Do You TEI?

Survey Findings of Text Encoding Practices in Libraries

Michelle Dalmau, Digital Collections Services, IU Libraries
Fall 2013 Digital Library Brown Bag Series
20 November 2013
Twitter: **@mdalmau;** Hashtag: **#dlbb**

TEI-C Special Interest Group: Libraries

- Conveners & Research Partners-in-Crime
  - Michelle Dalmau, Digital Collections Services, Indiana University Libraries
  - Kevin Hawkins, Michigan Publishing, University of Michigan Libraries (co-author)

Do You TEI?
A Survey of Text Encoding Practices in Libraries

http://www.surveymonkey.com/s/tei_in_libraries

Background

Following on papers, presentations and discussions that resulted from the theme of the 2009 Conference and Members' Meeting of the TEI Consortium, "Text Encoding in the Era of Mass Digitization," the launch of the AccessTEI program in 2010, and the recent release of the "Best Practices for TEI in Libraries" in 2011, it behooves us—stewards of text encoding initiatives in academic libraries—to better understand if and how text encoding practices have changed as a result of mass digitization initiatives, declining budgets, and an increased emphasis on "productionizing" digital library services.

Methodology

* Online survey

* Posted from **November 4, 2012 to January 31, 2013**

* No more than **30 minutes to complete**

* Consists of study information, determination of eligibility, background info about the institution (type of library, size, attitudes) & text encoding practices

* Personal information **is not being collected** in this survey

* Recruiting **librarians and libraries staff** via listservs, social media, etc.

* Seeking a non-probability sample of respondents following a **quota sample of 200 respondents**

* Targeting communities of practice rather than individuals to minimize bias that may occur
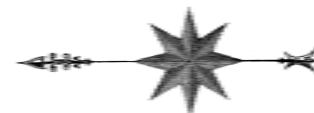
* **Take our survey!**

Goals
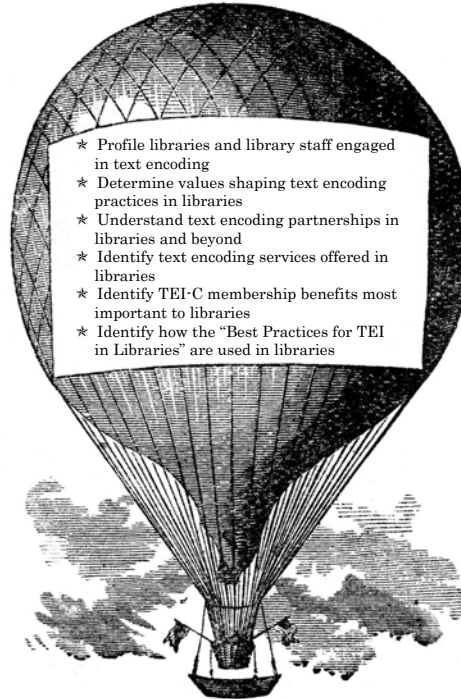
* Profile libraries and library staff engaged in text encoding
* Determine values shaping text encoding practices in libraries
* Understand text encoding partnerships in libraries and beyond
* Identify text encoding services offered in libraries
* Identify TEI-C membership benefits most important to libraries
* Identify how the "Best Practices for TEI in Libraries" are used in libraries

Benchmarks

The following data points gathered from raw data sources or publications will be used to compare/contrast, and, if possible, correlate findings from the online survey:

* TEI-C membership profile of library institutions from 2005–2012

* Evolution/devolution of electronic text centers within libraries from as early as 2000 to present

* Study by Siemens et al., "Understanding the TEI-C Community: A Study in Breadth and Depth, Toward Membership and Recruitment," presented at the 2008 TEI Conference and Members' Meeting, and subsequent publication in the *Journal of the Text Encoding Initiative*, "The Apex of Hipster XML GeekDOM" (http://jtei.revues.org/210)

Michelle Dalmau, Indiana University Digital Library Program, mdalmau@indiana.edu
Kevin S. Hawkins, MPublishing, University of Michigan, kevin.s.hawkins@ultraslavonic.info

Scope in brief

- Developed a SurveyMonkey survey with yes-no, multiple-choice, and free-response questions.

- Announced online on November 4, 2013 and closed on January 31, 2013.

- Survey participants had to answer "yes" to saying they work in a library.

- We encouraged responses from more than one person at the same institution.

- 138 began the survey; 112 "completed" it

Survey Structure

- Study Information
- Determination of Eligibility
- Background Information About the Institution (type of library, size, attitudes)
- Text Encoding Practices
  - Standards
  - Collaborations/Partnerships
  - Types of Text Encoding Projects
  - And more!

History and Hypothesis

Select milestones in the library digitization of text

- 1992: Founding of IATH at Virginia
- 1997: Founding of GDZ at SUB Göttingen
- 1999: Text Creation Partnership formed
- 2004: Google announced mass-digitization partnership with leading research libraries
- 2005: Open Content Alliance created
- 2008: Formation of HathiTrust
- 2010: TEI-C launches AccessTEI
- 2011: Release of a complete rewriting of the Best Practices for TEI in Libraries

The Question of all Questions

How has advocating for such wide-ranging library objectives —from digital access and preservation to digital literacy and scholarship; f r o m   s u p p o r t i n g   n o n e x p r e s s i v e / nonconsumptive research practices to research practices rooted in the markup itself —informed the evolution or devolution of text encoding projects in libraries?
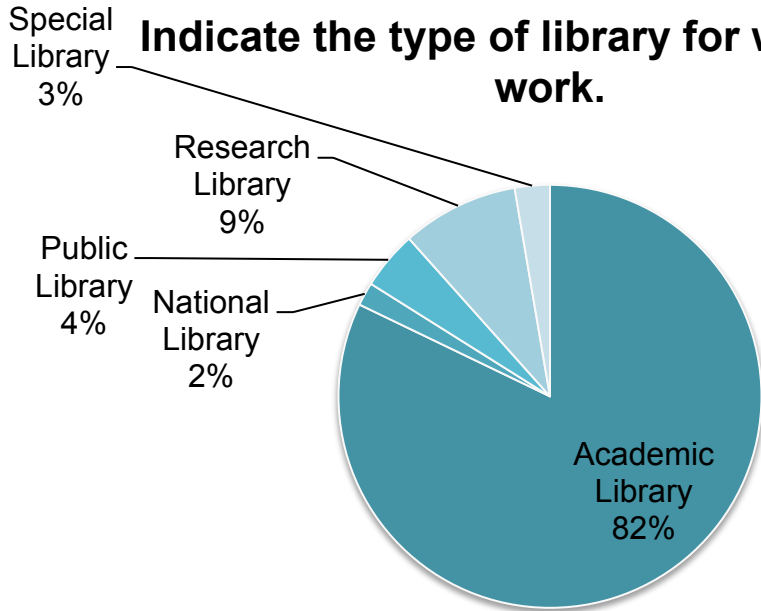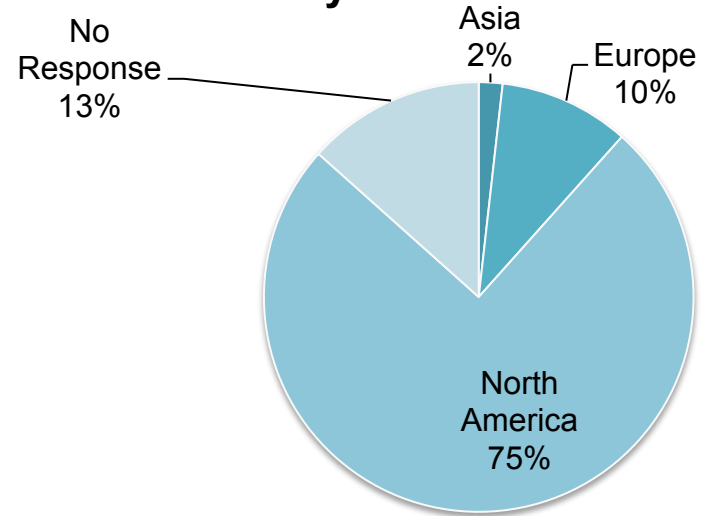
Profile of Survey Respondents

Who responded?

- Of the 112 respondents, we can see from IP addresses that:
  - 55 are clearly affiliated with an institution; 41 of which are unique institutions
  - 57 are unidentifiable due to off-site internet connections (via ISPs)

Indicate the type of library for which you work.

Special Library 3%

Research Library 9%

Public Library 4%

National Library 2%

Academic Library 82%

(n=112)

Where is your institution located?

No Response 13%

Asia 2%

Europe 10%

North America 75%

What is the size of your academic institution based on student enrollment *(or patrons served)*?

Up to 5,000 7%

5,000-10,000 14%

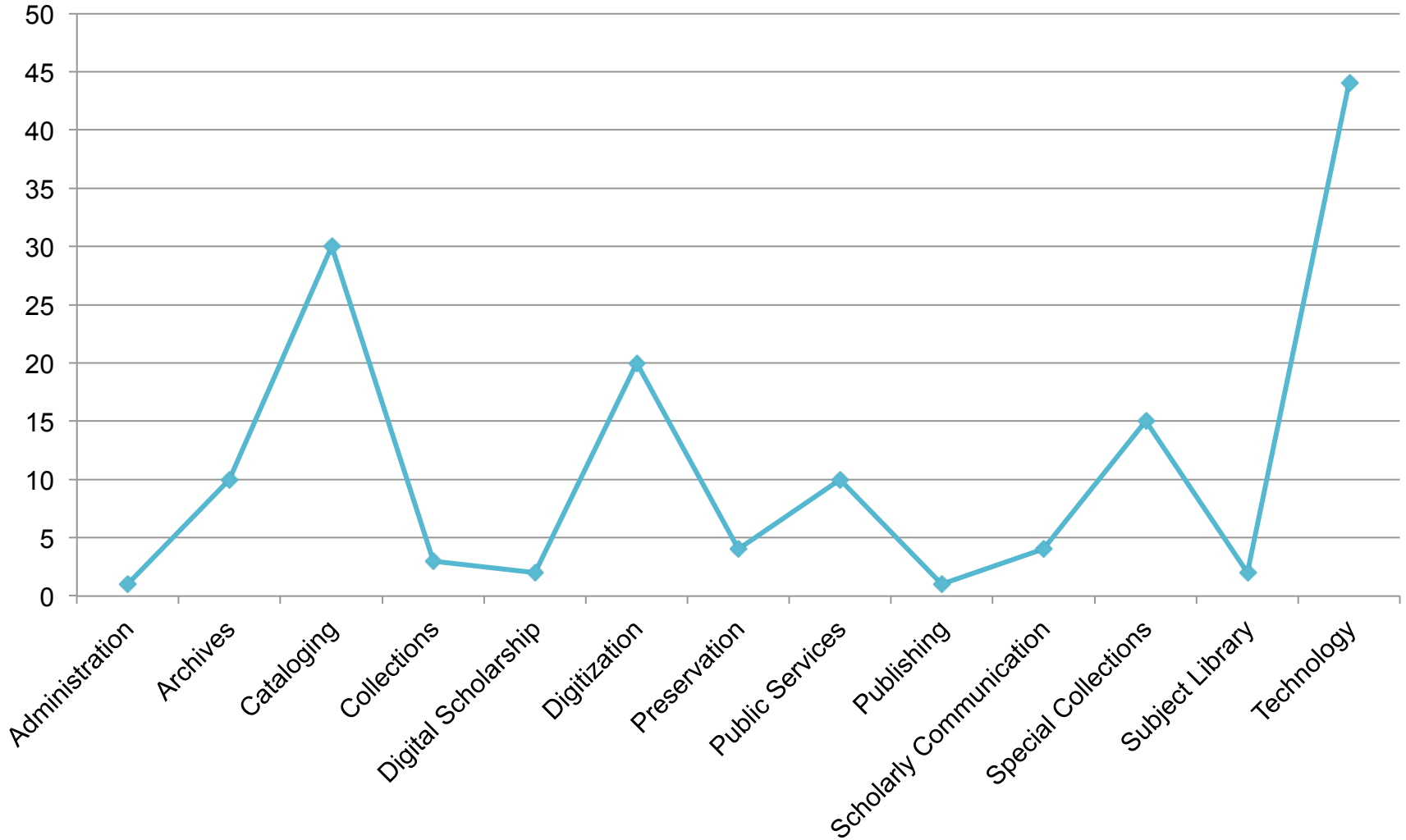No Response 22%

Over 40,000 13%

10,000-25,000 24%

25,000-40,000 20%
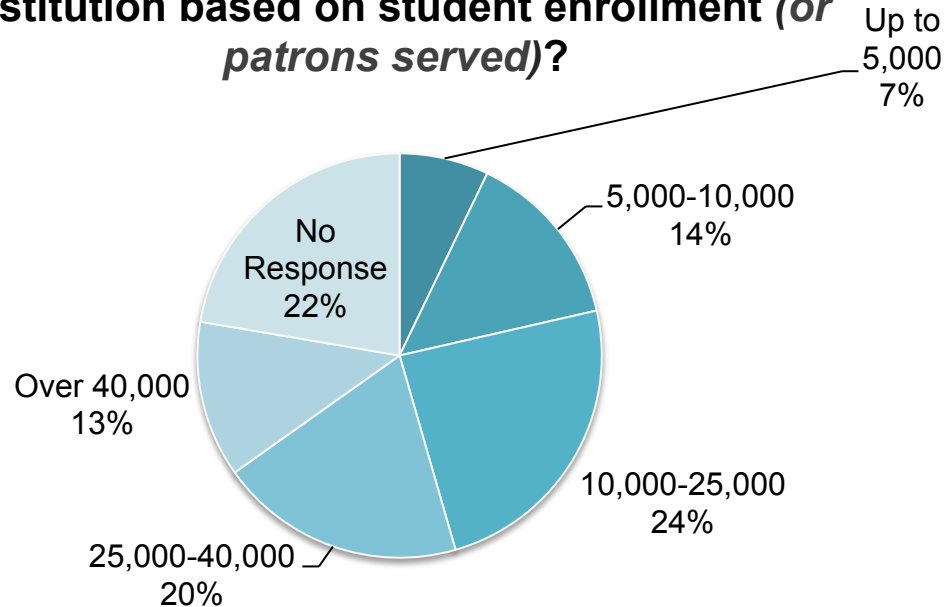
What is the name of your unit or branch library?
(n=99 reported only one unit; n=9 reported more than one unit; n=4 no response)

Administration 3%
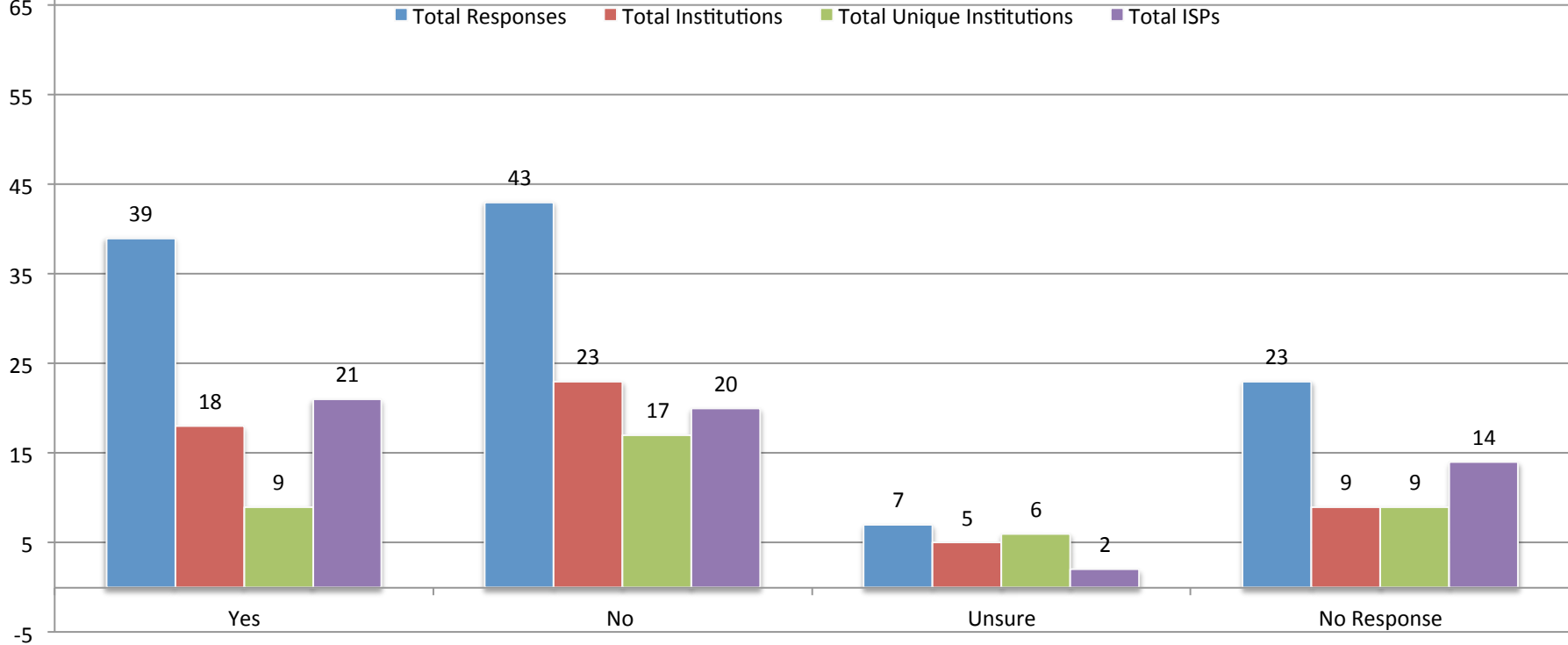Archives 6%
Cataloging 11%
Collections 4%
Digital Scholarship 5%
General Library 14%
Preservation 1%
Public Services 2%
Publishing 3%
Special Collections 11%
Subject Library 8%
Technology 32%

Twitter: @mdalmau #dlbb
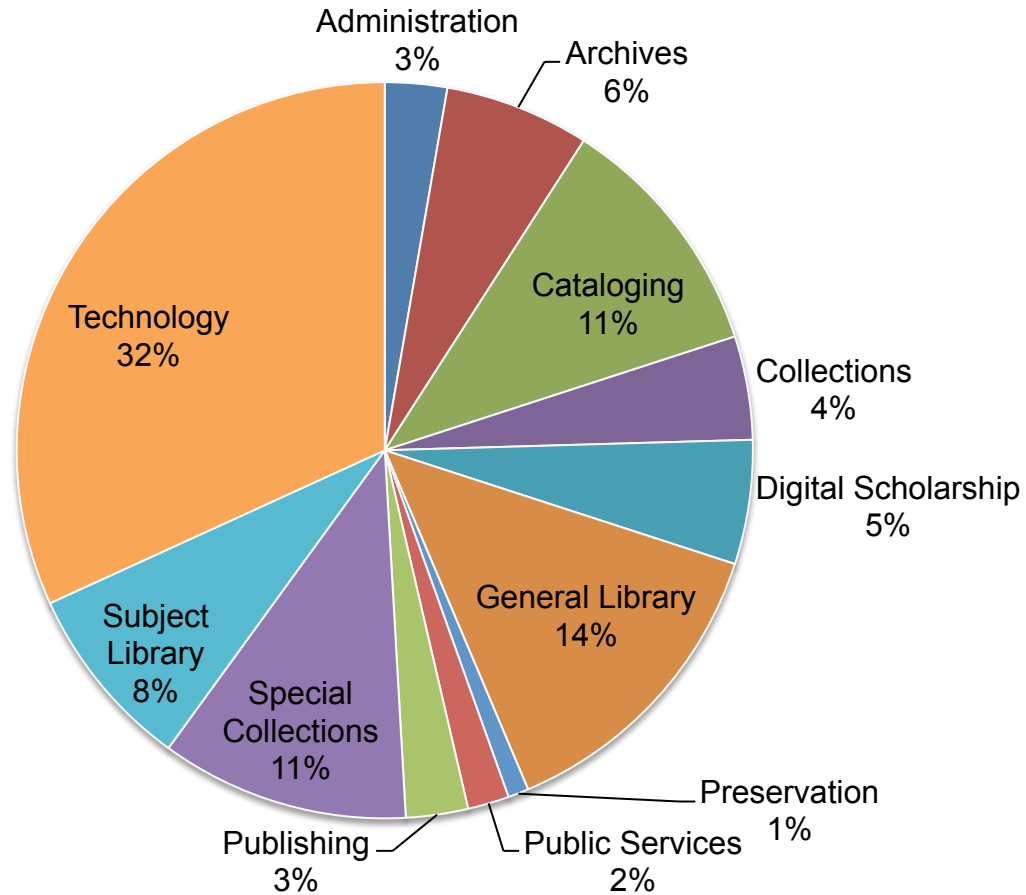
List the units or people (in terms of roles) in your library with which/ whom you partner (n=58)

Twitter: @mdalmau #dlbb

TEI-C Membership

Is your institution a member of the TEI Consortium? (n=112)

Legend: Total Responses, Total Institutions, Total Unique Institutions, Total ISPs

	Total Responses	Total Institutions	Total Unique Institutions	Total ISPs
Yes	39	18	9	21
No	43	23	17	20
Unsure	7	5	6	2
No Response	23	9	9	14

TEI-C Member Institutions: 2005-2013
(missing 2012 data)

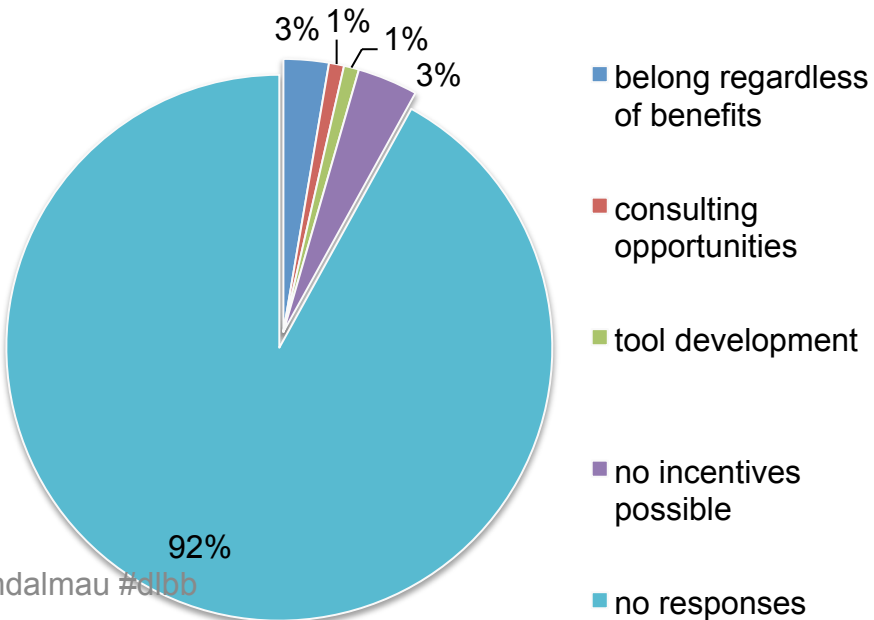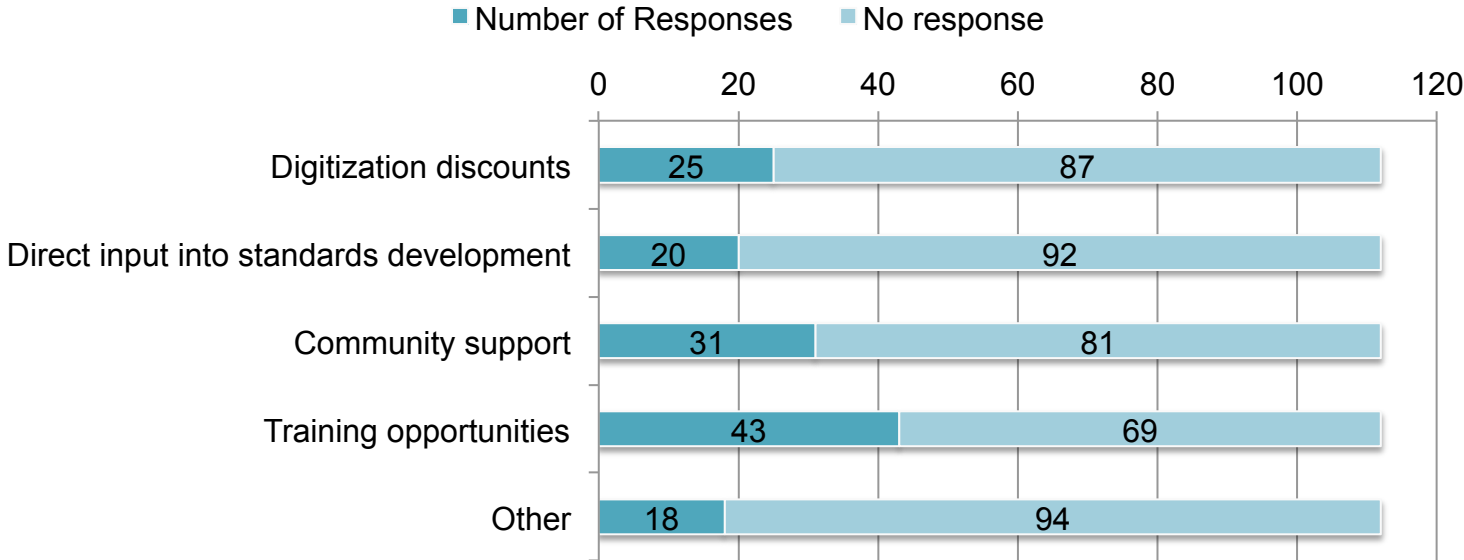■ Libraries   ■ Non-Libraries   ■ Combination   ■ Unsure
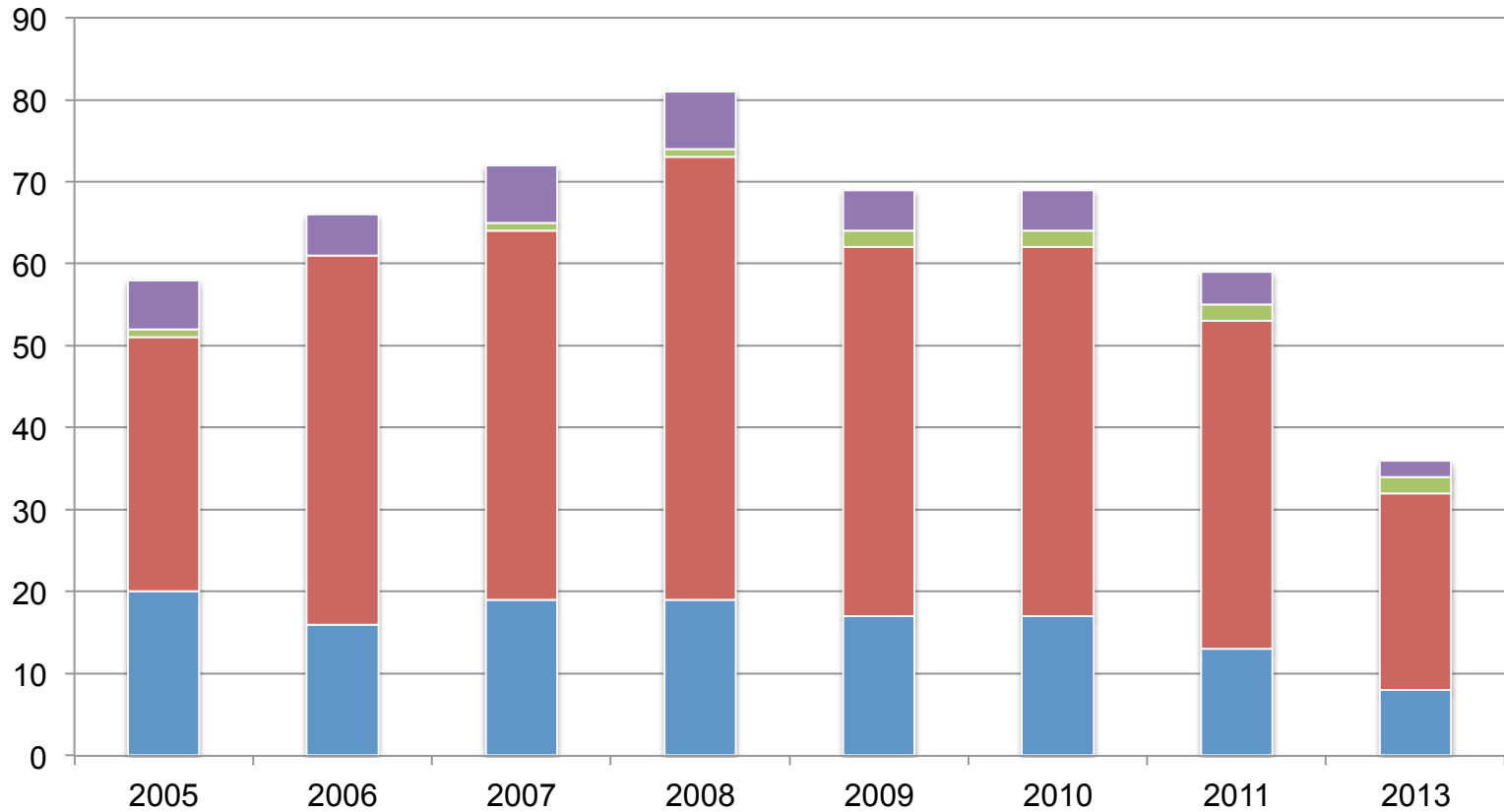
*Membership data (2005-2011) provided by Martin Mueller; Coded by Kevin Hawkins

Twitter: @mdalmau #dlbb

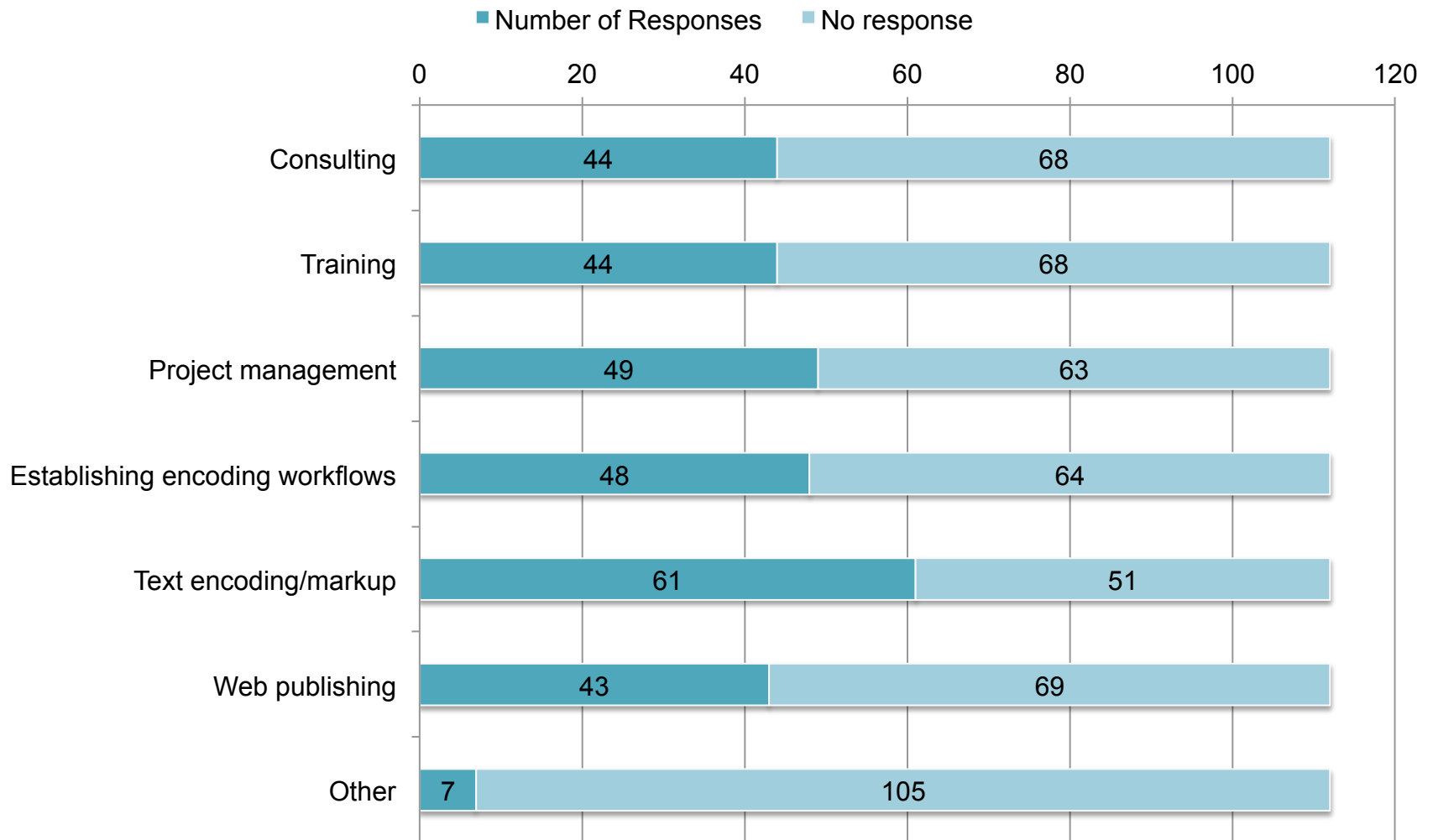# Would your institution become or remain a TEI Consortium member if doing so would allow for: (n=112)

■ Number of Responses ■ No response

Category	Number of Responses	No response
Digitization discounts	25	87
Direct input into standards development	20	92
Community support	31	81
Training opportunities	43	69
Other	18	94

Pie chart:
- belong regardless of benefits: 3%
- consulting opportunities: 1%
- tool development: 1%
- no incentives possible: 3%
- no responses: 92%

Twitter: @mdalmau #dlbb

Text Encoding Practices & Partnerships in Libraries

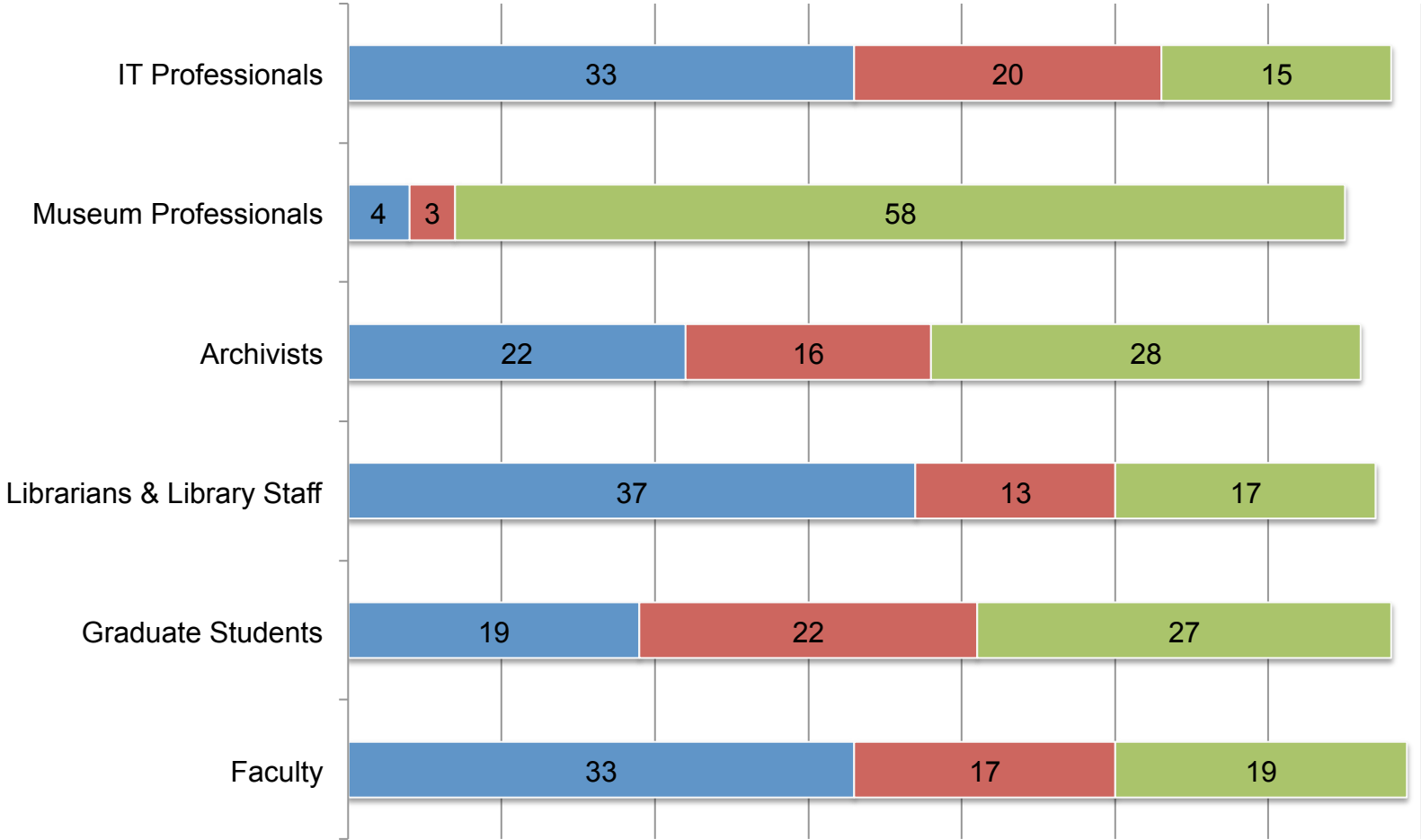# How many text encoding projects have you or members of your unit participated in over the years? (n=112)

None 4%

1-5 40%

6-10 13%

11-15 7%

16-20 2%

More than 30 12%

No Response 22%

Twitter: @mdalmau #dlbb
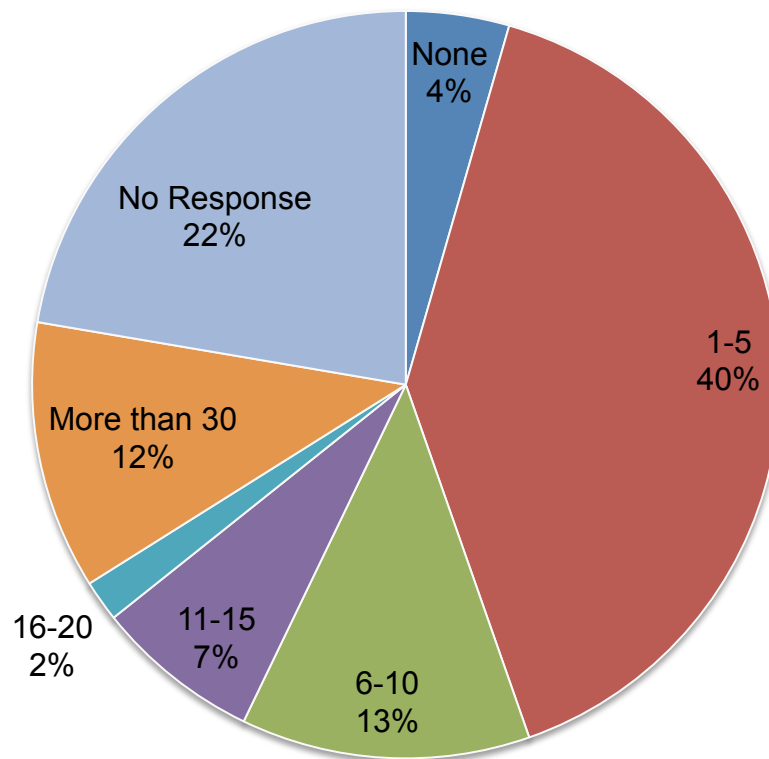
In what ways do you or members of your unit support text encoding projects (select all that apply)? (n=112)

■ Number of Responses ■ No response

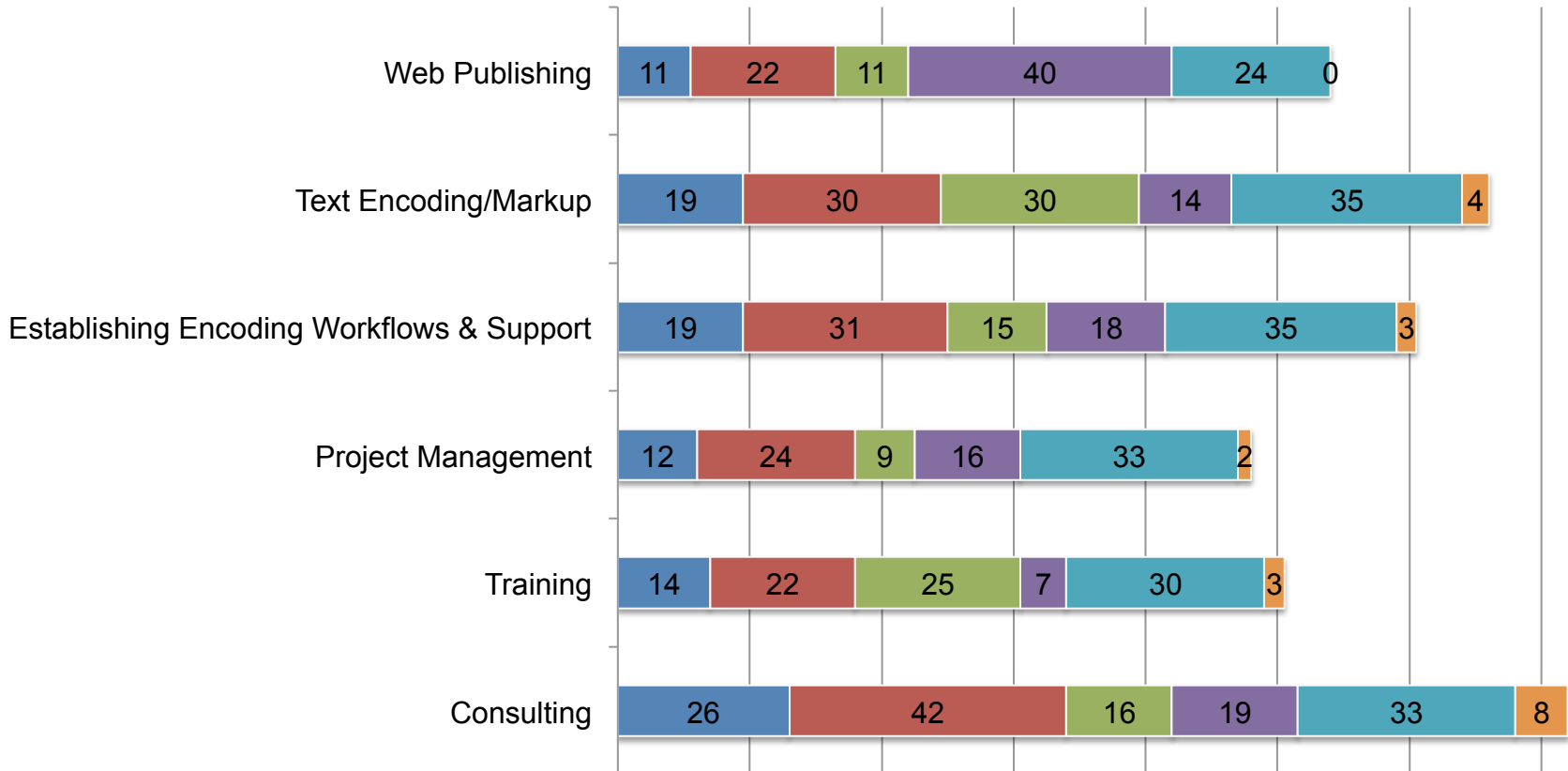| Category	Number of Responses	No response
Consulting	44	68
Training	44	68
Project management	49	63
Establishing encoding workflows	48	64
Text encoding/markup	61	51
Web publishing	43	69
Other	7	105

Twitter: @mdalmau #dlbb

Indicate how frequently you partner with the following people when undertaking a text encoding project:

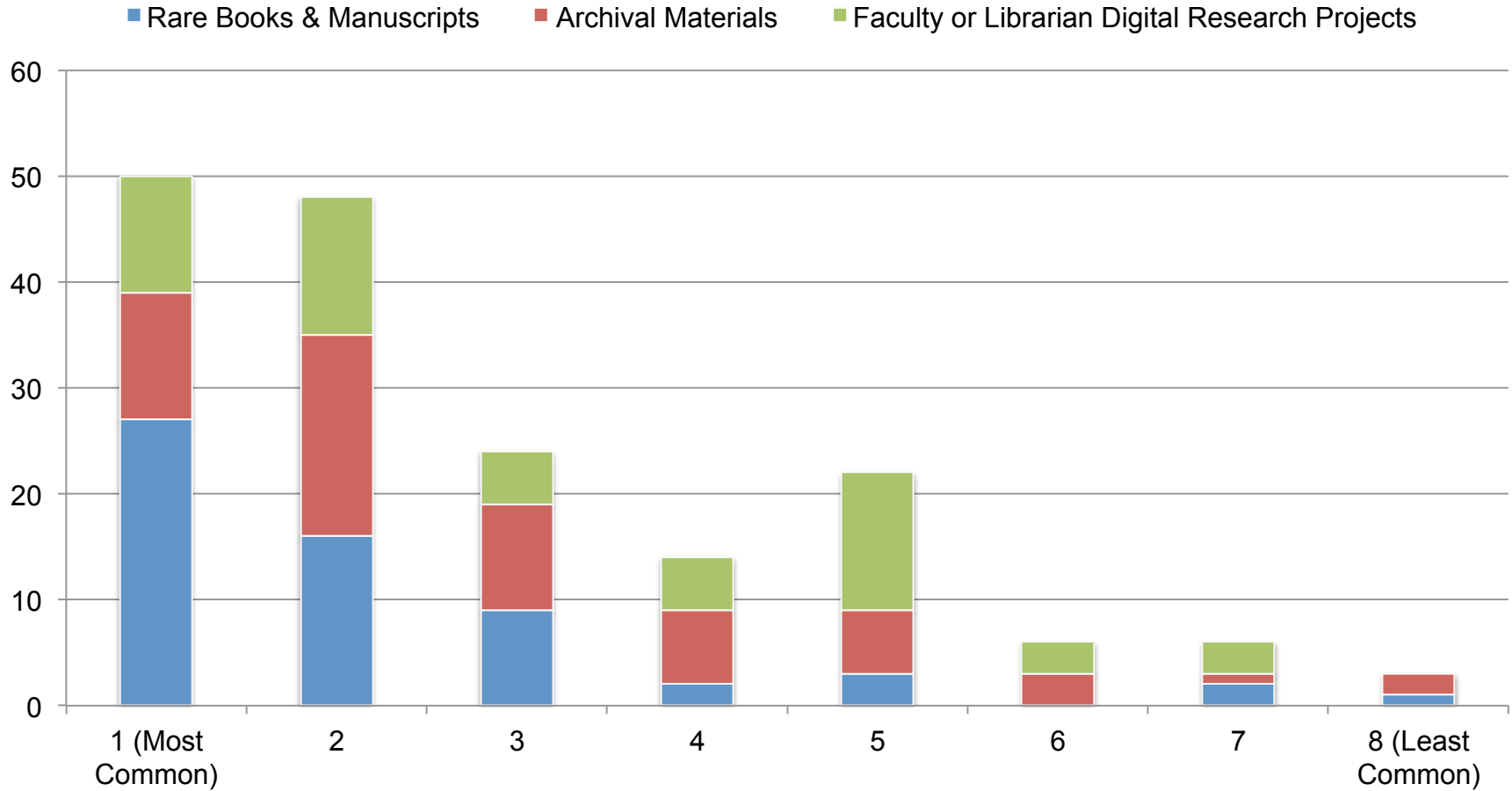■ Almost Always/Often   ■ Sometimes   ■ Seldom/Never

	Almost Always/Often	Sometimes	Seldom/Never
IT Professionals	33	20	15
Museum Professionals	4	3	58
Archivists	22	16	28
Librarians & Library Staff	37	13	17
Graduate Students	19	22	27
Faculty	33	17	19

Twitter: @mdalmau #dlbb

Indicate how you collaborate on text encoding projects with the following categories of people:

■ Archivists ■ Faculty ■ Graduate Students ■ IT Professionals ■ Librarians & Library Staff ■ Museum Professionals
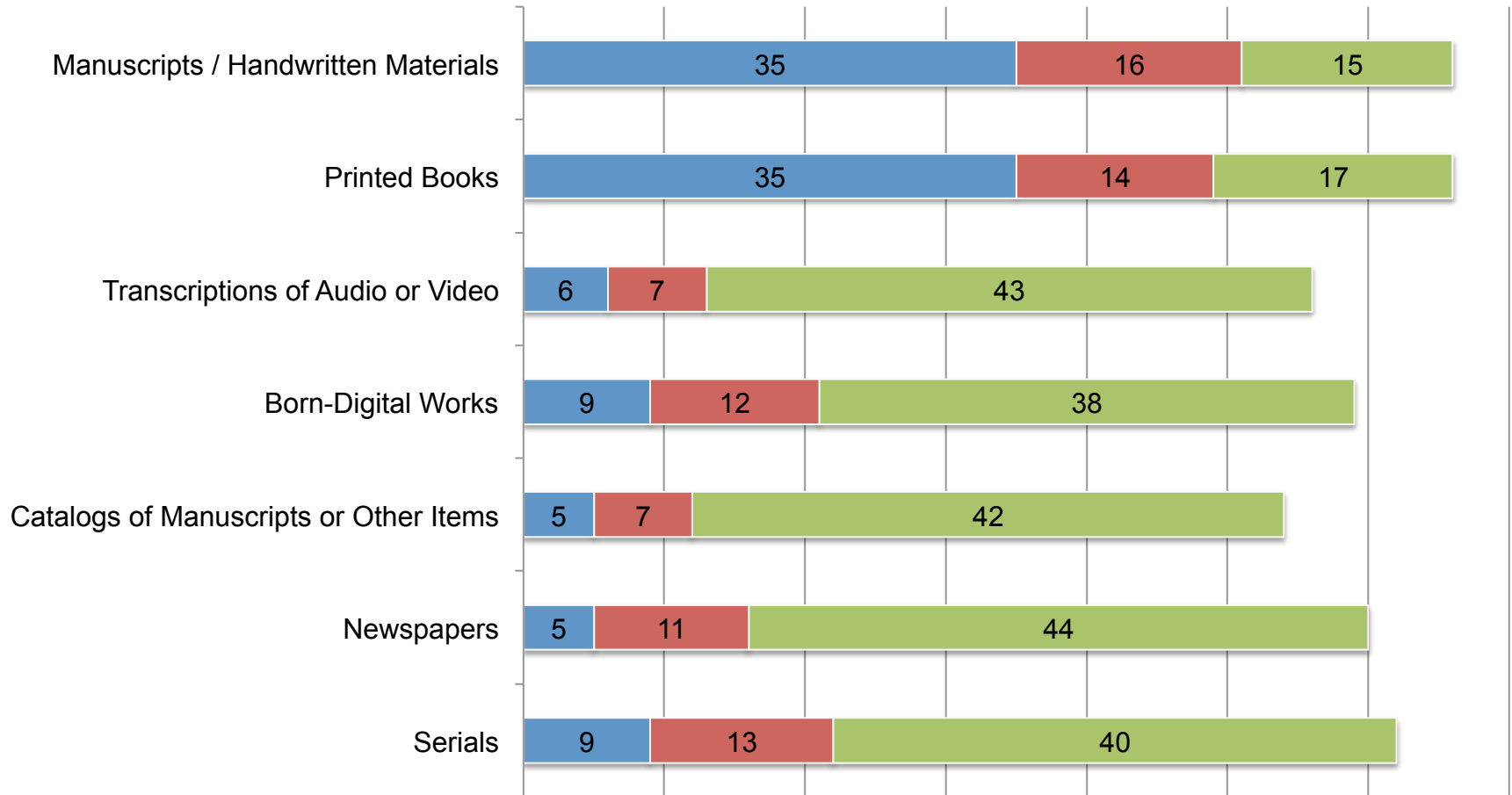
Category	Archivists	Faculty	Graduate Students	IT Professionals	Librarians & Library Staff	Museum Professionals
Web Publishing	11	22	11	40	24	0
Text Encoding/Markup	19	30	30	14	35	4
Establishing Encoding Workflows & Support	19	31	15	18	35	3
Project Management	12	24	9	16	33	2
Training	14	22	25	7	30	3
Consulting	26	42	16	19	33	8

Rank the nature of your text encoding projects by "dragging and dropping" each option into place (1 is most common, 8 is least common)
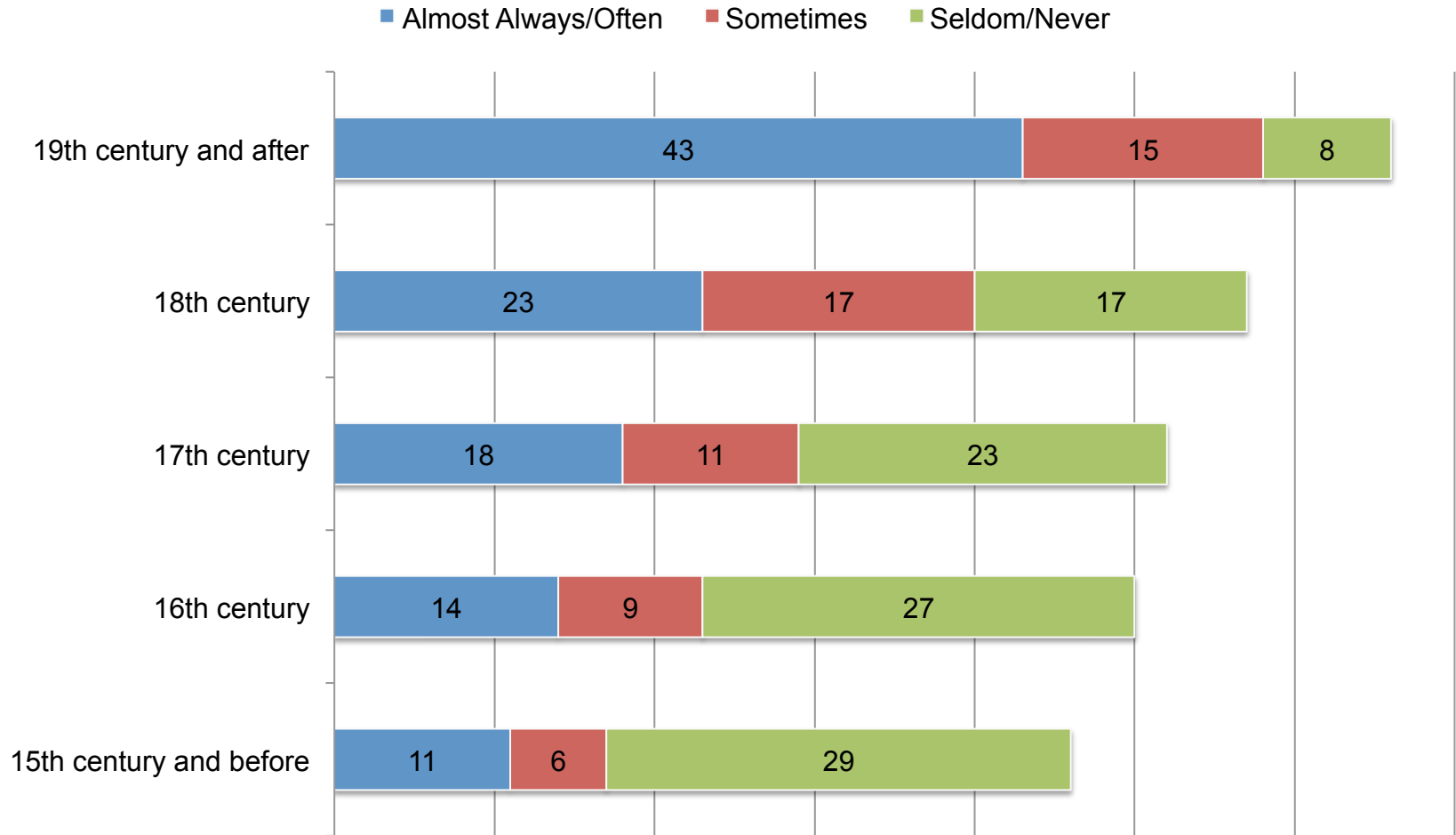
■ Rare Books & Manuscripts   ■ Archival Materials   ■ Faculty or Librarian Digital Research Projects

What type of materials do you digitize and encode? Indicate frequency.

■ Almost Always/Often   ■ Sometimes   ■ Seldom/Never

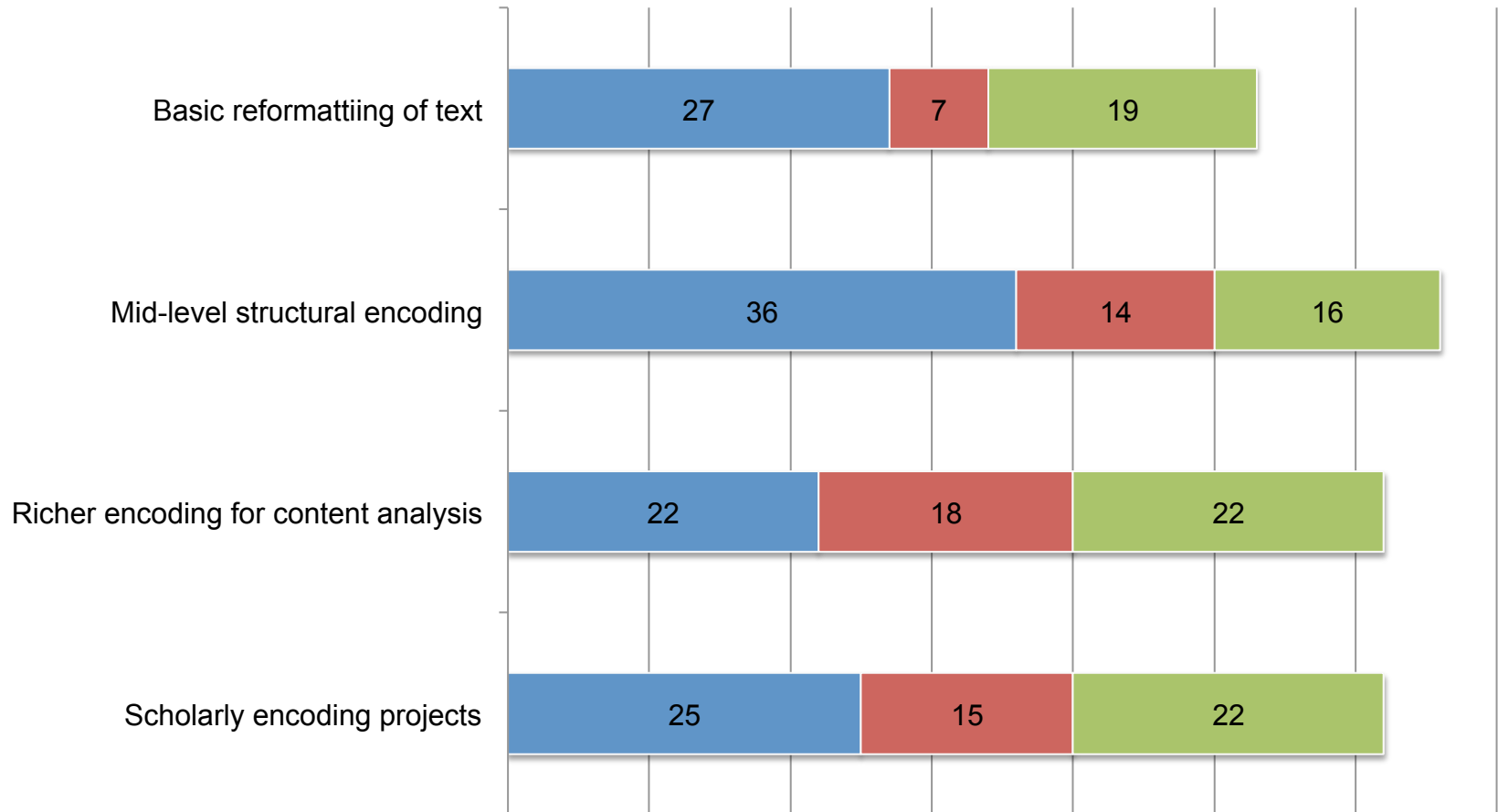| Material	Almost Always/Often	Sometimes	Seldom/Never
Manuscripts / Handwritten Materials	35	16	15
Printed Books	35	14	17
Transcriptions of Audio or Video	6	7	43
Born-Digital Works	9	12	38
Catalogs of Manuscripts or Other Items	5	7	42
Newspapers	5	11	44
Serials	9	13	40

Twitter: @mdalmau #dlbb

Indicate the frequency in which materials from the following time periods are encoded:

- **Almost Always/Often**
- **Sometimes**
- **Seldom/Never**

Time period	Almost Always/Often	Sometimes	Seldom/Never
19th century and after	43	15	8
18th century	23	17	17
17th century	18	11	23
16th century	14	9	27
15th century and before	11	6	29

Twitter: @mdalmau #dlbb

Describe the types and frequency of encoding projects you undertake based on the following levels of encoding:

■ Almost Always/Often   ■ Sometimes   ■ Seldom/Never

Level	Almost Always/Often	Sometimes	Seldom/Never
Basic reformattiing of text	27	7	19
Mid-level structural encoding	36	14	16
Richer encoding for content analysis	22	18	22
Scholarly encoding projects	25	15	22

Twitter: @mdalmau #dlbb

Levels of Encoding (Q89-Q91) v. Number of Encoding Projects (Q24)

Legend: 0, 1-5, 6-10, 11-15, 16-20, More than 30

Basic Reformatting of Text (Q89): Almost Always, Often, Sometimes, Seldom, Never

Mid-level Structural Encoding (Q90): Almost Always, Often, Sometimes, Seldom, Never

Richer Encoding for Content Analysis (Q91): Almost Always, Often, Sometimes, Seldom, Never

Scholarly Encoding Projects (Q92): Almost Always, Often, Sometimes, Seldom, Never

Twitter: @mdalmau #dlbb

Levels of Encoding (Q89-Q91) v. Number of Encoding Projects (Q24)
(n=40)

Legend: ■ 1-5 ▲ 6-10 ● More than 30

Basic Reformatting of Text (Q89): Almost Always, Often, Sometimes, Seldom, Never
Mid-level Structural Encoding (Q90): Almost Always, Often, Sometimes, Seldom, Never
Richer Encoding for Content Analysis (Q91): Almost Always, Often, Sometimes, Seldom, Never
Scholarly Encoding Projects (Q92): Almost Always, Often, Sometimes, Seldom, Never

Twitter: @mdalmau #dlbb

Level of Encoding v. Types of Materials (top two items)

Legend: Manuscripts, Printed Books

Y-axis: 0, 5, 10, 15, 20, 25, 30

X-axis categories:
- Basic Reformatting of Text (Q89): Almost Always/Often, Sometimes, Seldom/Never
- Mid-level Structural Encoding (Q90): Almost Always/Often, Sometimes, Seldom/Never
- Richer Encoding for Content Analysis (Q91): Almost Always/Often, Sometimes, Seldom/Never
- Scholarly Encoding Projects (Q92): Almost Always/Often, Sometimes, Seldom/Never

Text Encoding Interests & Attitudes in Libraries

How would you rate the level of interest in text encoding by members of your library as whole? (n=112)

No Response
5%

Extremely Interested
1%

Not Applicable
4%

Very Interested
9%

Not at all Interested
14%

Moderately Interested
32%

Slightly Interested
35%

Twitter: @mdalmau #dlbb
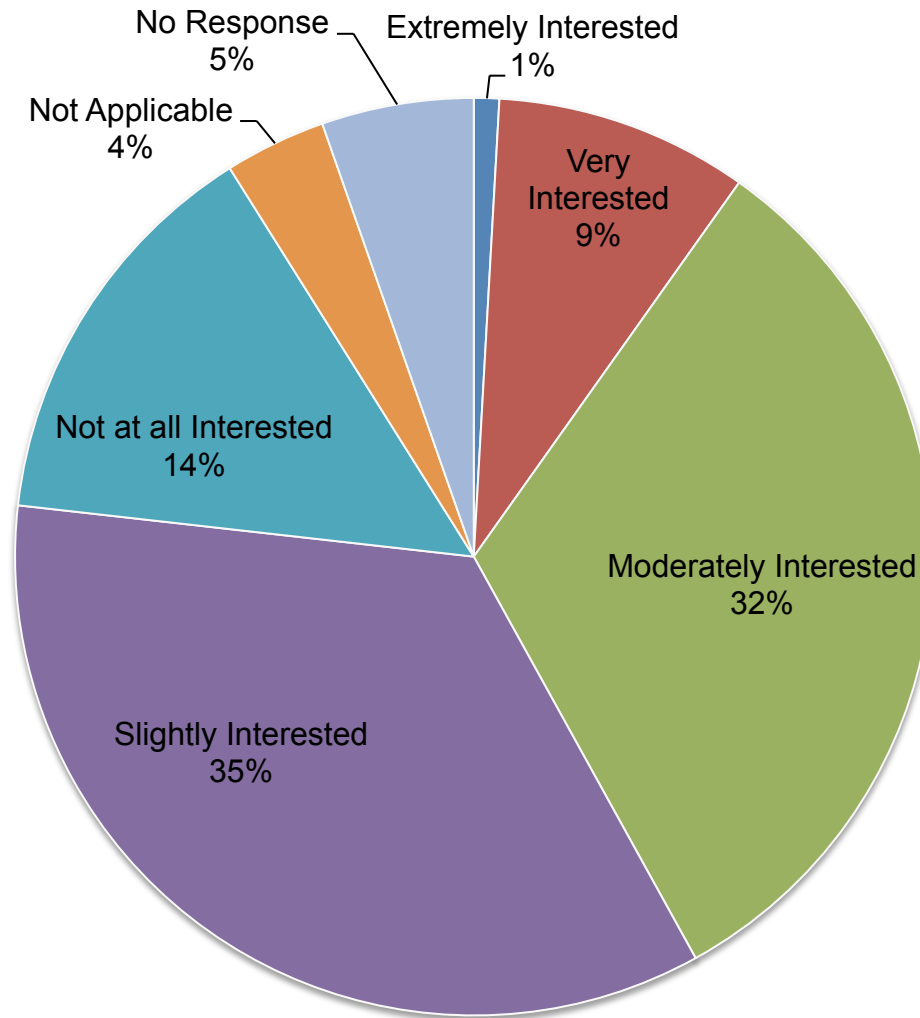
How would you rate your library's administrative support for text encoding projects today? (n=112)

Extremely Supportive
4%

Very Supportive
19%

Moderately Supportive
33%

Slightly Supportive
21%

Not at all Supportive
12%

Not Applicable
5%

No Response
6%

Q5 * Q6 Crosstabulation

Count

Q5	Q6				
	Extremely Interested		Moderately Interested	N/A	Not at all Interested
Q5	6	0	0	0	0
Extremely Supportive	0	1	3	0	1
Moderately Supportive	0	0	17	1	3
N/A	0	0	0	3	0
Not at all Supportive	0	0	3	0	5
Slightly Supportive	0	0	4	0	4
Very Supportive	0	0	9	0	3
Total	6	1	36	4	16

Q5 * Q6 Crosstabulation

Count

Q5	Q6		Total
	Slightly Interested	Very Interested	
Q5	1	0	7
Extremely Supportive	0	0	5
Moderately Supportive	13	3	37
N/A	3	0	6
Not at all Supportive	5	0	13
Slightly Supportive	13	2	23
Very Supportive	4	5	21
Total	39	10	112

Legend: Extremely Supportive, Very Supportive, Moderately Supportive, Slightly Supportive, Not Supportive
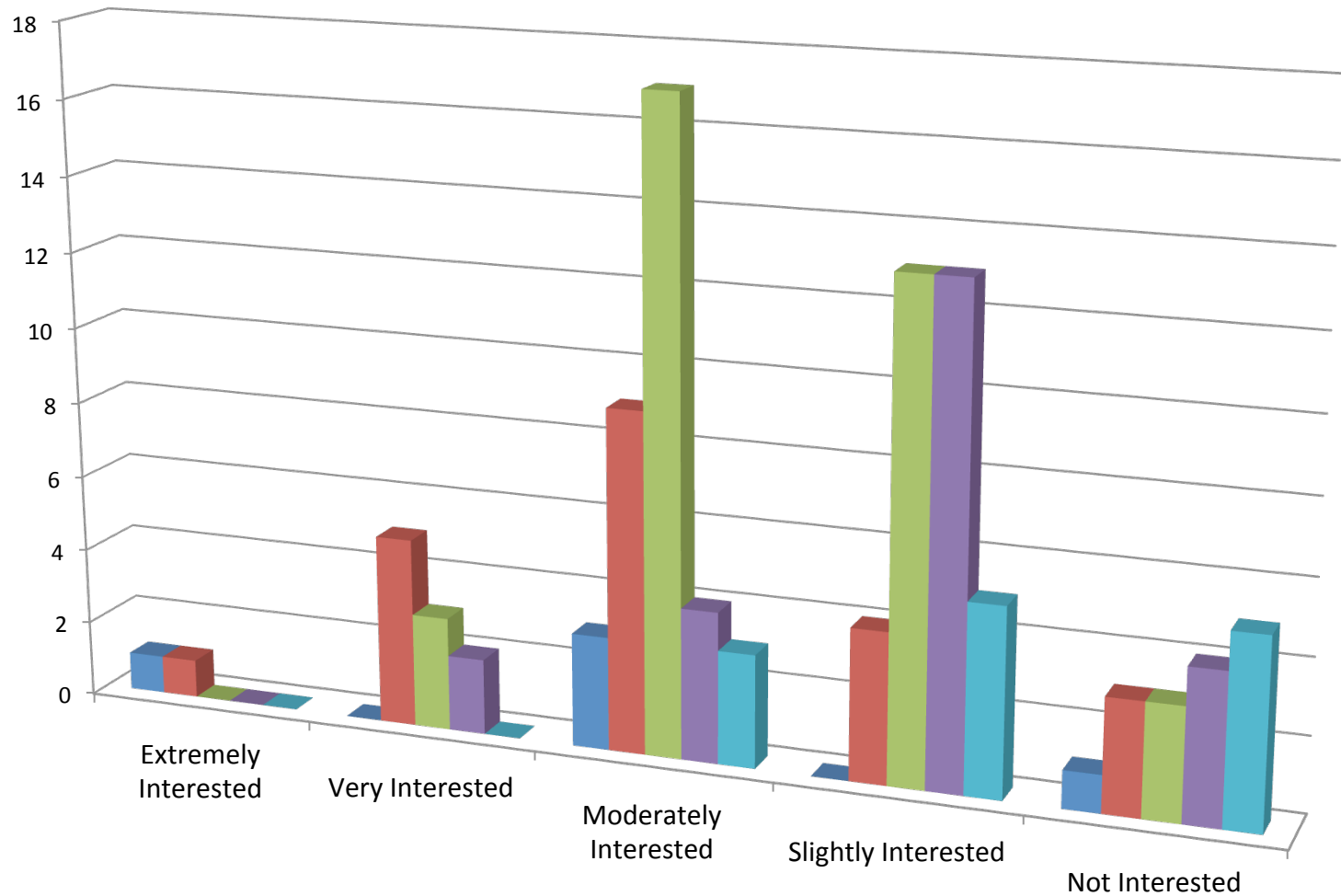
x-axis: Extremely Interested, Very Interested, Moderately Interested, Slightly Interested, Not Interested

Twitter: @mdalmau #dlbb

In a few sentences, could you describe how you see the state of and attitudes toward text encoding in your library today? (n=63)

Neutral
35%

Negative
44%
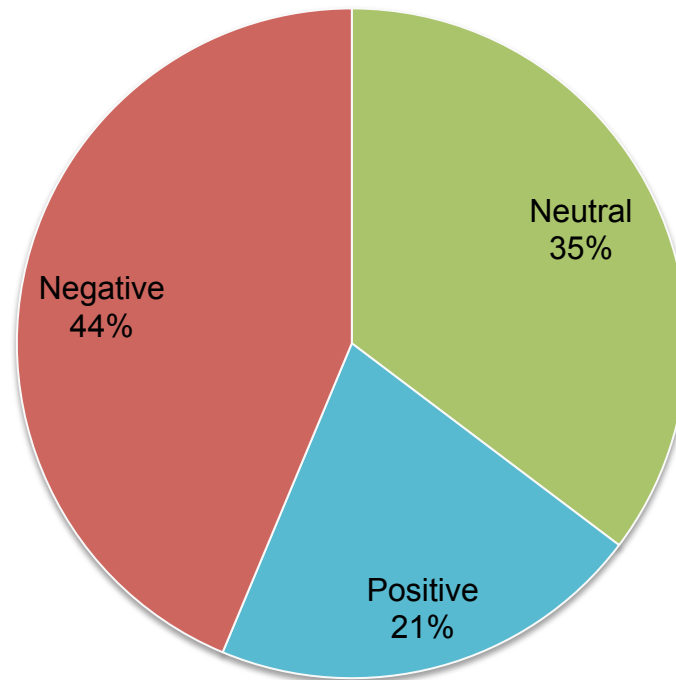
Positive
21%

In a few sentences, could you describe how you see the state of and attitudes toward text encoding in your library today?
Coded Positive
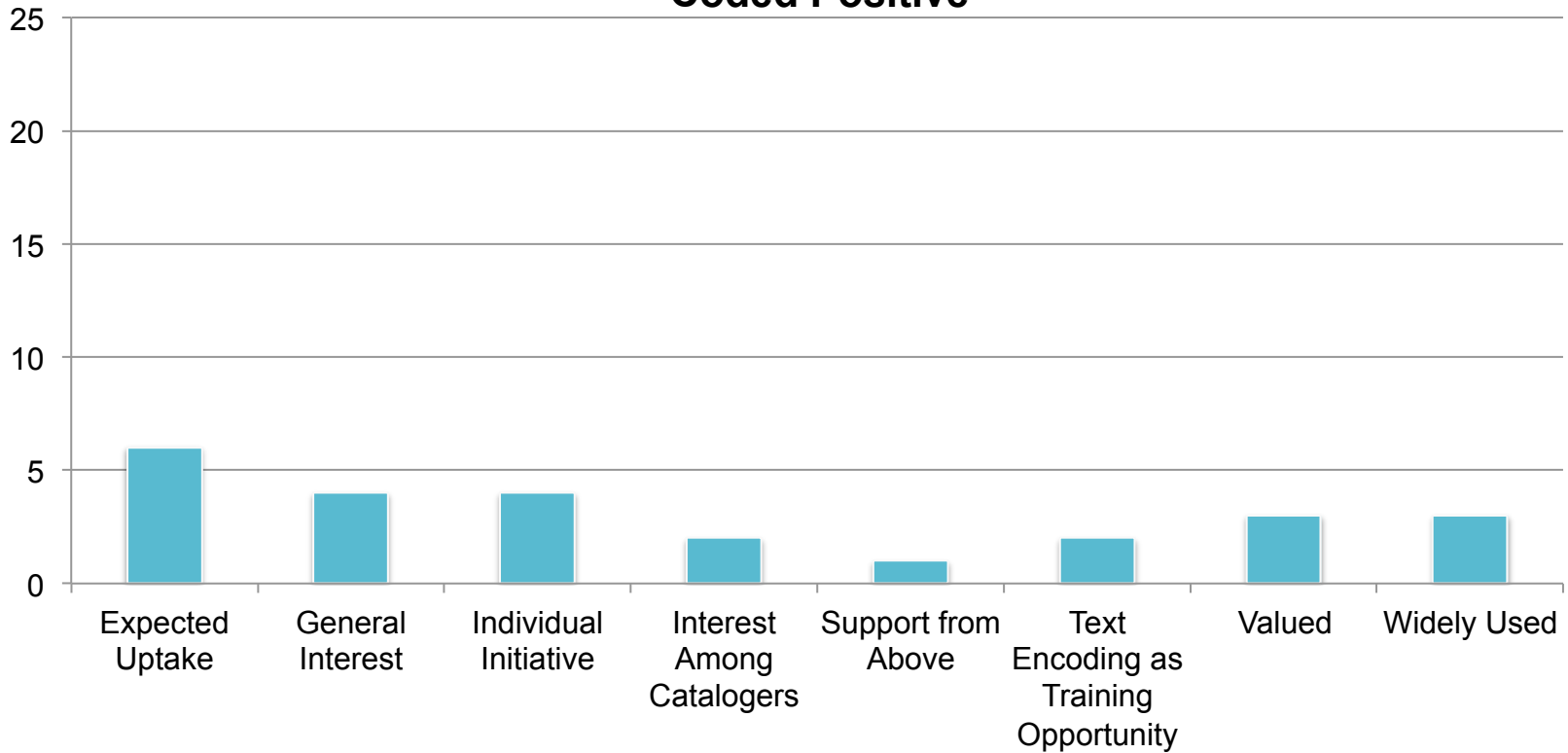
Twitter: @mdalmau #dlbb

In a few sentences, could you describe how you see the state of and attitudes toward text encoding in your library today?
Coded Negative
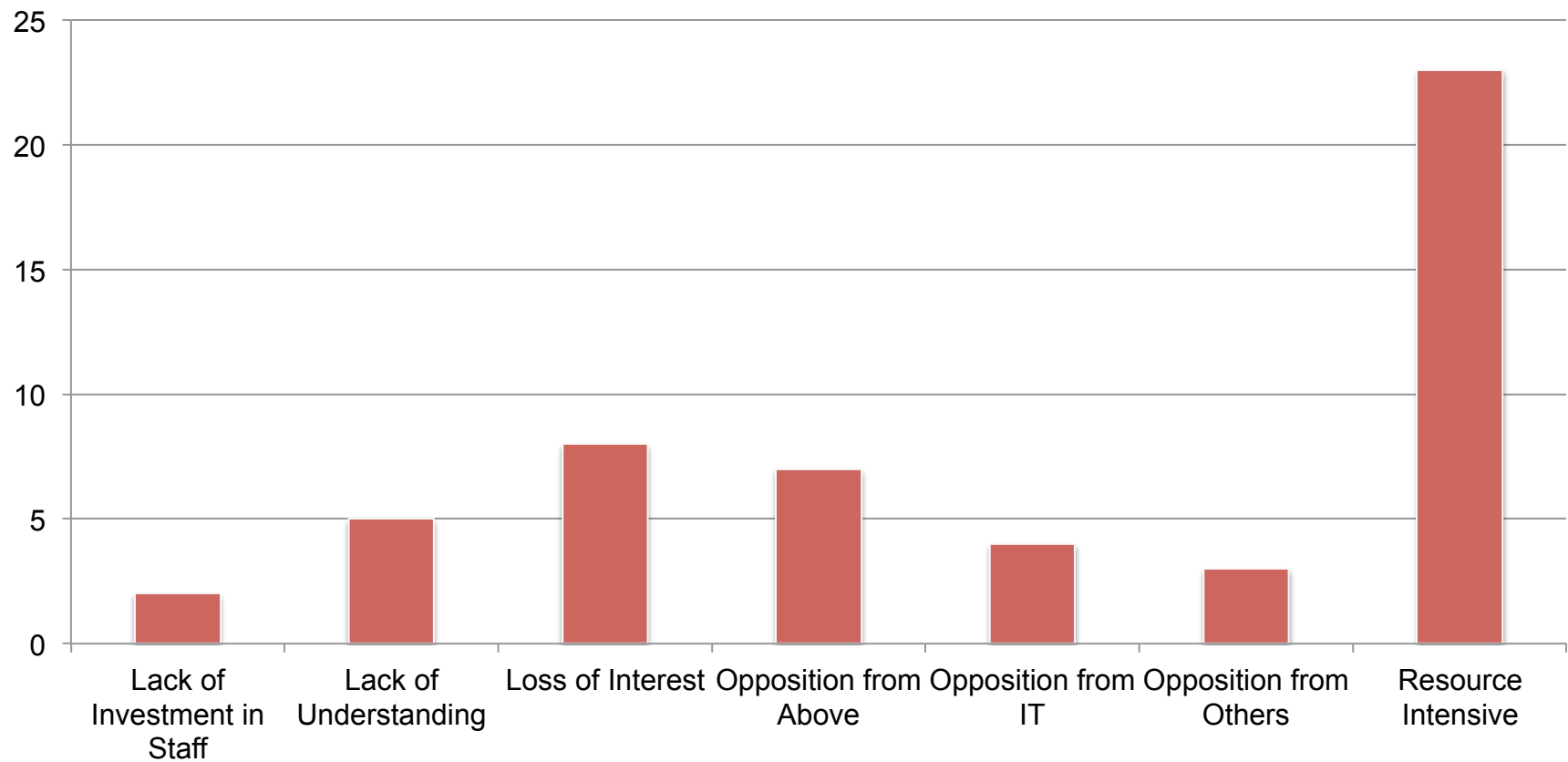
Twitter: @mdalmau #dlbb

In a few sentences, could you describe how you see the state of and attitudes toward text encoding in your library today?
Coded Neutral
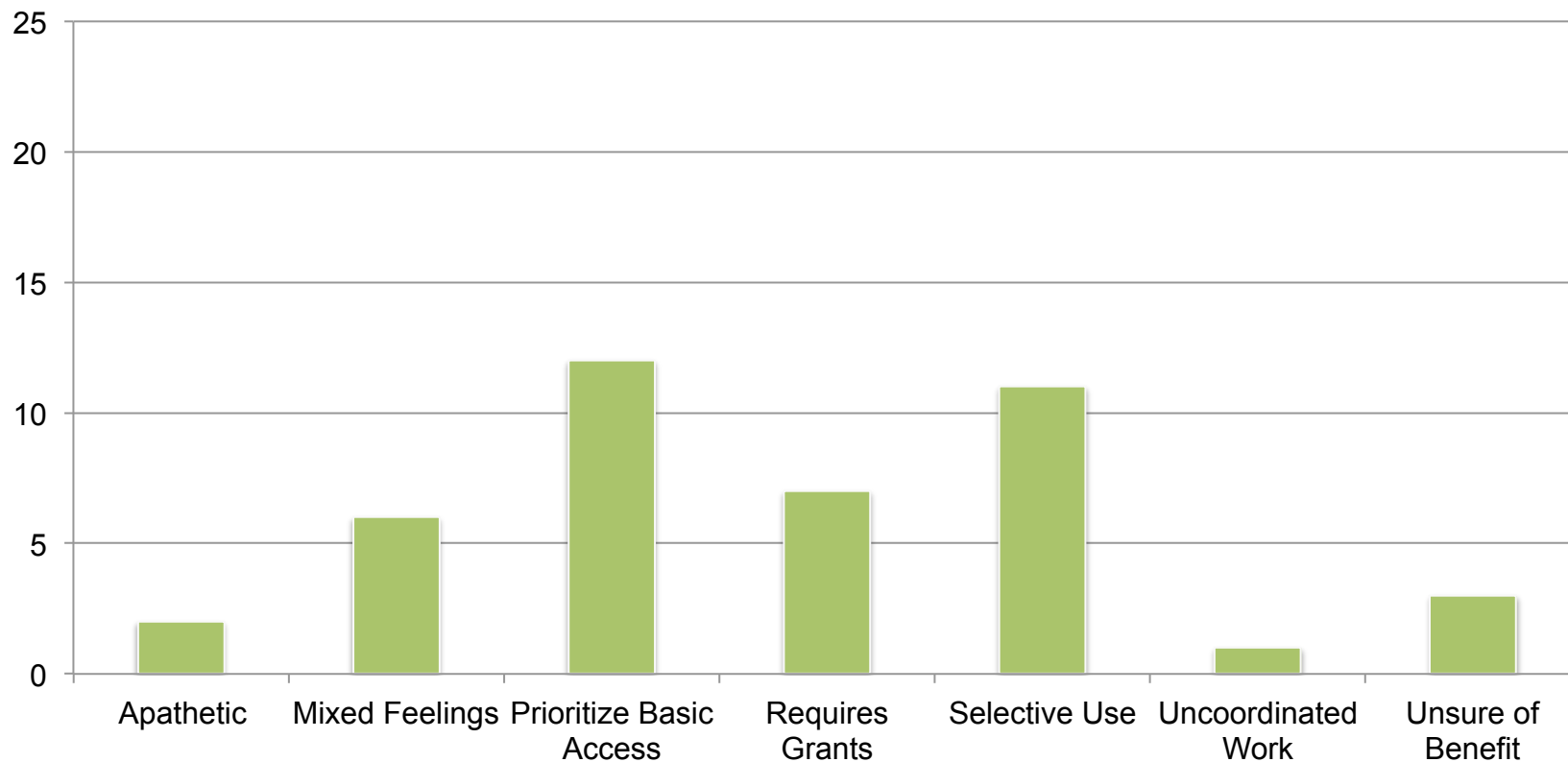
What's Next?

You Ask.

Center for Scholarly Communication & Digital Curation
Northwestern University Library

{ About } { Resources } { Website } { Twitter } { Subscribe }

DIGITAL HUMANITIES

TEI-Nudge or Libraries and the TEI

by JOSH on Oct 1, 2013 • 10:22 am   3 Comments

We are very pleased to welcome this guest post by **Martin Mueller**, *Professor Emeritus of English and Classics at Northwestern University. Read more from Martin at his* Scalable Reading *blog.*

I have put the following proposal to the Board and Technical Council of the TEI. The proposal sits in a tag cloud of Library, Interoperability, Customization, and Outreach. I start with the details of the proposal (which I formulate in terms of 'is' rather than 'should be') and follow them up with some broader reflections.

1. The TEI Consortium assumes responsibilities for the "Best Practices for TEI in Libraries" document (BP hereafter), integrates it fully into the TEI Guidelines, and makes the maintenance of BP a central and continuing responsibility of the Technical Council.
2. BP abandons its current five levels and focuses on the current "Level 4," using as its model a P5 version of the EEBO schema, which has been battle tested with some 50,000 texts from three centuries of Western print culture. (The EEBO schema operates at pretty much the same level of granularity as the Level 4 schema, as do the various Chadwyck-Healey collections.)
3. Maintenance of BP by the TEI goes beyond maintaining the rules of the schema and involves the development and maintenance of a "cradle to grave processing model" or set of default procedures that support the display and querying

ABOUT

Northwestern University Library's Center for Scholarly Communication and Digital Curation supports Northwestern scholars through active exploration of new models for disseminating research results, and by developing programs to support long-term retention and reuse of the scholarly output of the University.

BLOGROLL

ARL Policy Notes
CenterNet
DH Now
HASTAC
Heather Morrison
James Grimmelmann

Martin Mueller, Northwestern University, TEI-Nudge Proposal
http://cscdc.northwestern.edu/blog/?p=872

Twitter: @mdalmau #dlbb

TAPAS: TEI Archiving, Publishing and Access Services
http://www.tapasproject.org

A TEI Project

Best Practices for TEI in Libraries

Author: TEI SIG on Libraries

Editors: Kevin Hawkins, Michelle Dalmau, and Syd Bauman

2011-10

Version 3.0 (October 2011)

Table of contents

Best Practices for TEI in Libraries
http://purl.org/TEI/teiinlibraries

Twitter: @mdalmau #dlbb

Questions?

Thanks!