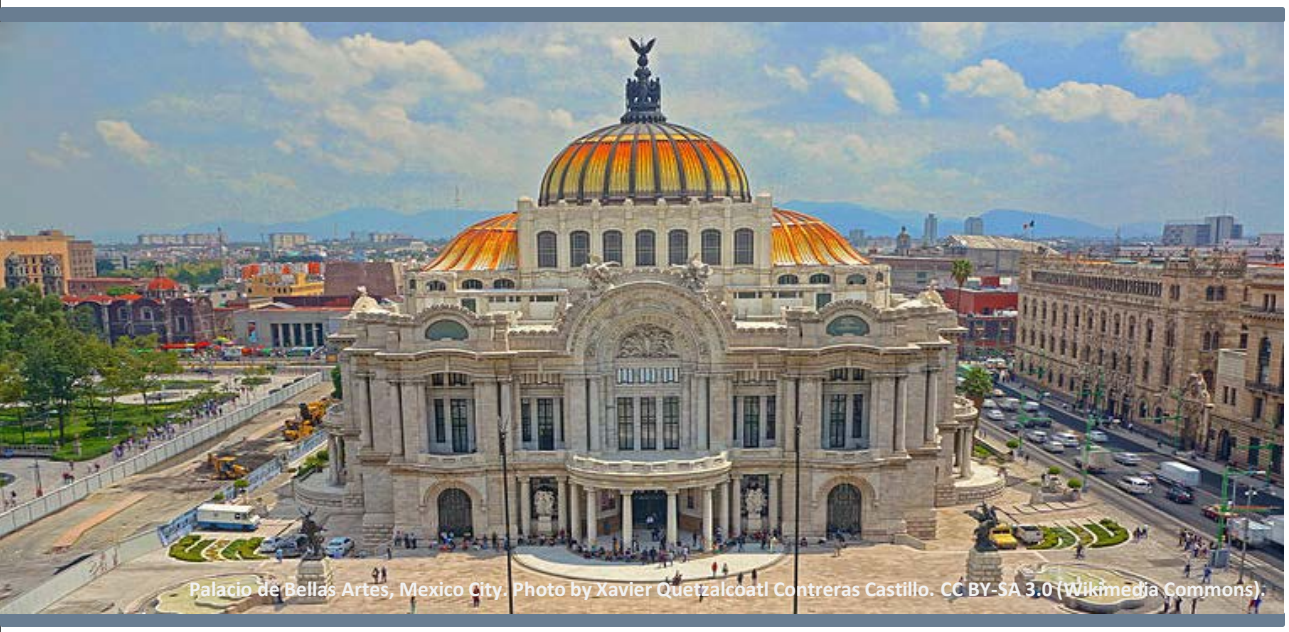
Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

OpenGLAM

OPEN KNOWLEDGE



Palacio de Bellas Artes, Mexico City. Photo by Xavier Quetzalcoatl Contreras Castillo. CC BY-SA 3.0 (Wikimedia Commons).

OpenGLAM Benchmark Survey: Preliminary Results from Finland, Poland, Switzerland and The Netherlands

Beat Estermann, 19 July 2015 – Wikimania, Mexico City

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

This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 License.

Welcome!

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

Introduction to the OpenGLAM Benchmark Survey Project

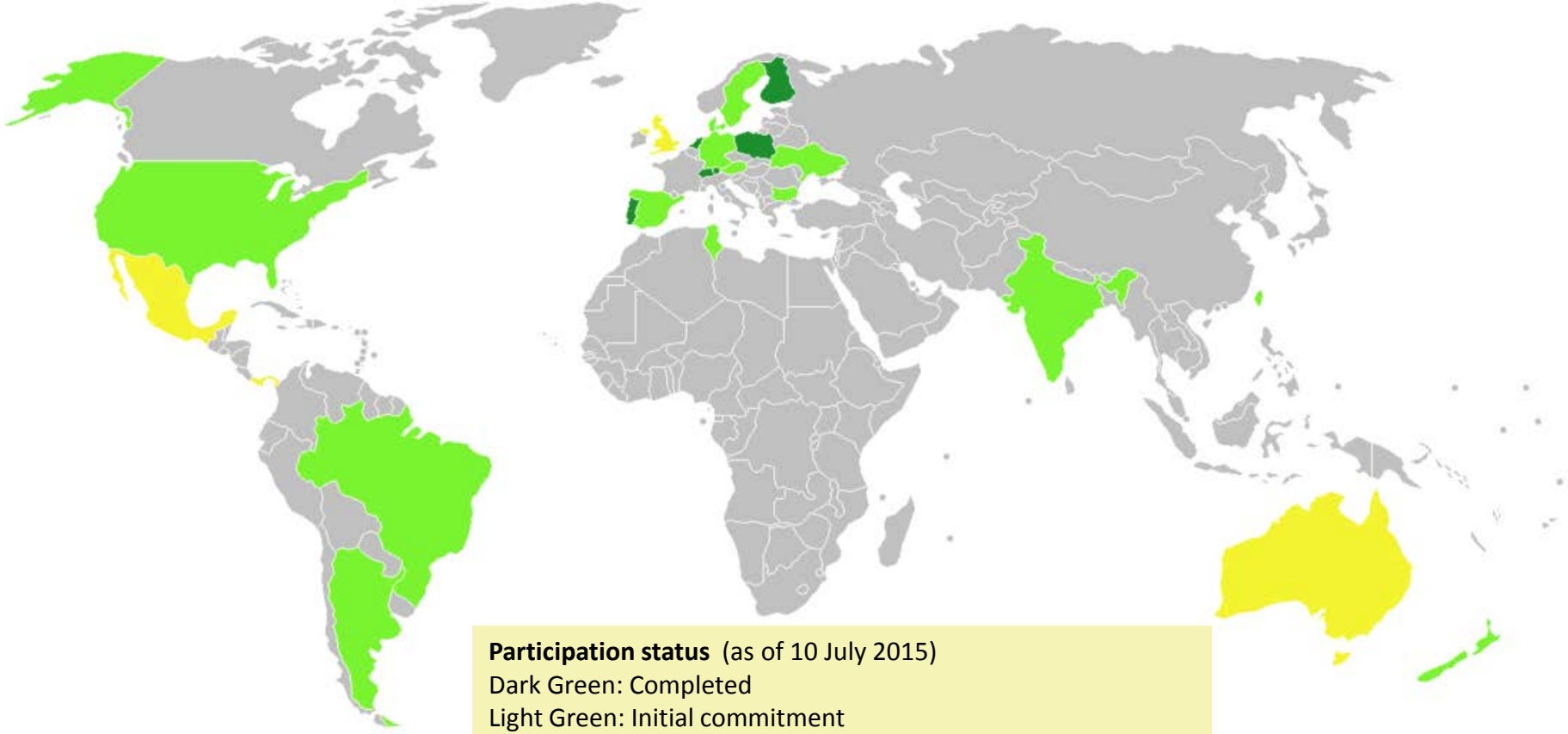
What is the OpenGLAM Benchmark Survey about?

- 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 10 July 2015)
Dark Green: Completed
Light Green: Initial commitment
Yellow: Clarification needed / presently insufficient support

Further countries are welcome to join; but hurry up!

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

Underlying graphic file: https://commons.wikimedia.org/wiki/File:BlankMap-World-v2.png Roke et al. (CC-by-sa)

Preliminary Results

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

Dutch, Finnish, Polish, and Swiss Data Samples

	CH (incl. FL)	FI	NL	PL
N institutions contacted[1]	1543 (11)	356	1393	669
N responses[2]	278 (2)	81	146	79
Response rate[3]	19%	26%	12%	12%
– Archives	27%	50%	16%	21%
– Libraries	29%	37%	15%	17%
– Museums	16%	21%	10%	10%
N inhabitants (2013)	8.1 mio.	5.4 mio.	16.8 mio.	38.5 mio.
GLAM density (GLAMs/mio. inhabitants)	189	66	83	17

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).

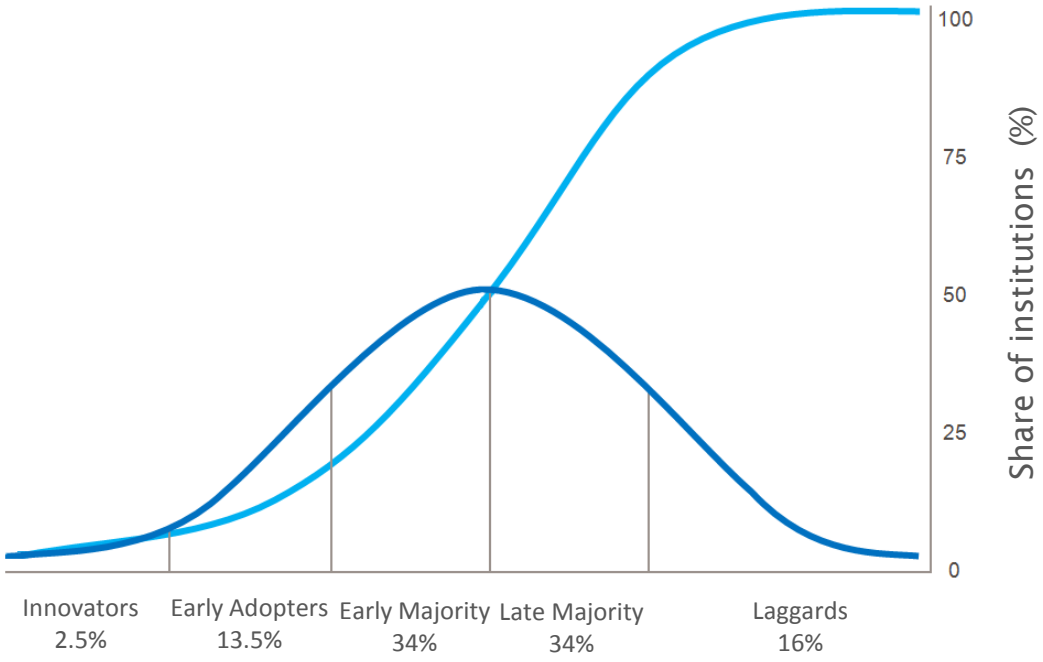
[2] Number of institutions which have completed the questionnaire (almost) in full, i.e. institutions which have responded to at least 20 out of the 24 non-conditional questions of the questionnaire.

[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…

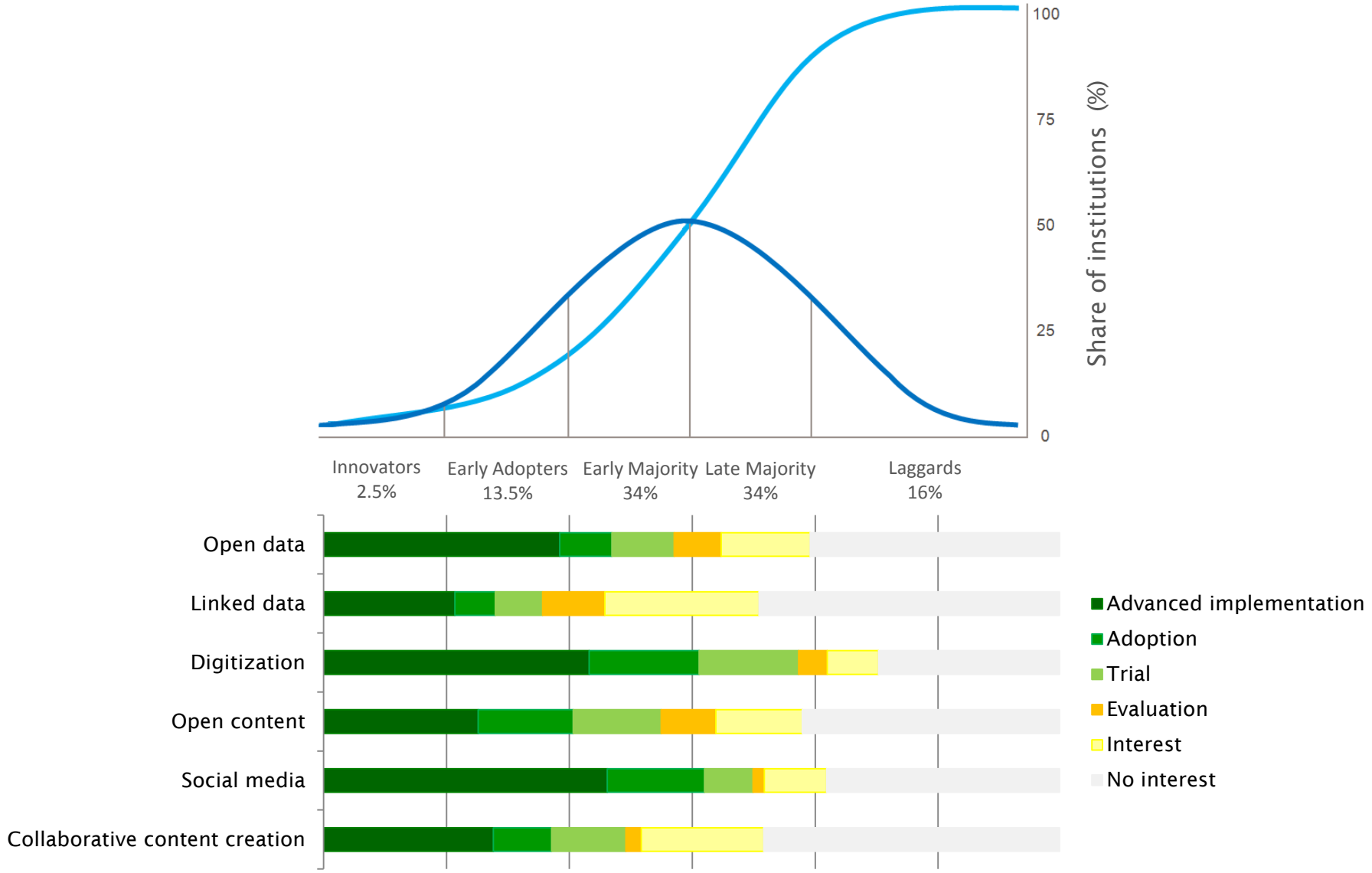
…Open Data?
…Linked Data / Semantic Web?
…Digitization
…Open Content?
…Engaging Audiences on the Internet
…Collaborative Content Creation

Innovation Diffusion Model,
Everett Rogers, 1962



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

Diffusion of Innovative Practices among GLAMs



Finland, Poland, Switzerland, The Netherlands, all institution types combined, N = 584.
Cases with «stagnation» / «discontinuance» have been ignored.

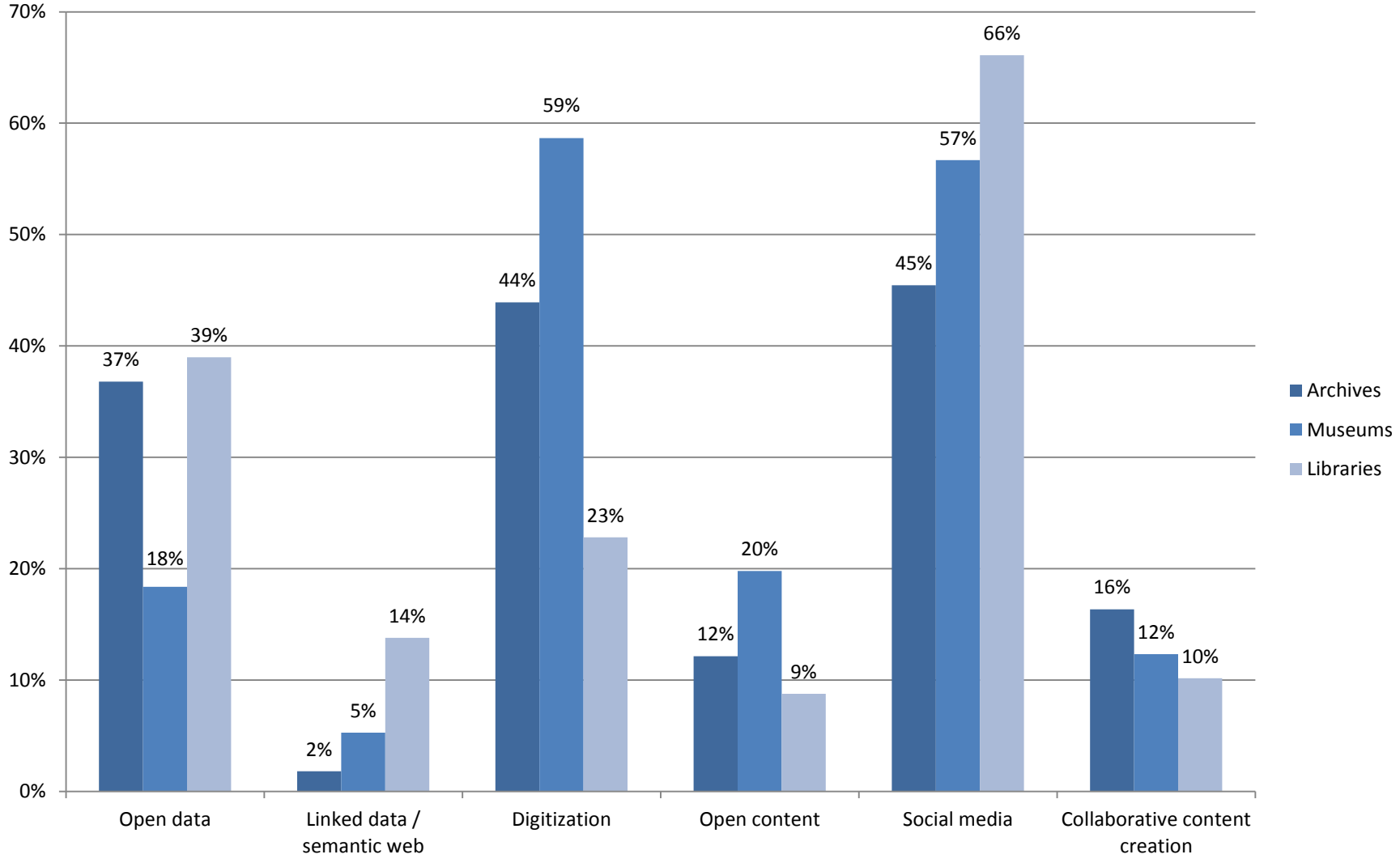
Documentation: Thresholds / Rules for the Assignment of Cases to Innovation Adoption Stages (1/2)

	No interest	Interest	Evaluation	Trial	Adoption	Advanced implementation	Stagnation / discontinuance
Open Data	OD_interest = **0** (ODLD_interest = 0)	OD_interest = **1** (ODLD_interest = 1)	OD_metadata_5yrs_avg >= **0.5%**	OD_metadata_pres_avg >= **0.5%**	OD_metadata_pres_avg > **10%** **AND** OD_metadata_5yrs_avg > OD_metadata_pres_avg + 10	metadata_pres_avg > **50%** **AND** OD_catalogues_pres > **50%**	OD_metadata_pres_avg >= **0.5%** **AND** OD_metadata_pres_avg **< 66%** **AND** OD_metadata_5yrs_avg **<=** OD_metadata_pres_avg **AND** imp_OD **<= 3**
Linked Data	LD_interest = **0** (ODLD_interest = 0)	LD_interest = **1** (ODLD_interest = 1)	LD_metadata_5yrs_avg >= **0.5%**	LD_metadata_pres_avg >= **0.5%**	LD_metadata_pres_avg > **10%** **AND** LD_metadata_5yrs_avg > OD_metadata_pres_avg + 10	LD_metadata_pres_avg > **50%** **AND** LD_catalogues_pres > **50%**	LD_metadata_pres_avg >=**0.5%** **AND** LD_metadata_pres_avg **< 66%** **AND** LD_metadata_5yrs_avg **<=** LD_metadata_pres_avg **AND** imp_LD **<= 3**
Digitization	DIGI_interest = **0**	DIGI_interest = **1**	DIGI_5yrs_avg >= **0.5%**	DIGI_pres_avg >= **0.5%**	(DIGI_pres_avg > **10%** **AND** DIGI_5yrs_avg > DIGI_pres_avg +5) **OR** (DIGI_pres_avg > **5%** **AND** DIGI_5yrs_avg > DIGI_pres_avg +10)	DIGI_pres_avg > **50%**	DIGI_pres_avg >= **0.5%** **AND** DIGI_pres_avg < **66%** **AND** DIGI_5yrs_avg <= DIGI_pres_avg **AND** imp_DIGI <= 3

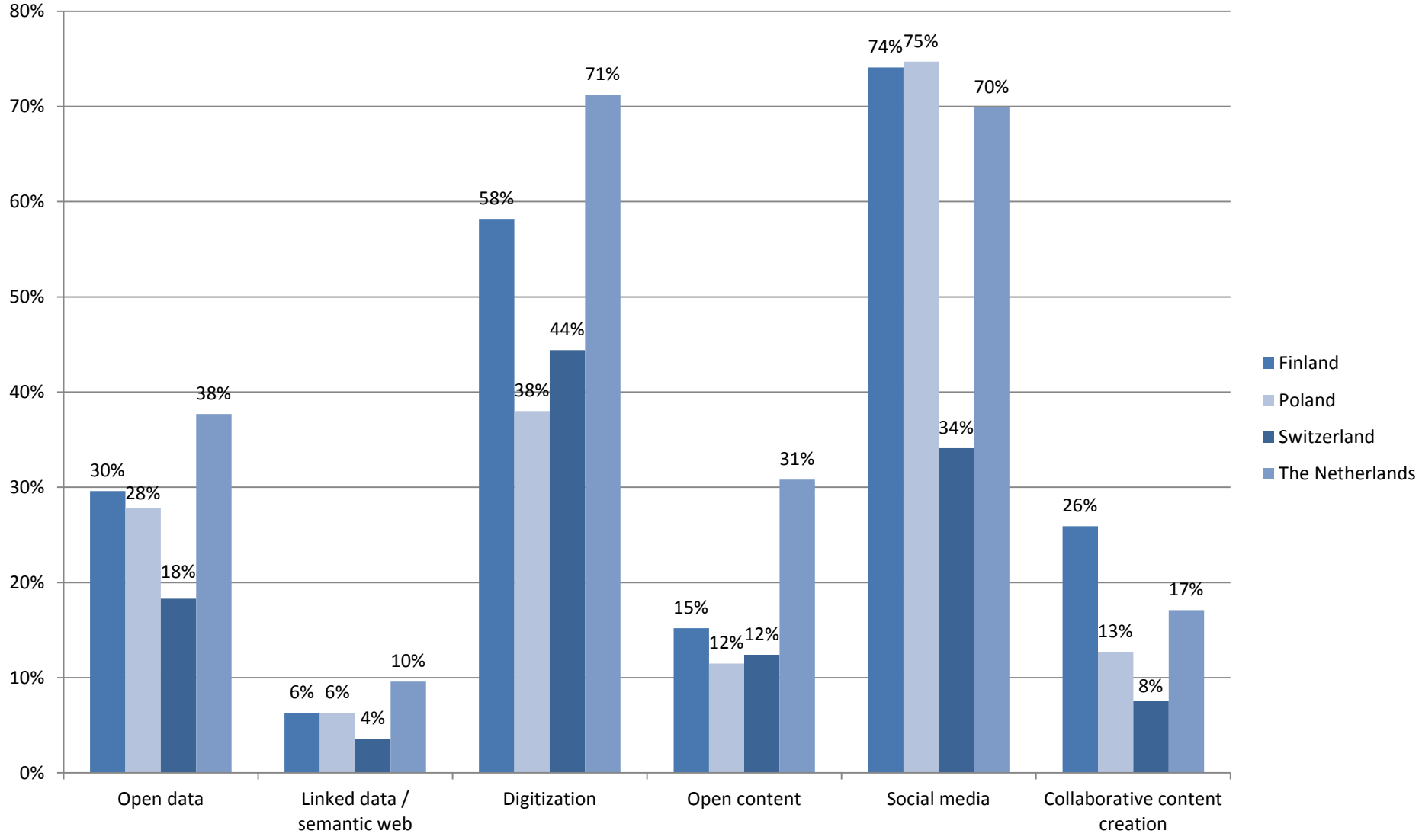
Documentation: Thresholds / Rules for the Assignment of Cases to Innovation Adoption Stages (2/2)

	No interest	Interest	Evaluation	Trial	Adoption	Advanced implementation	Stagnation / discontinuance
Open Content	OC_interest = **0**	OC_interest = **1**	OC_5yrs_avg >= **0.5%**	OC_pres_avg >= **0.5%**	(OC_pres_avg > **10%** **AND** OC_5yrs_avg > OC_pres_avg +5) **OR** (OC_pres_avg >**5%** **AND** OC_5yrs_avg > OC_pres_avg +10)	OC_pres_avg > **50%**	OC_pres_avg >= **0.5%** **AND** OC_pres_avg < **66%** **AND** OC_5yrs_avg <= OC_pres_avg **AND** imp_OC <= 3
Social Media	SM_interest = **0**	SM_interest = **1**	SM_types_1yr_num > 0	SM_types_pres_num > 0	(SM_types_pres_num > 0 **AND** SM_types_1yr_num > SM_types_pres_num) **OR** (SM_types_pres_num > 1 **AND** SM_types_1yr_num >= SM_types_pres_num)	SM_types_pres_num > 3	SM_types_pres_num > 0 **AND** SM_types_pres_num < 4 **AND** SM_types_1yr_num < SM_types_pres_num **AND** imp_ENGAGE <= 3
Collaborative Content Creation	WPWM_interest = **0**	WPWM_interest = **1**	COLLAB_types_1yr_num > 0	COLLAB_types_pres_num > 0) **OR** Collab_prof_num > 0	(COLLAB_types_pres_num > 0 **AND** COLLAB_types_1yr_num > COLLAB_types_pres_num) **OR** (COLLAB_types_pres_num > 1 **AND** COLLAB_types_1yr_num >= COLLAB_types_pres_num)	COLLAB_types_pres_num > 2	COLLAB_types_pres_num > 0 **AND** COLLAB_types_pres_num < 3 **AND** COLLAB_types_1yr_num < COLLAB_types_pres_num **AND** imp_COLLAB <= 3

Adoption Rates According to Institution Type



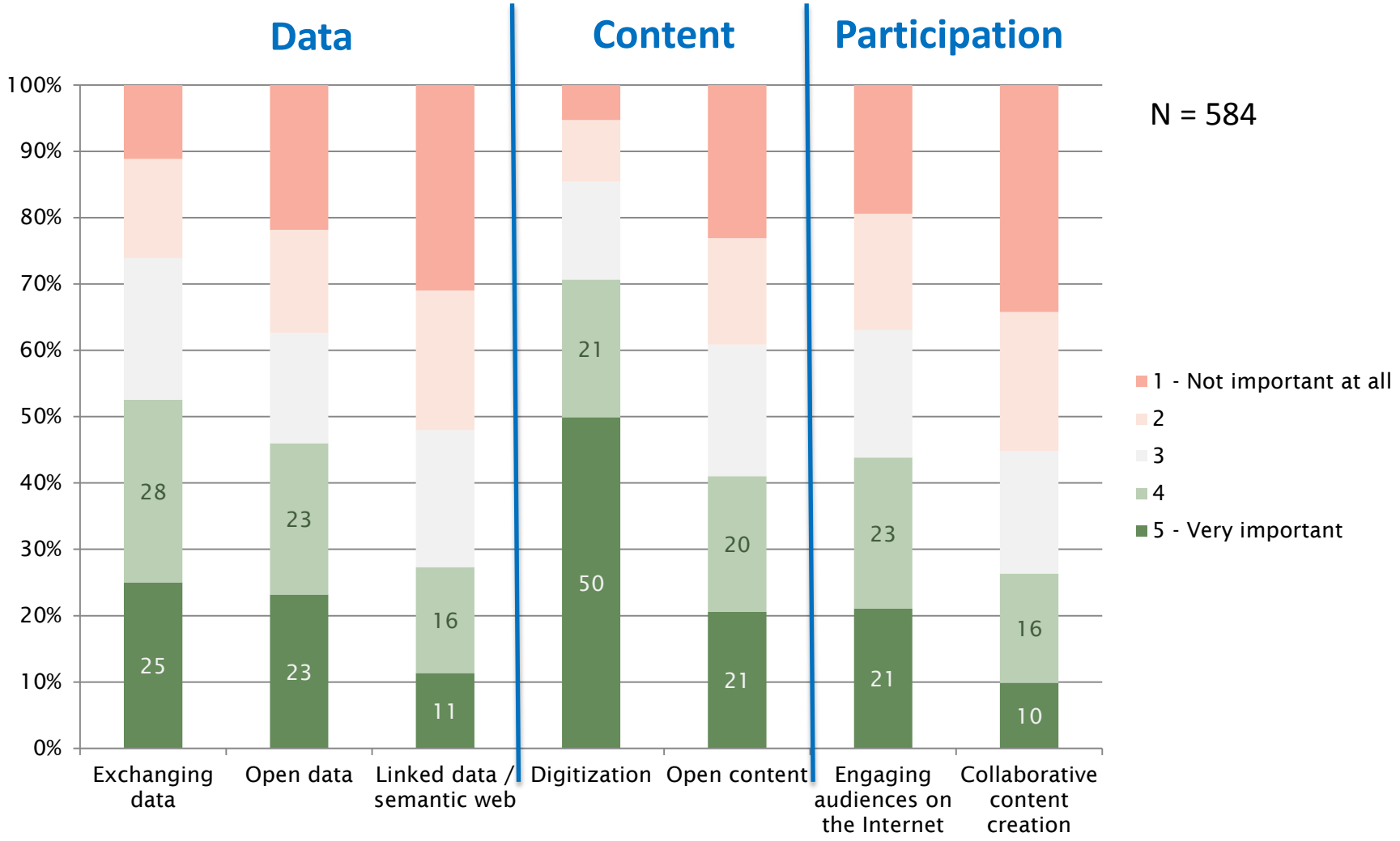
Adoption Rates – Country Comparison



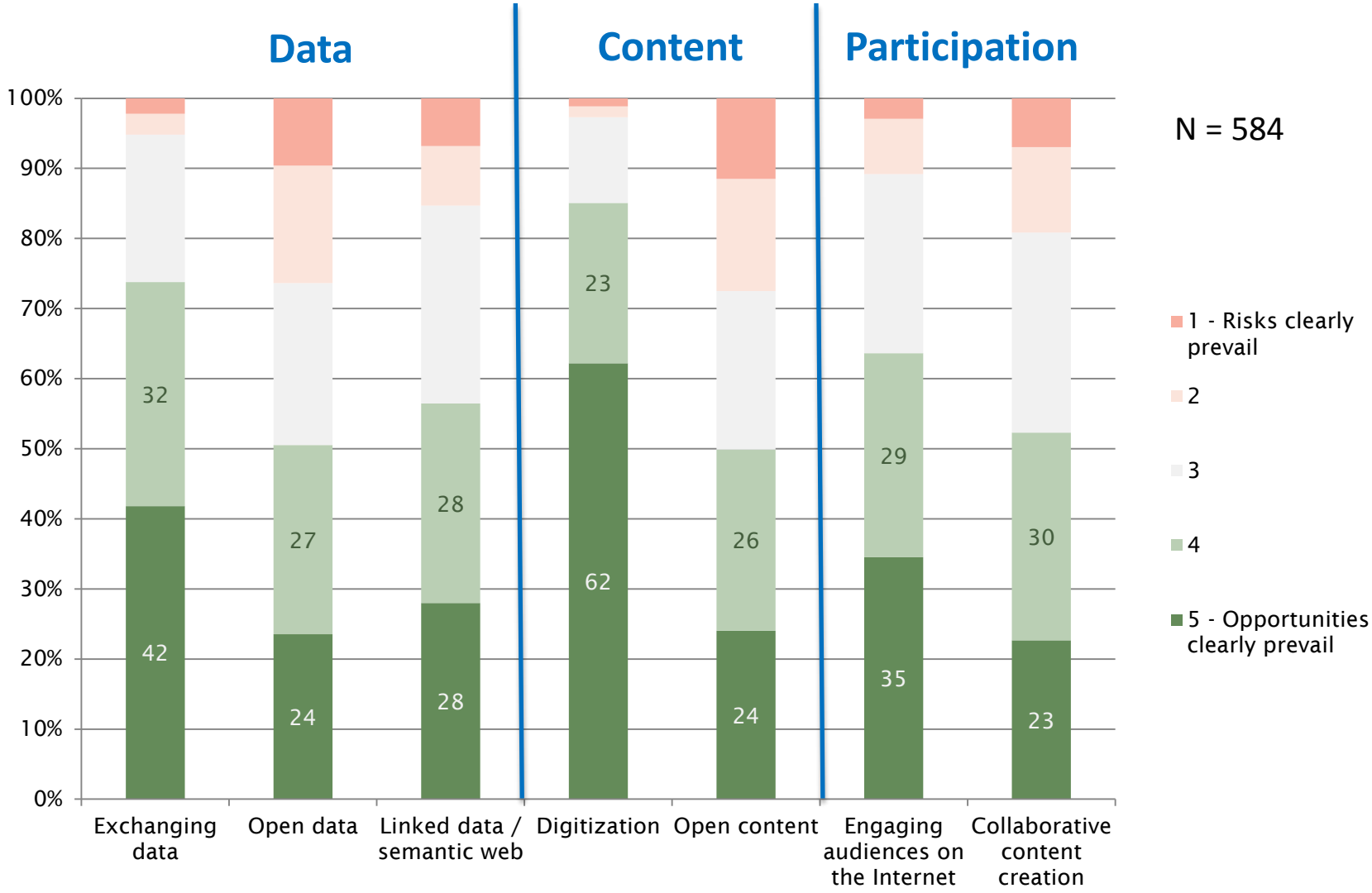
Adoption Rates – Further Insights

- Heritage institutions with a **local/regional focus** are **less likely** to use **social media** than other institutions
- The **institutions' size** **is** **positively correlated with** the adoption of:
  - **open data**
  - **linked data**
  - **social media use**
  - **collaborative content creation**
- However, the **institutions' size** **is** **not related to** their adoption of:
  - **digitization**
  - **open content**
- **Higher levels of public funding are:**
  - **positively related to** the adoption of **open data**
  - **negatively related to** the adoption of **open content**

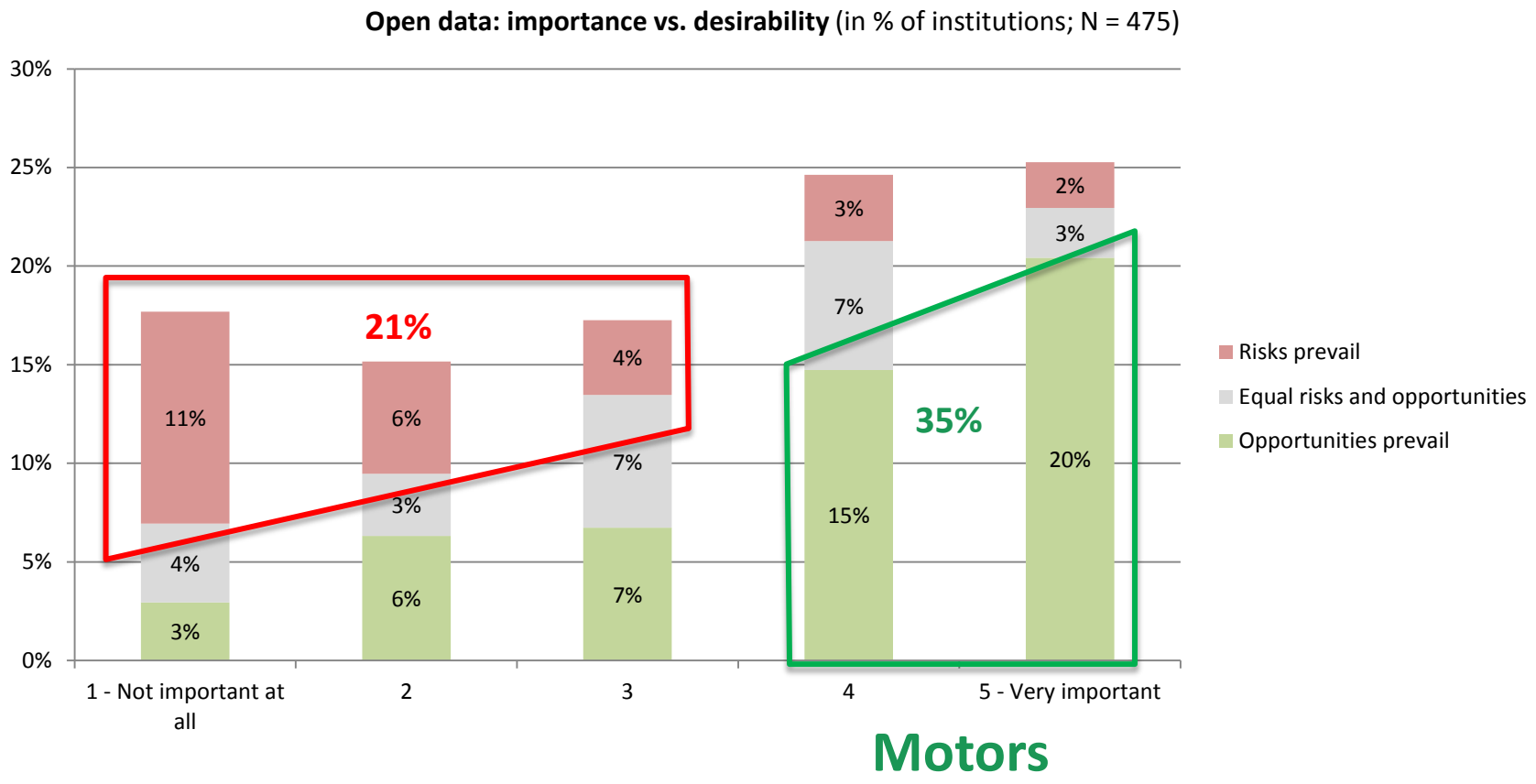
Importance of Various Practices



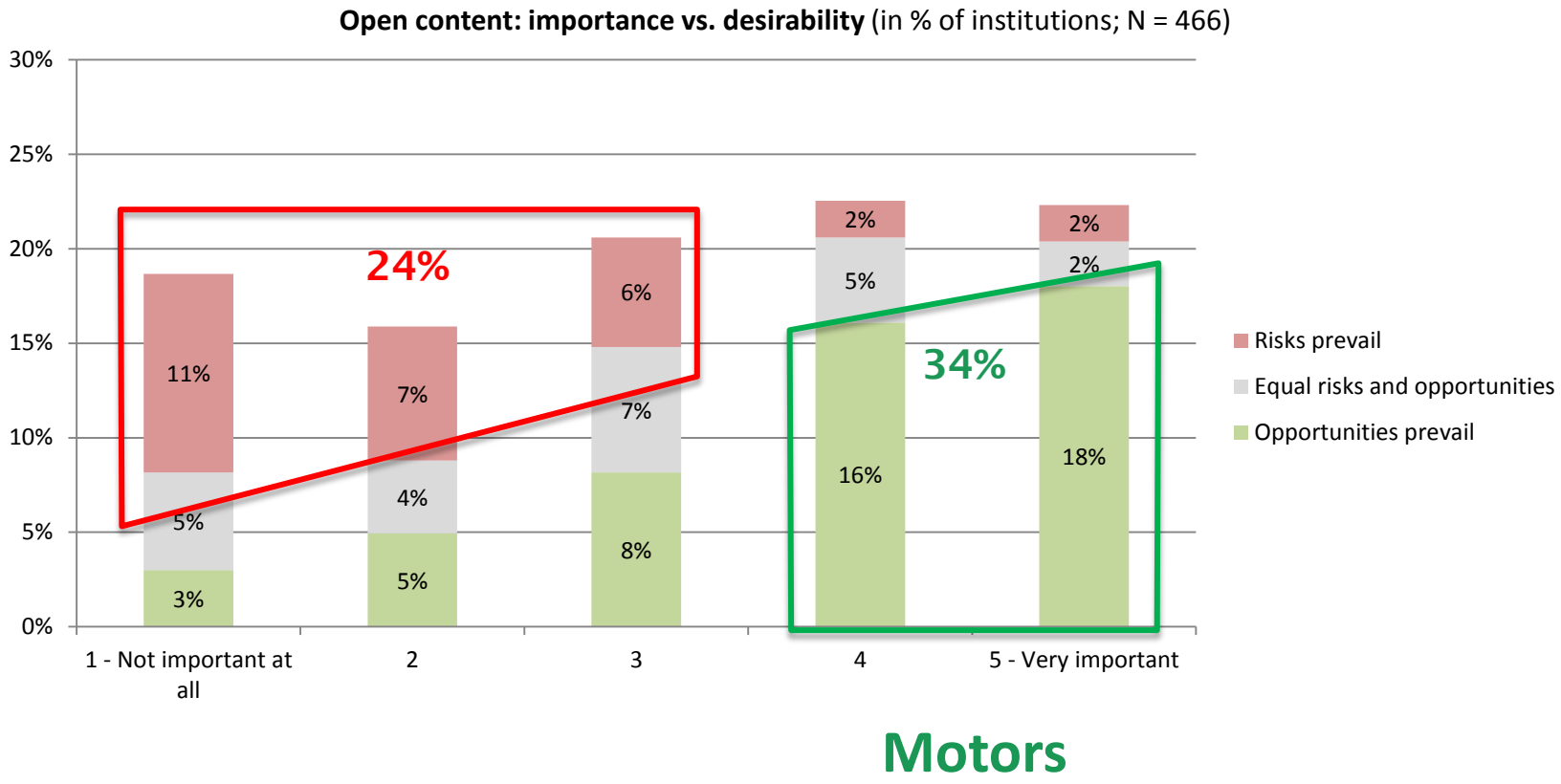
Desirability of Various Practices (Opportunities vs. Risks)



Open Data: Importance vs. Desirability

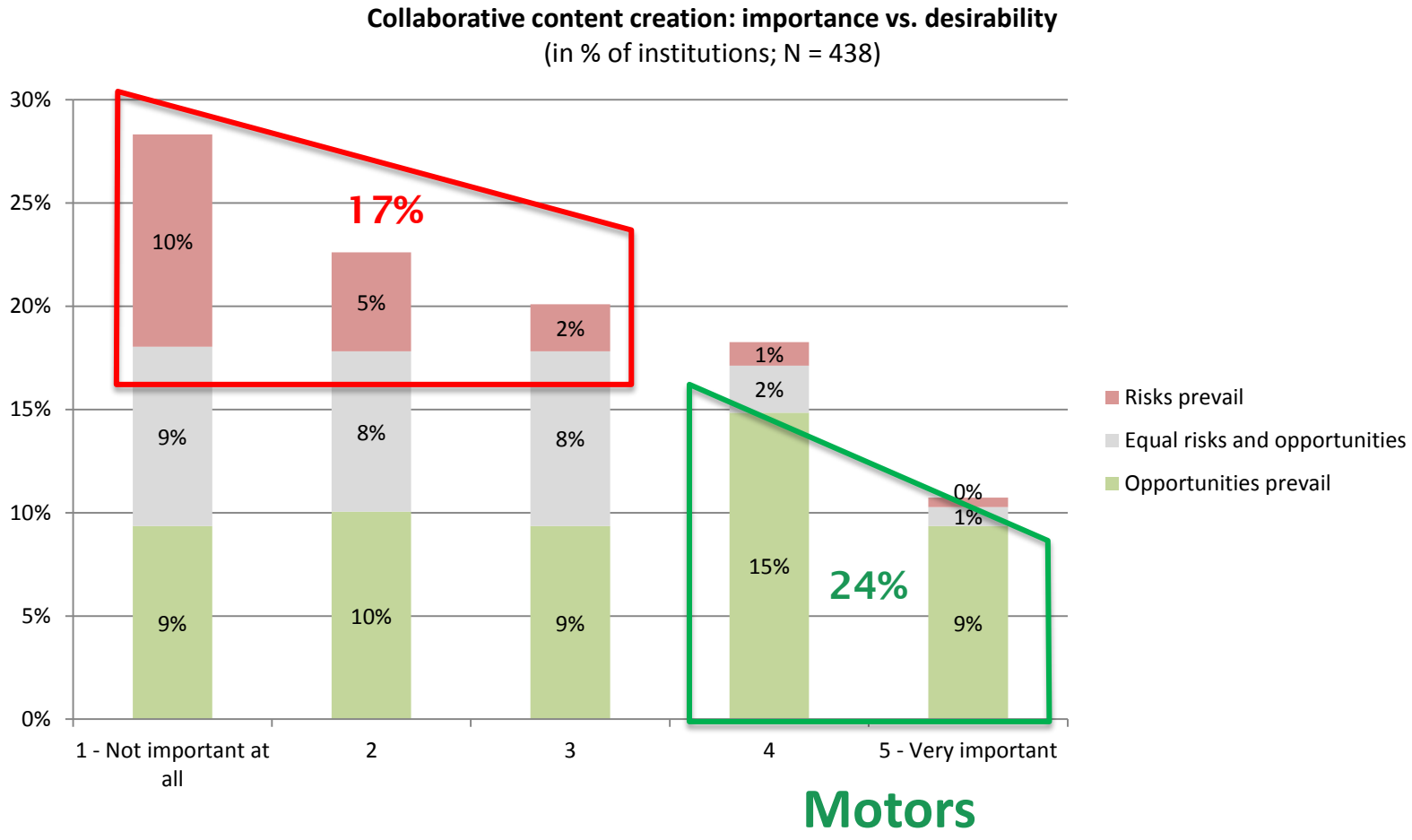


Open Content: Importance vs. Desirability

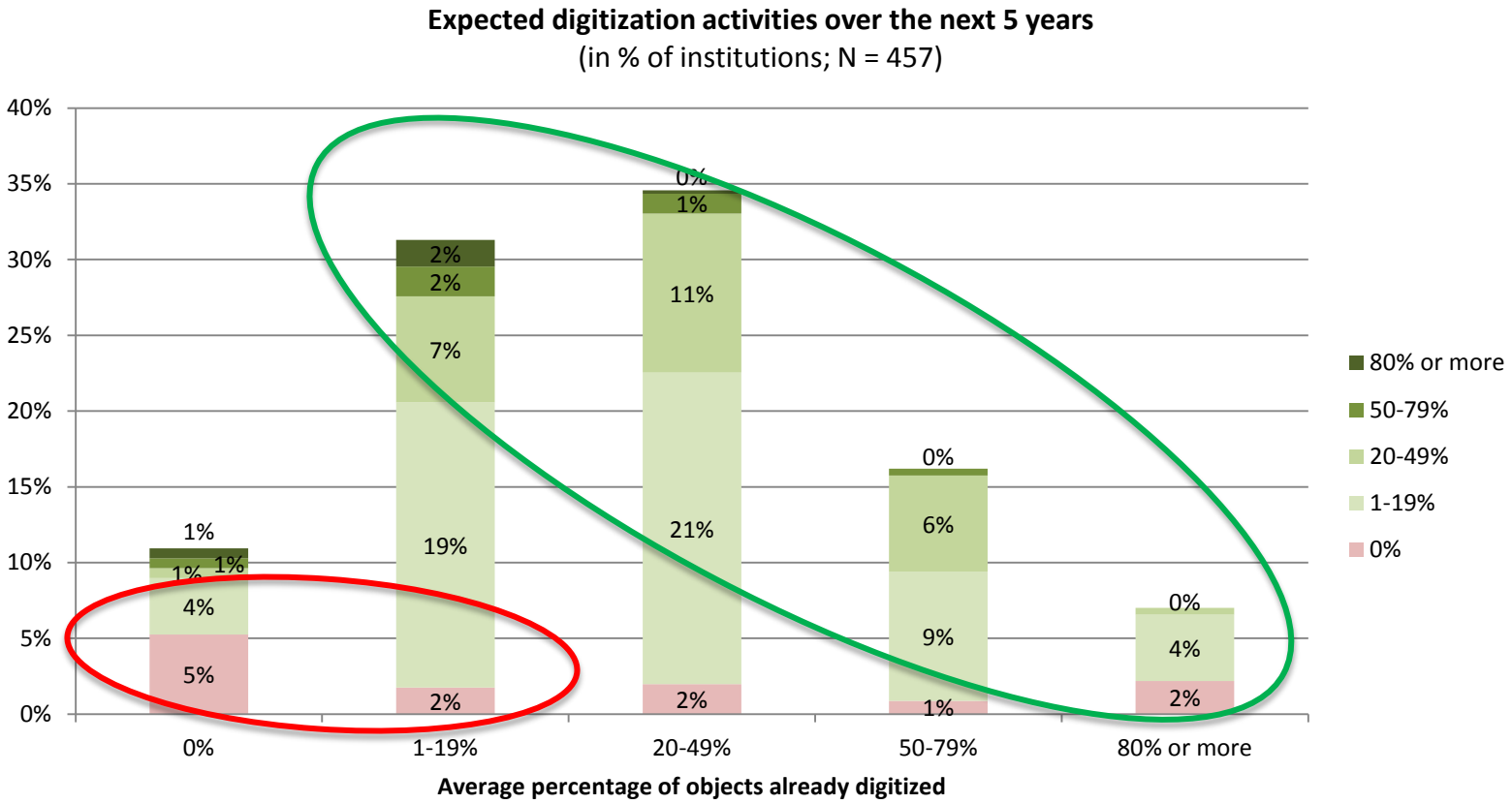


Motors

Collaborative Content Creation: Importance vs. Desirability



Digitization as a Prerequisite for Open Content



Digitization: Different types of objects (today / in 5 years)

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	366	357	282	314	47	120	221	543
	Missing	218	227	302	270	537	464	363	41
Mean		17.2350	33.1997	14.9610	34.6153	25.1277	32.5258	32.3385	30.2547
Median		5.0000	20.0000	2.0000	20.0000	5.0000	10.0000	10.0000	25.0000
Percentiles	25	.0000	2.0000	.0000	.0750	.0000	.0000	.0000	5.0000
	50	5.0000	20.0000	2.0000	20.0000	5.0000	10.0000	10.0000	25.0000
	75	20.0000	60.0000	10.0000	70.0000	50.0000	67.5000	60.0000	49.0000

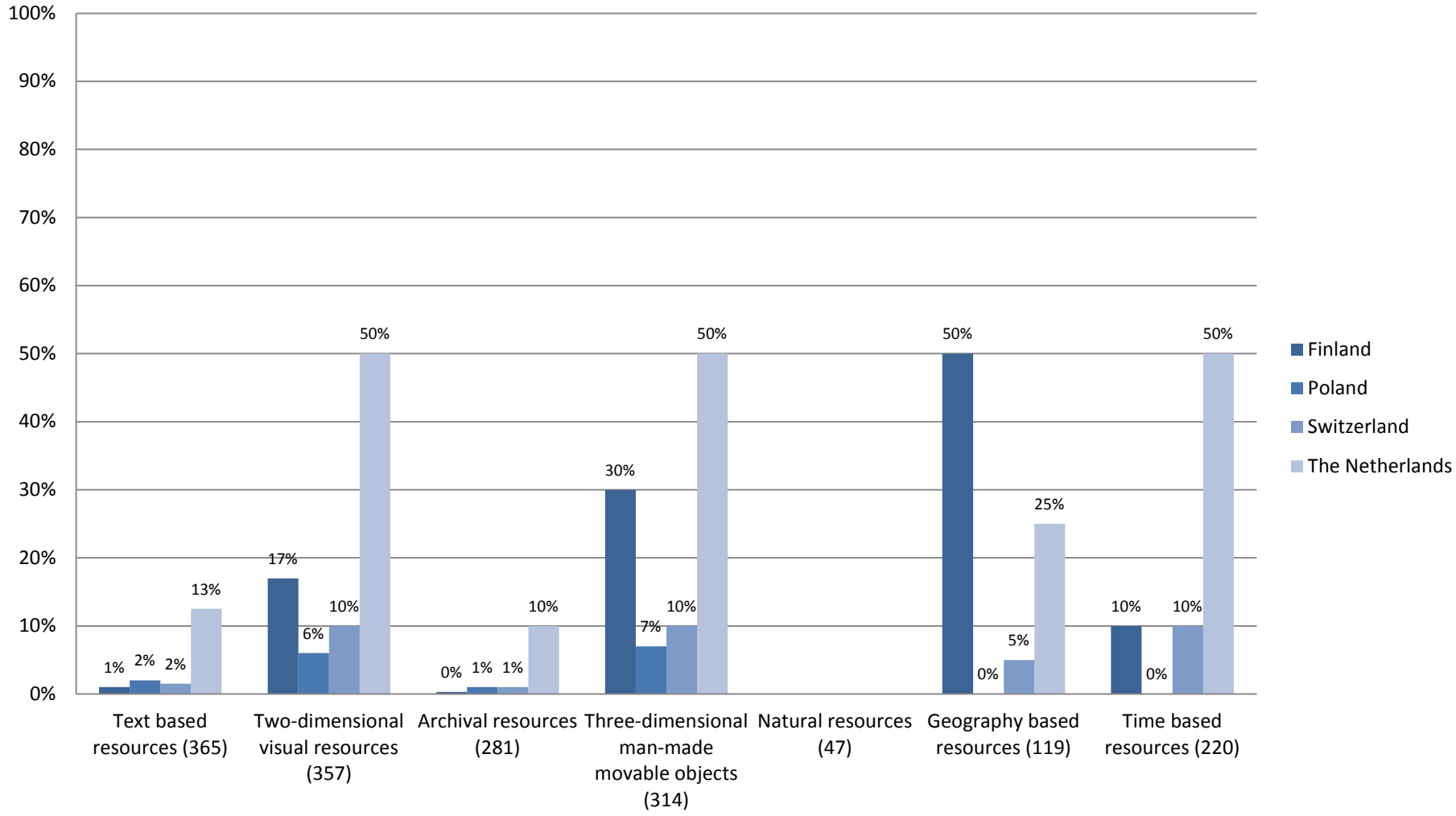
Statistics

		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	319	317	252	279	39	100	197	489
	Missing	265	267	332	305	545	484	387	95
Mean		32.1520	51.7565	30.2012	53.9631	33.1538	44.9410	50.0706	44.3729
Median		20.0000	50.0000	15.0000	50.0000	10.0000	45.0000	50.0000	40.0000
Percentiles	25	2.0000	15.0000	2.0000	15.0000	2.0000	4.2500	10.0000	20.0000
	50	20.0000	50.0000	15.0000	50.0000	10.0000	45.0000	50.0000	40.0000
	75	50.0000	90.0000	50.0000	95.0000	60.0000	88.7500	90.0000	66.6667

Digitization Today – Country Comparison

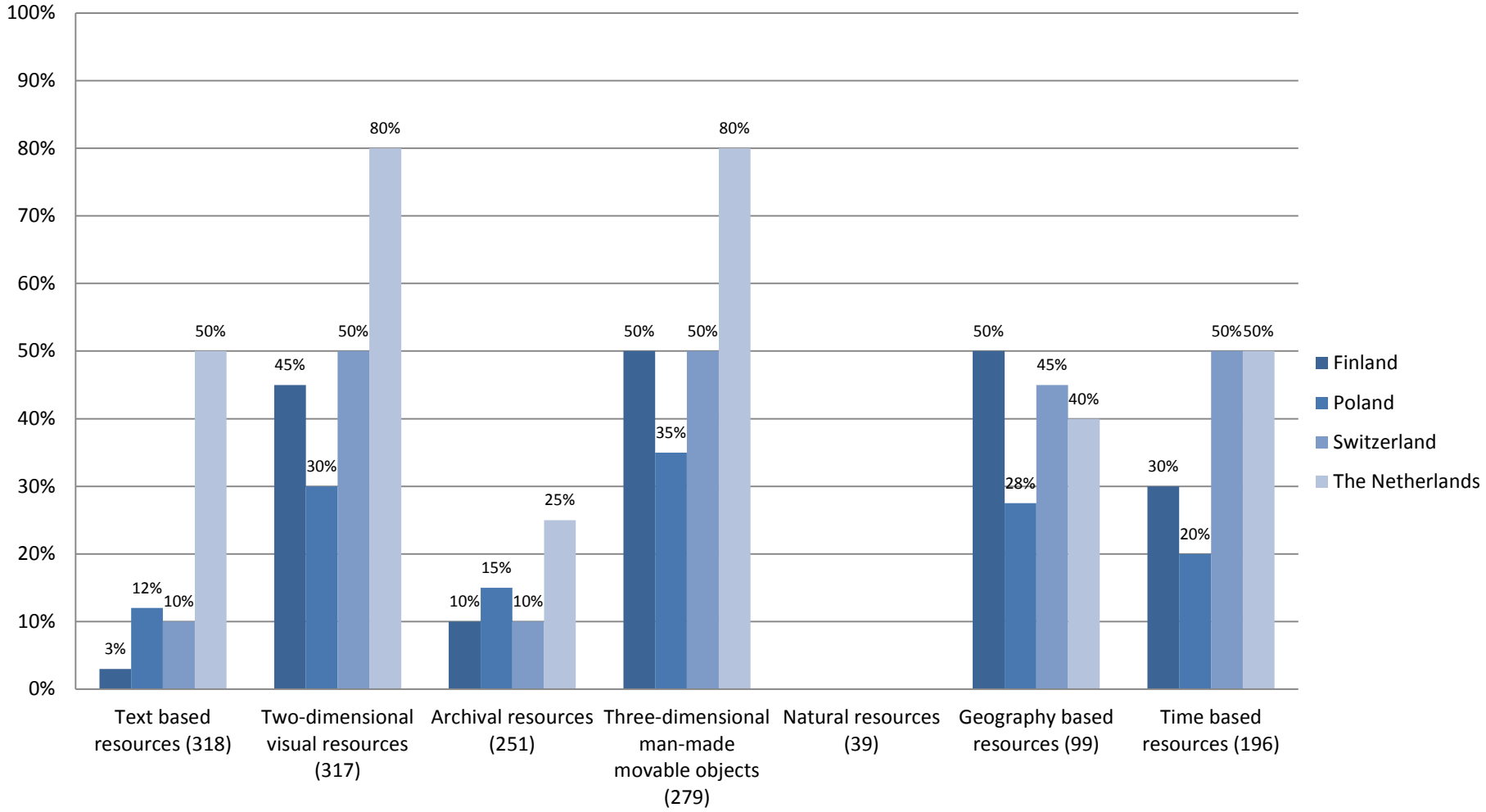
Percentage of resources already digitized by the average institution (median)

(N is indicated for each object type; the values for natural resources are not shown due to the small sample size)



Digitization in 5 Years – Country Comparison

Percentage of resources expected to be digitized in 5 years by the average institution (median)
(N is indicated for each object type; the values for natural resources are not shown due to the small sample size)



Open Content: Different types of objects (today / in 5 years; in % of digitized objects)

Statistics

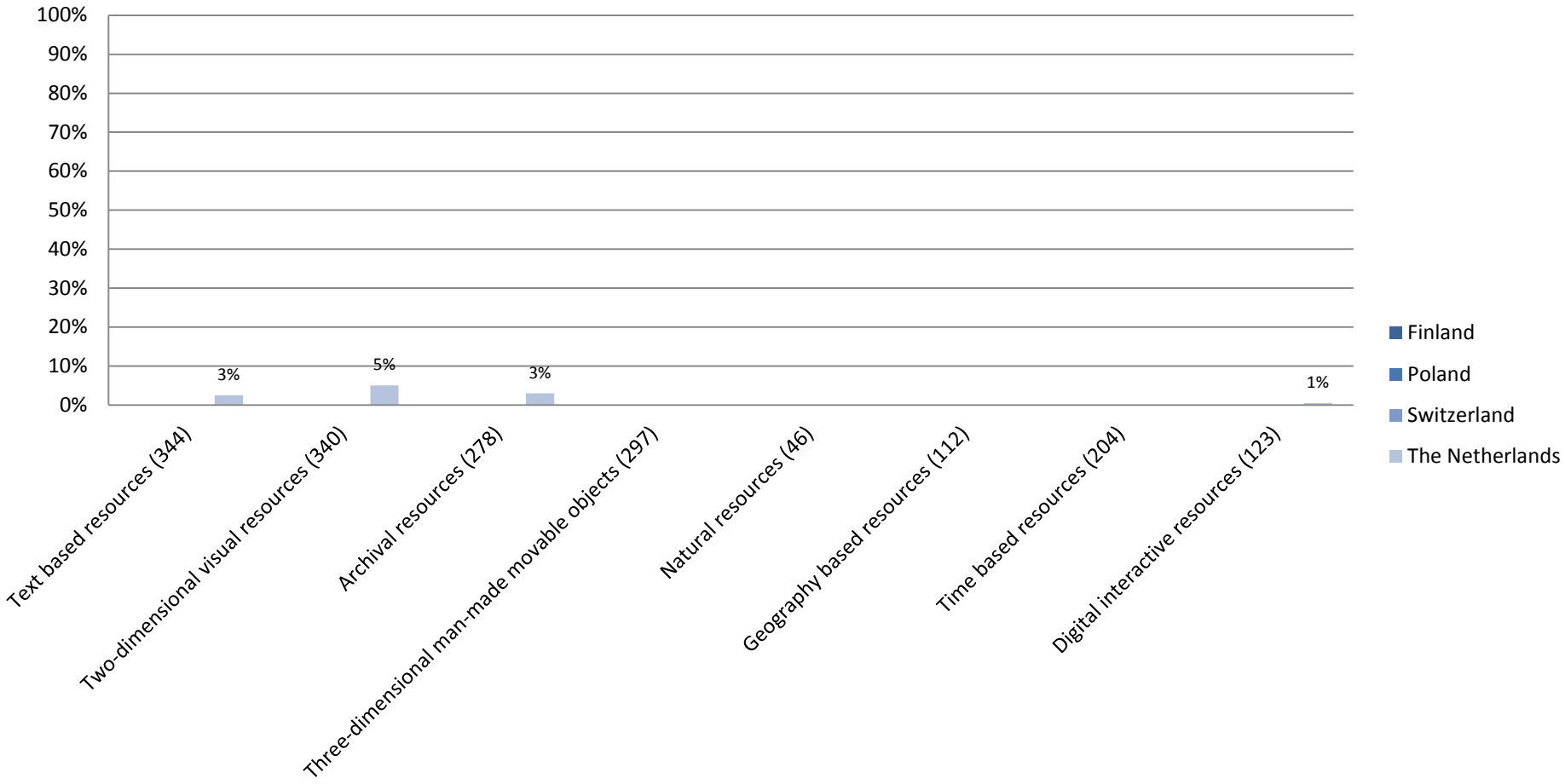
		Text based resources: available as open content (% of digitized)	Two-dimensional visual resources: available as open content (% of digitized)	Archival resources: available as open content (% of digitized)	Three-dimensional man-made movable objects: available as open content (% of digitized)	Natural resources: available as open content (% of digitized)	Geography based resources: available as open content (% of digitized)	Time based resources: available as open content (% of digitized)	Digital interactive resources: available as open content (% of digitized)	Average percentage of objects presently available as "open content" (in % of digitized objects)
N	Valid	222	257	160	203	22	64	143	124	430
	Missing	362	327	424	381	562	520	441	460	154
Mean		39.1164	29.2038	37.1921	20.9547	29.6591	38.4321	22.9816	13.4040	23.7521
Median		6.3492	.0000	.0000	.0000	.0000	5.0000	.0000	.0000	4.1479
Percentiles	25	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	50	6.3492	.0000	.0000	.0000	.0000	5.0000	.0000	.0000	4.1479
	75	100.0000	66.6667	100.0000	20.0000	70.0000	100.0000	20.0000	10.0000	44.3254

Statistics

		Text based resources: available as open content in 5 years (% of digitized)	Two-dimensional visual resources: available as open content in 5 years (% of digitized)	Archival resources: available as open content in 5 years (% of digitized)	Three-dimensional man-made movable objects: available as open content in 5 years (% of digitized)	Natural resources: available as open content in 5 years (% of digitized)	Geography based resources: available as open content in 5 years (% of digitized)	Time based resources: available as open content in 5 years (% of digitized)	Digital interactive resources: available as open content in 5 years (% of digitized)	Average percentage of objects available as "open content" in 5 years (in % of digitized objects)
N	Valid	211	234	166	188	28	64	144	109	390
	Missing	373	350	418	396	556	520	440	475	194
Mean		47.2222	42.6961	44.7670	39.2118	45.6349	60.3398	38.1857	28.1385	36.0464
Median		40.0000	30.6250	25.0000	22.2222	45.0000	81.6667	20.0000	5.0000	25.8681
Percentiles	25	.0000	.0000	.0000	.0000	.0000	11.6667	.0000	.0000	1.0000
	50	40.0000	30.6250	25.0000	22.2222	45.0000	81.6667	20.0000	5.0000	25.8681
	75	100.0000	100.0000	100.0000	88.5526	100.0000	100.0000	95.8333	50.0000	60.0000

Open Content Today – Country Comparison

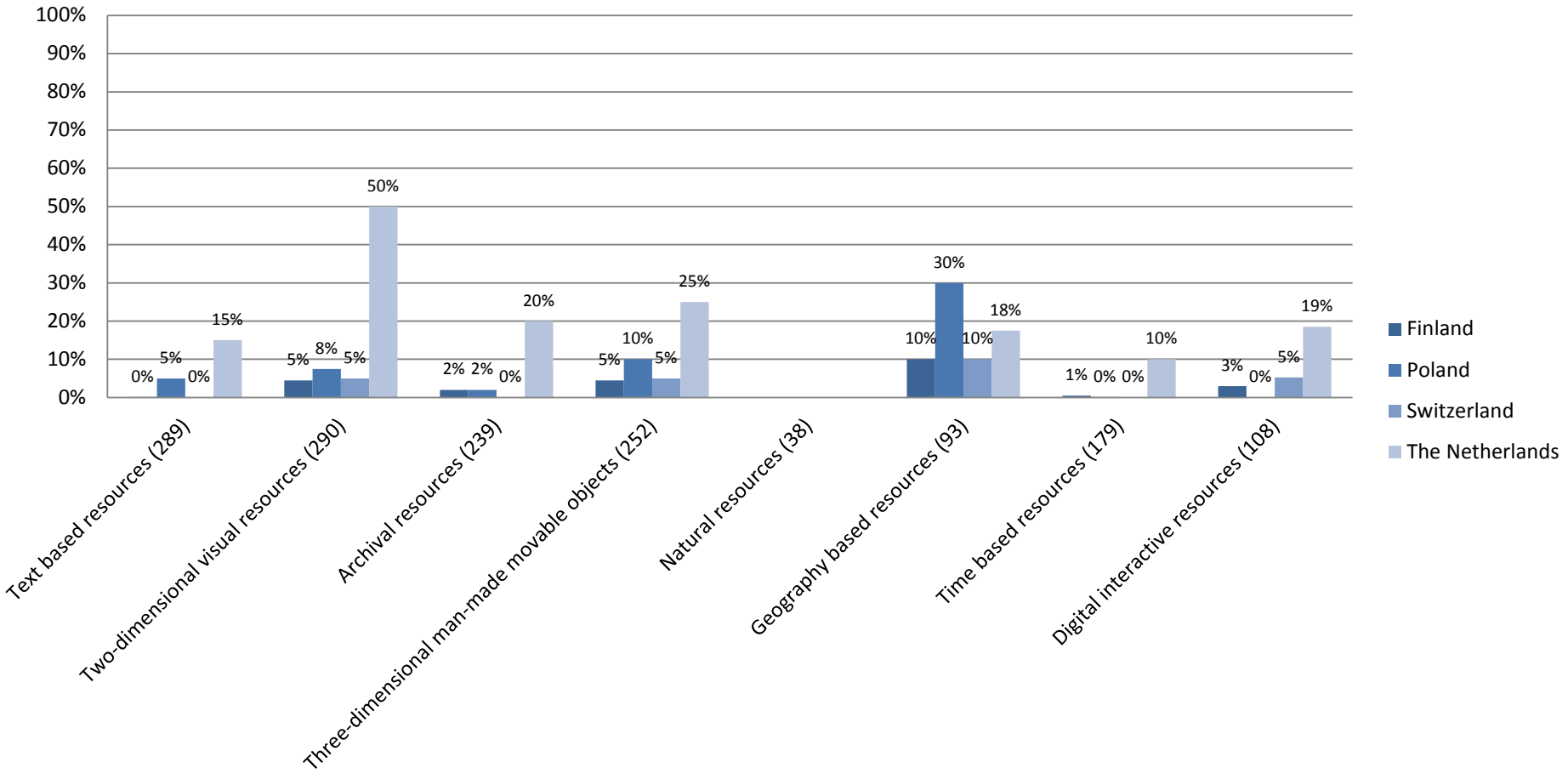
Percentage of resources already made available as open content by the average institution (median)
(N is indicated for each object type; the values for natural resources are not shown due to the small sample size;
"0%" value tags have been suppressed)



Open Content in 5 Years – Country Comparison

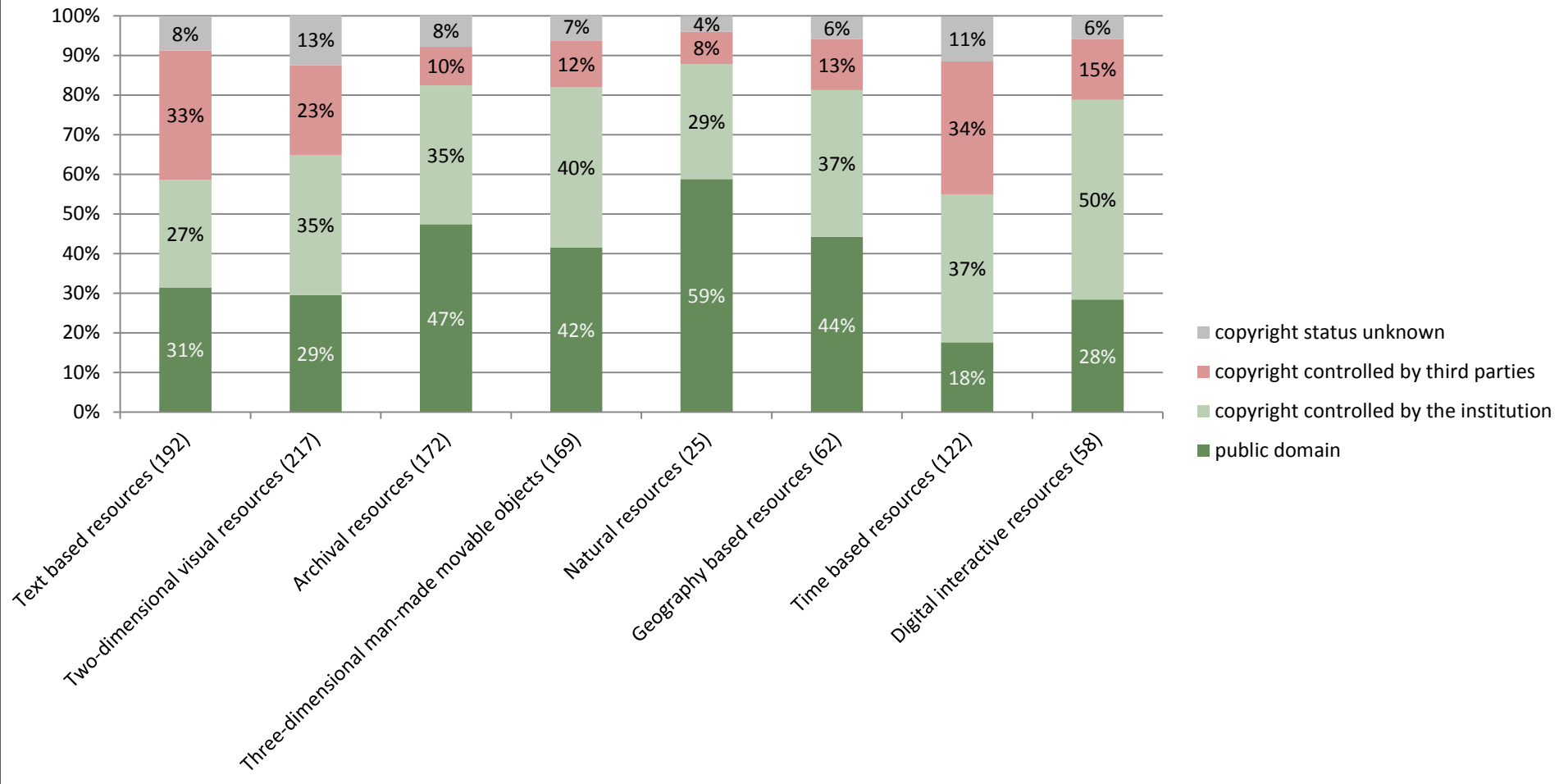
Percentage of resources expected to be made available as open content in 5 years by the average institution (median)

(N is indicated for each object type; the values for natural resources are not shown due to the small sample size)



Copyright Clearance as a Prerequisite for Open Content

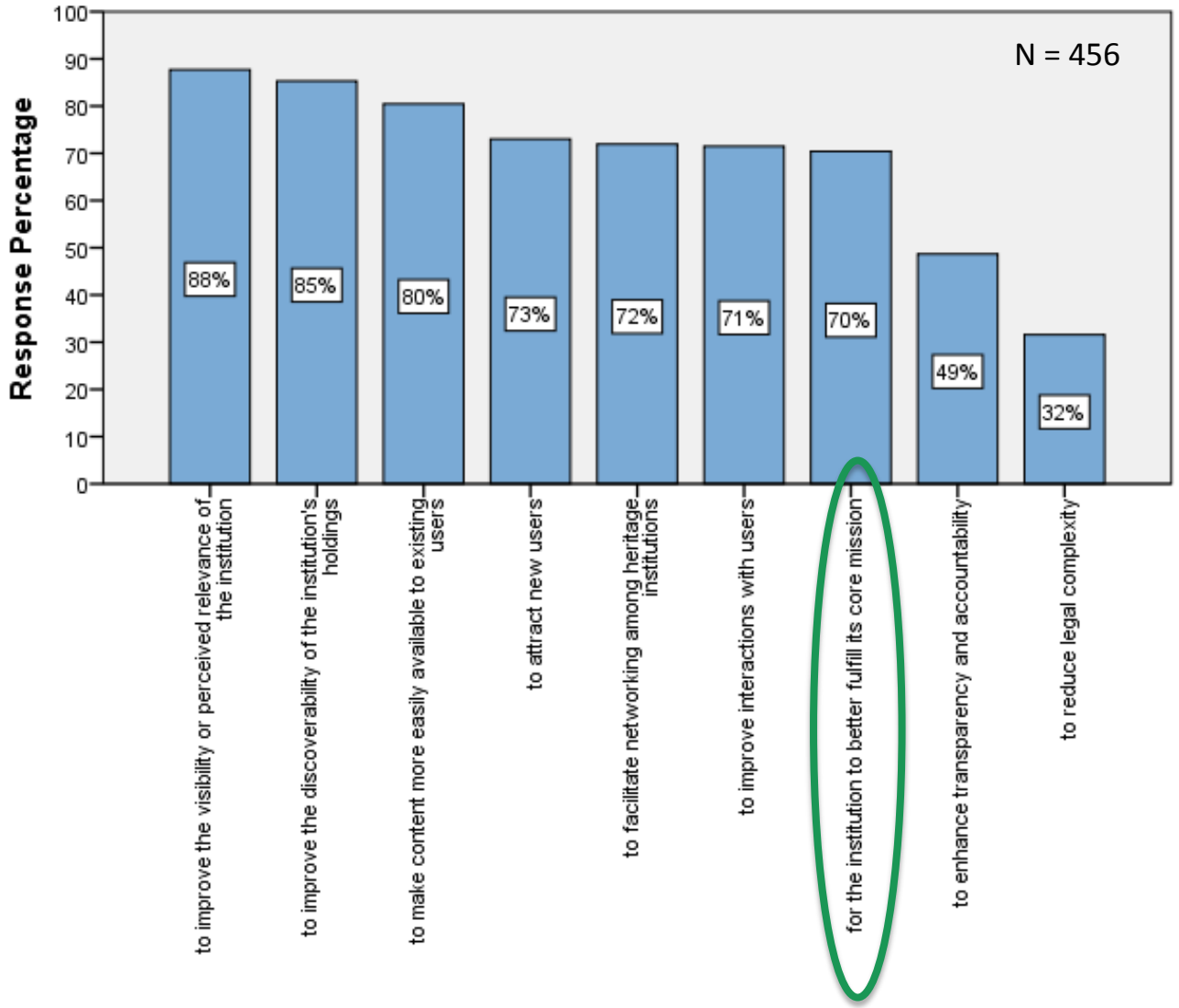
What is the copyright situation of the objects in your collection?
(average percentages for each object type; N is indicated for each object type)



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.*

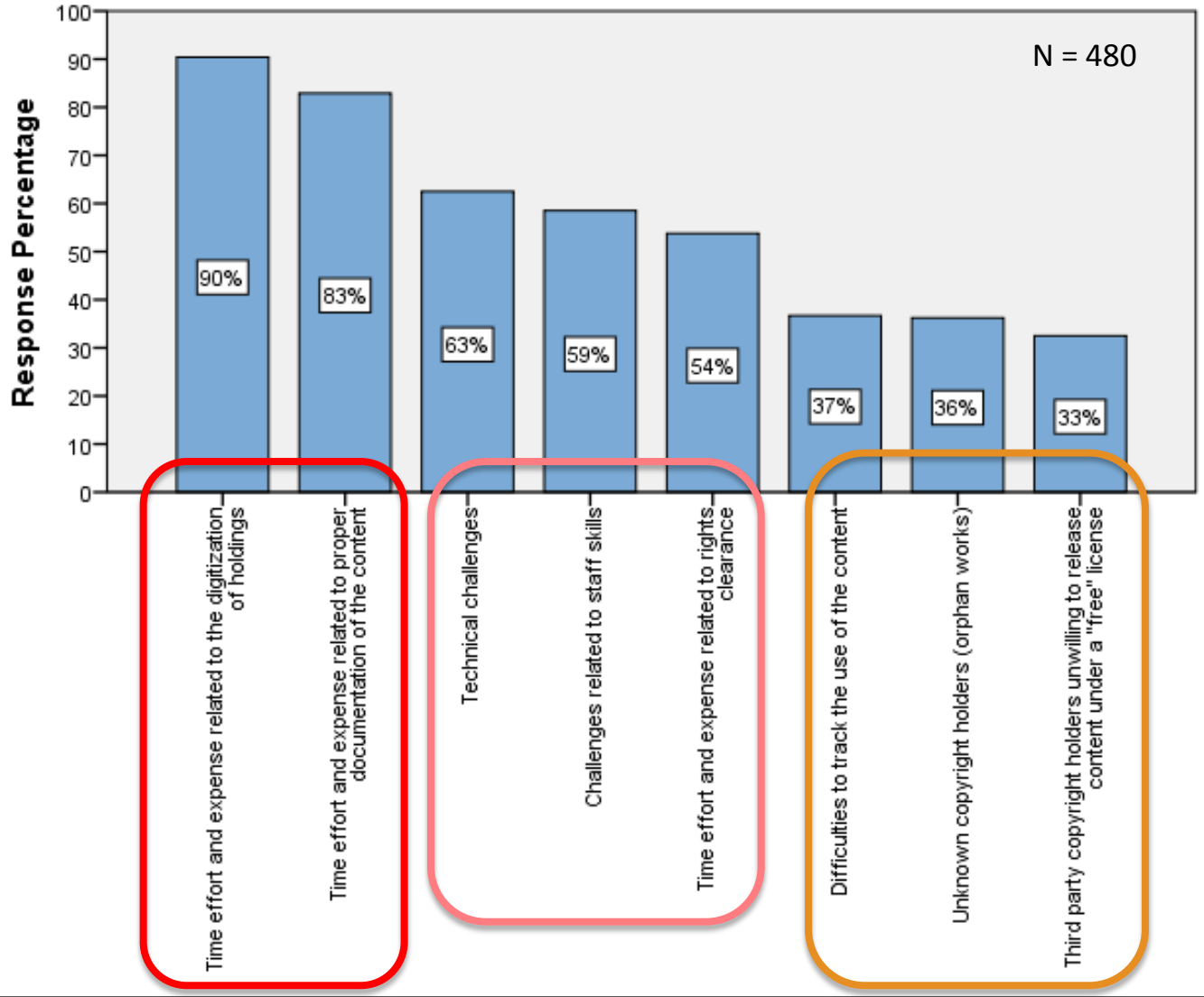
Benefits of Open Content

D7: In the case of my institution, ‘open content’ is an important means…



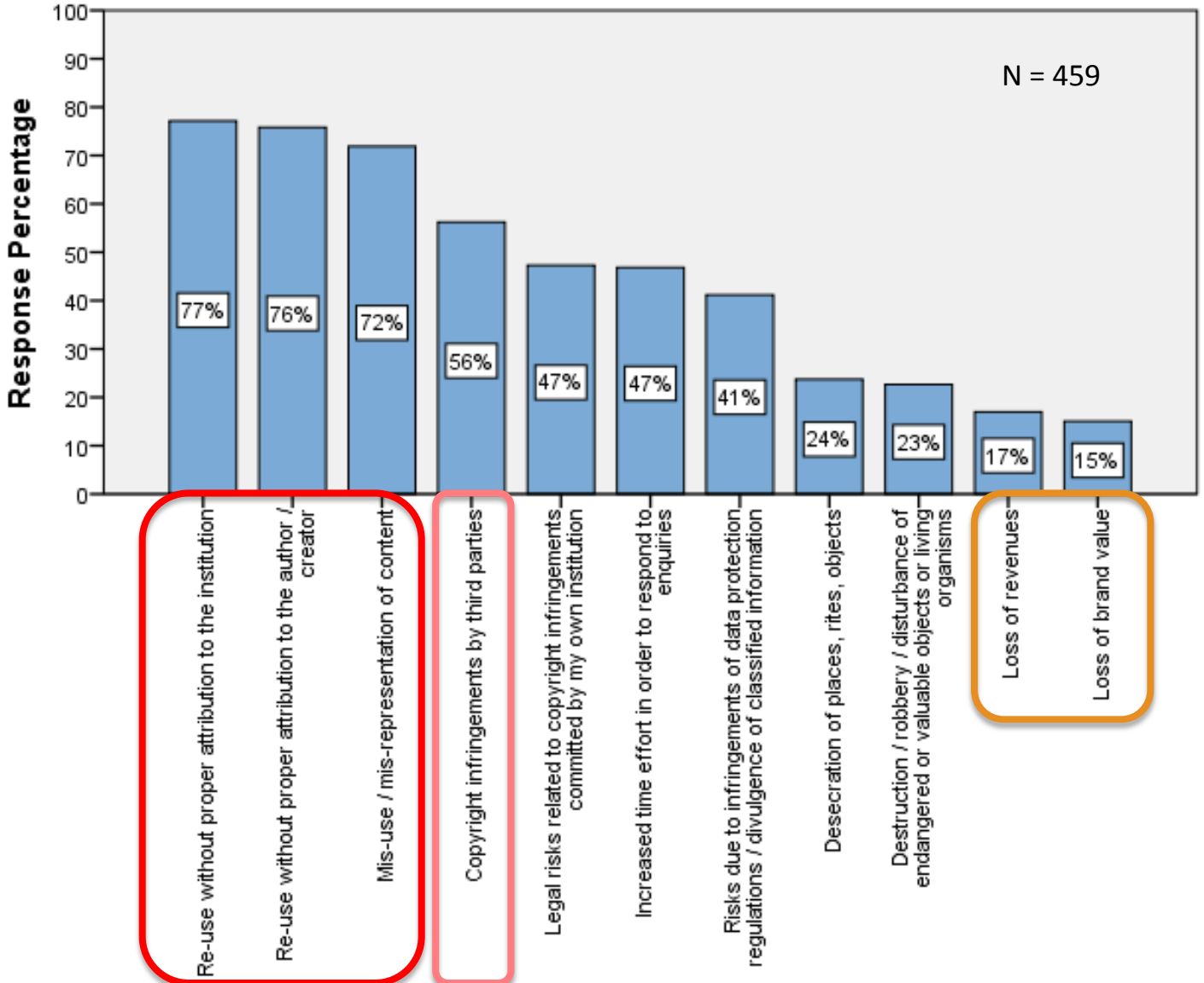
Challenges Related to Open Content

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

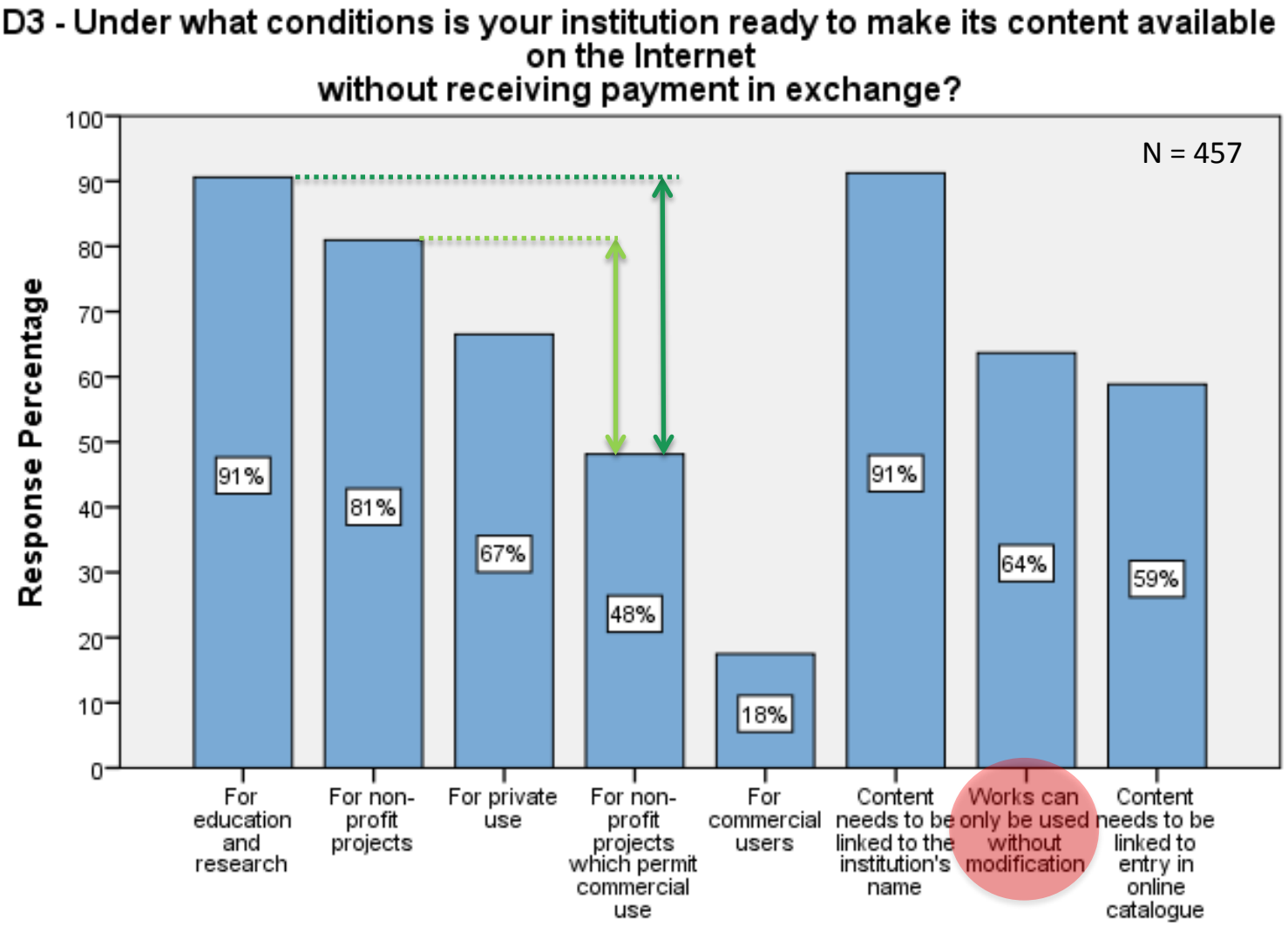


Risks Related to Open Content

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



Conditions for Releasing Content

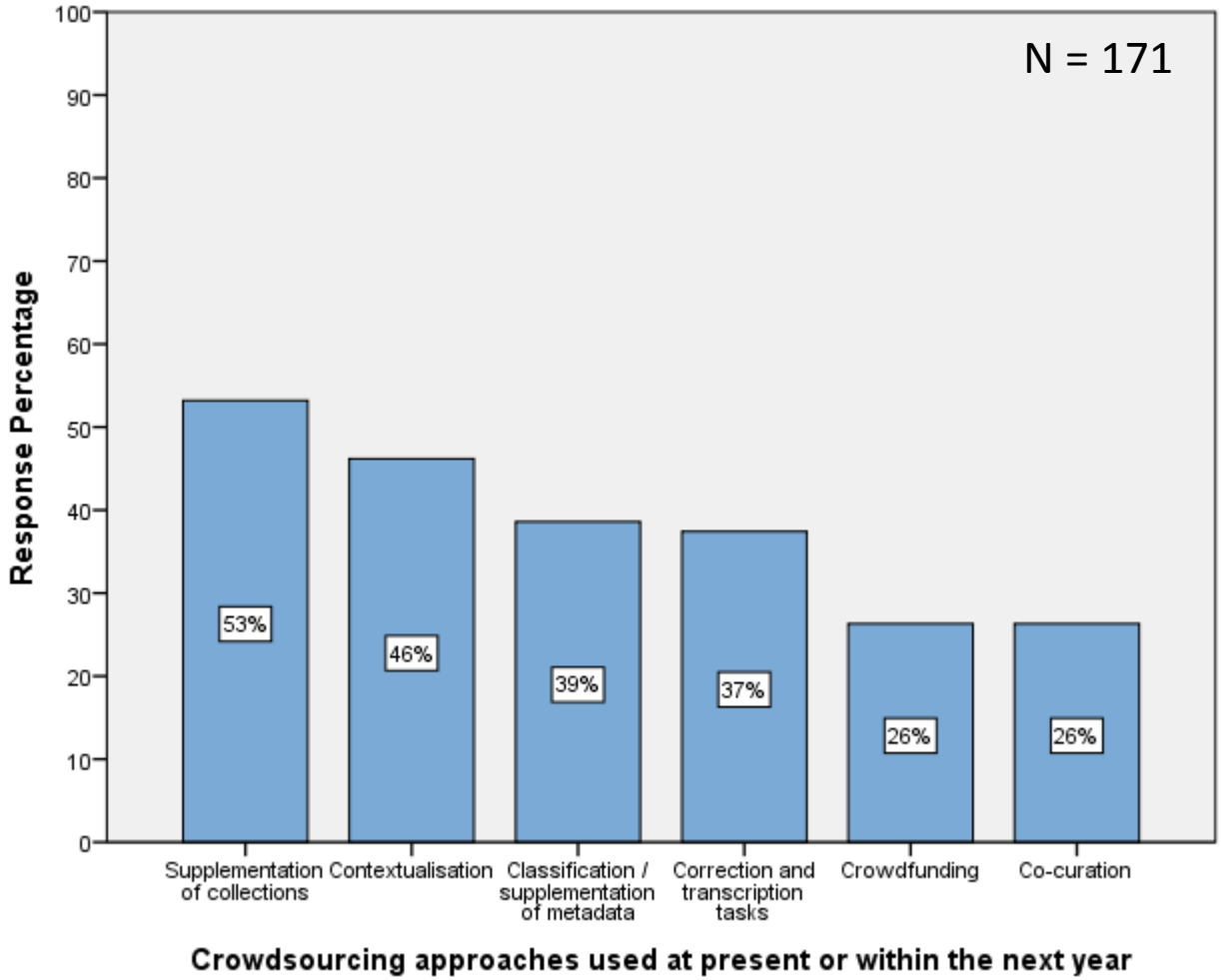


GLAM Involvement in Wikipedia

- 14% of responding GLAMs have staff members who are involved in Wikipedia **as part of their job**
  (**28%** have staff members who are involved in Wikipedia **either as part of their job or during leisure time**)
- Wikimedia Commons: 4%
- Wikidata: 2%

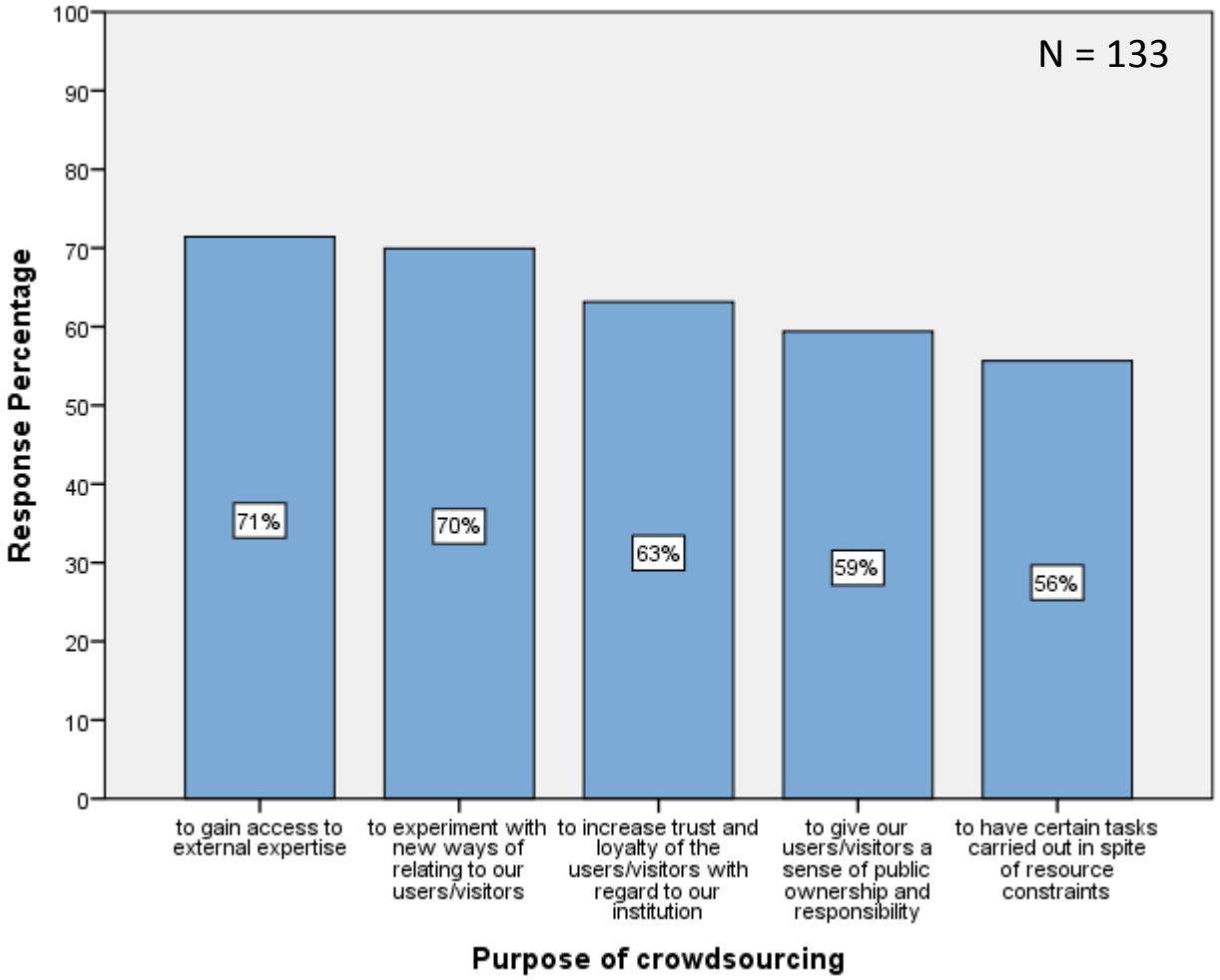
Most popular crowdsourcing approaches

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.*



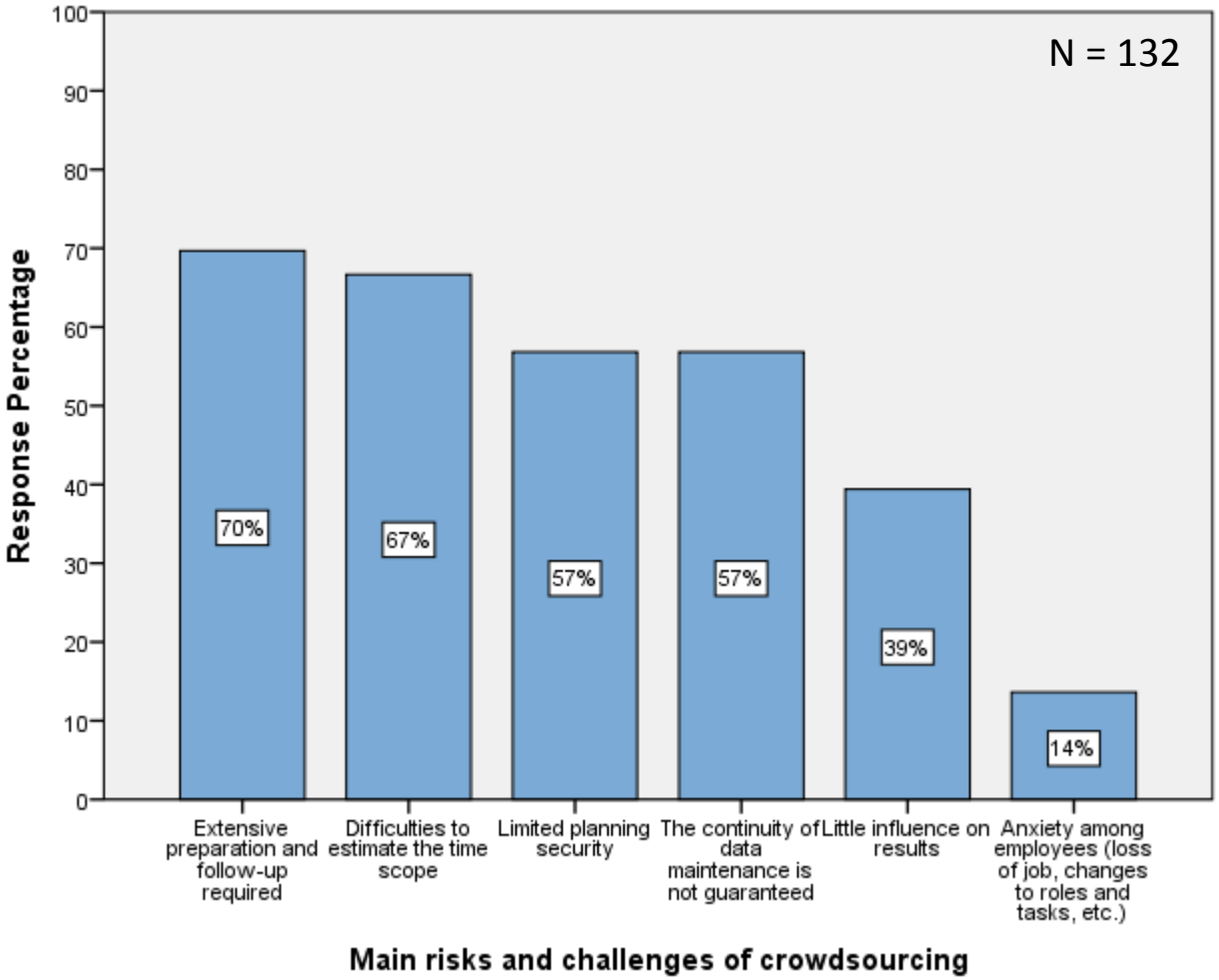
Purpose of crowdsourcing

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** is a widespread practice in the Dutch and Finnish GLAM sector, with adoption rates well above 50%.
  (Switzerland and Poland are lagging behind; there are considerable differences between countries and institution types).
- The **Use of Social Media** is a widespread practice in the Finnish, Polish, and Dutch GLAM sector, with adoption rates of 70% and more.
  (Switzerland is lagging behind; there are considerable differences between countries and institution types).
- **Open Data** and **Open Content** are about to diffuse within the GLAM sector; for Open Data we are observing an overall adoption rate of ca. 26%; for Open Content the adoption rate is about 17%.
  (The Netherlands are ahead of the other three countries; there are considerable differences between countries and institution types).
- **Collaborative Content Creation** is diffusing slightly more slowly than Open Data / Open Content
  (There are considerable differences between countries).
  - These findings are reflected in the **importance accorded to the various practices**; over 40% of respondents consider Open Data and Open Content as important; ca. 25% say so with regard to Collaborative Content Creation.

Prerequisites for Open Content

- Over the next 5 years we will see **leaps in digitization activities and in freely licensed content**. – Are we ready?
- 70% 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, lack of staff skills as well as technical challenges.

- The **greatest risks associated with open content** are re-use of content without proper attribution or mis-use / mis-interpretation of the content. Another important risk are copyright infringements.

- Roughly one third of the institutions is worried about **issues related to rights clearance** and **tracking the use of the content.**

- When it comes to opening up content, loss of revenues, or diminished brand value **are only minor preoccupations.**

Involvement in Wikipedia

- Ca. 24% of GLAMs consider **Collaborative Content Creation** as important and believe that the opportunities prevail over the risks
- Ca. 14% of GLAMs are already involved in **Wikipedia**.
- **Wikimedia Commons** and **Wikidata** still seem to play a minor role for GLAMs (with 4% and 2% of institutions involved).

Outlook

What's next?

- Data collection in further countries – Your support is appreciated… http://survey.openglam.ch
- Context analysis (searching for explanations for country differences)
- 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

What else would you like to know?

- Discussion…

Thank you for your attention!

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

- **Project Portal:**
  - http://survey.openglam.ch