder and leader of the Image Spatial Data Analysis Group at the National Center for Supercomputing Applications, University of Illinois, and led project planning and served as co-Principal Investigator for the NSF-funded portion of the project.
- **Steve Cohen** (Michigan State University, US) is an evaluation specialist who helped with project assessment throughout the grant.
- **Matthew Geimer** (Michigan State University, US) is a computer scientist who contributed technical and analytical expertise to the project.
- **Jennifer Guiliano** (formerly University of Illinois Urbana Champaign, now Assistant Director for the Maryland Institute for Technology in the Humanities, University of Maryland) served as project manager for the NSF-funded portion of the grant and also contributed subject expertise as a cultural historian and digital humanist.
- **Rob Kooper** (University of Illinois Urbana Champaign, US) is a computer scientist and Senior Research Programmer for the Image Spatial Data Analysis Group at the National Center for Supercomputing Applications. He served as co-Principal Investigator for the NSF-funded portion of the project.
- **Michael Meredith** (University of Sheffield, UK) contributed computer science expertise and served as developer for the JISC-funded portion of the project.
- **Dean Rehberger** (Michigan State University, US) is Director of MATRIX, the Center for Humane Arts, Letters, and Social Sciences Online at Michigan State University and History Adjunct Curator of the MSU Museum and served as Principal Investigator for the NEH-funded portion of the project and contributed subject expertise in the digital humanities generally as well as expertise specific to his involvement with the Quilt Index.
- **Justine Richardson** (Michigan State University, US) served as project manager for the NEH-funded portion of the project based at MATRIX, Michigan State University. She also contributed subject expertise in cultural history and digital humanities as well as expertise specific to her involvement with the Quilt Index.
- **Michael Simeone** (University of Illinois Urbana Champaign, US) contributed as a subject expert in historical cartography as well as served as project manager for the NSF-funded portion of the project based at the Institute for Computing in Humanities, Arts, and Social Science (I-CHASS), University of Illinois.

Contributing additional expertise in computer science

Wayne Dyksen (Michigan State University, US)
Alhad Gokhale (Independent Researcher)
Zach Pepin (Michigan State University, US)
William Punch (Michigan State University, US)
Tenzing Shaw (University of Illinois Urbana Champaign, US)

Contributing additional expertise in quilt making and quilt history

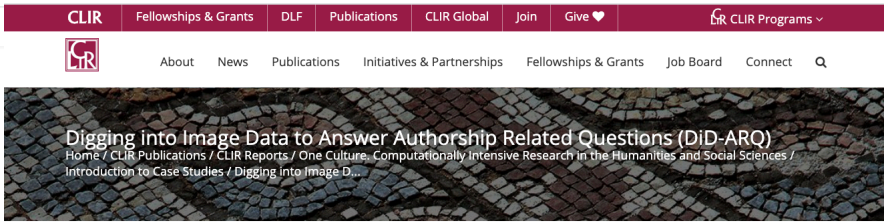
Beth Donaldson (Michigan State University Museum, US)
Amy Milne (Alliance for American Quilts, US)
Marsha MacDowell (Michigan State University and MSU Museum, US)
Amanda Silkarskie Michigan State University, US)
Mary Worrall (Michigan State University Museum and Quilt Index Project, US)

Other consulting quilt experts

Karen Alexander, Barbara Brackman, Janneken Smucker, Merikay Waldvogel, Jan Wass and members of the American Quilt Study Group email discussion list.

Contributing art historical and other expertise related to medieval manuscripts

Heather Tennyson (University of Illinois Urbana Champaign, US)
Colin Dunn (Scriptura Limited, University of Oxford, UK)
Godfried Croenen (University of Liverpool, UK)



Digging into Image Data to Answer Authorship Related Questions (DiD-ARQ)

<<Previous Case Study | Next Case Study>>

Jump to: Project Participants | Project Outcomes

Project Outcomes

Project website

Peer-reviewed publications

Ainsworth, Peter and Meredith, Michael. "Breaching the Strongroom: a Pervasive Informatics Approach to Working with Medieval Manuscripts," *Proceedings of the KMIS 2011 International Conference on Knowledge Management and Information Sharing*, Joachim Felipe and Kecheng Liu, eds. 2011: Setúbal, Portugal. pp. 264-71. ISBN 978-989-8425-81-2.

Ainsworth, Peter. "Digital Attraction: from the real to the virtual in manuscript studies," *Forum : University of Edinburgh Postgraduate Journal of Culture & The Arts*, issue on Authenticity (May 2011), 14 p. http://www.forumjournal.org/site/issue/12/peter-ainsworth

Simeone, Michael, Jennifer Guiliano, Rob Kooper, and Peter Bajcsy. "Digging into data using new collaborative infrastructures supporting humanities-based computer science research." *First Monday* 16.5 (2 May 2011).

Presentations and posters

Ainsworth, Peter, Presentation of the DID and Online Froissart projects, seminar on "Temporality and Value at the Intersection of the Arts and Humanities," University of Southampton, UK, 12 April 2012.

Bajcsy, Peter. Presentation at Wolfram Technology conference in IL; October 13, 2010, http://www.wolfram.com/events/techconf2010/speakers.html.

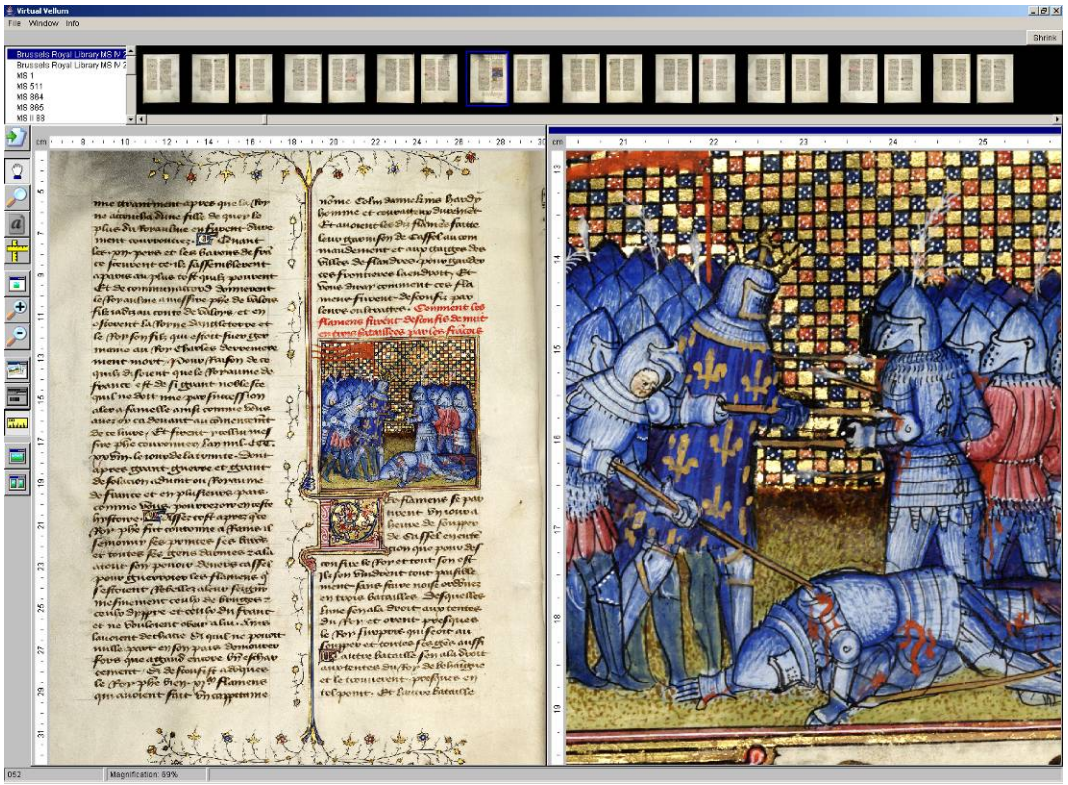
—. Presentations at Imaging at Illinois workshop in IL, October 14-15, 2010, http://www.imaging.beckman.illinois.edu/imaging2010/.

—. Presentation at the Gordon Challenge in Data-Intensive Discovery conference in CA, October 26-29, 2010, http://www.sdsc.edu/gordongrandchallenge/.

—. Presentation at the Supercomputing Conference 2010, NSF funded panel on Grand Challenges in Humanities, Arts and Social Sciences, New Orleans, Louisiana , November 14-16, 2010; http://sc10.supercomputing.org/schedule/event_detail.php?evid=stpan108.

—, Rob Kooper, Luigi Marini, Tenzing Shaw, Jennifer Guiliano, Anne D. Hedeman, Robert Markley, Michael Simeone, Natalie Hanson, "Supporting Scientific Discoveries to Answer Art Authorship Related Questions Across Diverse Disciplines and Geographically Distributed Resources," Microsoft Research eScience Workshop, October 11–13 in Berkeley, CA, http://research.microsoft.com/en-us/events/escience2010/default.aspx (accepted as poster August 2010)

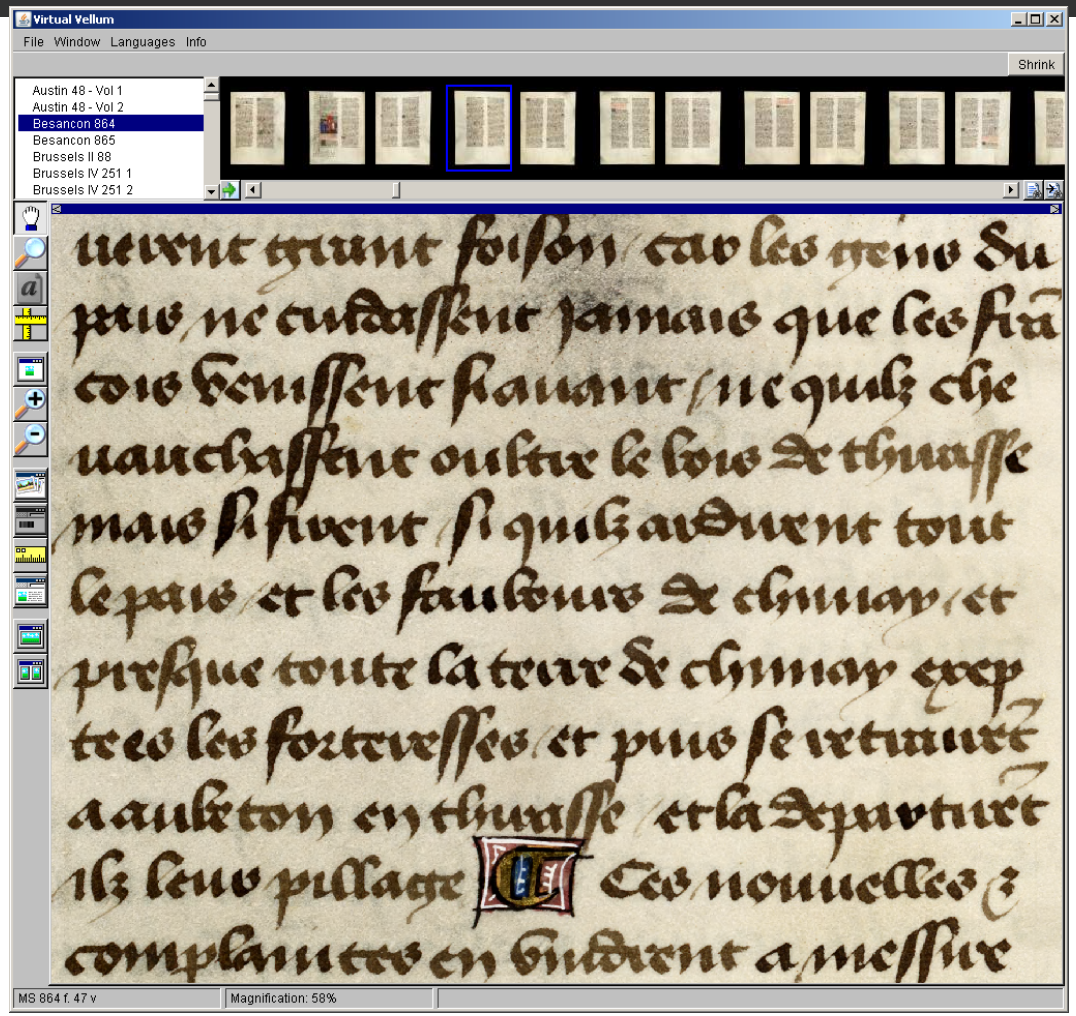
Froissart's Chronicles



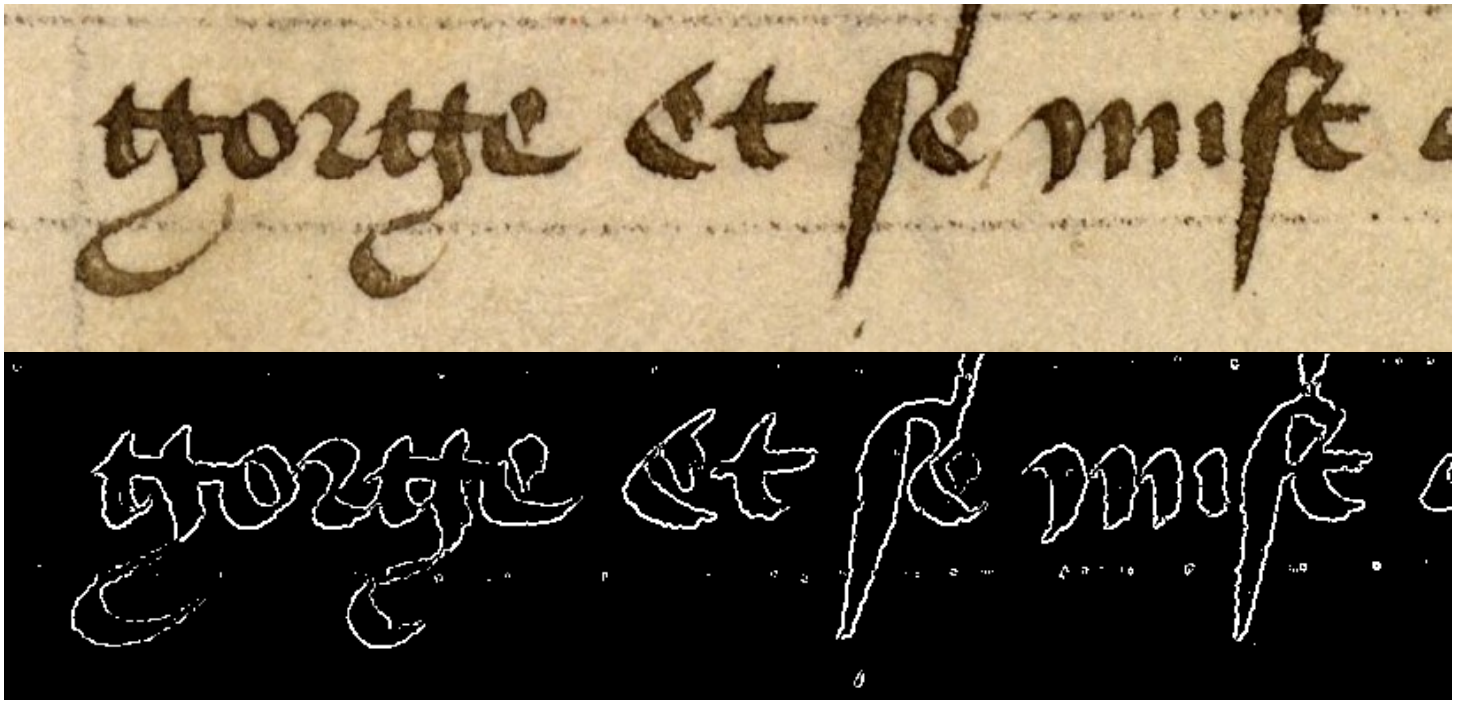
Images © Brussels Royal Library and Scriptura Ltd

University of Sheffield, UK

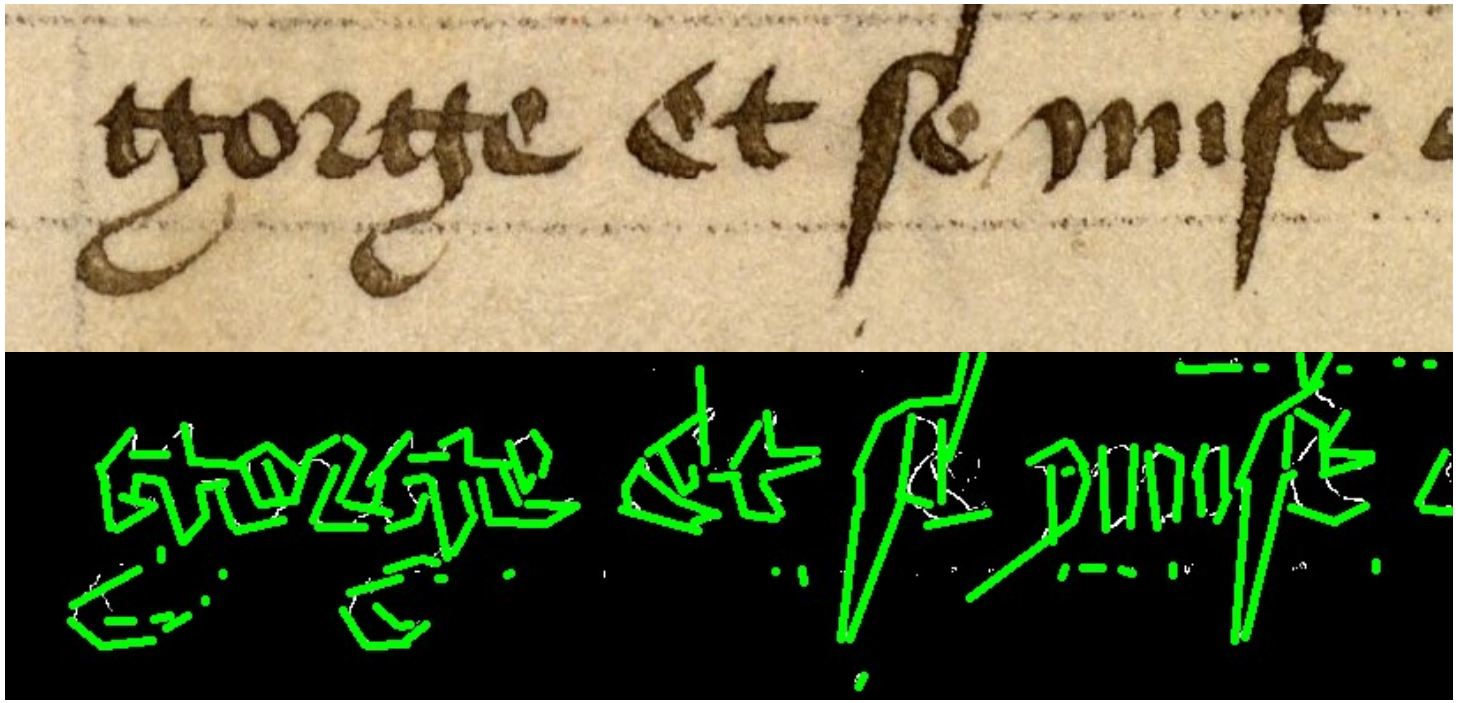
Froissart's Chronicles



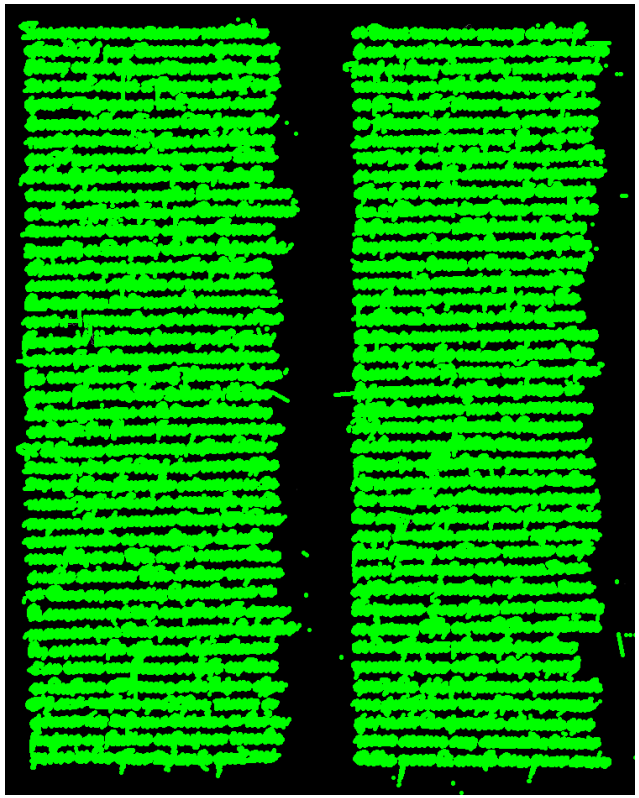
Extracting a Digital Fingerprint using Polygonal Models and Shape Recognition

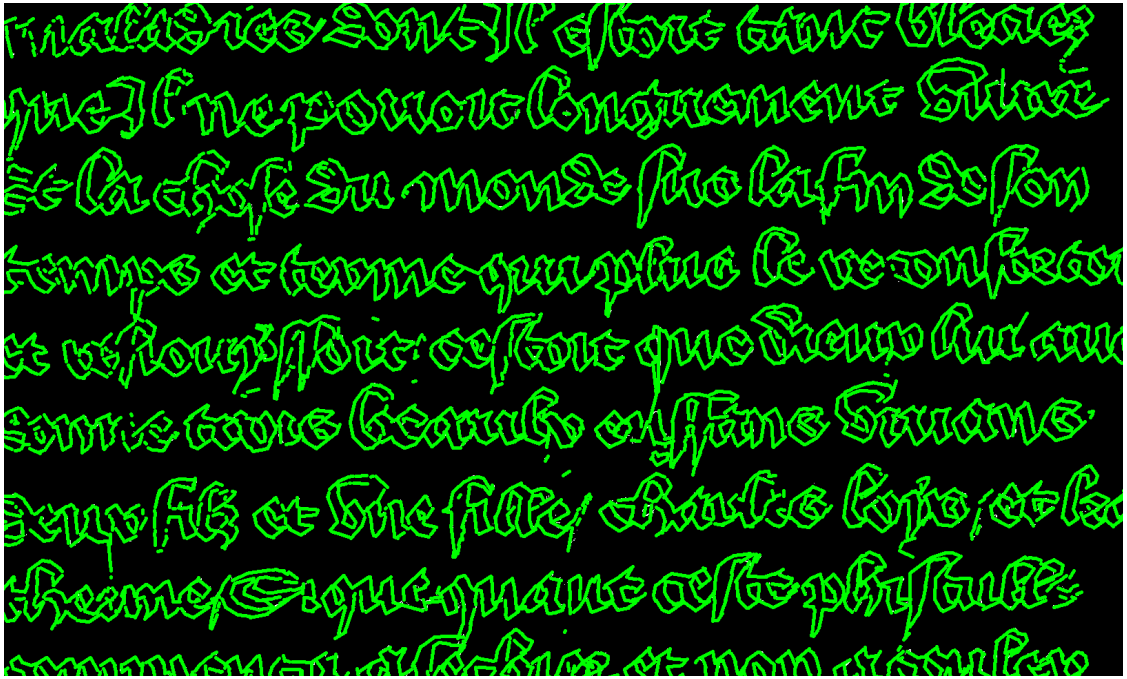


Extracting a Digital Fingerprint using Polygonal Models and Shape Recognition



Extracting a Digital Fingerprint using Polygonal Models and Shape Recognition





Extracting a Digital Fingerprint using Polygonal Models and Shape Recognition

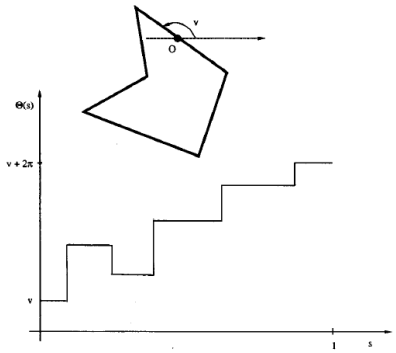
1) Apply Sobel edge detection to source image – from the Image2Learn library
2) Fit line segments to edge map data using EM algorithm – designed to run on multiple cores
3) **Apply shape recognition algorithms to polygonal models to identify similar letters, words, symbols and patterns**

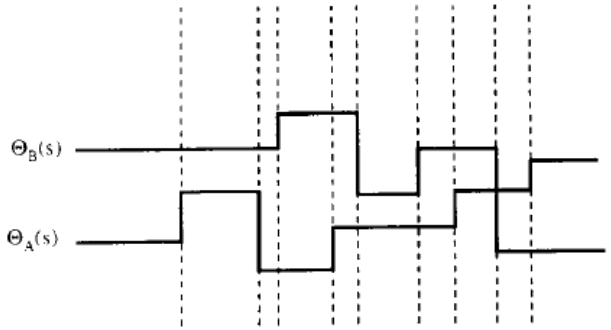
Turning Angle Function Comparison

Comparison is invariant to scale, rotation and starting reference point

Shape represented by a series of turning angles and lengths made between itself and previous segment

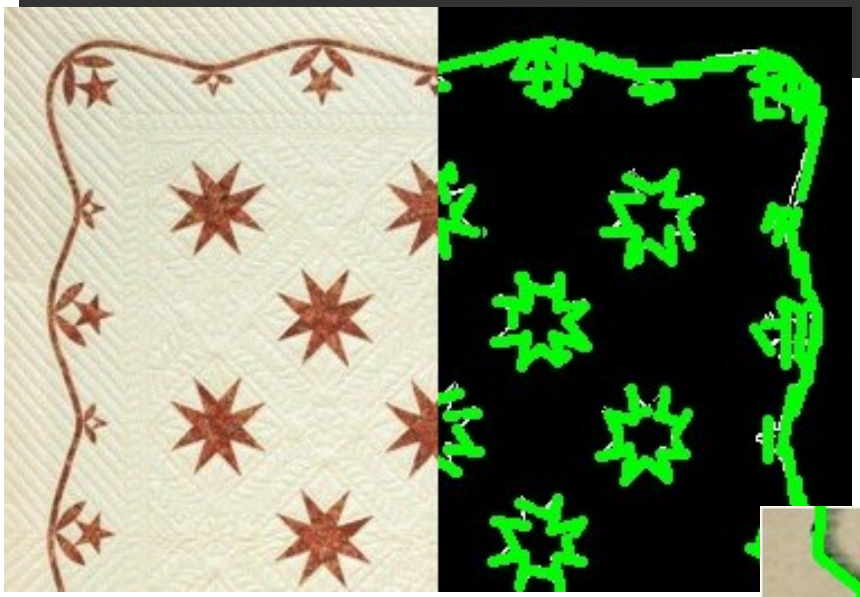
Find the minimum distance between shapes by comparing their turning angle functions with respect to vertical and horizontal shifts (starting reference point and rotation respectively)



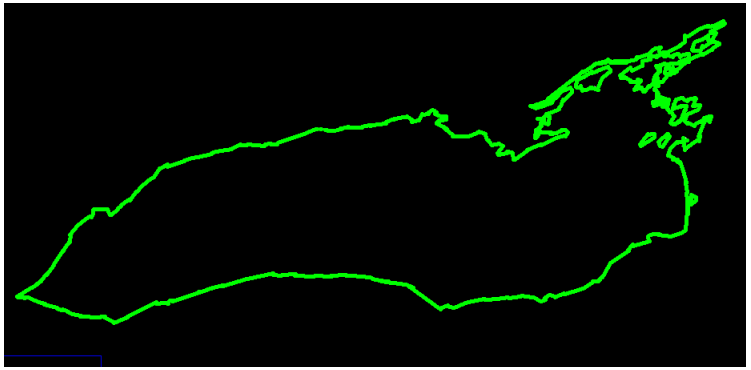


Illustrations from Arkin, E.M., Chew, L.P., Huttenlocher, D.P., Kedem, K., Mitchell, J.S.B., "An Efficiently Computable Metric for Comparing Polygonal Shapes," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13(3), 1991

Applying algorithms across the different collections



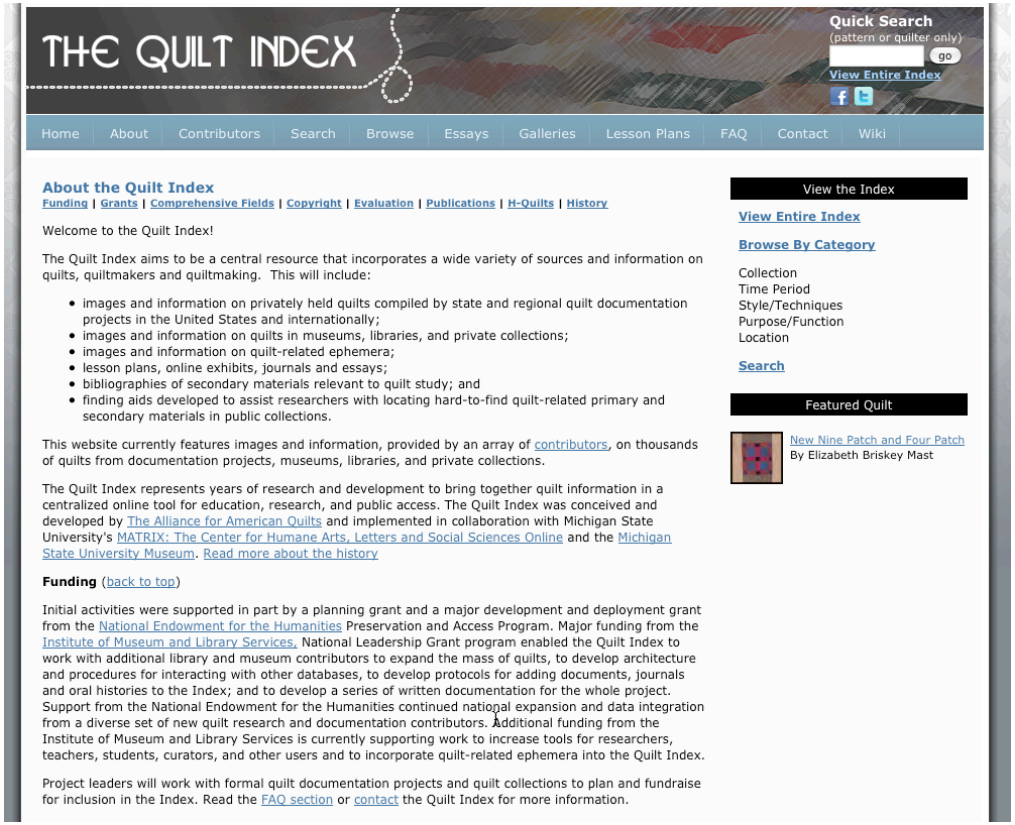
Applying algorithms across the different collections



Polygon fitting

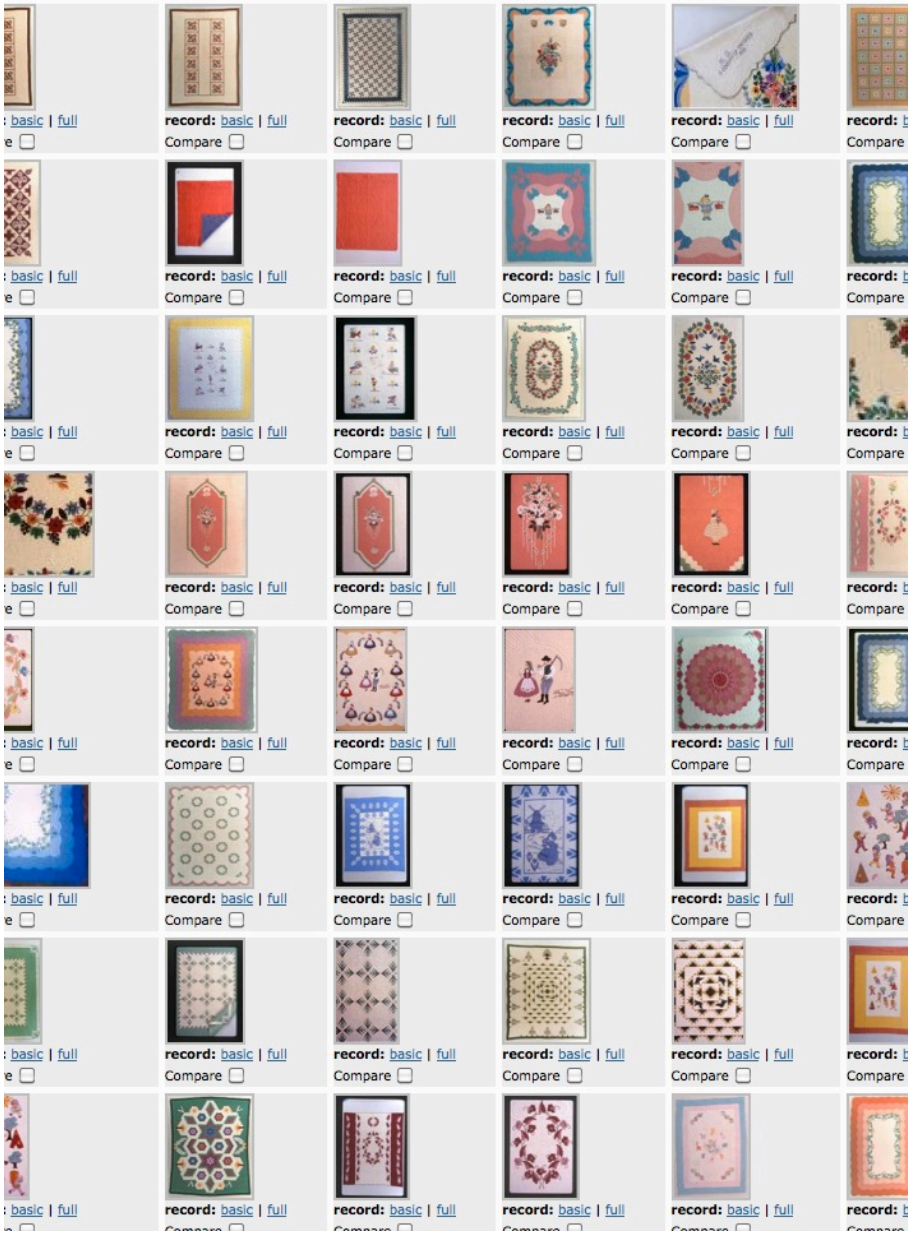
DiD-ARQ: 19th and 20th Century Quilts

- Quilt Index: Online database and management system for thematic collections that are housed in repositories of diverse size, focus, and mission.
- A digital library for researching and teaching.



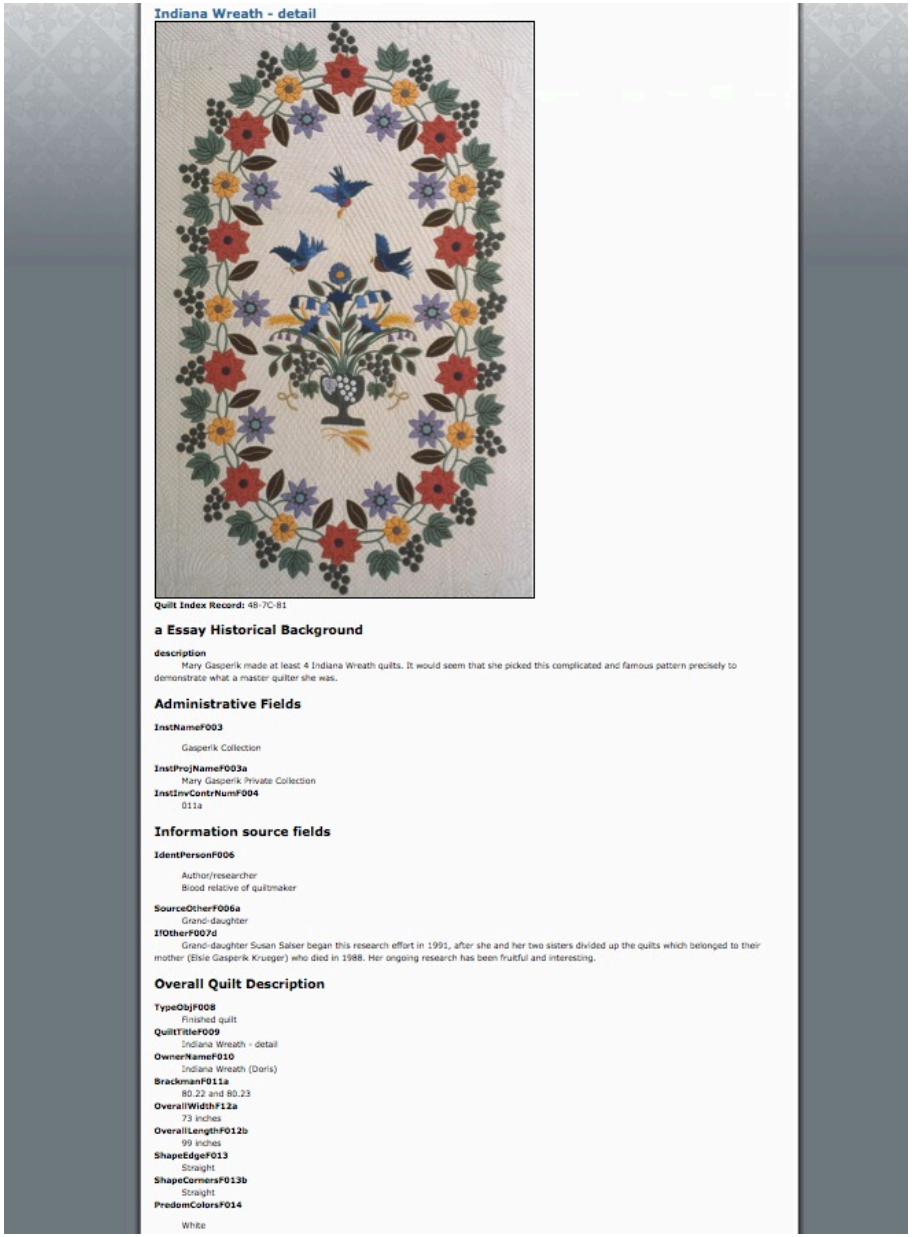
Starting Points

- > 50,000 Images



Starting Points

- Good Testbed
- Rich Documentation
- 129 Metadata Fields



Indiana Wreath - detail

Quilt Index Record: 48-7C-81

a Essay Historical Background

description
Mary Gasperik made at least 4 Indiana Wreath quilts. It would seem that she picked this complicated and famous pattern precisely to demonstrate what a master quilter she was.

Administrative Fields

InstNameF003
Gasperik Collection

InstProjNameF003a
Mary Gasperik Private Collection

InstInvContrNumF004
011a

Information source fields

IdentPersonF006
Author/researcher
Blood relative of quiltmaker

SourceOtherF006a
Grand-daughter

IfOtherF007d
Grand-daughter Susan Salser began this research effort in 1991, after she and her two sisters divided up the quilts which belonged to their mother (Elsie Gasperik Krueger) who died in 1988. Her ongoing research has been fruitful and interesting.

Overall Quilt Description

TypeObjF008
Finished quilt

QuiltTitleF009
Indiana Wreath - detail

OwnerNameF010
Indiana Wreath (Doris)

BrackmanF011a
80.22 and 80.23

OverallWidthF12a
73 inches

OverallLengthF012b
99 inches

ShapeEdgeF013
Straight

ShapeCornersF013b
Straight

PredomColorsF014
White

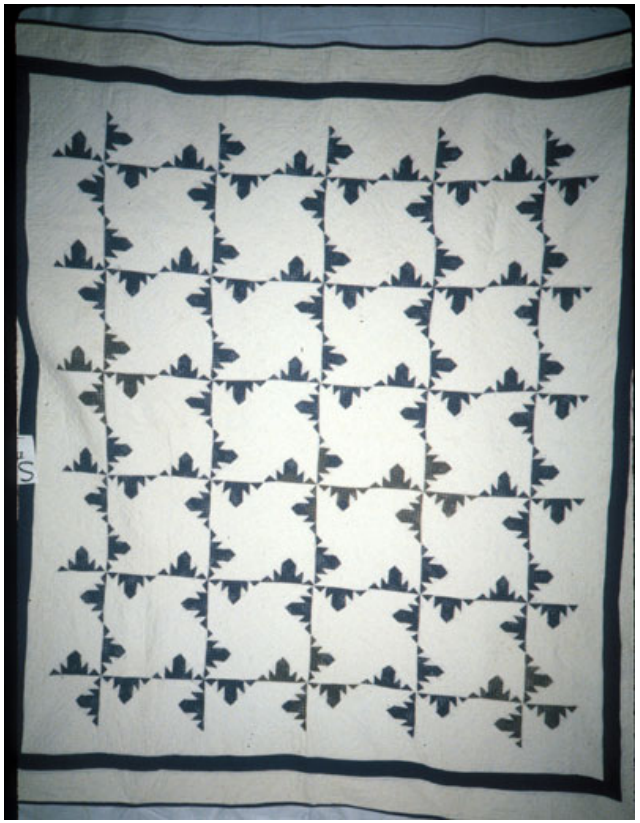
Challenges

- Images not standardized
- Few color calibrated
- Metadata varies in quality and vocabulary

Authorship

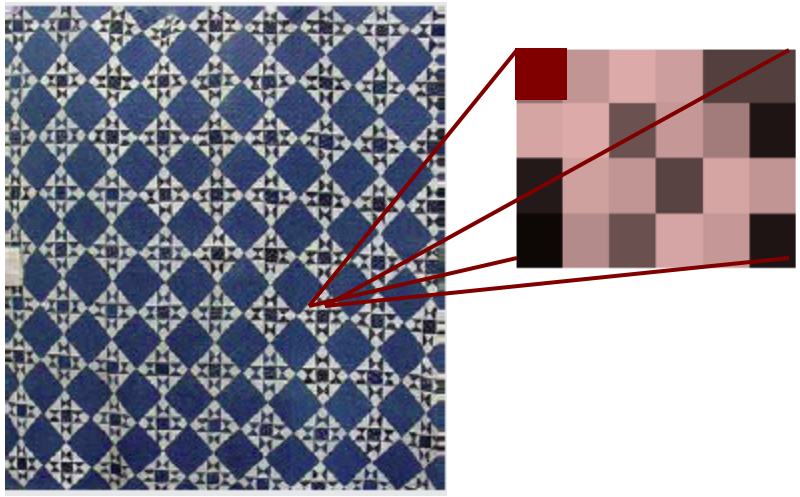
- Corporate
- Community-based
- Individual
- Handcrafted
- Machine-crafted

Indigo Quilts



Indigo Quilts

Convert image to HSL representation and look at it one pixel at a time



Current Ideal

vs = .001233

Repeat for entire picture, summing the differences for each pixel comparison

Return the summed differences and associate with that quilt image

Rank results based on values for quilts; smaller number means more like pure indigo

Quilt Analysis Algorithms - Approach

- **Project was interesting to external people**
  - **Alhaad Gokhale from India was interested in working on the quilt project under the Google Summer of Code funding umbrella at NCSA. He was supervised by Peter Bajcsy to complete his BS thesis project while addressing the crazy vs. non-crazy classification problem.**
  - **Resulting algorithm delivered to MSU**

- **Crazy Quilts**
  - Segment quilt based on color
  - Assign each segment a class using perimeter and area metrics
  - Merge similar regions
  - Generate signatures for distinct regions
  - Classify using SVM

Crazy Quilt

Crazy Quilt

Segment the image into regions. The regions are not squares; below is for illustration only

Compare regions and merge if similar

vs

< 1, 3.4, -10, 15, 4… >

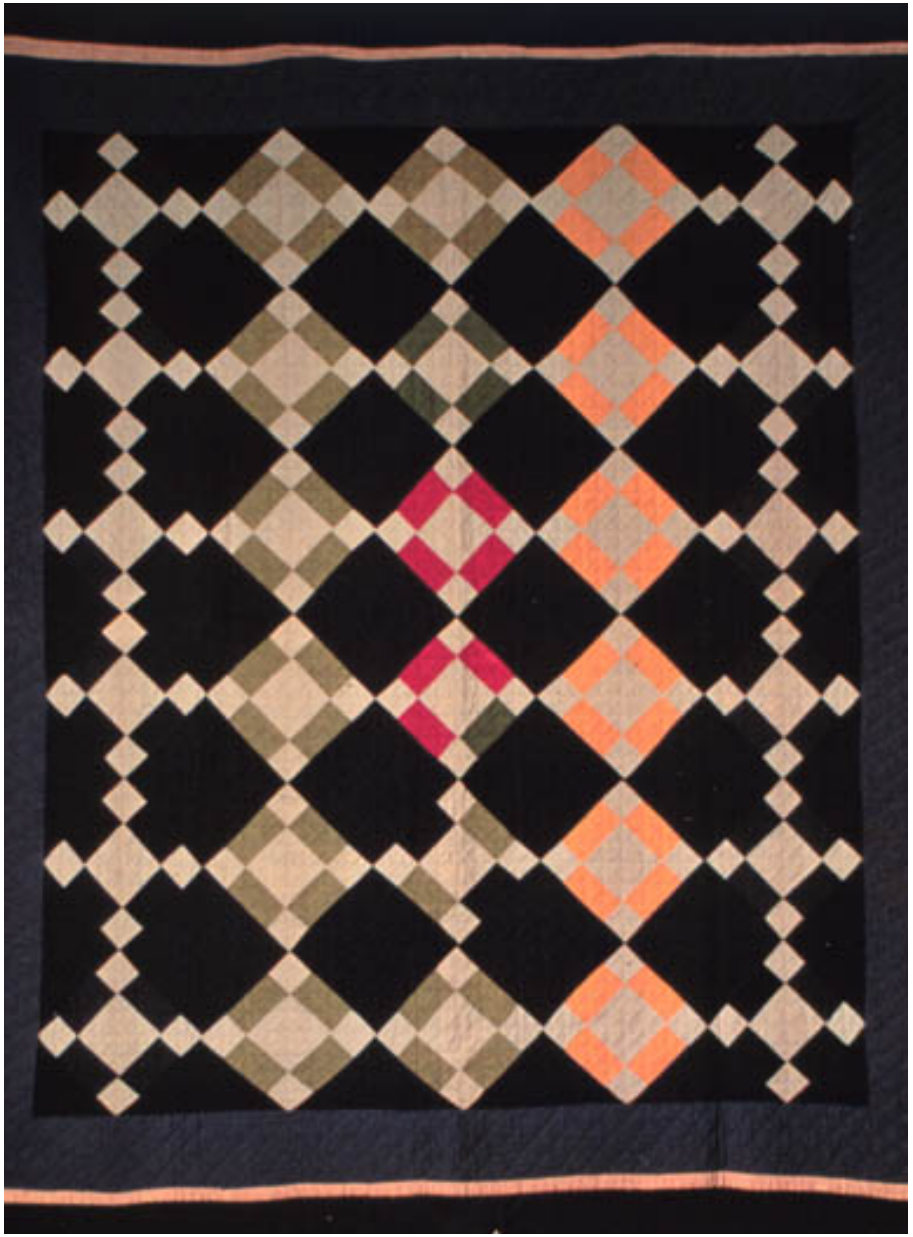
Compute signature for quilt using regions as metrics. Assign this signature to the quilt.

Feed signatures into Trained SVM Classifier

SVM Classifier → YES

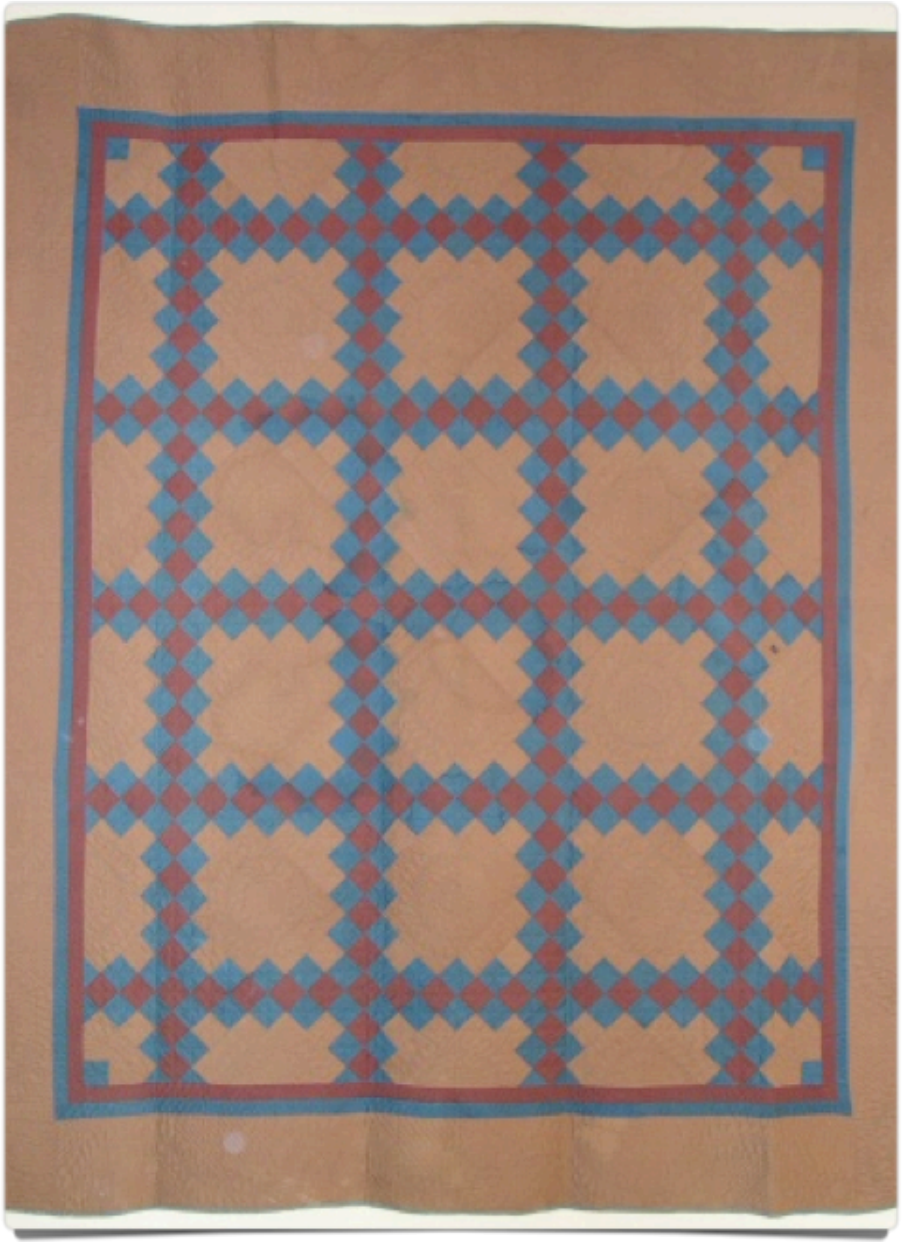
Get Yes/No result from SVM

- Amish Quilts

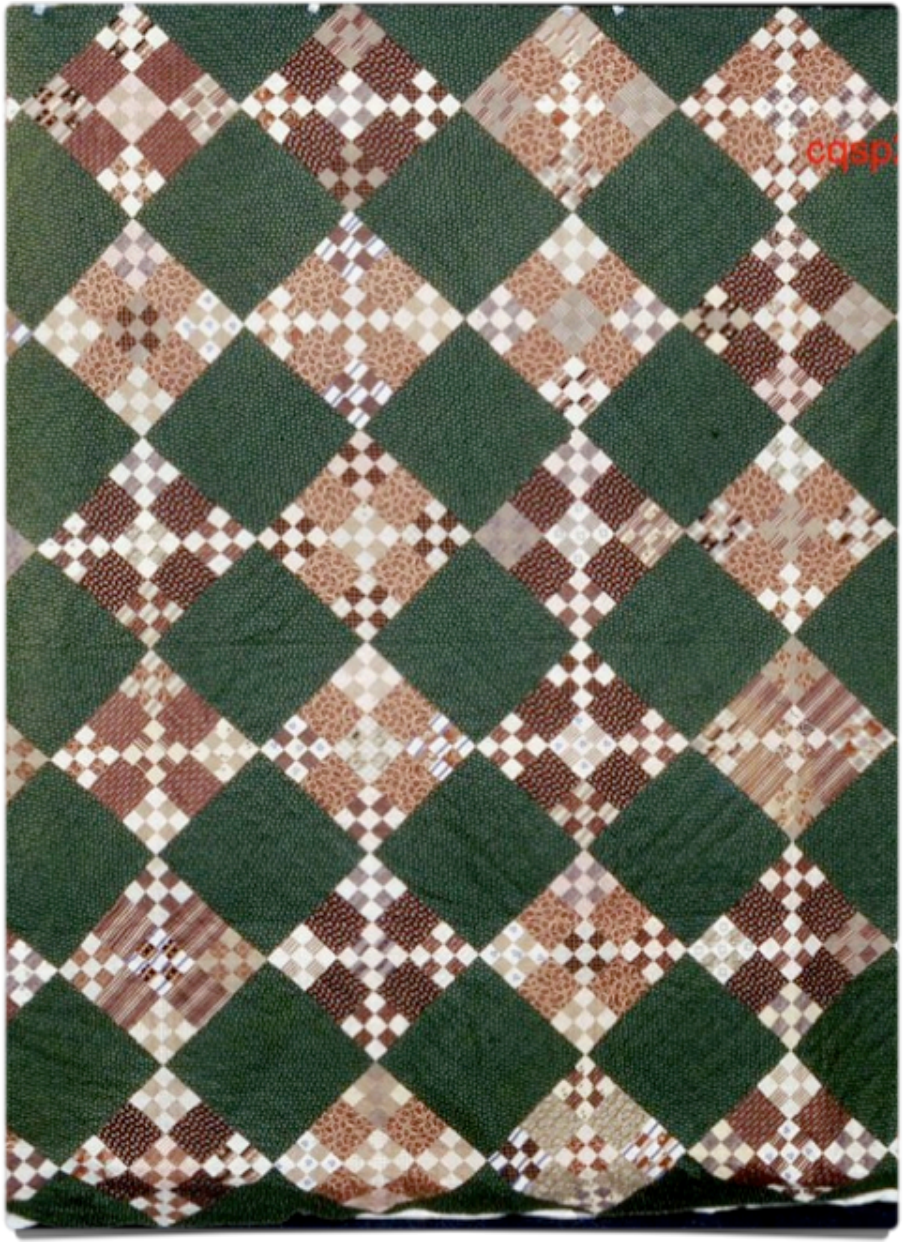


- Amish Quilt
- Nine Patch
- c.1850

- Amish Quilt
- Chained Nine Patch
- 1890



- Amish Quilt
- Nine Patch Variation
- c.1876-1900



- Amish Quilt
- Bear Paw
- 1900

- Amish Quilt
- Bear Paw
- 1910

- Amish Quilts
- Bear Paw
- 1910


QUILT
INDEX

Slave Biographies

SLAVE BIOGRAPHIES:
ATLANTIC DATABASE NETWORK

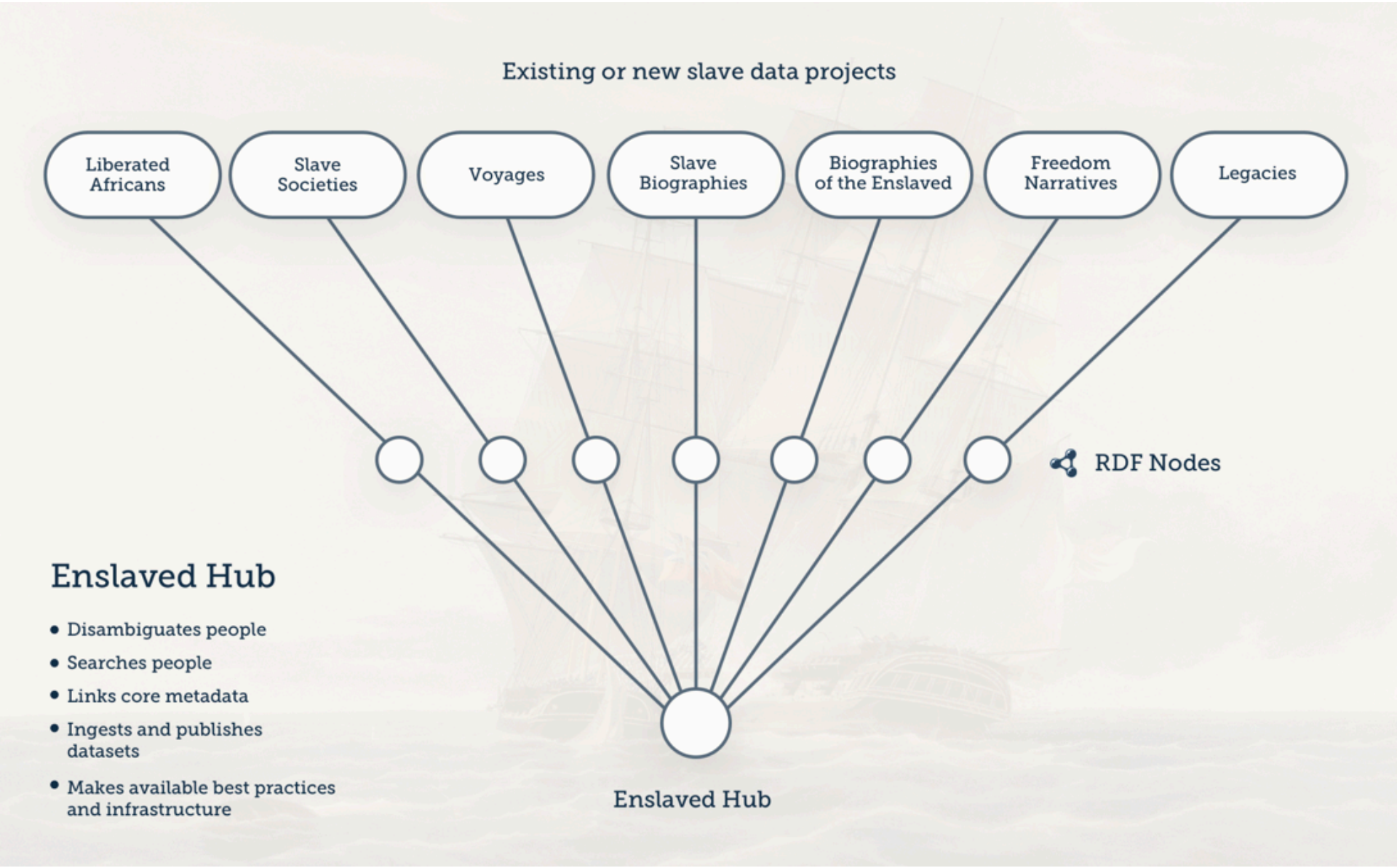
HOME ABOUT THE PROJECT DATA BLOG CONTACT ACKNOWLEDGEMENTS

Slave Biographies: The Atlantic Database Network is a database of information on the identities of enslaved people in the Atlantic World.

Slave Biographies collates data on individual slaves meticulously collected by researchers over the past 20 years. Reviewed by an Advisory Board of experts, datasets include among other information the names, ethnicities, skills, occupations, and illnesses of slaves. The collections reveal much about slave life in the New World and about African slaves' lives in parts of the Old World.

The initial phase of *Slave Biographies* will establish a best practice methodology for how to structure the database to handle datasets containing descriptions of slaves. Phase I will culminate in a freely

show a video or image.



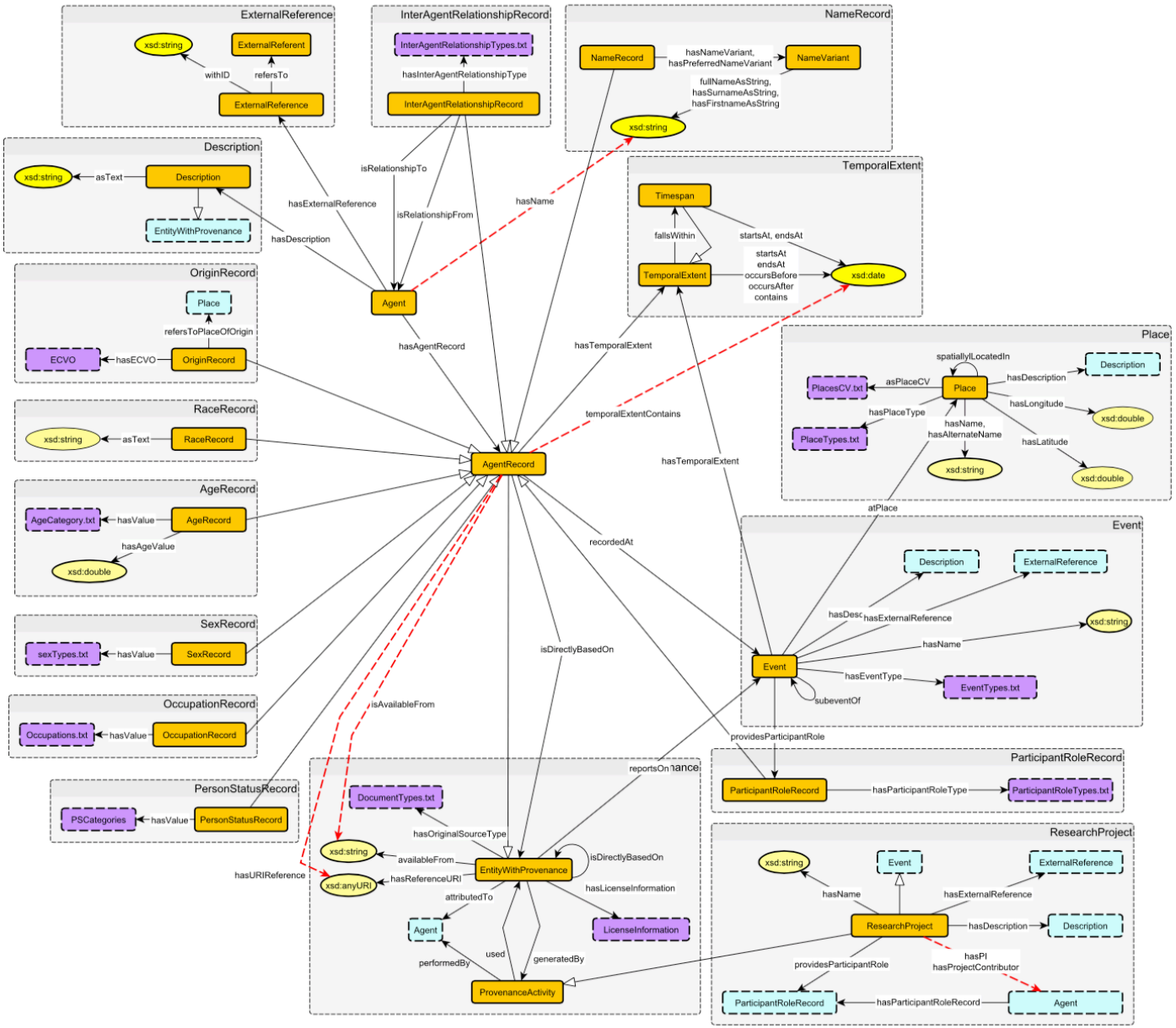
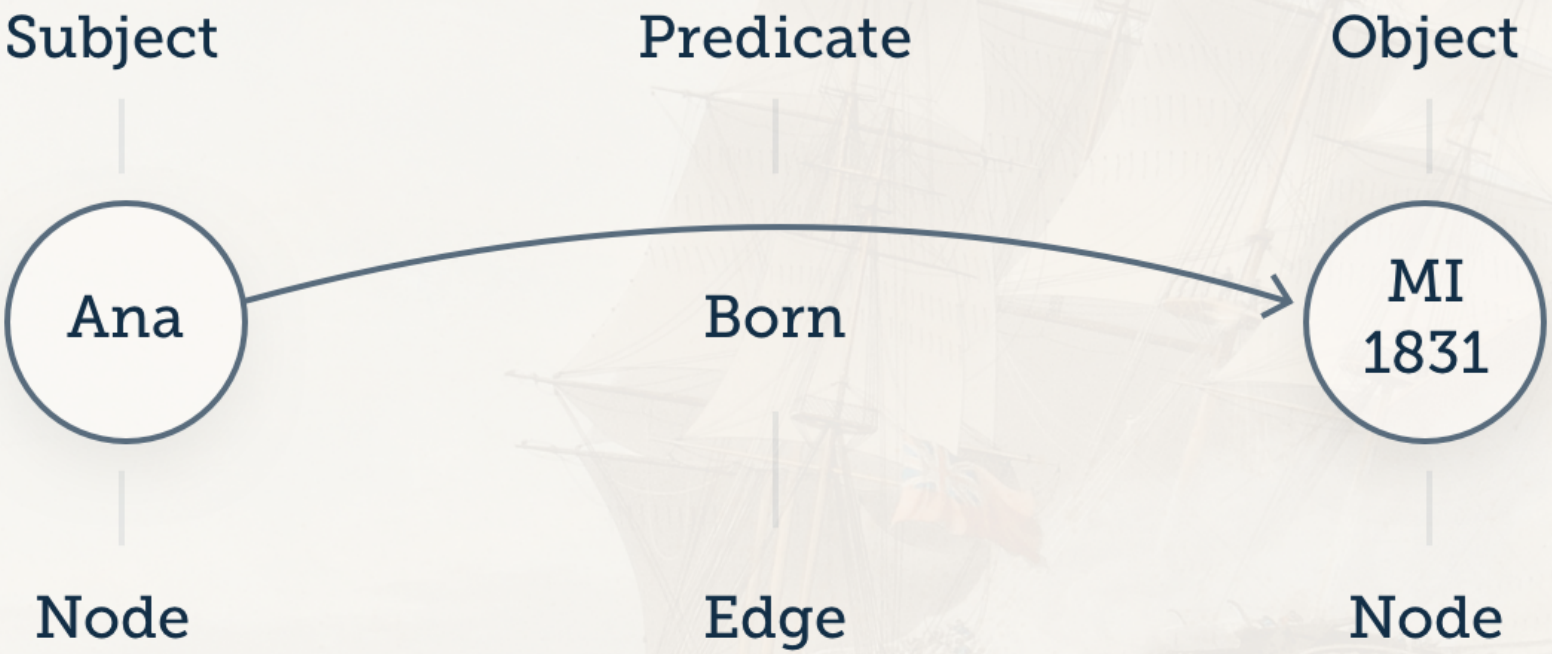
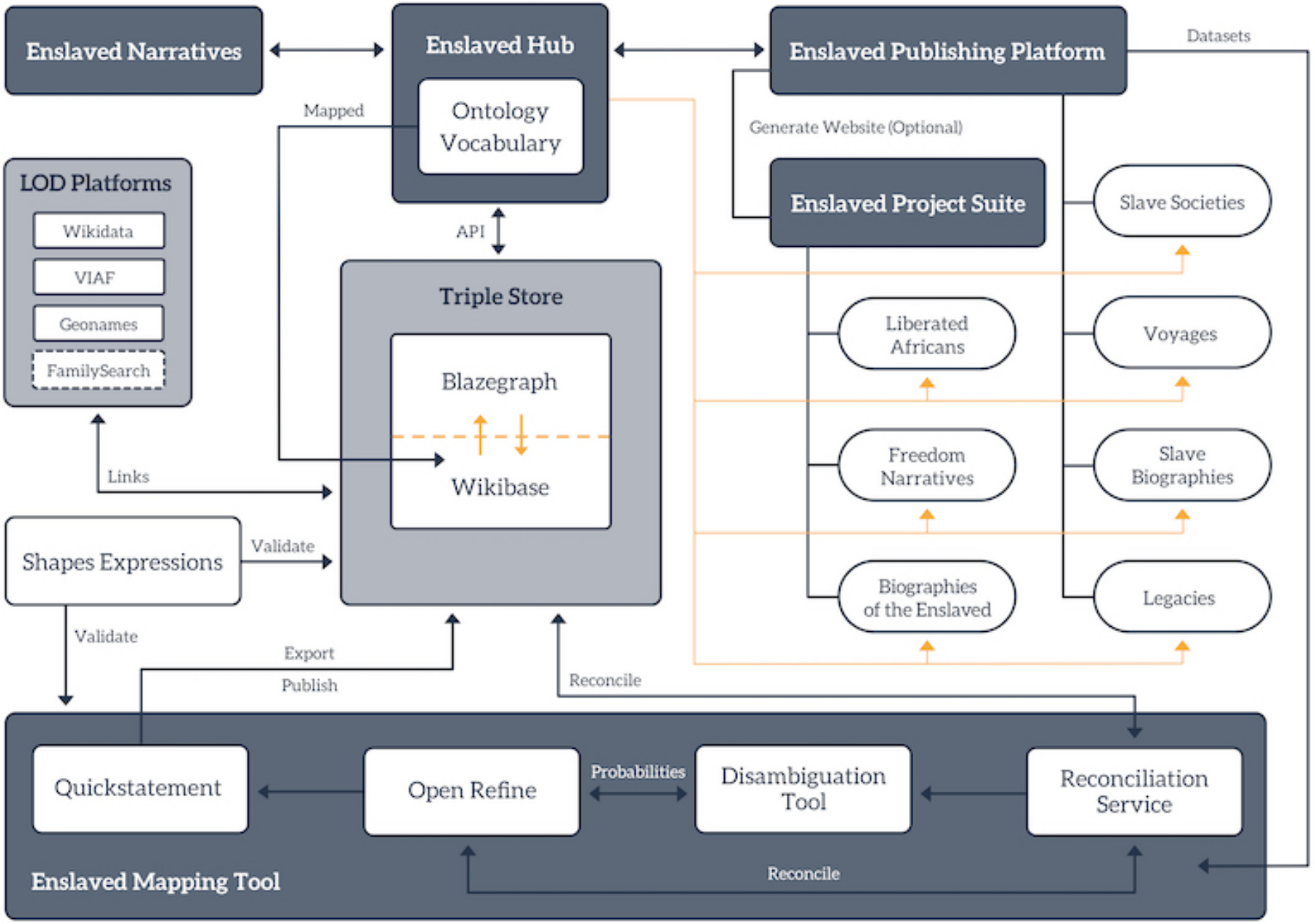


Figure 3.1: Schema Diagram for the Enslaved Ontology.

Triple



Enslaved


Enslaved Narratives
Enslaved Hub
Ontology
Vocabulary
Enslaved Publishing Platform
Datasets
Mapped
Generate Website (Optional)
LOD Platforms
Wikidata
VIAF
Geonames
FamilySearch
API
Enslaved Project Suite
Slave Societies
Triple Store
Blazegraph
Wikibase
Liberated Africans
Voyages
Freedom Narratives
Slave Biographies
Biographies of the Enslaved
Legacies
Links
Shapes Expressions
Validate
Validate
Export
Publish
Reconcile
Quickstatement
Open Refine
Probabilities
Disambiguation Tool
Reconciliation Service
Enslaved Mapping Tool
Reconcile

Digging into Quilt Data

Dean Rehberger

Director, Matrix

rehberge@msu.edu

Twitter: deanreh