



OpenGLAM Benchmark Survey Preliminary Results from Finland and Poland

Beat Estermann, 11 April 2015 - GLAM-Wiki Conference, Den Haag

► Bern University of Applied Sciences | E-Government Institute || Open Knowledge | OpenGLAM Working Group

Welcome!

- ▶ **Introduction to the OpenGLAM Benchmark Survey Project**
- ▶ **Preliminary Results: Finnish and Polish Data Sets**
- ▶ **Implications for GLAM Outreach**
- ▶ **Outlook**

Introduction to the OpenGLAM Benchmark Survey Project

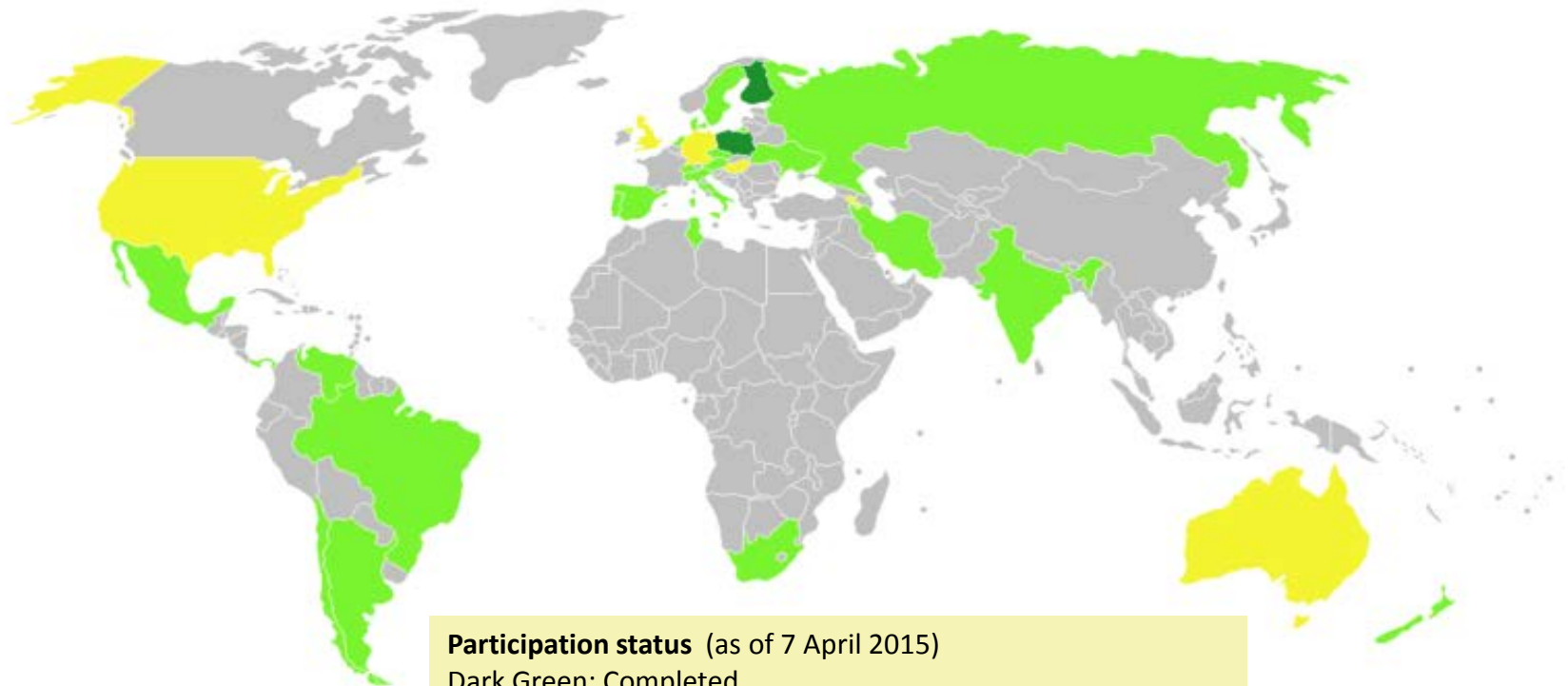
What is the OpenGLAM Benchmark Survey about?

- ▶ Online survey conducted among heritage institutions throughout the world in 2014/2015.
- ▶ Focusing on questions related to digitization, exchange of metadata, open data/open content, semantic web, social media, crowdsourcing
- ▶ Inspired by an earlier pilot survey, carried out in Switzerland in 2012
- ▶ Organized in a federative manner, which means that the organization depends on volunteers and partners in each country

What do we want to achieve?

- ▶ **Measure the state of advancement** of OpenGLAM in the participating countries
- ▶ **Inform the GLAM community** about the latest developments in the area of OpenGLAM
- ▶ **Identify potential partners** for open data and/or crowdsourcing projects
- ▶ Use the study report as a communication instrument to **promote OpenGLAM**
- ▶ Provide **international comparisons**:
 - Allowing each country to see where it stands compared to other countries.
 - Provide the international OpenGLAM community with a tool that helps it better understand the particularities of each country

Overview of participating countries



Participation status (as of 7 April 2015)
Dark Green: Completed
Light Green: Initial commitment
Yellow: Clarification needed / presently insufficient support

Further countries are welcome to join!

Overview:

https://outreach.wikimedia.org/wiki/GLAM/OpenGLAM_Benchmark_Survey/Participating_countries

Preliminary Results

- Finnish and Polish Data Sets
- Focus on Open Content and Collaborative Content Creation / Crowdsourcing

Finnish and Polish Data Samples

	Finland	Poland
N institutions contacted ^[1]	356	669
N responses ^[2]	81	79
Response rate ^[3]	26%	12%
– Archives	50%	21%
– Libraries	37%	17%
– Museums	21%	10%

Notes:

[1] Number of institutions which have been contacted (each institution is counted once, even if it has been contacted via several email addresses or through different channels, e.g. follow-up phone calls in addition to e-mail reminders).

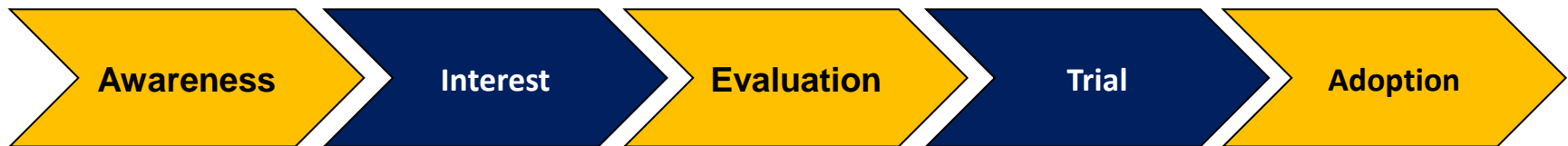
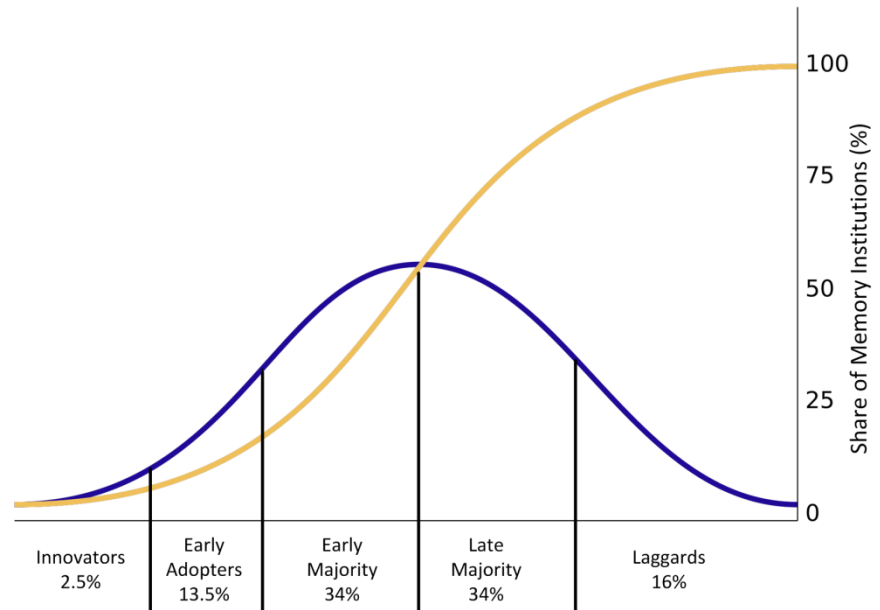
[2] Number of institutions which have completed the questionnaire (almost) in full, i.e. institutions which have responded to at least 20 out of the 24 non-conditional questions of the questionnaire.

[3] For the purpose of the calculation of the response rate, institutions which have started to fill in the questionnaire but dropped out after question A2 as they did not correspond to the survey's definition of heritage institutions, are counted as responses.

Where do GLAMs stand with regard to...

- ...Open Data?
- ...Linked Data / Semantic Web?
- ...Digitization
- ...Open Content?
- ...Engaging Audiences on the Internet
- ...Collaborative Content Creation

Innovation Diffusion Model,
Everett Rogers, 1962

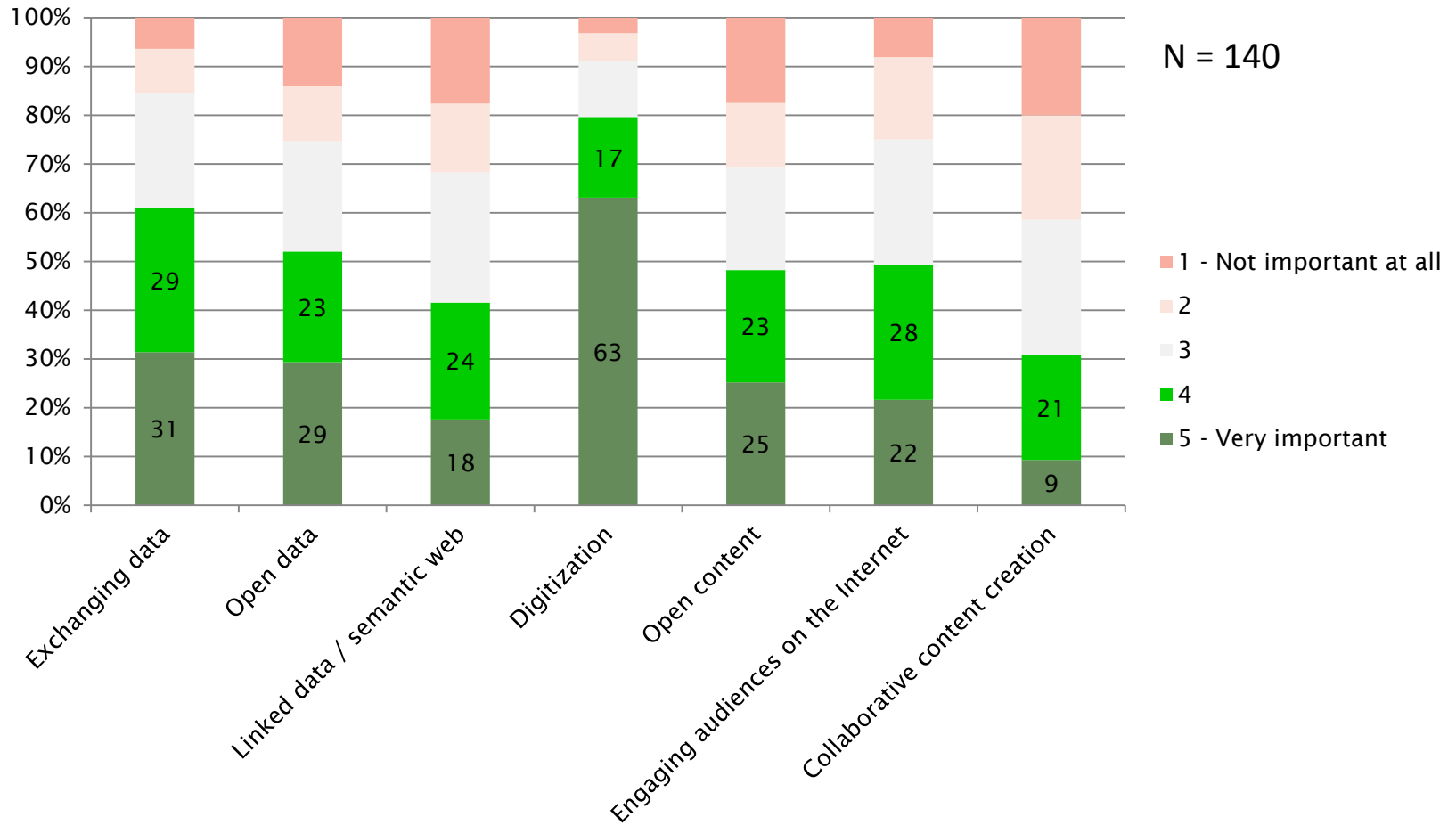


What are the perceived risks and opportunities? (drivers vs. hindering factors)
What are the expected benefits?

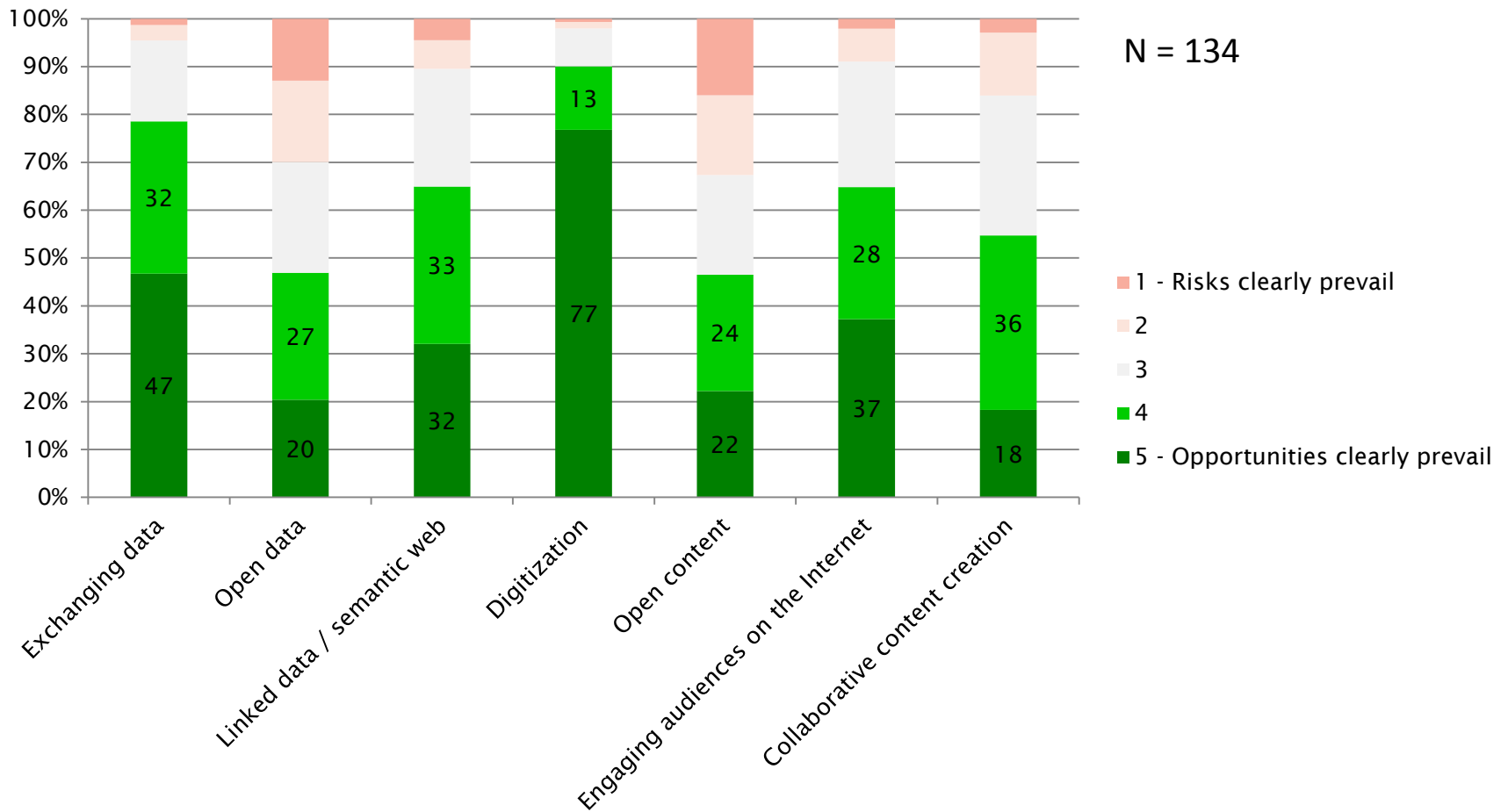
Tentative Visualization of a Simple Diffusion Model

	Interest	Evaluation	Trial	Adoption	Advanced Implementation
	<i>interested in obtaining further information / training / consulting</i>	<i>future practice intended</i>	<i>at least some practice at present / so far</i>	<i>medium level of practice at present / so far</i>	<i>advanced level of practice at present / so far</i>
				<i>20% or more</i>	<i>50% or more</i>
Open Data	83%	54%	43%	29%	20%
Linked Data / Semantic Web	83%	25%	13%	8%	6%
				<i>20% or more</i>	<i>50% or more</i>
Digitization	76%	81%	83%	54%	18%
Open Content	82%	63%	42%	12%	4%
				<i>20% or more</i>	<i>50% or more</i>
Engaging Audiences on the Internet	74%	92%	89%	53%	23%
				<i>at least 3 social media types</i>	<i>at least 5 social media types</i>
Collaborative Content Creation	78%	38%	29%	16%	8%
				<i>at least 2 crowdsourcing approaches</i>	<i>at least 3 crowdsourcing approaches</i>

Importance of Various Practices

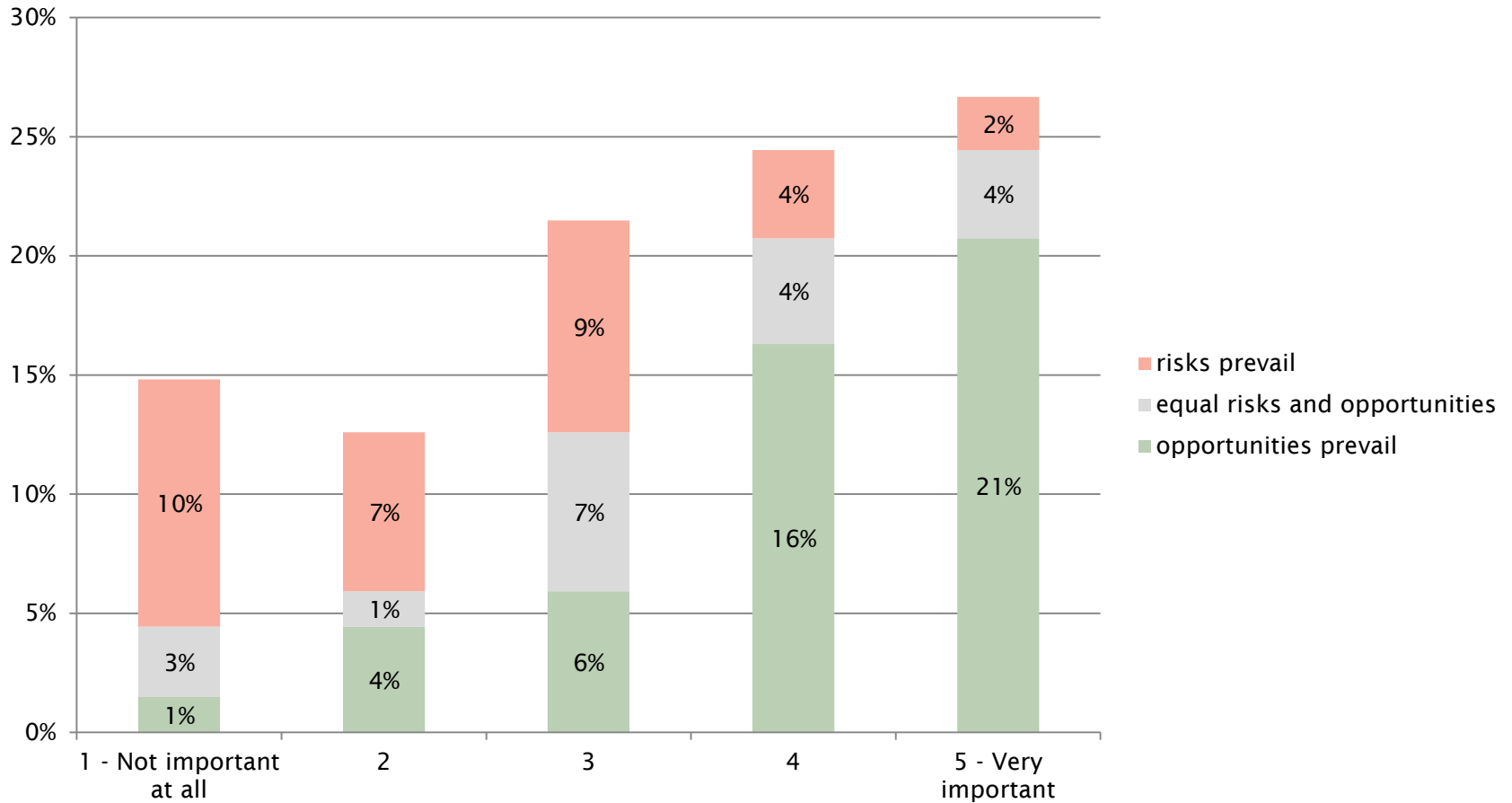


Desirability of Various Practices (Opportunities vs. Risks)



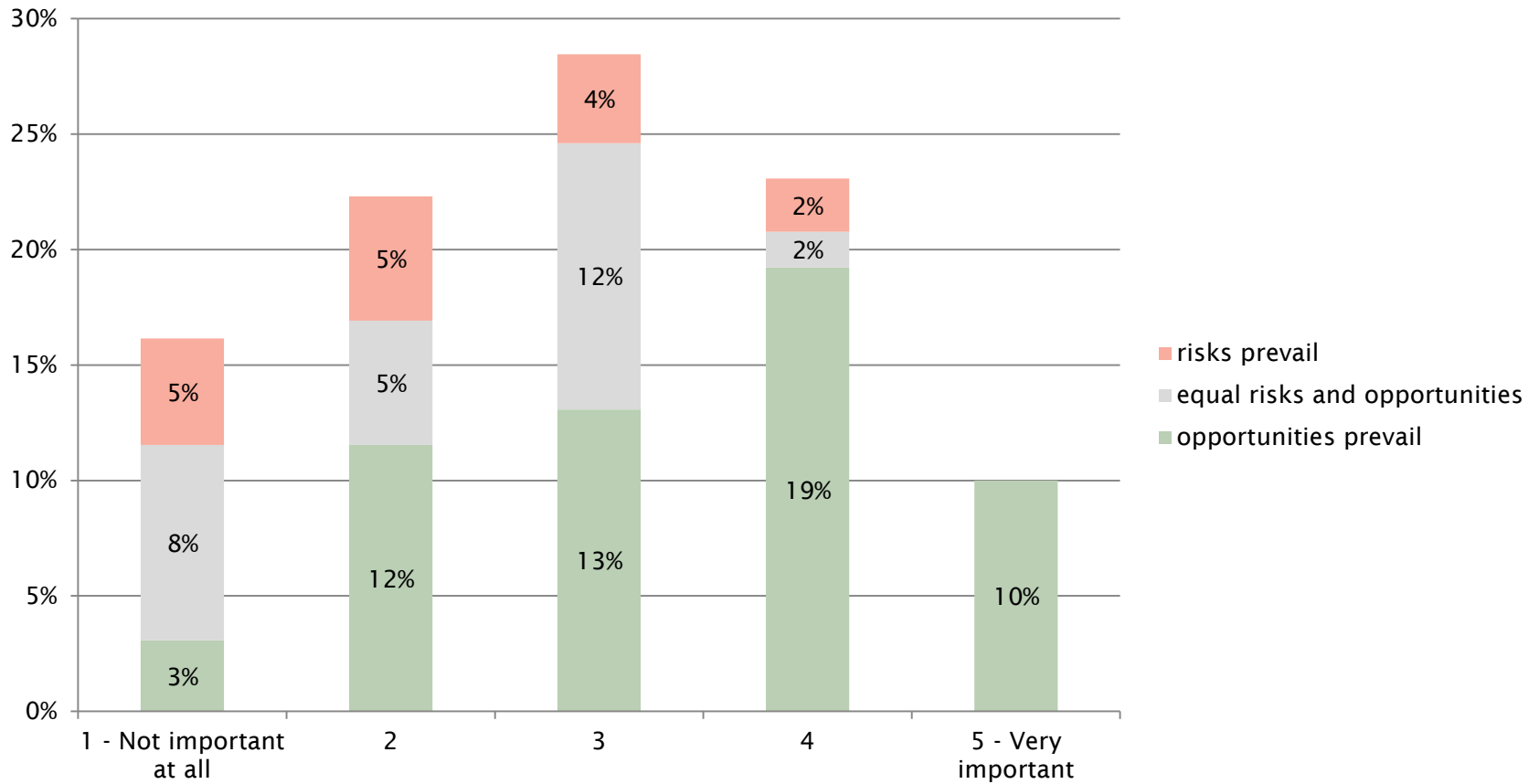
Open Content: Importance vs. Desirability

Importance and desirability of "open content" (in % of institutions; N=135)



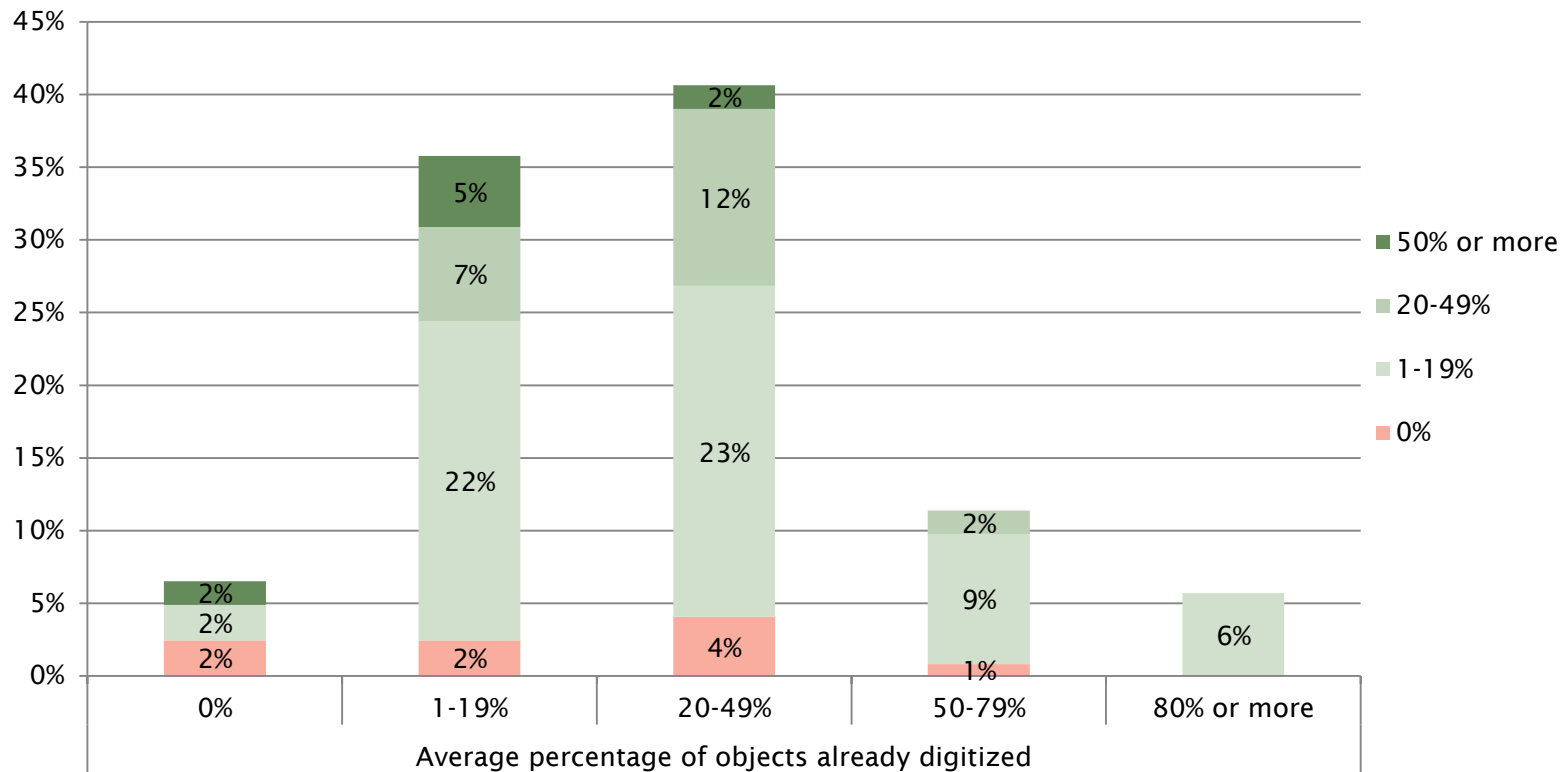
Collaborative Content Creation: Importance vs. Desirability

Importance and desirability of "collaborative content creation"
(in % of institutions; N=130)



Digitization as a Prerequisite for Open Content

Expected digitization activities over the next 5 years
(in % of institutions; N=123)



Digitization: Different types of objects

Statistics

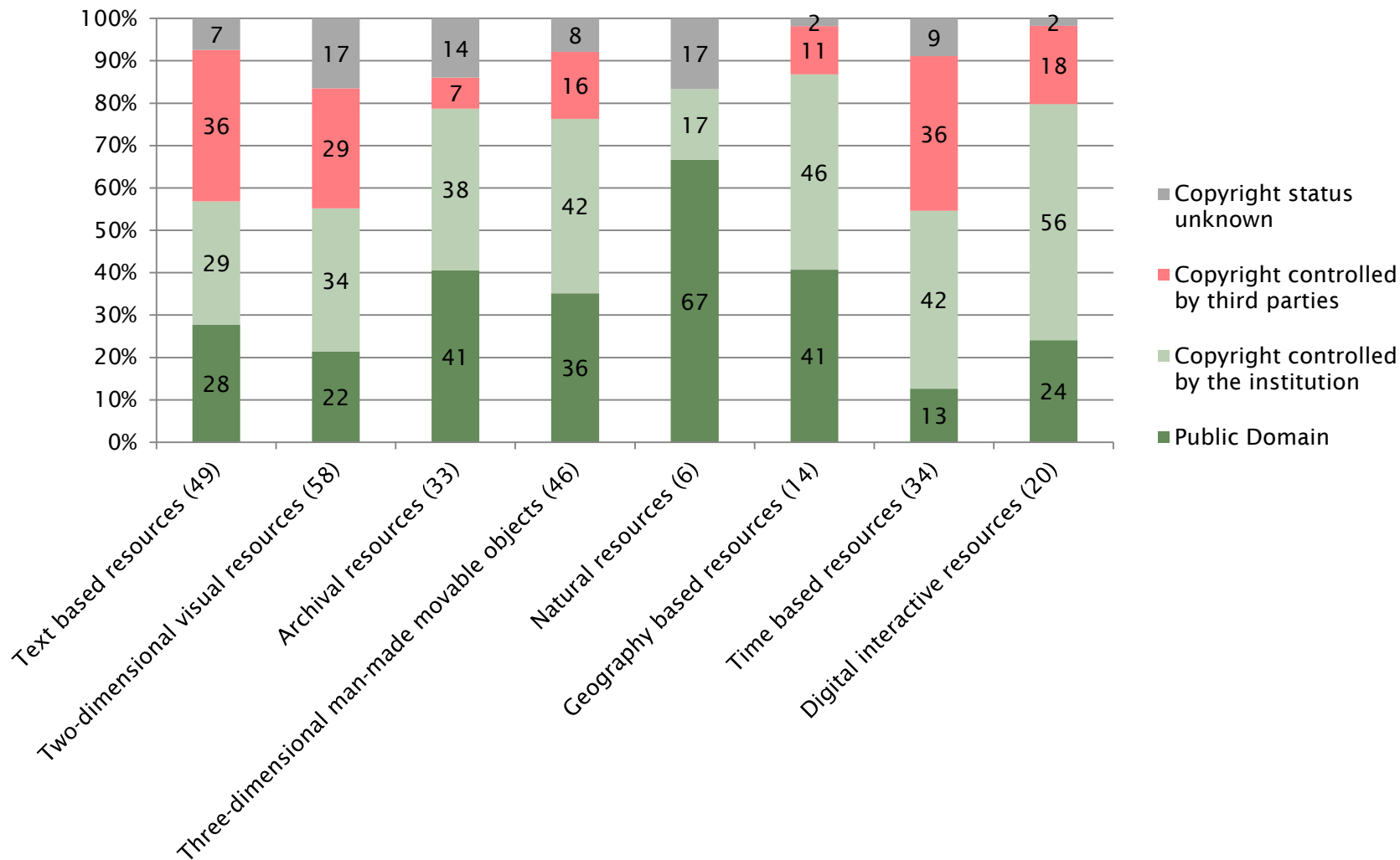
		Text based resources: already digitized	Two-dimensional visual resources: already digitized	Archival resources: already digitized	Three-dimensional man-made movable objects: already digitized	Natural resources: already digitized	Geography based resources: already digitized	Time based resources: already digitized	Average percentage of objects already digitized
N	Valid	109	108	65	90	13	35	63	151
	Missing	51	52	95	70	147	125	97	9
Mean		11.2505	27.7741	10.4292	31.1689	31.5385	37.6029	27.2333	26.6036
Median		1.0000	10.0000	.5000	20.0000	5.0000	25.0000	5.0000	23.0000
Percentiles	25	.0000	.6250	.0000	2.7500	.0000	.0000	.0000	3.3333
	50	1.0000	10.0000	.5000	20.0000	5.0000	25.0000	5.0000	23.0000
	75	10.0000	50.0000	10.0000	56.2500	65.0000	80.0000	55.0000	38.0000

		Text based resources: digitized in 5 years	Two-dimensional visual resources: digitized in 5 years	Archival resources: digitized in 5 years	Three-dimensional man-made movable objects: digitized in 5 years	Natural resources: digitized in 5 years	Geography based resources: digitized in 5 years	Time based resources: digitized in 5 years	Average percentage of objects expected to be digitized in 5 years
N	Valid	95	100	59	85	9	29	56	136
	Missing	65	60	101	75	151	131	104	24
Mean		23.0579	42.2480	24.4424	49.9376	40.3333	46.0724	39.6232	39.8012
Median		10.0000	35.0000	10.0000	45.0000	15.0000	50.0000	22.5000	33.7500
Percentiles	25	1.5000	10.0000	1.0000	17.5000	3.0000	10.0000	3.2500	17.5759
	50	10.0000	35.0000	10.0000	45.0000	15.0000	50.0000	22.5000	33.7500
	75	30.0000	78.7500	50.0000	90.0000	90.0000	85.0000	80.0000	56.5625

Copyright Clearance as a Prerequisite for Open Content

What is the copyright situation of the objects in your collections?

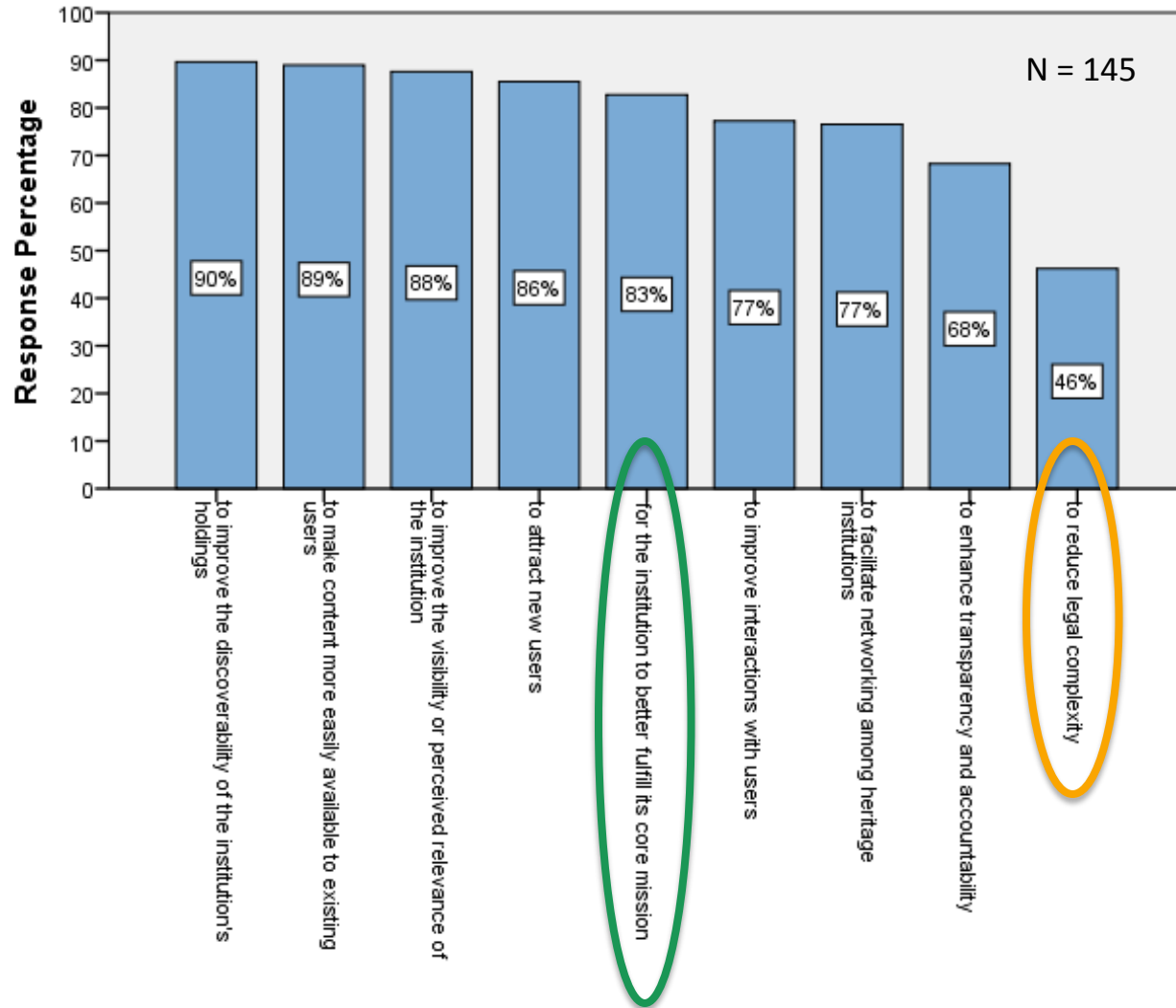
(average percentages for each object type; N is indicated for each object type)



Note: The data may be somewhat biased as it does - by definition - not include the data from institutions who weren't able to provide numbers regarding the copyright status of their objects.

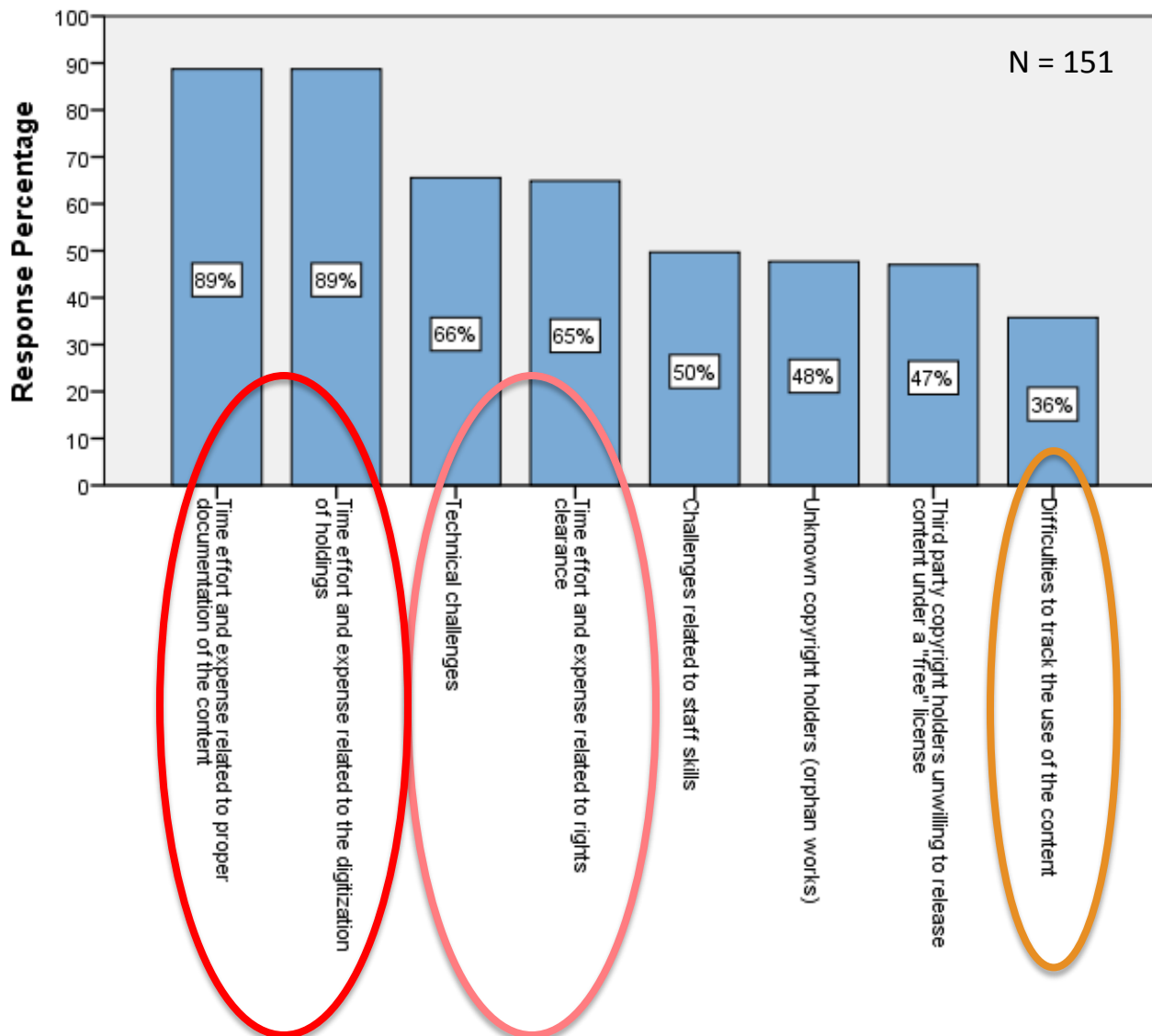
Benefits of Open Content

D7: In the case of my institution, 'open content' is an important means...



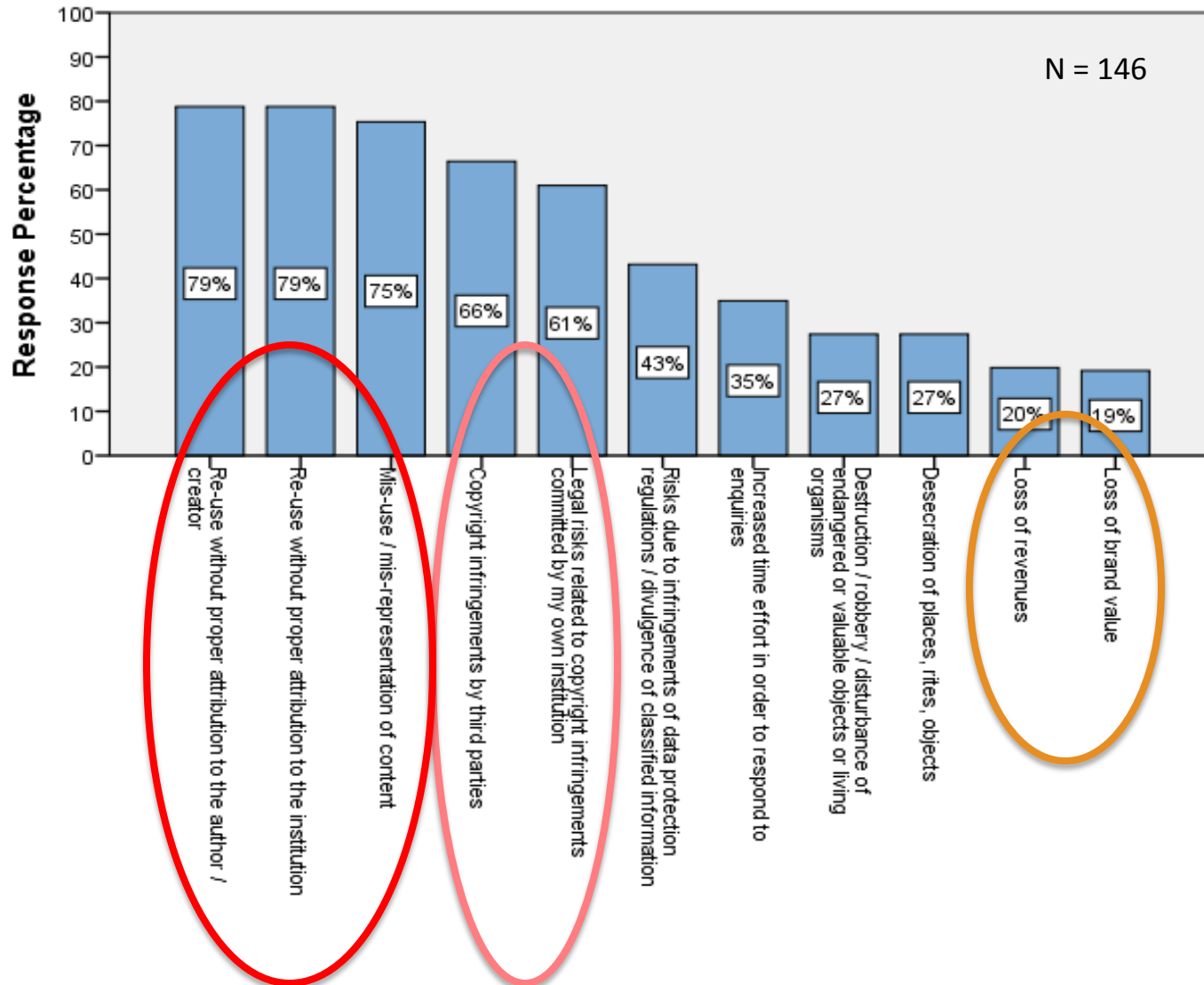
Challenges Related to Open Content

D8: For my institution, these are important challenges related to 'open content'...



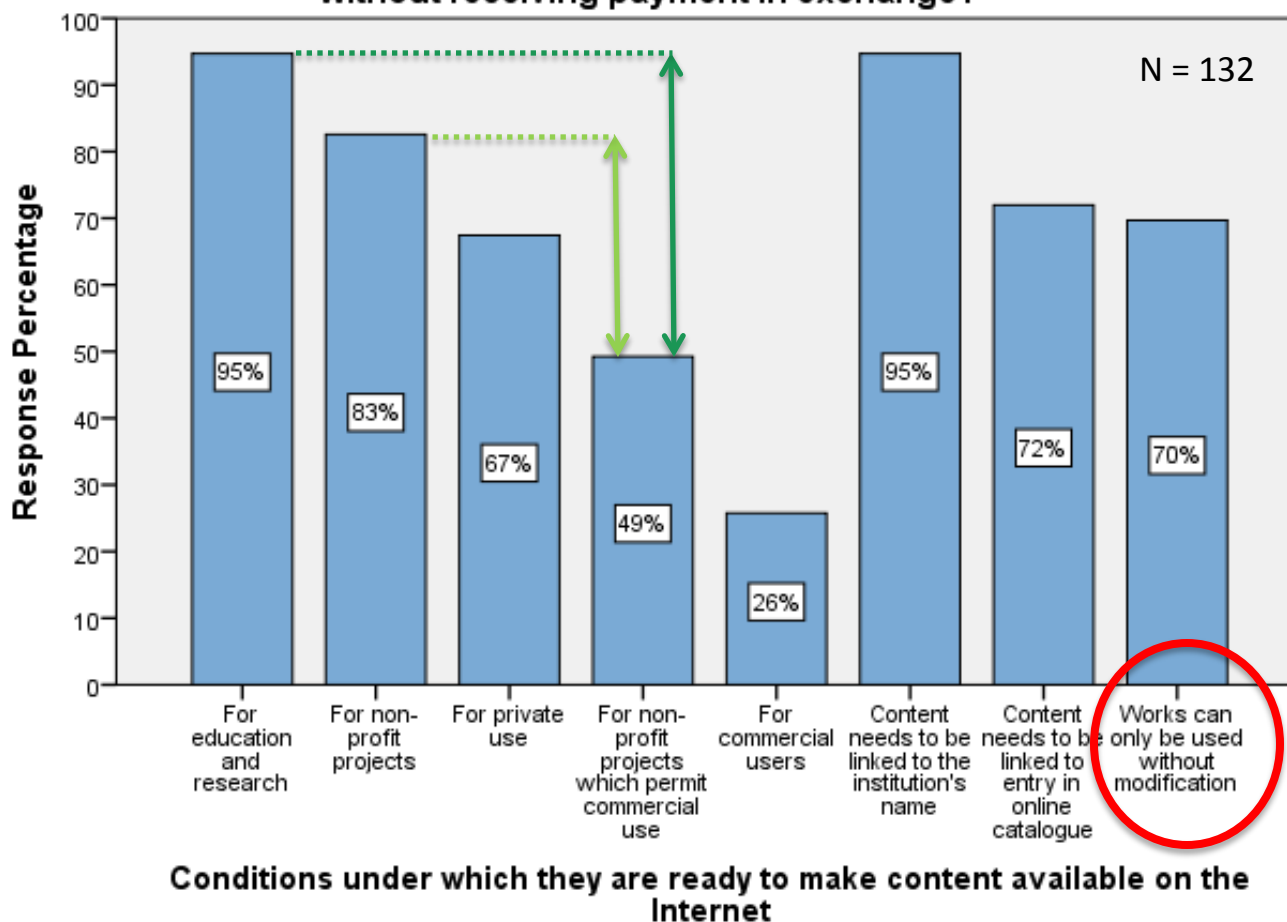
Risks Related to Open Content

D9: For my institution, these are important risks related to 'open content'...



Conditions for Releasing Content

D3 - Under what conditions is your institution ready to make its content available on the Internet without receiving payment in exchange?

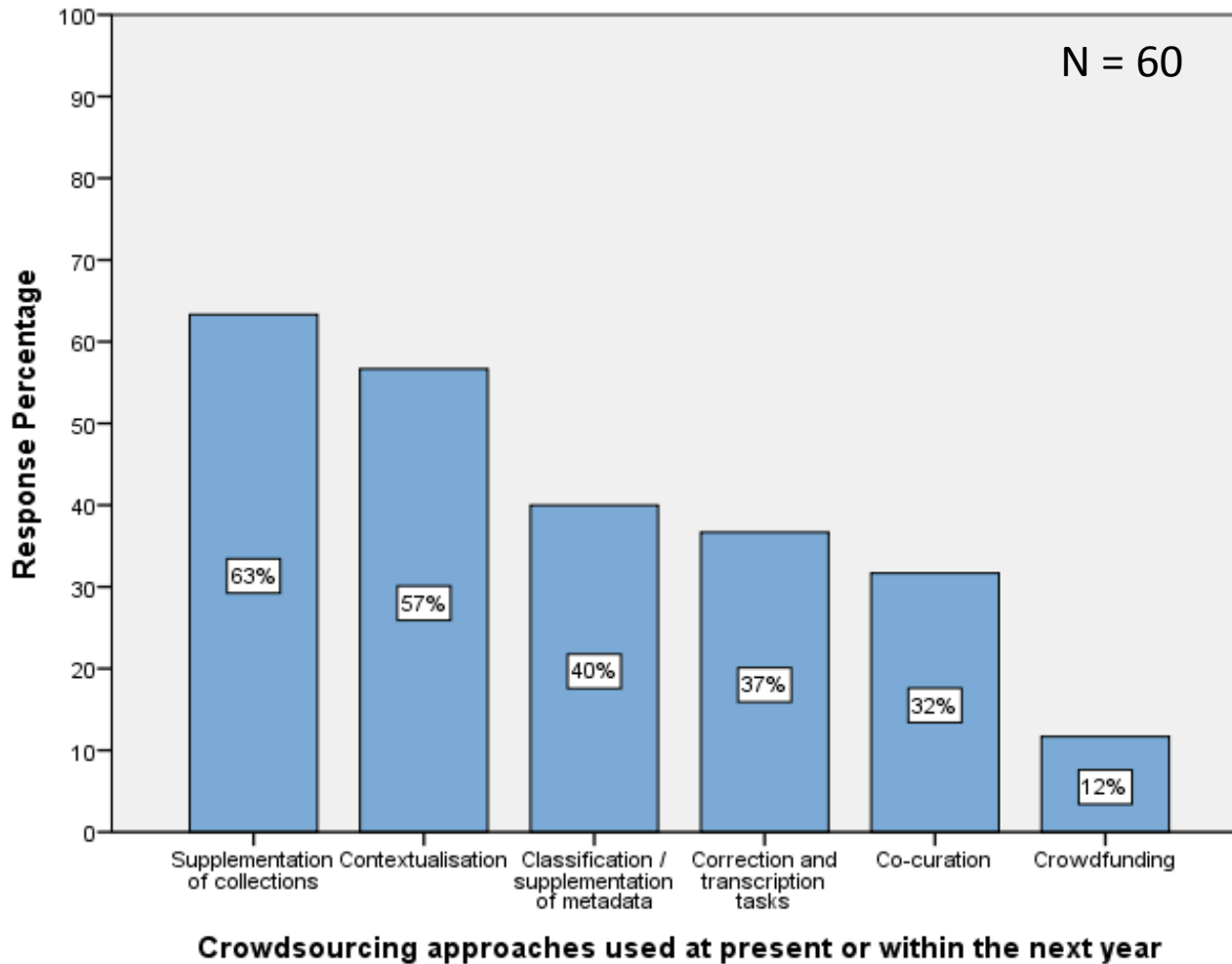


GLAM Involvement in Wikipedia

- ▶ 17% of responding GLAMs have staff members who are involved in Wikipedia as part of their job
- ▶ Wikimedia Commons: 3%
- ▶ Wikidata: 2%

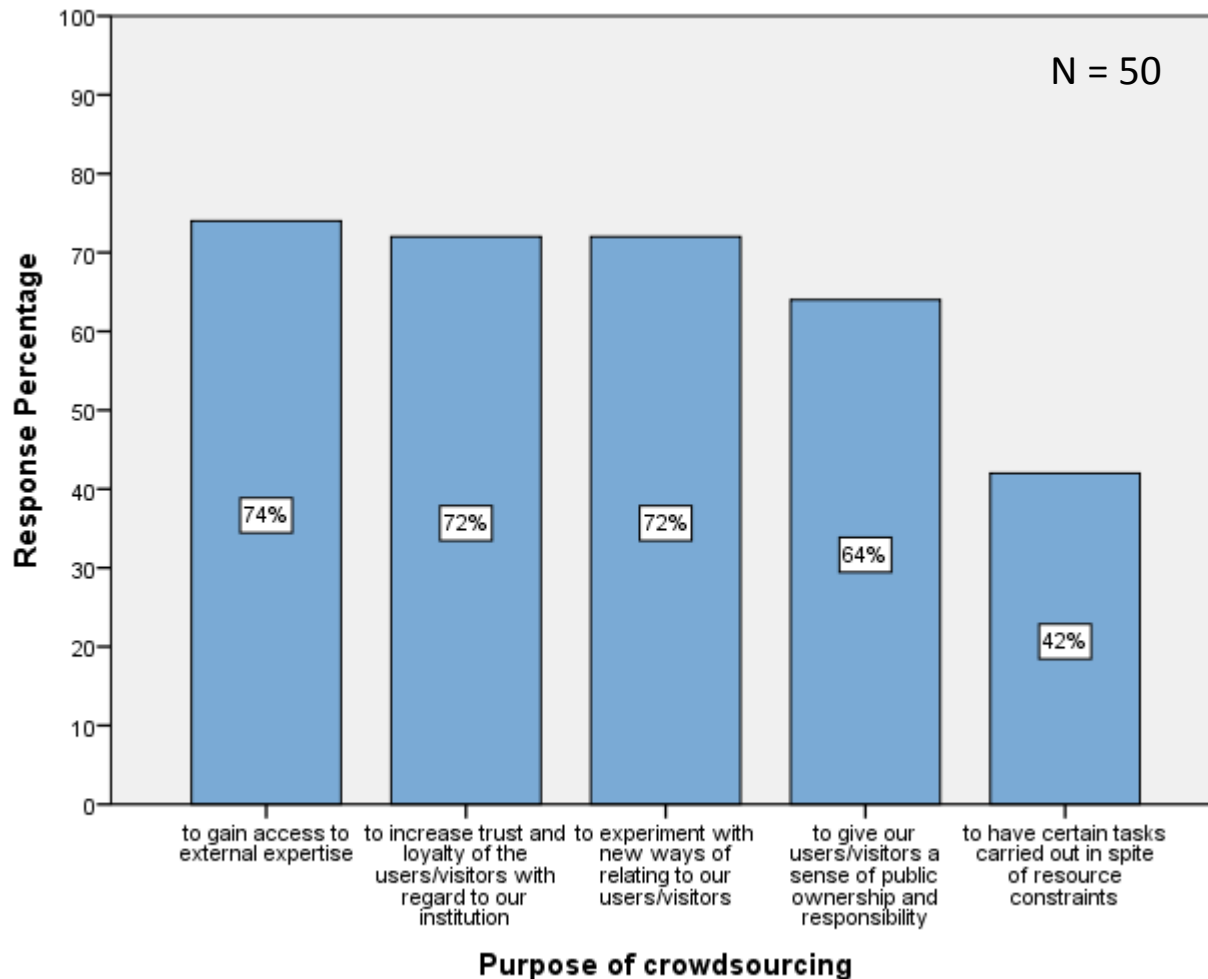
Most popular crowdsourcing approaches

Note: The graph summarizes the data from the institutions which indicated that they are presently using at least one crowdsourcing approach or are planning to use one within the next year.



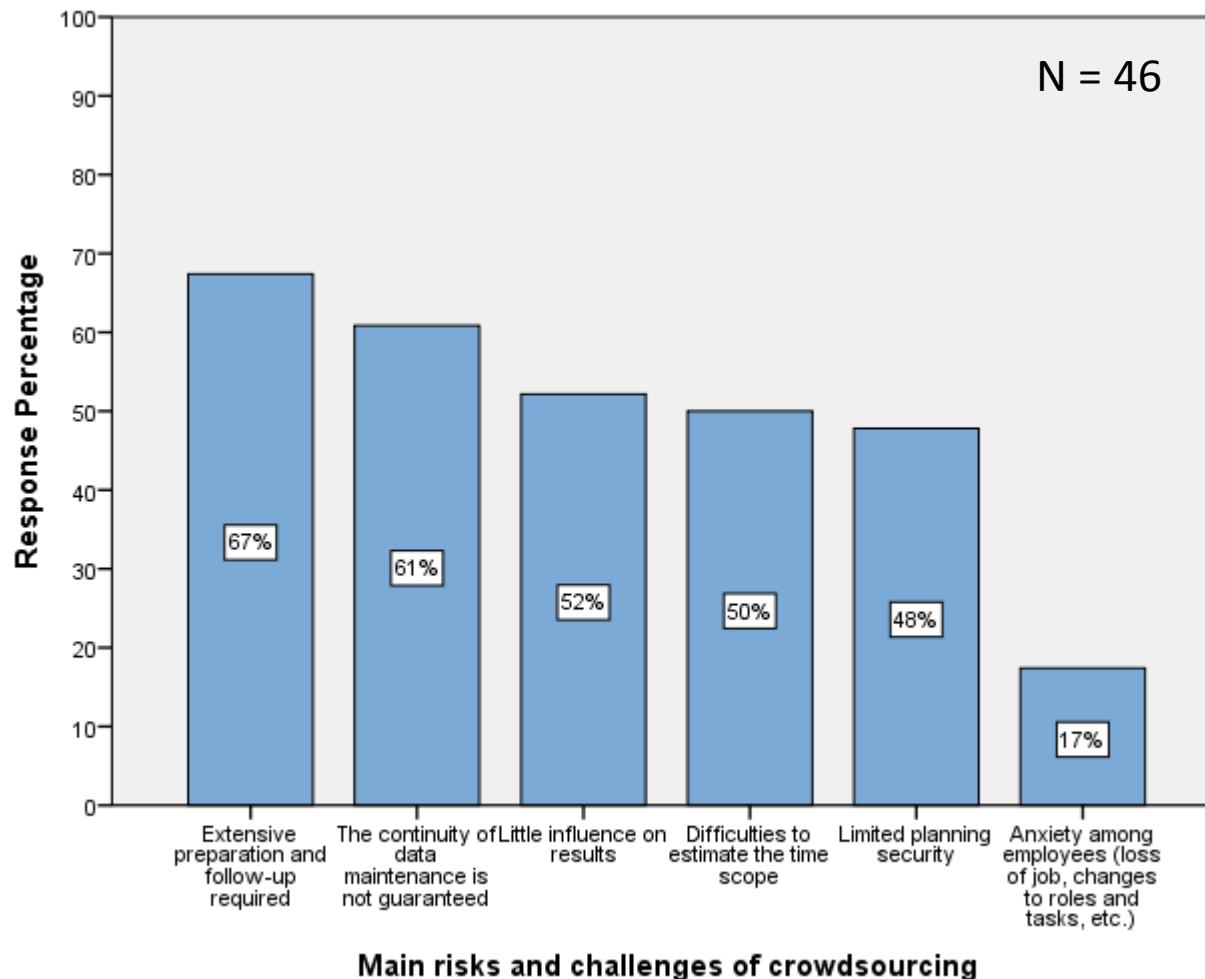
Purpose of crowdsourcing

Note: This question was asked only to institutions which indicated that they are presently using at least one crowdsourcing approach or are planning to use one within the next year.



Risks and Challenges of Crowdsourcing

Note: This question was asked only to institutions which indicated that they are presently using at least one crowdsourcing approach or are planning to use one within the next year.



What are the Implications for GLAM Outreach?

Summing up the main findings...

Diffusion of Various Practices

- ▶ **Digitization and the Use of Social Media** are widespread practices in the Finnish/Polish GLAM sector, with adoption rates above 50%.
- ▶ **Open Data and Open Content** are about to diffuse within the FI/PL GLAM sector; for Open Data we are observing an adoption rate of ca. 30%; for Open Content the adoption rate is just above 10%.
- ▶ **Collaborative Content Creation** is diffusing more slowly than Open Data / Open Content
- ▶ These findings are reflected in the **importance accorded to the various practices**; ca. 50% of respondents consider Open Data and Open Content as important; ca. 30% say so with regard to Collaborative Content Creation

Prerequisites for Open Content

- ▶ Over the next 5 years we will see **leaps in digitization activities**.
- ▶ Ca. 80% of institutions think that opening up content helps them better **fulfill their core mission**.
- ▶ In general, issues related to **copyright clearance** may be an excuse, but not a show-stopper regarding the opening up of content; **across all object types, institutions may release at least 50% of their holdings as open content**.
- ▶ **When it comes to promoting the opening up of content, the present mindset of GLAMs poses two major challenges: their aversion against the prospect of commercial use of open content, and their reluctance to let third parties modify the content.**

Challenges / Risks related to Open Content

- ▶ The **greatest challenges regarding open content** are the **time effort and expense related to the digitization and the documentation of content**. Other important challenges comprise the **time effort and expense related to rights clearance** as well as **technical challenges**.
- ▶ The **greatest risks associated with open content** are **re-use of content without proper attribution or mis-use of the content**. Other important risks are **copyright infringements**.
- ▶ When it comes to opening up content, **loss of revenues, diminished brand value, or problems with tracking the use are only minor preoccupations**.

Involvement in Wikipedia

- ▶ Ca. 30% of GLAMs consider **Collaborative Content Creation** as important and believe that the opportunities prevail over the risks
- ▶ Ca. 20% of GLAMs are already involved in **Wikipedia**.
- ▶ **Wikimedia Commons** and **Wikidata** still seem to play a minor role for GLAMs (with 3% and 2% of institutions involved).

Outlook

What's next?

- ▶ Data collection in further countries – **Your support is appreciated...**
- ▶ Country comparisons / Context analysis
- ▶ Comparisons between different types of GLAMs
- ▶ More detailed analyses – **Feel free to join!**
- ▶ Publication of country reports
- ▶ Publication of an international report in the first half of 2016
- ▶ Explore further uses of the GLAM inventory as a by-product of the survey

Thank you for your attention!

▶ **Contact Details:**

- Beat Estermann
E-mail: beat.estermann@bfh.ch
Phone: +41 31 848 34 38

▶ **Project Portal:**

- http://outreach.wikimedia.org/wiki/GLAM/OpenGLAM_Benchmark_Survey