

Preliminary Findings: A Comparative Study of User- and Cataloger-Assigned Subject Terms

Hannah Marie Marshall

Metadata Librarian for Image Collections

Cornell University

Overview

- Methodology
 - Survey design & data collection
- Background of the Arts and Sciences Images for Teaching Collection
 - Study population
- Research Questions
- Findings
- Conclusions

Survey

Control



Please review the image and then enter terms that you think describe the image in the spaces provided. Please limit your terms to one- or two-word phrases and provide as many terms as you think are relevant, but feel free to leave spaces blank.



Term 1	<input type="text"/>
Term 2	<input type="text"/>
Term 3	<input type="text"/>
Term 4	<input type="text"/>
Term 5	<input type="text"/>
Term 6	<input type="text"/>
Term 7	<input type="text"/>
Term 8	<input type="text"/>
Term 9	<input type="text"/>

>>

Survey Powered By [Qualtrics](#)

Variable



Please review the image and then enter terms that you think describe the image in the spaces provided. Please limit your terms to one- or two-word phrases and provide as many terms as you think are relevant, but feel free to leave spaces blank. When deciding on descriptive terms for each image, keep the following three questions in mind:

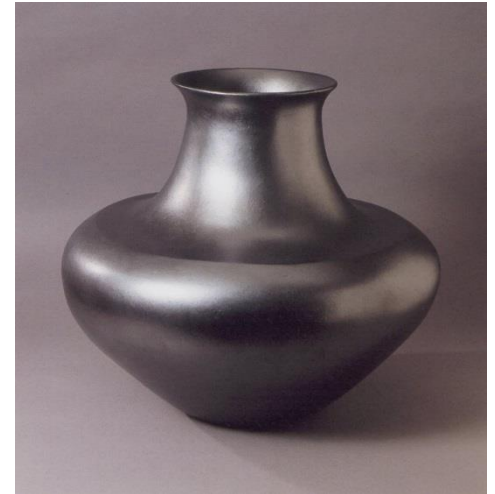
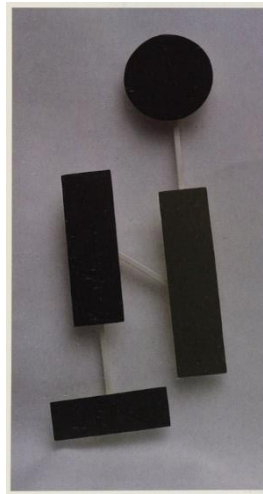
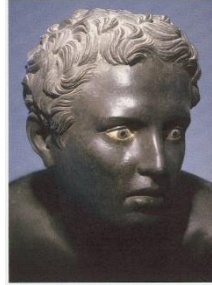
1. What is the image of?
2. What is the image about?
3. What is the image a good example of?



Term 1	<input type="text"/>
Term 2	<input type="text"/>
Term 3	<input type="text"/>
Term 4	<input type="text"/>
Term 5	<input type="text"/>
Term 6	<input type="text"/>
Term 7	<input type="text"/>
Term 8	<input type="text"/>
Term 9	<input type="text"/>

>>

Survey Images



The Arts and Sciences Images for Teaching Collection

- 43,233 images
- 30,001 works
- April 2013 backlog = over 5,000 images
- Backlog eliminated May 2014
- Collection development driven primarily by faculty in Art History, Classics, Anthropology
- Images of art, architecture, cultural artifacts
- Available in Artstor and Luna
- Cataloged in PiCtor

The screenshot displays the Knight Visual Resources Collection website. At the top, it says "KNIGHT VISUAL RESOURCES COLLECTION" and "Cornell University Library". Below this is a navigation bar with "Collections", "Explore", "Create", "Share This", and "Help". A search bar is also present. The main content area is titled "Browse Knight Visual Resources Collection" and features a "Featured Media" section with a grid of image thumbnails. On the left, there is a "Collection Selection" sidebar with a list of various collections. Below the featured media, there is a "Browse" section with a "Select a collection type" dropdown and a list of collection types including "Artstor Digital Library" and "Shared Shelf Institutional Collections". A list of collections is shown below, with "Cornell Knight Visual Resources Collection" highlighted. On the right side, there is a "Log out" button and a "Select Language" dropdown. At the bottom right, there is a small image of a building and a text box describing the collection.

The Arts and Sciences Images for Teaching Collection

Image Cataloging

- Structure based on VRA Core 4.0
- CCO (Cataloging Cultural Objects)
- Work records allow up to 9 subject terms
- Image records allow up to 5 subject terms
- Subject cataloging practice has been inconsistent over time
- Current practice is to do full descriptive and subject cataloging for all images and to use the work/image relationship whenever possible
 - Getty AAT
 - LCSH
 - (Iconclass)

The screenshot displays the image cataloging software interface, showing metadata for two objects: the Arnolfini Portrait and an Aristonothos Krater.

Arnolfini Portrait Metadata:

Field	Data
Creator	Eyck, Jan van (Flemish, ca 1390-1441) Gender: Male, (painter)
Culture	Flemish
Title	Arnolfini Portrait Portrait of Giovanni (?) Arnolfini and His Wife Arnolfini Marriage Detail
Work Type	Classification: Paintings Oil painting; Portrait; Painting
Date	1434
Material	Oil on oak
Measurements	82.2 x 60 cm
Style Period	Renaissance
Repository	London: National Gallery
Accession Number	NG186
Subject	Portraits ; Weddings ; Arnolfini, Giovanni ; Cenami, Giovanna ; Marriage ; Dogs ; Slippers ; Mirrors ;
Collection	Knight Visual Resources Image Collection
ID Number	Record/Work: 3359 Image/File: 1044 File Names: U0566_0095
Source	Publication: RISD Museum of Art Data from: Cornell University
Rights	The KVRIF digital image collection is available only to members of the Cornell University Community for noncommercial educational and scholarly purposes (instruction, study, research, and scholarship); this includes classroom projection, display on computer monitors, and use in class assignments. These images may not be downloaded for use in electronic or print publications (including scholarly publication and websites), exhibitions, or broadcasts without

Aristonothos Krater Metadata:

Heading	Ceramics-Greek
Title	Aristonothos Krater
Creator	Aristonothos
Original Entity	
Title	
FigSite	
Dates	ca 680-650 BCE
Location	Rome: Mus. Capitolino
Object No	172
Material	Ceramic painted krater
Dimensions	36.3 cm high
Culture	Greek
Period	Orientalizing
View Type	Overall
Description	Dv.: side B: Blinding of the Giant Polyphemus
CallNo1	Q-05.5
CallNo2	DR AristoPtr
CallNo3	
CallNo4	
CallNo5	
ShellList	Q-05.5 OR AristoPtr 1
1	Polyphemus [Greek mythology]
2	
3	
4	
5	

The interface also shows a 'Work' record for the Arnolfini Portrait with the following details:

- Culture:** Culture 2 (Brazilian), Culture 3 (Citation)
- Period:** Modernist
- Country:** Brazil
- Continent:** South America
- Region:**
- StartYear:** 1933
- EndYear:** 1933
- Date Qualifier:**
- WorkDates:** 1933-1933
- DisplayDates:** 1933
- Era:**

The bottom of the screen shows a status bar with the following information:

Created: HB Entered: 4/7/2006 Modifier: MNV Modified: 5/6/2009 Reviewer: Reviewed * No
 Record: 133 of 43233 Search

Study Population

- Study Population
 - Undergraduate students enrolled in Art History and Classics courses, Fall 2014

	Participants
Responses	80
Response rate	20%
Completion Rate	33.8%
Avg. # of terms assigned per image	2.45



Cornell University
Department of Classics



Cornell University
Department of History of Art and Visual Studies



Research Questions

“Is this useful?”

**“This” = subject cataloging for images
“Useful” = improving the search utility
of this content & facilitating
successful image retrieval by users**

Do users search for images using the same terms we use to describe them?

What is the level of correspondence between the existing subject terms for these images and the user-assigned subject terms?

Do users search for images using the same types of terms we use to describe them?

What is the level of correspondence in the types of subject terms assigned by users and those in the existing metadata?

Can the search utility of images be improved by teaching users to think more like catalogers?

Does providing users with a formula for analyzing the subjects of images change the nature and content of their responses when asked to perform descriptive tasks?

Research Question # 1

Do users search for images using the same terms we use to describe them?

What is the level of correspondence between the existing subject terms for these images and the user-assigned subject terms?

IMAGE 7					
Cataloger		Control		Variable	
	code		code		code
Naples (Italy)	primary	Art	worktype	19th century	period
Dusk (Evening)	primary	art	worktype	2/3 rule	tertiary
Landscapes	primary	art	worktype	animals	primary
Dancing	primary	beautiful	value judgment	art	worktype
Crowds	primary	bleak	secondary	art	worktype
Boats	primary	Blue	primary	Beach	primary
bonfire	primary	bonfire	primary	beach	primary
Bays	primary	bright	tertiary	blue	primary
		captives	primary	boats	primary
		cave	primary	bonfire	primary

“Bonfire” – literal match for the control group and the variable group (successful image retrieval)

“Boats” – literal match for the variable group but not the control group (successful image retrieval)

“Bays” – non-match (unsuccessful image retrieval)

Research Question # 2

Do users search for images using the same types of terms we use to describe them?

What is the level of correspondence in the types of subject terms assigned by users and those in the existing metadata?

IMAGE 7					
Cataloger	code	Control	code	Variable	code
Naples (Italy)	primary	Art	worktype	19th century	period
Dusk (Evening)	primary	art	worktype	2/3 rule	tertiary
Landscapes	primary	art	worktype	animals	primary
Dancing	primary	beautiful	value judgment	art	worktype
Crowds	primary	bleak	secondary	art	worktype
Boats	primary	Blue	primary	Beach	primary
bonfire	primary	bonfire	primary	beach	primary
Bays	primary	bright	tertiary	blue	primary
		captives	primary	boats	primary
		cave	primary	bonfire	primary

Non-Subject terms ie. worktype, creator, style/period, culture, materials/techniques, etc.

Tertiary – demonstrates an awareness of the work as a cultural document reflecting a time and place

- “What is the image a good example of?” / “How does the image communicate?”
- Identifies devices
- ie. “symbolism, “abstraction”, “chiaroscuro”

Secondary – incorporates cultural and iconographic knowledge

- “What is the image about?”
- Interprets figures and gestures

Primary – perception of the work’s pure form

- “What is the image of?” / “What does the image include?”
- Identifies figures and gestures

Research Question # 3


Can the search utility of images be improved by teaching users to think more like catalogers?

Does providing users with a formula for analyzing the subjects of images change the nature and content of their responses when asked to perform descriptive tasks?


IMAGE 7					
Cataloger		Control		Variable	
	code		code		code
Naples (Italy)	primary	Art	worktype	19th century	period
Dusk (Evening)	primary	art	worktype	2/3 rule	tertiary
Landscapes	primary	art	worktype	animals	primary
Dancing	primary	beautiful	value judgment	art	worktype
Crowds	primary	bleak	secondary	art	worktype
Boats	primary	Blue	primary	Beach	primary
bonfire	primary	bonfire	primary	beach	primary
Bays	primary	bright	tertiary	blue	primary
		captives	primary	boats	primary
		cave	primary	bonfire	primary

Control

Variable

 Cornell University

Please review the image and then enter terms that you think describe the image in the spaces provided. Please list your terms to one- or two-word phrases and provide as many terms as you think are relevant, but feel free to leave spaces blank.



Term 1

Term 2

Term 3

Term 4

Term 5


Term 6

Term 7

Term 8


Term 9

Survey Powered By [Qualtrics](#)

 Cornell University

Please review the image and then enter terms that you think describe the image in the spaces provided. Please list your terms to one- or two-word phrases and provide as many terms as you think are relevant, but feel free to leave spaces blank. When deciding on descriptive terms for each image, keep the following three questions in mind:

1. What is the image of?
2. What is the image about?
3. What is the image a good example of?



Term 1

Term 2

Term 3

Term 4

Term 5

Term 6

Term 7

Term 8

Term 9

Survey Powered By [Qualtrics](#)

A Comparative Study of Cataloger- and User-assigned Subject Terms

Hannah Marie Marshall, Metadata Librarian for Image Collections
Cornell University Library

Research Question #1

What is the search utility of our images to our users?
What is the level of concordance between cataloger- and user-assigned subject terms for images?

- Based on evidence that there appears to be a strong link between search behavior and the search utility of our image collections
- This was accomplished by comparing the literal terms assigned by each group and analyzing the frequency with which each group assigned the same terms to the same image
- The level of literal correspondence between the control subject terms and those assigned by the experimental group was only 38.2%
- Of the terms added to the primary level of subject analysis, 70% were secondary terms, 30% were tertiary, and 0% were non-subject terms

Literal correspondence between responses and existing metadata

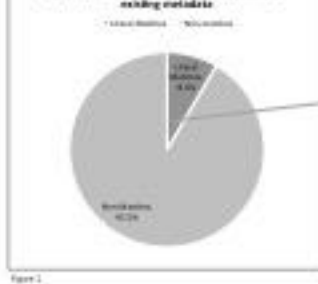


Figure 1

Corresponding literal terms broken down by type

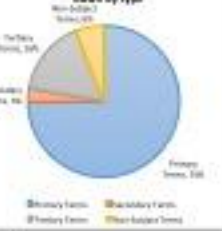


Figure 2

Literal correspondence between control/variable groups and existing metadata

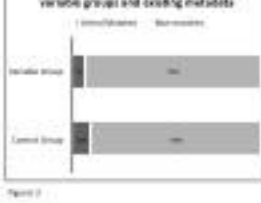


Figure 3

Concordance of types of terms assigned by the control group and variable group

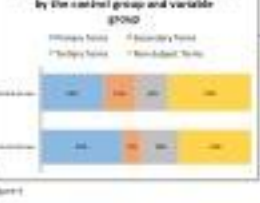


Figure 4

Research Question #3

By asking users to contemplate the primary, secondary and tertiary levels of a subject meaning, can we improve the accessibility of our image collections?

- Does providing users with a set of questions about the images change the nature and content of their responses when asked to perform descriptive tasks?
- This study is the design of a variable group who were asked to contemplate the following three questions as they performed the descriptive tasks on the survey:
 - What is the image about?
 - What is the image about?
 - What is the image a good example of?

	Control	Variable	Non-Subject
Responses	30	31	30
Completion rate	71%	81.0%	83.0%
Average time to complete image	2:21	2:11	2:21

Research Question #2

What types of terms does each group assign, and in what ratios, do they assign them?
What is the level of concordance in the type of subject levels being performed by each group?

- When the terms assigned by the respondents differed from the existing metadata (97.3% of the time), do they differ because each group is performing functionally different kinds of analysis or because they are using different levels of interpretation?
- The control analyses of images has traditionally been conducted on a graduated scale that identifies different levels of interpretation and analysis. In what degree does each group use them?
- The definitions for each category were arrived at through a review of the literature on the subject and prior research
- Figure 5 presents the ratios of primary, secondary, tertiary, and non-subject terms assigned by each group
- Figure 6 presents the same data adjusted to include non-subject terms
- Figure 7 breaks the findings down by the type of image - images of 2D works of art and images of 3D objects

Types of Subject Terms

Primary: definition of the work (person, place, thing, etc.)
"Man in the image of..." "Man in the image of..."
"Queen Elizabeth II..." "Queen Elizabeth II..."

Secondary: descriptive subject terms
"Portrait of..." "Portrait of..."
"Man in the image of..." "Man in the image of..."

Tertiary: descriptive subject terms
"Man in the image of..." "Man in the image of..."
"Queen Elizabeth II..." "Queen Elizabeth II..."

Non-Subject: descriptive subject terms
"Man in the image of..." "Man in the image of..."
"Queen Elizabeth II..." "Queen Elizabeth II..."

Comparison of all types of terms assigned by each participant group



Percentages of subject and non-subject terms assigned by each participant group



Comparison of subject terms assigned by each participant group



Figure 5

Types of terms assigned to images of 2D and 3D works by each participant group

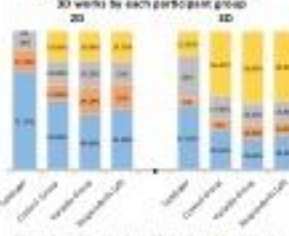


Figure 6

Findings!

Methodology

Design
One set of images were two database queries - one for metadata prepared one for the variable group. Each query contained the image ID number (which was used to find the image in the database) and the image ID number. The variable group was asked to assign subject terms to the image. The control group was asked to assign subject terms to the image. The control group was asked to assign subject terms to the image. The control group was asked to assign subject terms to the image.

Recruitment & Participants

The survey was distributed primarily to students of anthropology, including students in various departments of the University. The survey was distributed to students in the College of Arts and Sciences and the Department of Anthropology. The survey was distributed to students in the Department of Anthropology. The survey was distributed to students in the Department of Anthropology.

Timeline

January 2014 Initial presentation to the faculty at Cornell
February 2014 Data collection begins
March 2014 Data collection ends
April 2014 Data collection ends
May 2014 Data collection ends
June 2014 Data collection ends
July 2014 Data collection ends
August 2014 Data collection ends
September 2014 Data collection ends
October 2014 Data collection ends
November 2014 Data collection ends
December 2014 Data collection ends

Participant Experience

The control and variable groups were asked to perform the same task. The control group was asked to assign subject terms to the image. The variable group was asked to assign subject terms to the image. The control group was asked to assign subject terms to the image. The variable group was asked to assign subject terms to the image.

Research Question #1

The control and variable groups were asked to perform the same task. The control group was asked to assign subject terms to the image. The variable group was asked to assign subject terms to the image. The control group was asked to assign subject terms to the image. The variable group was asked to assign subject terms to the image.

Participant Experience

The control and variable groups were asked to perform the same task. The control group was asked to assign subject terms to the image. The variable group was asked to assign subject terms to the image. The control group was asked to assign subject terms to the image. The variable group was asked to assign subject terms to the image.

Research Question #2

The control and variable groups were asked to perform the same task. The control group was asked to assign subject terms to the image. The variable group was asked to assign subject terms to the image. The control group was asked to assign subject terms to the image. The variable group was asked to assign subject terms to the image.

Participant Experience

The control and variable groups were asked to perform the same task. The control group was asked to assign subject terms to the image. The variable group was asked to assign subject terms to the image. The control group was asked to assign subject terms to the image. The variable group was asked to assign subject terms to the image.

Research Question #3

The control and variable groups were asked to perform the same task. The control group was asked to assign subject terms to the image. The variable group was asked to assign subject terms to the image. The control group was asked to assign subject terms to the image. The variable group was asked to assign subject terms to the image.

Conclusions

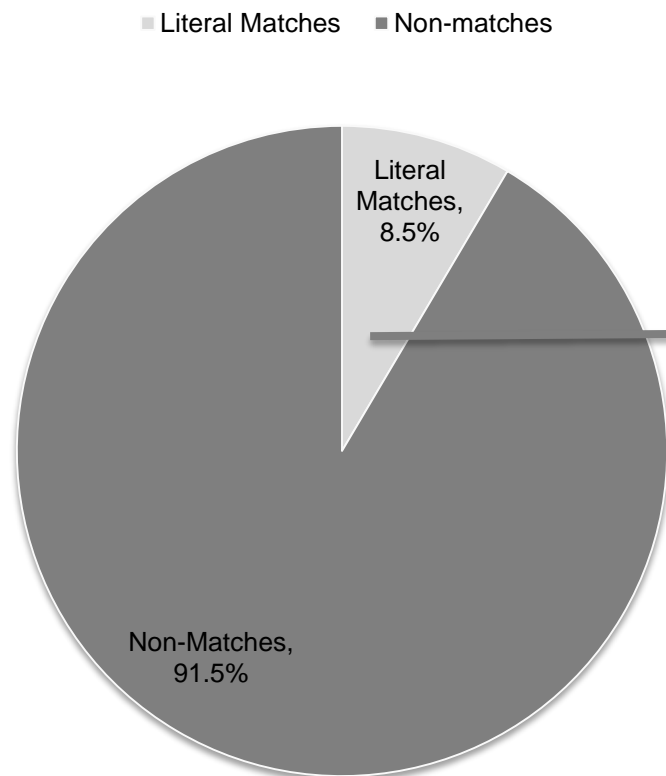
- Primary terms yield the greatest search utility and higher levels of successful image retrieval
- The higher the level of the subject term, the higher the search utility. The higher the level of the subject term, the higher the search utility. The higher the level of the subject term, the higher the search utility.
- High numbers of non-subject terms among the responses indicate that subject meaning did not automatically affect the nature and content of their responses. Secondary level subject terms were more influential in the control group.
- Primary participants using secondary levels on the different levels of subject meaning did not automatically affect the nature and content of their responses. Secondary level subject terms were more influential in the control group.



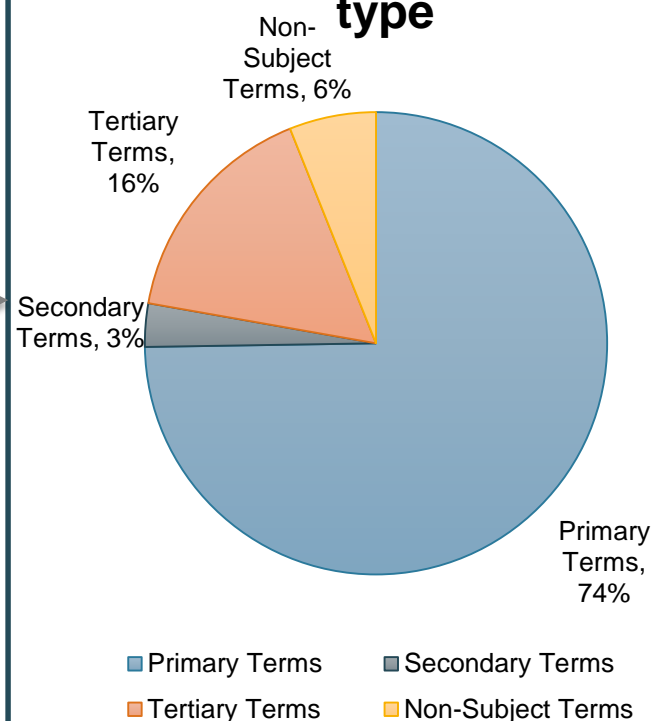
Findings

Literal Matches

Literal correspondence between responses and existing metadata



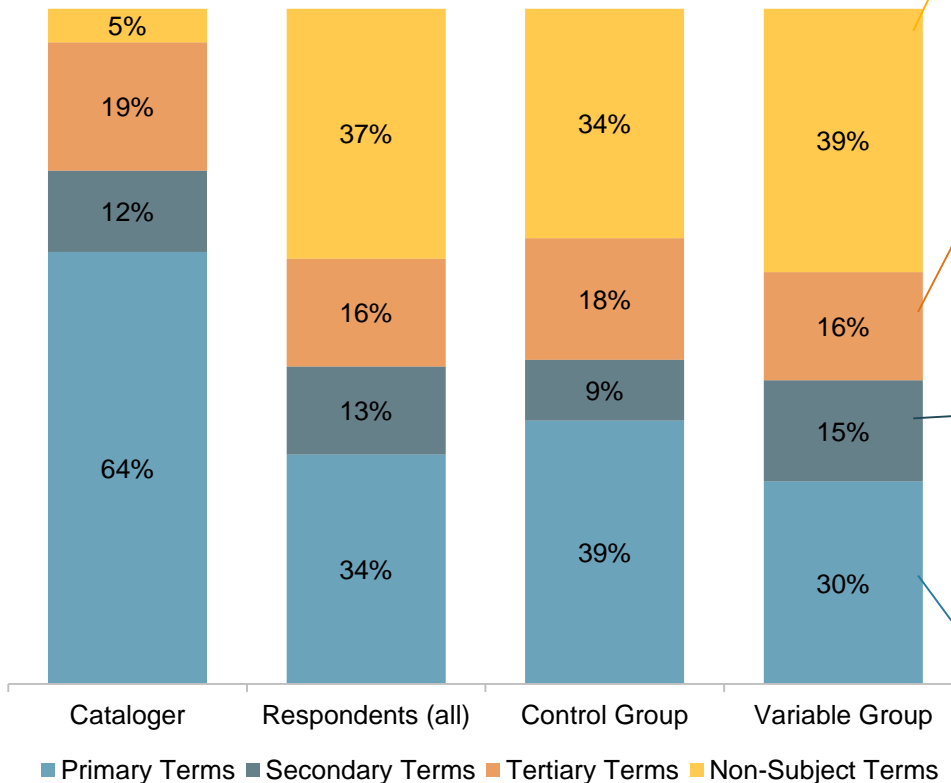
Corresponding literal terms broken down by type



Findings

Types of Terms

Comparison of all types of terms assigned by each participant group



Non-Subject terms – descriptive terms addressing aspects of a work that are not related to its subject

- i.e. worktype, creator, style/period, culture, materials/techniques, etc.

Tertiary – demonstrates an awareness of the work as a document of cultural activity that reflects a time and place

- “What is the image a good example of?” / “How does the image communicate?”
- Identifies devices
- i.e. “symbolism, “abstraction”, “chiaroscuro”

Secondary – incorporates cultural and iconographic knowledge

- “What is the image about?”
- Interprets figures and gestures
- i.e. “Christ”, “banishing”, “prayer”

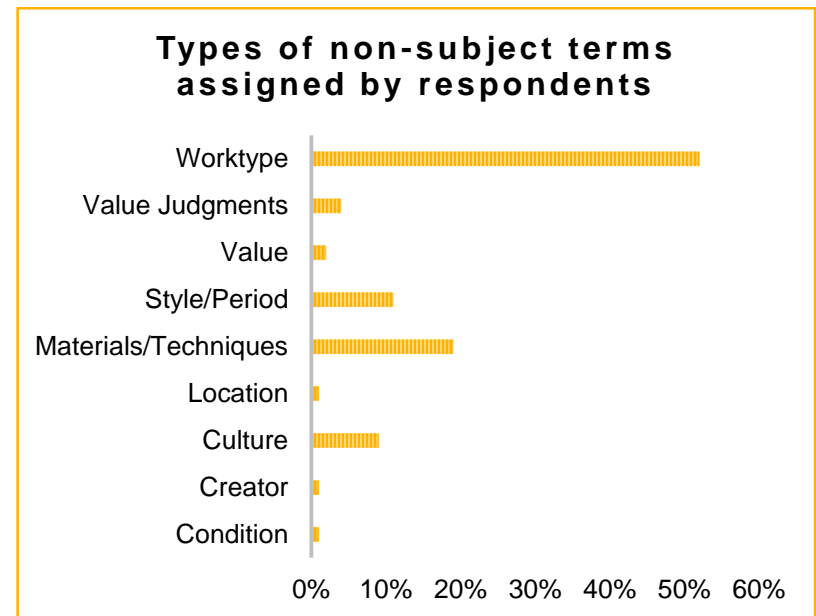
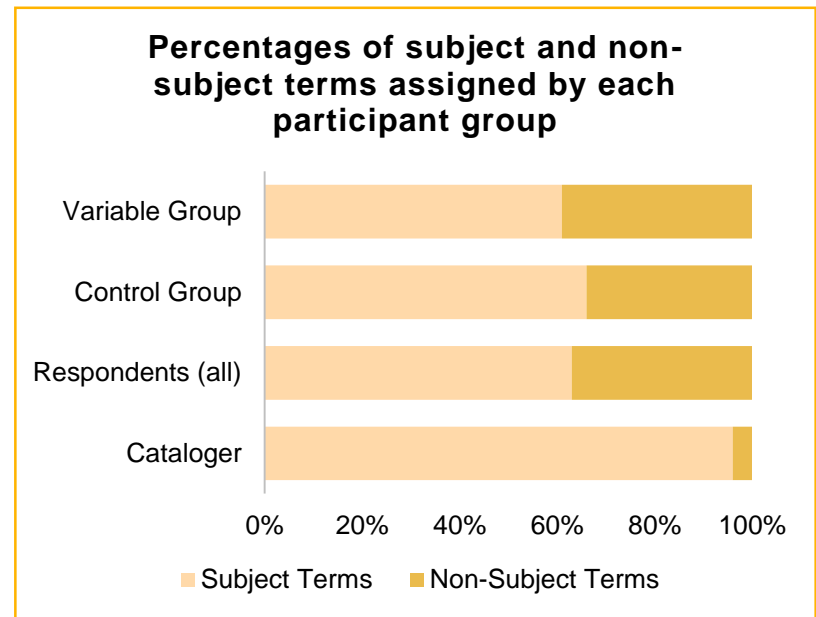
Primary – perception of the work’s pure form

- “What is the image of?” / “What does the image include?”
- Identifies figures and gestures
- i.e. “man”, “pointing”, “clasped hands”, “inscription”

Findings

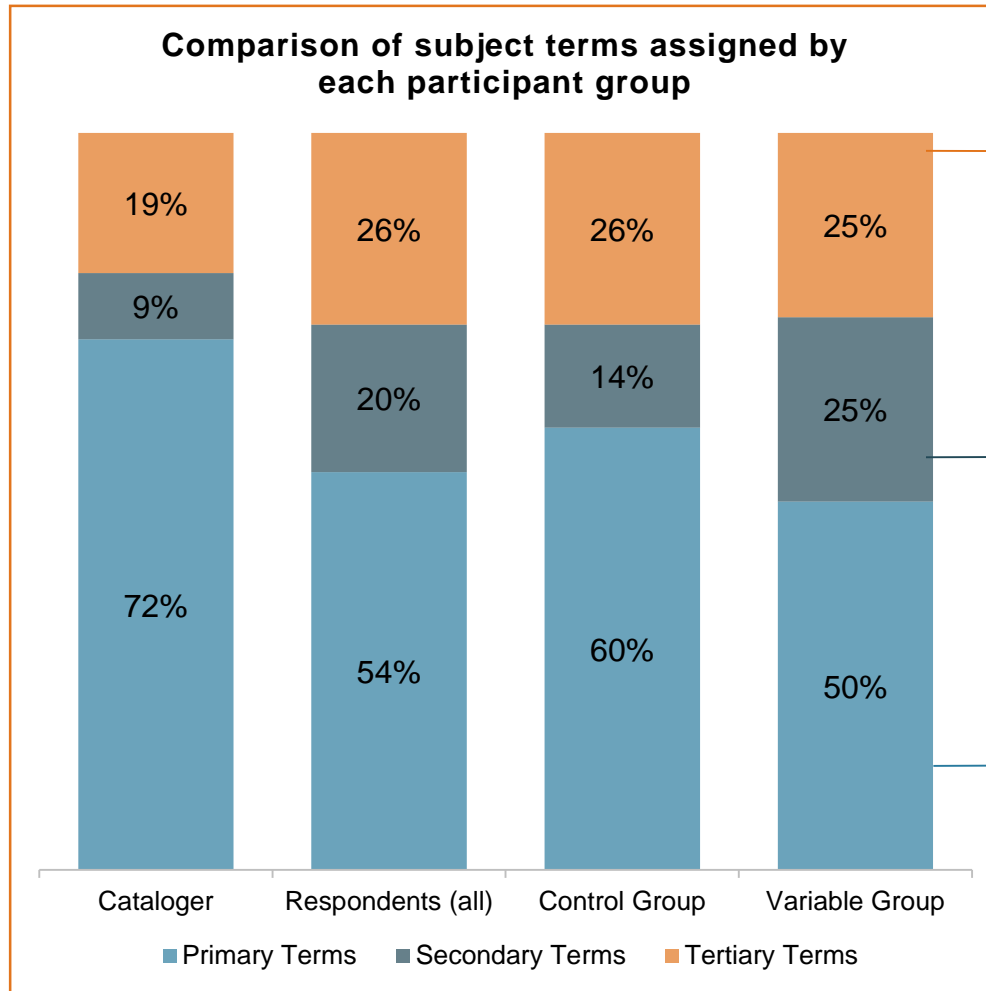
Non-Subject Terms

- Over 1/3 of the participant responses were non-subject terms while less than 5% of the existing metadata were non-subject terms
- Of the non-subject terms assigned by participants, about $\frac{3}{4}$ were terms describing the physical properties of the work depicted in the image
- “Value judgments” refers to when the respondent was expressing their personal opinion about the work of art depicted (ie. “weird”, “pointless”, “confusing”)
- “Value” refers to cases where the participant was speculating as to the market value of the work (ie. “priceless”, “expensive”, “cheap”)



Findings

Types of subject terms



Tertiary – demonstrates an awareness of the work as a document of cultural activity that reflects a time and place

- “*What is the image a good example of?*” / “*How does the image communicate?*”
- Identifies devices
- ie. “symbolism”, “abstraction”, “chiaroscuro”

Secondary – incorporates cultural and iconographic knowledge

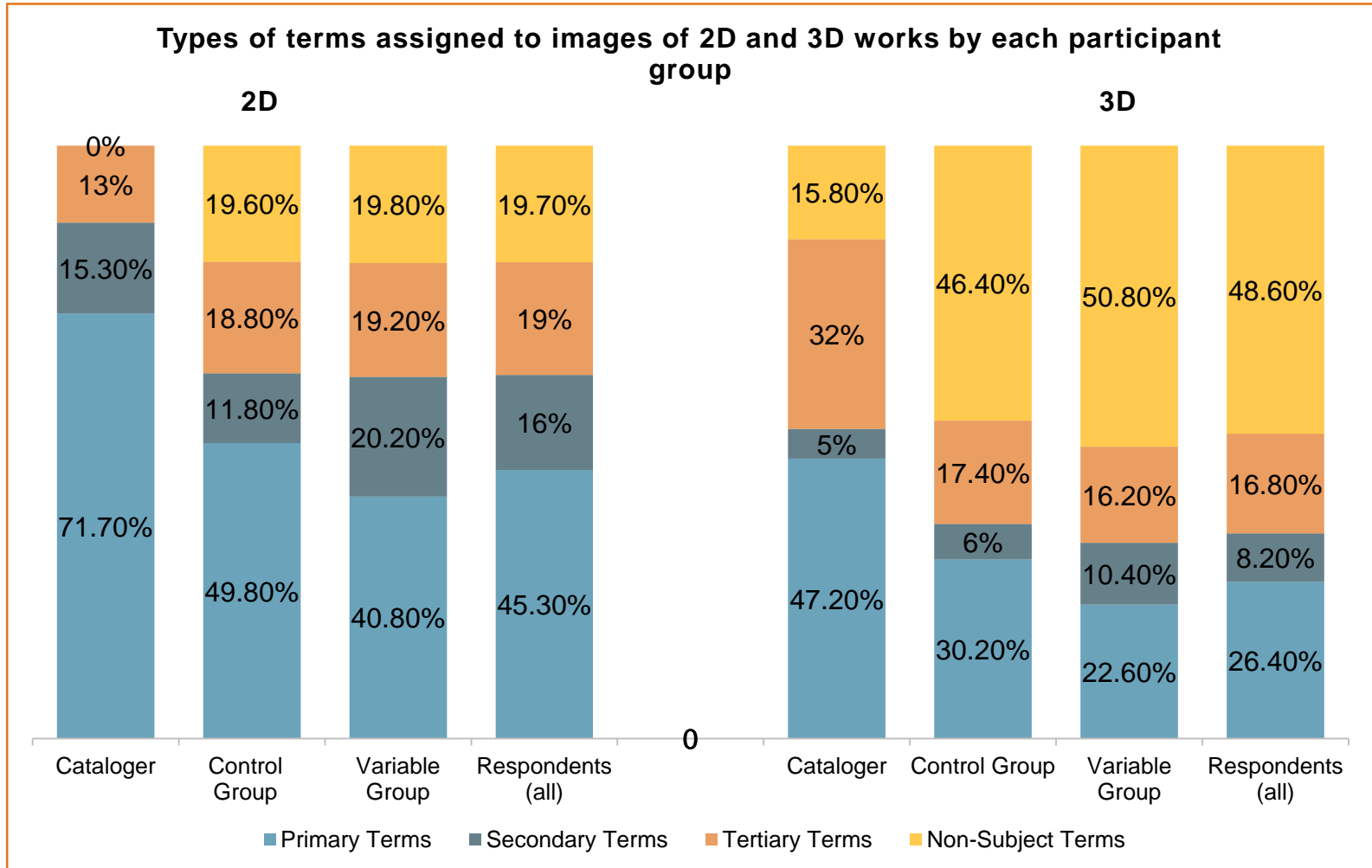
- “*What is the image about?*”
- Interprets figures and gestures
- ie. “Christ”, “banishing”, “prayer”

Primary – perception of the work’s pure form

- “*What is the image of?*” / “*What does the image include?*”
- Identifies figures and gestures
- ie. “man”, “pointing”, “clasped hands”, “inscription”

Findings

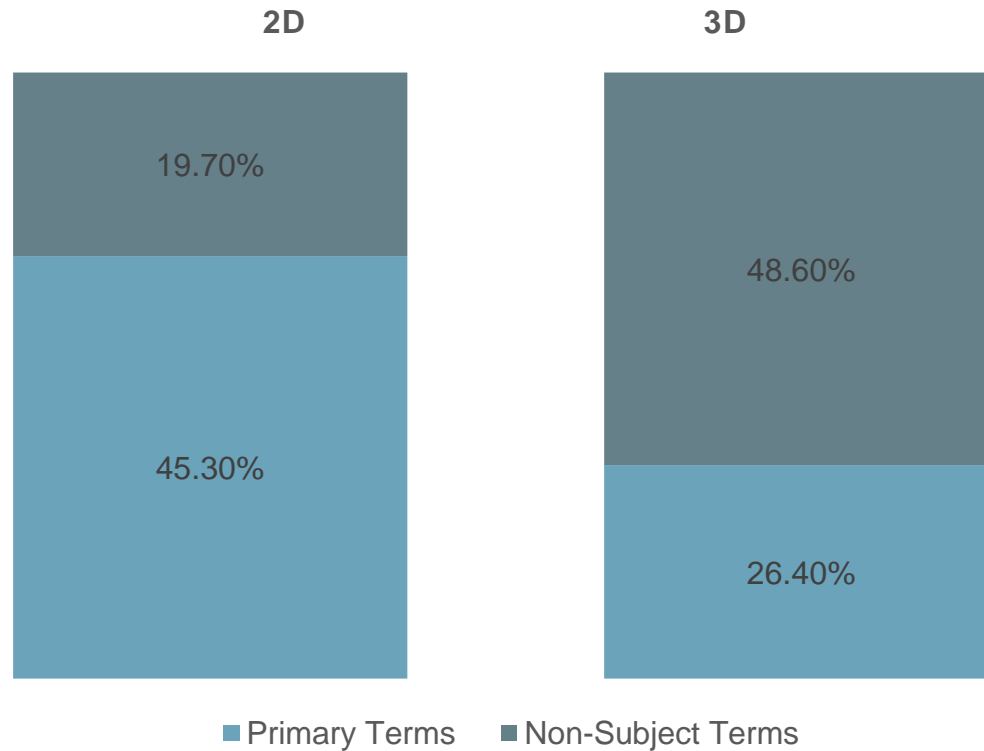
Images of 2D vs. 3D Works



Findings

Images of 2D vs. 3D Works

Participant-assigned primary and non-subject terms assigned to 2D and 3D works



Findings

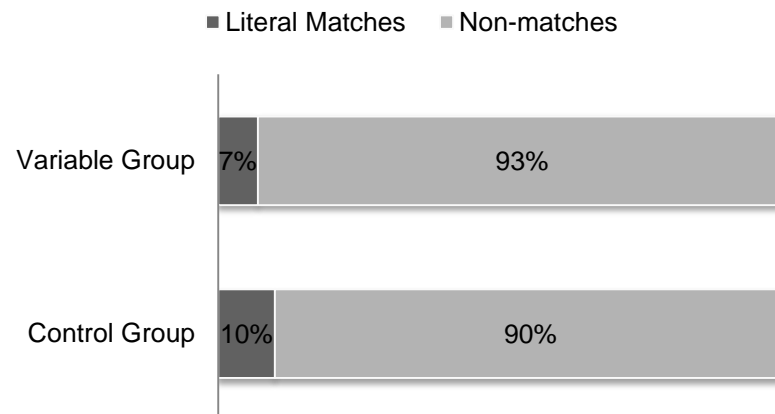
Control vs. Variable Group

No dramatic findings between the control and variable groups

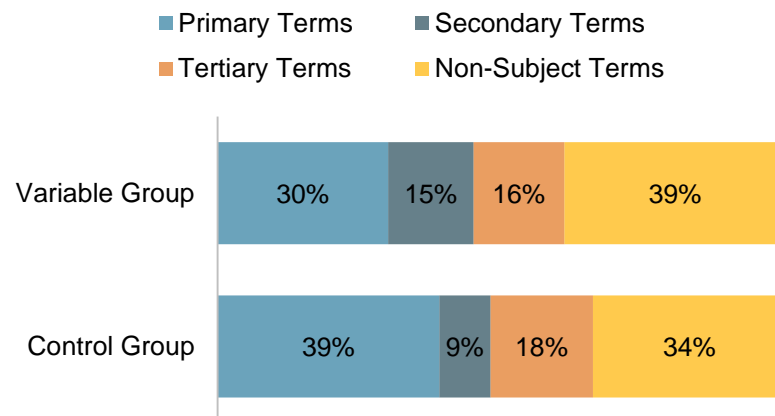
Relative to the control group, the variable group had:

- 9% fewer primary terms
- 6% more secondary terms
- 2% fewer tertiary terms
- 5% more non-subject terms

Literal correspondence between control/variable groups and existing metadata



Comparison of types of terms assigned by the control group and variable group



(Preliminary) Conclusions

Conclusion	Potential Applications
Primary terms yield the greatest search utility and higher levels of successful image retrieval.	Focus cataloging resources on assigning high rates of primary terms
High numbers of non-subject terms applied to images of 3D works suggest that subject metadata is a weak access point for 3D works	Forego subject cataloging for images of 3D works to focus on other descriptive access points
Priming participants using questions based on the different types of subject meaning did not dramatically effect the nature and content of their responses.	
Poor image retrieval seems to be due more to problems of vocabulary than to fundamentally different approaches to subject analysis	This could account for the lack of significant differences between the control and variable group

Thank You!