TEI Tite Survey Questionnaire Data Analysis

1. Are you part of a TEI member organization or project?

57.8 % of respondents are a part of a TEI member organization or project. Statistically this is approximately half.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	N	35	42.2	42.2	42.2
	Y	48	57.8	57.8	100.0
	Total	83	100.0	100.0	

	N	Minimum	Maximum	Sum	Mean		Std. Deviation
					Statistic	Std. Error	
Q1	83	-1	1	13	.16	.109	.994

H_o: (#Y) – (#N) = 0 (or average = 0)

One-Sample t-Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Q1	1.436	82	.155	.157	-.06	.37

Chi-Squared

	Q1
Chi-Square	2.036[a]
df	1
Asymp. Sig.	.154

	Observed N	Expected N	Residual
-1	35	41.5	-6.5
1	48	41.5	6.5
Total	83		

3. Does your organization convert print or manuscript materials to machine-readable text?

84.3 % (a statistically significant proportion, which is to be expected by the nature of the participant pool) do convert materials to machine-readable text.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	N	13	15.7	15.7	15.7
	Y	70	84.3	84.3	100.0
	Total	83	100.0	100.0	

	N	Minimum	Maximum	Sum	Mean		Std. Deviation
					Statistic	Std. Error	
Q3	83	-1	1	57	.69	.080	.731

H_o: (#Y) – (#N) = 0 (or average = 0)

One-Sample t-Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Q3	8.555	82	.000	.687	.53	.85

Chi-Squared Test

	Observed N	Expected N	Residual
-1	13	41.5	-28.5
1	70	41.5	28.5
Total	83		

	Q3
Chi-Square	39.145a
df	1
Asymp. Sig.	.000

3.a. If YES, which of the following methods does it primarily use for full text generation?
(1 = Manual Keyboarding; 2 = OCR Scanning; 3 = OCR Scanning and Manual Keyboarding; 4 = Digitization Is Locally Outsourced)

3.b. If NO, is it because of any of the following? (check all that apply)
(1 = Too Expensive; 2 = Lack of Expertise; 3 = Lack of Partnerships)

Manual Keyboarding is the most common method for full text generation. Only 14.1% of respondents, or 10 respondents total, currently locally outsource digitization.

The most common reason for not converting text was Lack of Expertise; however, this is not significantly more than the other two options.

Statistics

		Q3a	Q3b1	Q3b2	Q3b3
N	Valid	71	11	14	10
	Missing	12	72	69	73

Q3a

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	25	30.1	35.2	35.2
	2	8	9.6	11.3	46.5
	3	28	33.7	39.4	85.9
	4	10	12.0	14.1	100.0
	Total	71	85.5	100.0	
Missing	System	12	14.5		
Total		83	100.0		

	Observed N	Expected N	Residual
1	25	17.8	7.2
2	8	17.8	-9.8
3	28	17.8	10.2
4	10	17.8	-7.8
Total	71		

	Q3a
Chi-Square	17.620b
df	3
Asymp. Sig.	.001

Q3b1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	11	13.3	100.0	100.0
Missing	System	72	86.7		
Total		83	100.0		

Q3b2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	14	16.9	100.0	100.0
Missing	System	69	83.1		
Total		83	100.0		

Q3b3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	10	12.0	100.0	100.0
Missing	System	73	88.0		
Total		83	100.0		

4. Would your organization become (or remain) a TEI member if doing so offset the cost of digitization?

A significant proportion (80.7% of respondents) indicated that they would become or remain a TEI member if doing so offset the cost of digitization.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		7	8.4	8.4	8.4
	N	9	10.8	10.8	19.3
	Y	67	80.7	80.7	100.0
	Total	83	100.0	100.0	

Descriptive Statistics

	N	Minimum	Maximum	Sum	Mean		Std. Deviation
					Statistic	Std. Error	
Q4	76	-1	1	58	.76	.075	.651

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Q4	10.228	75	.000	.763	.61	.91

Chi-Squared

	Observed N	Expected N	Residual
-1	9	38.0	-29.0
1	67	38.0	29.0
Total	76		

	Q4
Chi-Square	44.263a
df	1
Asymp. Sig.	.000

5. How many pages of primary material does your organization digitize in a year?

Although the variation is wide, with anywhere from 8 million to zero, the average number of pages of primary materials digitized by the organization reporting is 176,001.

Descriptive Statistics

	N	Range	Minimum	Maximum	Sum	Mean		Std. Deviation
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic
Q5Converted	66	8000000	0	8000000	11616070	176001.06	122432.220	994644.061

Q6. Which markup does your organization use for text encoding?
(& in gray if available) Why has it selected this(these) markup(s)?

Total entries (n): 68
times TEI or TEI-Lite appears: 41
times XML appears: 17
Others: HTML (2); MEP DTD (1); METS (2); OAI (1); EAD (2); Dublin core (1); Text Class (1); DJVU (1)

TEI	
Most appropriate to the content, poetry	
xml	
XML via CONTENTdm	
Standardization	
TEI lite-based (custom)	
unclear: i.e. markup language? TEI-conformant XML	
none	
unknown/various	
P5 full	
more flexibility in representing data	
TEI Lite P4	
TEI Lite is suitable for vast majority of our requirements, and we haven't yet the time / resource to convert to P5	
XML	
Industry standard	
MEP DTD	
We were part of a grant-funded project to develop it.	
TEI P5	
METS, TEI, EAD, OAI	
For practical reasons. Handcoded material is encoded in TEI, OCRd texts have simpler encodings.	
TEI P4	
It fits best our needs.	
TEI P5	
De facto standard	
XML, HTML	
convenience, flexibilty	
TEI	
TEI	
durability, scientific hygiene	
xml	
current technology	
tei	
TEI P5 XML	
Because it's best suited for our material	
html, tei-xml	
None	
TEI	
It's the standard, it's got a chance for survival, and we've developed a set of tools for online, interoperable publication around it	
TEI	

It's the clear standard for humanities projects

TEI level 4-lite

Work started a few years ago, although we're considering switching to level 5 encoding

xml/ fedora

what other projects that we collaborate with use

xml

private

To provide students the means of collating, marking, and hypertextualizing text.

TEI

We are TEI fans.

XML

standardization of document file formats; multiple uses (data separated from processing); independence of local platforms; well-structured data division

personal xml

convenient to import into SQL database

currently TEI P4

We haven't moved our recently completed materials to P5; current development uses the same schema.

TEI

Because using the TEI we can identify off the textual structures that we need for scholarly editions.

various

don't know how to do it.

METS + MODS inside

To be interoperable with the German national infrastructure

TEI

de facto standard in humanities

FUTURE, TEI LITE

OUR NEW DIGITAL PARTNER IS A LIBRARY AND THEY ARE ALREADY USING TEI LITE

TEI

staff knowledge; discipline appropriate (humanities materials)

TEI

TEI

TEI

Standard for humanities, flexibility.

TEI P5

Quite accurate for Renaissance books, but must be adapted to specificities of old prints

TEI P5

TEI mostly

Established standard; flexible and suitable for wide variety of texts

TEI and EAD

TEI

it is well supported

SGML/XML following TEI guidelines

Well suited for the historical and literary source material

TEI P5

outstanding community participation and support

TEI, EAD, none

TEI P5

standard

dublin Core

standard and easy to use

TEI P4

TEI P 5

latest standard

TEI

Semantic richness & status of TEI as lingua franca.

Text Class (derived from TEI Lite P3)

This was developed in the '90s for use in our delivery system, and we continue to use it because our system requires it.

TEI-Lite, and internal schemes

We see TEI as both standards-based, and as something we have long experience using it. Non-standard schemes were used both before TEI was available, and since we've introduced new digitazation workflows that don't come with TEI capabilities out-of-box.

XML

the projects we are participating in have required so

TEIlite

suits workflow

XML + TEI

platform-independent, machine-readable, international standard,...

TEI-XML

de facto standard for humanities / social sciences

TEI

because it's the standard, and it works.

Various formats and markups

Pragmatic reasons, efficiency, purpose of digitized text.

Xmetal

TEI P4 and EAD

library practices

um..., TEI

Moving towards TEI/xml from html

XML-ABBYY 6.0 and DJVU

Because our digitizing partner uses them.

7. Does your organization outsource digitization?

Approximately half of the respondents outsource digitization currently.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	1	1.2	1.2	1.2
	N	46	55.4	55.4	56.6
	Y	36	43.4	43.4	100.0
	Total	83	100.0	100.0	

Descriptive Statistics

	N	Minimum	Maximum	Sum	Mean		Std. Deviation
					Statistic	Std. Error	
Q7	82	-1	1	-10	-.12	.110	.999

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Q7	-1.106	81	.272	-.122	-.34	.10

	Observed N	Expected N	Residual
-1	46	41.0	5.0
1	36	41.0	-5.0
Total	82		

	Q7
Chi-Square	1.220[a]
df	1
Asymp. Sig.	.269

8. Does your organization have a regular outsourcing workflow for digitization projects?

Half of the respondents answering this question have a regular outsourcing workflow.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		36	43.4	43.4	43.4
	N	23	27.7	27.7	71.1
	Y	24	28.9	28.9	100.0
	Total	83	100.0	100.0	

Descriptive Statistics

	N	Minimum	Maximum	Sum	Mean		Std. Deviation
					Statistic	Std. Error	
Q8	47	-1	1	1	.02	.147	1.011

One-Sample t-Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Q8	.144	46	.886	.021	-.28	.32

	Observed N	Expected N	Residual
-1	23	23.5	-.5
1	24	23.5	.5
Total	47		

	Q8
Chi-Square	.021a
df	1
Asymp. Sig.	.884

8.a. If YES, would it alter its practices for a price reduction in full text generation and TEI (Tite) markup?

Approximately 2/3 (statistically more than half) of the respondents would alter their practice for a price reduction

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		53	63.9	63.9	63.9
	N	8	9.6	9.6	73.5
	Y	22	26.5	26.5	100.0
	Total	83	100.0	100.0	

Descriptive Statistics

	N	Minimum	Maximum	Sum	Mean		Std. Deviation
					Statistic	Std. Error	
Q8a	30	-1	1	14	.47	.164	.900

One-Sample t-Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Q8a	2.841	29	.008	.467	.13	.80

	Observed N	Expected N	Residual
-1	8	15.0	-7.0
1	22	15.0	7.0
Total	30		

	Q8a
Chi-Square	6.533a
df	1
Asymp. Sig.	.011

8. b. If NO or maybe, please explain.

survey needs not known or it depends category: we do both in-house and outsourced work, usually though not always with minimal mark-up for textual sources. We have no institutional stance, at present, as to what's actually needed and cost-effective.

We have requirements additional to those expressed in the TEI Lite DTD, which concern preference for certain tags, treatment of certain material etc, and we need to ensure that these guidelines of ours are followed.

We're currently experimenting with our first outsourcing contracts.

Department of Old Bulgarian Literature, Institute of Literature, Bulgarian Academy of Sciences has no outsourcing at the moment, but it is looking for. Bulgarian libraries with collections of medieval manuscripts have very limited budget and no financial possibility to support digization and presetvation of written cultural heritage. We hope it will be possible in the future, with EC funding.

No reliable service ... too many problems with the texts produced. It's proved better to train project staff who work page by page.

The real answer is: we do most of the digitization inside the institution, but this is becoming more and more difficult, so occasionally we have some outsourcing, which, again, is considerably expensive in my country.

Possibly, bu TEI is overkill for the type of markup we need. Yes, many libraries do rely on TEI for a lot of their markup, but fewer of us are doing bound volumes now.   We're heavily focused on newspapers and archival manuscripts.

See 5a

we might alter practices for specific projects

We outsource mostly scanning and rarely keyboarding. Either would need to be cheaper than our current costs to make it worthwhile to switch.

Outsourcing workflows are currently negotiated and designed on a per-project basis.

no or maybe - our current vendor offers a range of associated servies and as a known long term client we benefit from that. However a full TEI outsource service may suit our needs

One project differs from another.

We might. We are a publisher, and most of our conversion from printed text (to PDF) is done for the purposes of making books available POD. We do not at this time convert text, printed or electronic files, to archival eformats.

I would have to see the workflows and their impact on our own.

9. At what contract price does your institution require a Request for Proposal (RFP) to be issued?

Taking only the smallest number into account when the participant gave a range of values the contract prices range from $30 to $75,000 with an average of $14,512, although some of these are in non-US currency and were not converted since there is no way to know for sure if any are in U.S. Dollars.

Descriptive Statistics

	N	Range	Minimum	Maximum	Sum Statistic	Mean Statistic	Mean Std. Error	Std. Deviation
Q9Converted	15	74970	30	75000	217680	14512.00	5399.427	20911.892

9.a. How, if at all, do discounts affect the price at which your institution requires an RFP to be issued?

Participants answering this question primarily did not know if or how a discount would affect the price at which the institution would require an RFP to be issued.

don't know

not at all

not

don't know

i'm not sure

I do not know yet

Not sure. Regardless of the discount, any job that costs more than $20,000 may be subject for bidding even if discounted rates have been pre-negotiated. Presumably the bid will go to the discounted vendor in the end anyway.

they do not afferct

we must go through a competitive bidding process for goods/services costing more than 5000 USD unless we are adding on to a previous contract

(unsure)

Unknown.

not sure

We don't digitize, but because we want to I'm filling out this section

Don't know

it doesn't

they do not

10. How many pages of digitization does your organization currently outsource in a year?

Taking only the smallest number into account when a range of values were given, pages of outsourced digitization per year currently ranges from 0 to 250,000 pages with an average value of 40,901.

Descriptive Statistics

	N	Range	Minimum	Maximum	Sum	Mean		Std. Deviation
						Statistic	Std. Error	
Q10Converted	27	250000	0	250000	1104320	40900.74	13487.204	70081.569

11. Given the following pricing options, how many pages of digitization would your organization outsource?
a. If it were free?   b. If the cost was $0.50 per page?
c. If the cost was $1.00 per page?   d. If the cost was $2.00 per page?

Answers to these questions were given in many formats and ranges making statistical analysis difficult. Submitted values are therefore listed below. It is clear that the value decreases (significantly in some cases) as the cost increases. The difference between 'Free' and '$0.50 / page" is often nothing though.

Free	a lot
$0.50 / page	10-15,00
$1.00 / page	5-10,000
$2.00 / page	1-5000
Free	all
Free	oh my, 3
$0.50 / page	20,000
$1.00 / page	10,000
$2.00 / page	10,000
Free	75,000
$0.50 / page	75,000
$1.00 / page	0
$2.00 / page	0
Free	unsure
$0.50 / page	nil
$1.00 / page	nil
$2.00 / page	nil
Free	1000
$0.50 / page	500
$1.00 / page	500
$2.00 / page	500
Free	10,000
Free	10000
$0.50 / page	10000
$1.00 / page	5000
$2.00 / page	0
Free	1000
$0.50 / page	500
$1.00 / page	300
$2.00 / page	300
Free	1000
$0.50 / page	500
$1.00 / page	250
$2.00 / page	150
Free	
$0.50 / page	1000
$1.00 / page	
$2.00 / page	
Free	1000
$0.50 / page	1000

$1.00 / page	1000
$2.00 / page	1000
Free	3000
$0.50 / page	300
$1.00 / page	
$2.00 / page	
Free	200.000
$0.50 / page	
$1.00 / page	
$2.00 / page	
Free	30000
$0.50 / page	9000
$1.00 / page	9000
$2.00 / page	none
Free	20,000
$0.50 / page	10,000
$1.00 / page	5,000
$2.00 / page	2,500
Free	30,000
$0.50 / page	AS NEEDE
$1.00 / page	0
$2.00 / page	0
Free	10,000 -
$0.50 / page	10,000
$1.00 / page	5,000
$2.00 / page	5,000
Free	all
$0.50 / page	all
$1.00 / page	5000
$2.00 / page	0
Free	all!
$0.50 / page	most
$1.00 / page	few
$2.00 / page	none
Free	30000
$0.50 / page	24000
$1.00 / page	12000
$2.00 / page	6000
Free	all proj
$0.50 / page	7500

$1.00 / page	<7500
$2.00 / page	<7500
Free	everythi
$0.50 / page	everythi
$1.00 / page	everythi
$2.00 / page	1/2
Free	everythi
$0.50 / page	0
$1.00 / page	0
$2.00 / page	0
Free	unknown
Free	5000
$0.50 / page	?
$1.00 / page	?
$2.00 / page	0
Free	100,000
$0.50 / page	40,000
$1.00 / page	25,000
$2.00 / page	25,000
Free	unlimite
$0.50 / page	4000
$1.00 / page	3000
$2.00 / page	1500
Free	300,000
$0.50 / page	20,000
$1.00 / page	10,000
$2.00 / page	
Free	50,000
$0.50 / page	10,000
$1.00 / page	don't kn
$2.00 / page	don't kn
Free	70,000-1
$0.50 / page	50,000-7
$1.00 / page	30,000
$2.00 / page	0
Free	all
Free	3000 pa
$0.50 / page	1000 pa
$1.00 / page	none
$2.00 / page	none

Free	more, if
$0.50 / page	400,000
$1.00 / page	200,000
$2.00 / page	100,000

12. Would an appropriately priced service encourage your organization to digitize more material in a given year?

A statistically significant proportion (75.9% of respondents) answered yes and would digitize more at the right price service.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		8	9.6	9.6	9.6
	N	12	14.5	14.5	24.1
	Y	63	75.9	75.9	100.0
	Total	83	100.0	100.0	

Descriptive Statistics

	N	Minimum	Maximum	Sum	Mean		Std. Deviation
					Statistic	Std. Error	
Q12	75	-1	1	51	.68	.085	.738

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Q12	7.978	74	.000	.680	.51	.85

	Observed N	Expected N	Residual
-1	12	37.5	-25.5
1	63	37.5	25.5
Total	75		

	Q12
Chi-Square	34.680a
df	1
Asymp. Sig.	.000

13. What is your organization's annual or project budget for digitization?

The values are presented in different currencies and in different forms. The values are listed below.

0
not broken down
1000
N/A
grant-based=varies
50,000+/-
contingent on grants
100,000
$15-25K
$1500
3000000 Euro
$150000
?
no fixed budget
funding dependent
20,000
unknown
None
$15K-$20K
$2,000
1000
relies upon grants
3000-5000 U$
2000 euro
varies
$300,000
Unbudgeted
as needed
approx. 100.000
depends
0
NOT YET BUDGETED
100,000
10000
$120,000 (NZ)
1000
40,000
40,000
12000 $
None
unset

0
grant based
varies
n/a
varies
35.000 euros
varies by project

+- 4000 USD
300
ca. $40K but varies
depends
none right now
$10,000
0
30,000-40,000
$200,000

14. What are the typical formats of materials your organization digitizes?

The most common format with 78.3% of respondents was Print.
Next was Handwritten documents at 56.6% of respondents.
Least was Microfilm at 27.7% of respondents.

Q14aPrint

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	65	78.3	100.0	100.0
Missing	System	18	21.7		
Total		83	100.0		

Q14bHandwritten

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	47	56.6	100.0	100.0
Missing	System	36	43.4		
Total		83	100.0		

Q14cMicrofilm

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	23	27.7	100.0	100.0
Missing	System	60	72.3		
Total		83	100.0		

15. Does your organization digitize pre-19th century materials?

Approximately half of the respondents (48.2%, 40 respondents) indicated that they do digitize pre-19th century materials.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		8	9.6	9.6	9.6
	N	35	42.2	42.2	51.8
	Y	40	48.2	48.2	100.0
	Total	83	100.0	100.0	

	N	Minimum	Maximum	Sum	Mean		Std. Deviation
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic
Q15	75	-1	1	5	.07	.116	1.004

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Q15	.575	74	.567	.067	-.16	.30

	Observed N	Expected N	Residual
-1	35	37.5	-2.5
1	40	37.5	2.5
Total	75		

	Q15
Chi-Square	.333a
df	1
Asymp. Sig.	.564

16. On average what price does your organization pay for digitization per page for each of the following types of material?
a. For post-19th century material?
b. For pre-19th century material?
c. For manuscript material?

Answers were given in different formats (e.g. per page or per document) and possibly in different currencies, so direct comparison, outlines were not conducted. The values given are listed below.

a. $3.50/page
b.
c. ?

a. $0.25/pg
b.
c.

a.   .15
b.
c.

a.
b. $1.50/page
c.

a.   NZD $0.65
b.
c.

a.
b.   don't know
c.

a.   don't know
b.   don't know
c.   don't know

a.   n/a
b.   n/a
c.   n/a

a.   0
b.   ?
c.   ?

a.   -
b.   -
c.   -

a. $12/hour
b. $15/hour
c.

a. $5/page

b.
c. $10/page

a.   300
b.
c.

a.
b.
c.   no price

a.   40 cents a page
b.   40 cents a page
c.   unpredictable

a.
b.
c. $20/shot

a. 0.10 - 0.30 Euros
b.
c.

a.   see comments
b.   see comments
c.   see comments

a.   UNKNOWN
b.
c.

a.   imaging 6$/page
b.   imaging 6$/page
c.   imaging 6$/page

a.
b.   2
c.

a.   .50
b.
c.   .75

a.
b.   2 $
c.   6 $

a. $0.25/page
b.   done in-house
c.   done in-house

a. $1.50 per page
b. $1.5 per page
c. $10/page

a.   n/a
b. $0.15
c.   n/a

a.   n/a
b.   n/a
c.   n/a

a.   ---
b.   ----
c.   ----

a. $1.25
b. $1.50
c.   NA

a. $0.50 per page
b.
c.   0

a.   1.00
b.   1.00
c.   varies

a.   .15, but only OCR
b.   we don't do this
c.   we don't do this

17. Does your organization digitize materials in non-Roman character sets?

Approximately half of the respondents (34 individual, 41%) digitize materials in non-Roman character sets.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	-1	38	45.8	52.8	52.8
	1	34	41.0	47.2	100.0
	Total	72	86.7	100.0	
Missing	System	11	13.3		
Total		83	100.0		

	N	Range	Minimum	Maximum	Sum Statistic	Mean Statistic	Mean Std. Error	Std. Deviation
Q17	72	2	-1	1	-4	-.06	.118	1.005

One-Sample t- Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower	Upper
Q17	-.469	71	.641	-.056	-.29	.18

	Observed N	Expected N	Residual
-1	38	36.0	2.0
1	34	36.0	-2.0
Total	72		

	Q17
Chi-Square	.222a
df	1
Asymp. Sig.	.637

18. Does your organization digitize materials in languages other than English?

A significant proportion (68.7% or 57 of 83 respondents) of respondents digitize materials in languages other than English

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		9	10.8	10.8	10.8
	N	17	20.5	20.5	31.3
	Y	57	68.7	68.7	100.0
	Total	83	100.0	100.0	

Descriptive Statistics

	N	Minimum	Maximum	Sum	Mean		Std. Deviation
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic
Q18	74	-1	1	40	.54	.098	.847

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Q18	5.489	73	.000	.541	.34	.74

	Observed N	Expected N	Residual
-1	17	37.0	-20.0
1	57	37.0	20.0
Total	74		

	Q18
Chi-Square	21.622a
df	1
Asymp. Sig.	.000

19. Does your organization need to digitize either of the following types of content?
a. Newspapers
b. Serials

Approximately half of the respondents (42.2% and 53%) need to digitize both types of content.

Q19aYesNo

	Frequency	Percent	Valid Percent	Cumulative Percent
	8	9.6	9.6	9.6
N	40	48.2	48.2	57.8
Y	35	42.2	42.2	100.0
Total	83	100.0	100.0	

Q19bYesNo

	Frequency	Percent	Valid Percent	Cumulative Percent
	9	10.8	10.8	10.8
N	30	36.1	36.1	47.0
Y	44	53.0	53.0	100.0
Total	83	100.0	100.0	

Descriptive Statistics

	N	Minimum	Maximum	Sum	Mean Statistic	Mean Std. Error	Std. Deviation
Q19a	75	-1	1	-5	-.07	.116	1.004
Q19b	74	-1	1	14	.19	.115	.989

One-Sample Test

	Test Value = 0 t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper
Q19a	-.575	74	.567	-.067	-.30	.16
Q19b	1.646	73	.104	.189	-.04	.42

Q19a

	Observed N	Expected N	Residual
-1	40	37.5	2.5
1	35	37.5	-2.5
Total	75		

Q19a

	Observed N	Expected N	Residual
-1	40	37.5	2.5
1	35	37.5	-2.5
Total	75		

Q19b

	Observed N	Expected N	Residual
-1	30	37.0	-7.0
1	44	37.0	7.0
Total	74		

	Q19a	Q19b
Chi-Square	.333a	2.649b
df	1	1
Asymp. Sig.	.564	.104

20. What is the most common organization of your originals?

1 = Sequential, Paginated
2 = Sequential, Unpaginated
3 = Nonsequential

The primary form of organization for originals was Sequential, Paginated, accounting for 72.3% of respondents.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		8	9.6	9.6	9.6
	Seq Pag	60	72.3	72.3	81.9
	SeqUnPag	13	15.7	15.7	97.6
	UnSeq	2	2.4	2.4	100.0
	Total	83	100.0	100.0	

21. What does your organization consider to be a reasonable turnaround time for digitizing materials through vendors?
1 = Less than 1 month
2 = 1 to 6 Months
3 = 6 Months to 1 Year

The turnaround time varied, but the majority of 59% was between 1 and 6 months. Only 4.8% or 4 of 83 indicated that it took more than 1 year.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		16	19.3	19.3	19.3
	<1Mon	14	16.9	16.9	36.1
	<1Year	4	4.8	4.8	41.0
	<6Mon	49	59.0	59.0	100.0
	Total	83	100.0	100.0	

22. Should TEI mediate with vendors on behalf of its subscribers, or would your organization prefer to deal with them directly in either of the following matters?

a. Quality control

　　　　Deal with TEI

　　　　Deal with Vendor

b. Aggregation

　　　　Deal with TEI

　　　　Deal with Vendor

In terms of quality control, approximately half of the respondents would deal with TEI, while the other half would prefer to deal directly with the vendor. Differently though, in terms of aggregation, a slightly significant proportion (44 individuals or 53% or total respondents) would prefer to work through TEI (Chi-Squared test was not significant at 95% confidence though).

Q22aTEIVendor

	Frequency	Percent	Valid Percent	Cumulative Percent
	11	13.3	13.3	13.3
N	38	45.8	45.8	59.0
Y	34	41.0	41.0	100.0
Total	83	100.0	100.0	

Q22bTEIVendor

	Frequency	Percent	Valid Percent	Cumulative Percent
	11	13.3	13.3	13.3
N	28	33.7	33.7	47.0
Y	44	53.0	53.0	100.0
Total	83	100.0	100.0	

	N	Minimum	Maximum	Sum	Mean		Std. Deviation
					Statistic	Std. Error	
Q22a	72	-1	1	-4	-.06	.118	1.005
Q22b	72	-1	1	16	.22	.116	.982

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Q22a	-.469	71	.641	-.056	-.29	.18
Q22b	1.920	71	.059	.222	.00	.45

Q22b

	Observed N	Expected N	Residual
-1	28	36.0	-8.0
1	44	36.0	8.0
Total	72		

Q22a

	Observed N	Expected N	Residual
-1	38	36.0	2.0
1	34	36.0	-2.0
Total	72		

	Q22a	Q22b
Chi-Square	.222[e]	3.556[e]
df	1	1
Asymp. Sig.	.637	.059

Additional Comments

One-person operation, independent.

I would like to see some customization available to the tite scheme.

As mentioned above, we tend to have additional encoding requirements, which vary from project to project, and we have to communicate closely with our suppliers to ensure that they are able to follow these when performing transcription for us. as such, I think a negotiated deal could work for us, as long as we could still communicate directly with the suppliers regarding delivery and turnaround of work and quality standards etc.

I think my project not really appropriate for vendor markup.

It is a wonderful idea for TEI to negotiate discounts with vendors. My organization would definitely be encouraged to increase its output if it could form a partnership with a vendor. My only concern would be the quality control of foreing language materials.

it's hard for me to answer some of these questions as I'm at a new organization but I know there could be quite a lot to digitize.

We are interested in vendors that could digitize also hebrew and arabic texts and texts in black letter.

I am working on a project as an individual scholar. My university funds these efforts only sporadically, but with more help in organizing the scanning & TEI, I can target grants both within my university and through larger organizations (e.g., NEH).

So far, I don't even know where to go to outsource my scanning & TEI. I need the most help in this way.

We are based in Taiwan and usually deal with Chinese material. However we are planning a project digitizing English, French and German print material next year.

It TEI could mediate with vendors in Europe it would be highly welcome.

The informational massive of the project REPRETORIUM of Old Bulgarian literature and letters (member of TEI) is analytical description of over 350 Slavic manuscripts from 11th-17th c. Description contains incipita, explicita and entire texts in Old Church Slavonic. The Repertorium is reserch project, not commercial. The outcome is free for schollars, students, citizens, etc.

I've used the services of two vendors and haven't been unsatisfied, but the results required proofing and encoding -- both expensive to do. I also use EEBO/TCP materials, but EEBO/TCP often does not digitize the most difficult texts because of their worn type or badly inked pages or many languages. However, I would use a vendor again if TEI backed it.

These answers are a bit scattered because we undertook a mass digitization project (from printed, bound material) several years ago and have been busily post-processing it, as well as encoding previously unpublished material. We'd like to digitize printed and manuscript sources as well as continue to encode newly edited material, but the upshot is that we don't see ourselves as digitizing materials *primarily*; our main goal currently involves born-digital content.

The idea is very good; such a service would help many small projects. As to the question 15, we digitize/transcribe all pre-20th century materials in our institution. Only 20th cent texts are outsourced.

The TEI organization itself has historically shown very little ability (or inclination) to demonstrate its value beyond a very small group of experts with specialized uses. While part of the leadership has made an effort to reach out, a significant part of the group seems to have limited interest in non-specialized uses. Until I see signs that TEI-C can actually begin to be a functional organization for a broader community, I would not suggest that my organization join.

We have 500+ books that we'd digitize if the price were right and if we had expert guidance in the undertaking.

We do very little of this.

Prices we negotiated at the moment:
- PDF (scan the books without unbinding them and save as plain image pdf file): Euro 0.08 per page

- PDF and raw OCR'ed text (in order to find the right PDF image) (a sample will be like the attached file. Ocr processed, not cleaned up but can search strings): 0.1 EUR per page
- accuracy of 99.95% keying: 0.35 EUR per 1000 keystrokes.
- accuracy of 99.99% keying: 0.60 EUR per 1000 keystrokes.
- digitizing microfilm to image: 0.075 EUR per image

WE ARE A UNIVERSITY PRESS. WE ARE PARTNERING WITH A LIBRARY ON A NEW PROJECT AND EXPECT TO START DIGITIZING USING TEI LITE IN THE NEAR FUTURE, FOR THE PROJECT AND BEYOND.

Our digitization to date has focused on image-only digitization (no machine-readable text) because of the cost of full text creation and encoding. A reduction in price and the ability to participate in a larger effort would reduce the cost for us to try out more text-based digitization.

The idea of using TEI as a middle man between projects and vendors is attractive. Many of us know what we want our documents to do, but don't have the vocabulary to communicate that to vendors.

the distinction of different digitization levels, should be done, corresponding to 4 different prices:
- high quality scanning
- manual keyboard typing
- OCR processing with manual post-correction
- TEI encoding

Apologies for the lack of numbers. My organization is at the cusp of beginning a digitization project. We believe we will secure digitization at 40-45 cents per 1000 characters delivered.

For the time being besides OCR-scanning of printed material our main concerns when it comes to data capture are conversion and encoding of previously digitalized material, legacy data migration and so on. Therefore some of the questions above are not quite applicable.

My project is very small -- it's just me but I'm currently in the process of looking for funding.

The Charles Brockden Electronic Archive and Scholarly Edition contracted with Aptara and keyed and coded 1,115 18th c periodical texts for about $13,000, at 99.95% accuracy. They handled professionally from A-Z, including site visit to discuss tagging requirements and processes. We were very pleased--glad to provide further contact information.

All best wishes for success in negotiating the discount!

#2a: We do all of these, and it's difficult to say which is primary.
#3: We do so much digitization with our existing workflows that investing in new workflows to handle outside content in a different format is a hurdle for us.
#10 and #15b: These assume bitonal scanning only!
#21: I'm unsure.
good luck

Much of our current projects involved more detailed markup than TEI Lite or TEI Tite provide. But we would digitize additional monographs using a simple DTD if the price were right.

I'm sorry, but after a few questions I simply had to abandon the survey. The provide possible answers do not allow me to give a good overview of the current state and practice of TEI/text digitization in the organization where I work. Rather than to provide I very skewed image, I thought it would be better to skip the survey in this case. In my/our case a questionaire with the possibility to enter free text would have been more adequate. But I fully understand the possible logistic nightmare that this would create for the organizers. Sorry for any inconvenience.

I can't answer most of these questions, and the logic of some is flawed.

We do not have significant out-of-copyright holdings, so our answers reflect a somewhat limited amount of available content.