

SODA Project

Social Sciences Data Archive

Report on a survey of the needs and practices
of Belgian researchers in social sciences
in terms of research data management



Authors of the report

Benjamin Peuch ^a, Jean-Paul Sanderson ^b, Laura Van den Borre ^c,
Rolande Depoortere ^a, Freya De Schamphelaere ^a, Samira Hajji ^a,
Johan Surkyn ^c, Aziz Naji ^d

^a State Archives of Belgium, *Dienst Digitale Archivering – Service Archivage digital* (DAD)

^b *Université catholique de Louvain* (UCLouvain), *Centre de recherche en démographie* (DEMO)

^c *Vrije Universiteit Brussel* (VUB), *Interface Demography* (ID)

^d Belgian Science Policy Office (BELSPO)

June 2019



Introduction

This document presents the findings of a survey investigating the needs and practices of Belgian social sciences researchers in terms of research data management. The survey consisted of a pre-test, conducted in August 2017, and was then launched in October 2018.

The survey took place within the framework of the Social Sciences Data Archive (SODA) project, which aims to set up a data archive for social sciences that will also act as a CESSDA (Consortium of European Social Science Data Archives) service provider in Belgium. The SODA project is financed by the Belgian Science Policy Office (BELSPO) under contract #FR/00/SO4.

This report is meant to disseminate the aggregated results of the survey, to provide feedback to all of the respondents, and to serve as a basis for a formal publication which will include further technical information.

Acknowledgments

The survey team would like to warmly thank all of the researchers who took the time to fill in our questionnaires, as well as the presidents, directors, coordinators, secretaries and other members of the research centers contacted who forwarded our request for participation to their fellow colleagues.

The fact that our survey elicited more than 200 reactions is a tremendous incentive for us. Not only that: it greatly helps us better understand the situation of Belgian social scientists today and how to tackle the challenges of setting up a data archive that addresses their needs.

The data of the survey will be made publicly accessible as soon as possible.

Structure of the report

First, the purpose and the methodology of the survey are presented, followed by a statement on the limitations of our work. After this, statistics based on the answers to each question are presented in bar charts.

The different subparts of this latter section are those of the questionnaires:

1. General information
2. Archiving research data
3. Documentation and metadata
4. Re-use of research data
5. Tools and software
6. Legal aspects

When several free text answers were provided by the respondents (typically when they selected “Other (*Please specify*)”), a short synthesis of the answers is given.

A list of the recipient research centers for each university, in alphabetical order, is placed at the end of the document.

Purpose of the survey

The goal of this survey was to provide the SODA project team with a good overview of the current practices of Belgian researchers in social sciences in terms of research data management.

- Are there well established data management codes and procedures in Belgian social science research centers?
- Is data sharing a widespread practice?
- Do Belgian researchers deposit their research data in international databases, like those managed by other CESSDA service providers (e.g. GESIS in Germany, DANS in the Netherlands, UKDS in the United Kingdom...)?

Such were the questions for which we sought answers, since they would help us better understand what Belgian social scientists know and do as well as what they need.

Methodology (1/2)

In the summer of 2017, the SODA project team started writing a questionnaire for researchers in social sciences in the Belgian context. We wanted to gather information on several fronts: research data archiving, data documentation, data re-use, tools and software, and legal questions related to research data. Because we feared that potential respondents might shun our questionnaire if it was too long or complicated, we did not include questions that required respondents to react with free text.

The original questionnaire was in French. Once a definitive version was achieved, it was translated in Dutch by Pascal Neckebrouck and in English by Benjamin Peuch. Then, in March 2018, the three questionnaires were encoded into Survey Monkey and sent to two Belgian research centers in social sciences: the *Centre de recherche en démographie* (DEMO) at the *Université catholique de Louvain* (UCLouvain), and Interface Demography (ID) at the *Vrije Universiteit Brussel*. The links to all three questionnaires were sent to both research centers. This constituted a pre-test for our survey.

The data collection phase lasted from 21/03/2018 to 11/04/2018.

Methodology (2/2)

When we analyzed the reactions of 24 respondents, it appeared that some questions were likely poorly worded. This pre-test allowed us to rework the questionnaires and improve their clarity.

After this testing phase, we sought to reach as many researchers who might be interested in using a data archive for the social sciences as possible. To this end, we conducted a review of the Belgian universities' websites in order to find contact information for research groups, research centers, and research institutes that corresponded to our target audience.

We then sent the new questionnaires to this larger array of potential respondents. Once again, the links to all three questionnaires were sent to all recipients, except for ID and DEMO. The data collection phase lasted from 05/10/2018 to 16/11/2018. With this broader sample frame, the survey elicited 205 reactions.

Answering each question was not mandatory to move forward with the questionnaire. For each of the 34 questions, the text of the question is given along with the total number of respondents who answered it and those who preferred to skip it.

Because we only rephrased certain questions or added a few more mostly in the legal section from the pre-test to the launch phase, we combined the results of both data collection instances.

Limitations (1/2)

This survey features several important limitations. First of all, as shown by the wide discrepancies in terms of participation between the various universities (cf. Q1), including the discrepancy between the number of respondents who chose to fill in the questionnaire in Dutch and those who chose to fill in the one in French, the data that we gathered are by no means representative of the Belgian research landscape in social sciences.

Furthermore, we made a selection of research centers and research institutes on the basis of a certain understanding of the concept of “social sciences”. As a point of comparison, the [Flemish Research Discipline Standard](#) (FRDS) developed by *Expertisecentrum O&O Monitoring* (ECOOM) defines the sector of “Social sciences” as “the group of disciplines that study society and the relationships between it and its members, including economics and business, law and legal studies, media and communications, pedagogical and educational sciences, political science, psychology and cognitive science, social and economic geography and sociology and anthropology.”

This aerial view corresponds well to what the SODA project team had in mind when selecting social science research centers; however, certain disciplines which can be found in the FRDS — such as animal experimental psychology, neuroimaging and neuropsychology, motor processes and action... — were left out.

Limitations (2/2)

A much more targeted approach would be required to ensure that each research center, each university, and each sub-discipline of the field of social sciences, too, be equally represented in the results of the survey. Our survey methodology is primarily quantitative; in order to make nuances appear and produce a more acute picture of the Belgian research landscape in social sciences, it would be opportune to complete our results with a more qualitative survey.

While a grand total of 90 research centers and research institutes were contacted, several did not participate in the survey. This could be because of a lack of interest for or of knowledge about the issues at hand. Furthermore, we cannot know with certainty whether the questionnaires were actually circulated within research groups or not.

Even though the translation was carefully performed and all questionnaires were proofread, the fact that the following statistics amalgamate questions and answers from questionnaires written in three different languages can induce semantic deviations.

Several respondents specified which research center they were affiliated to, but not all did. It is therefore difficult to estimate to what extent respondents from certain research groups felt more concerned by our survey and chose to reply to it while others did not.

General information

Q1. Your title (researcher, professor, doctoral student, post-doc...)

doctorandus, doctoranda, doctorant·e, PhD student...	90
professor, professeur·e, prof, gastprofessor	56
onderzoeker, chercheur·se, researcher...	16
assistent, assistant, maitre assistant	11
postdoctoraal onderzoeker, postdoc, post-doc	11
doctor, docteur, PhD, dr	4
hoofddocent	3
gastprofessor	2
projectmanager, onderzoeksleider	2
étudiante	1
vrijvillig medewerker	1
hoogleraar	1
[Some answers left out]	

Total: 227

Skipped: 2

Q2. Your institution and your research group / project / institute

Universiteit Gent	73
Universiteit Antwerpen	35
Katholieke Universiteit Leuven	24
Université catholique de Louvain	23
Université libre de Bruxelles	20
Université Saint-Louis – Bruxelles	9
Vrije Universiteit Brussel	6
Institut national de Criminalistique et de Criminologie – Nationaal Instituut voor Criminalistiek en Criminologie (INCC-NICC)	6
Université de Liège	5
Université de Namur	1
Nuclear Science and Technology Studies (SCK-CEN)	1
Haute École Bruxelles-Brabant (HE2B)	1
[Some answers left out]	

Total: 227

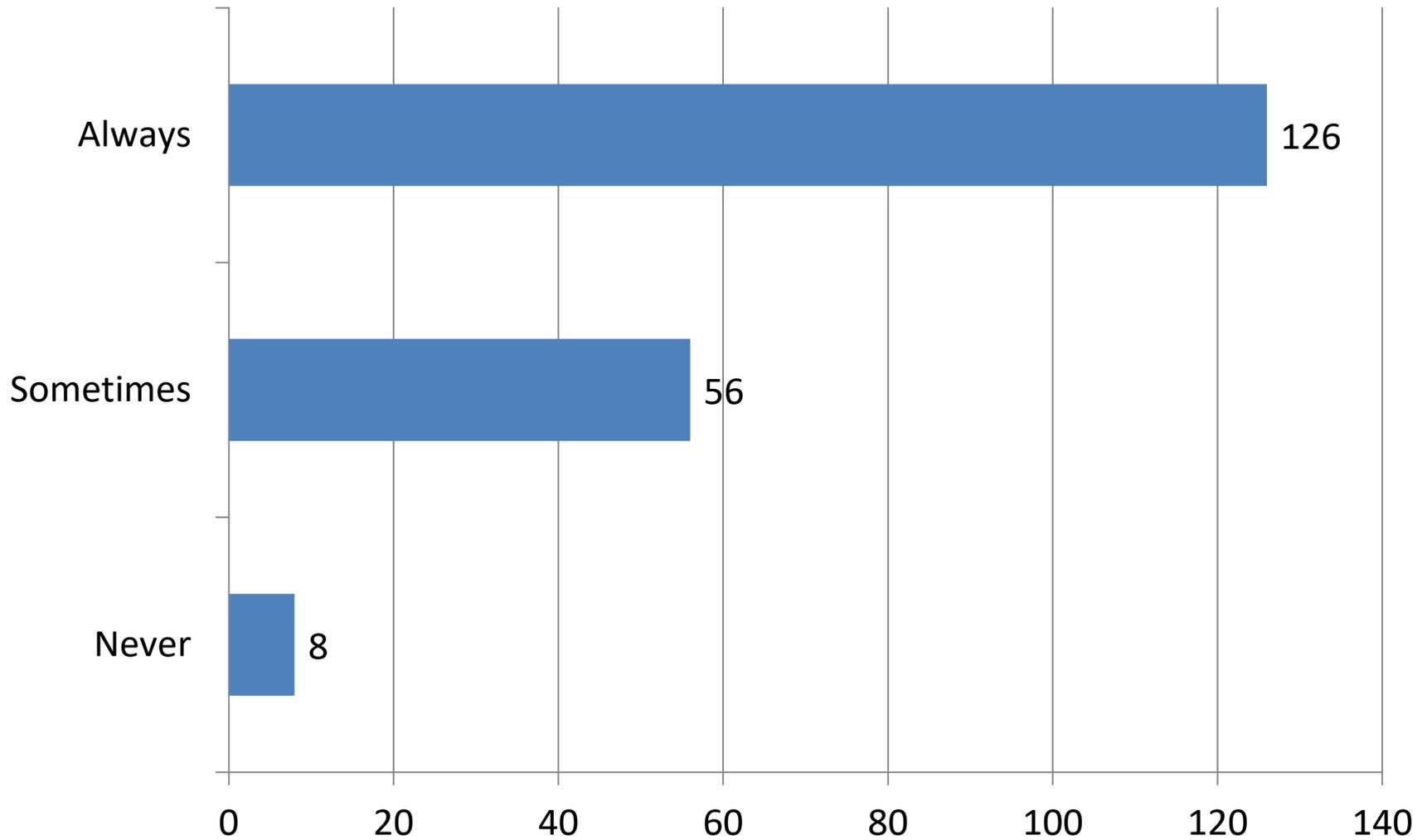
Skipped: 2

Archiving research data

A specification was added under the title of this section:

“Note: ‘Research data’ hereafter refers to all original data which you collected in the course of your research.”

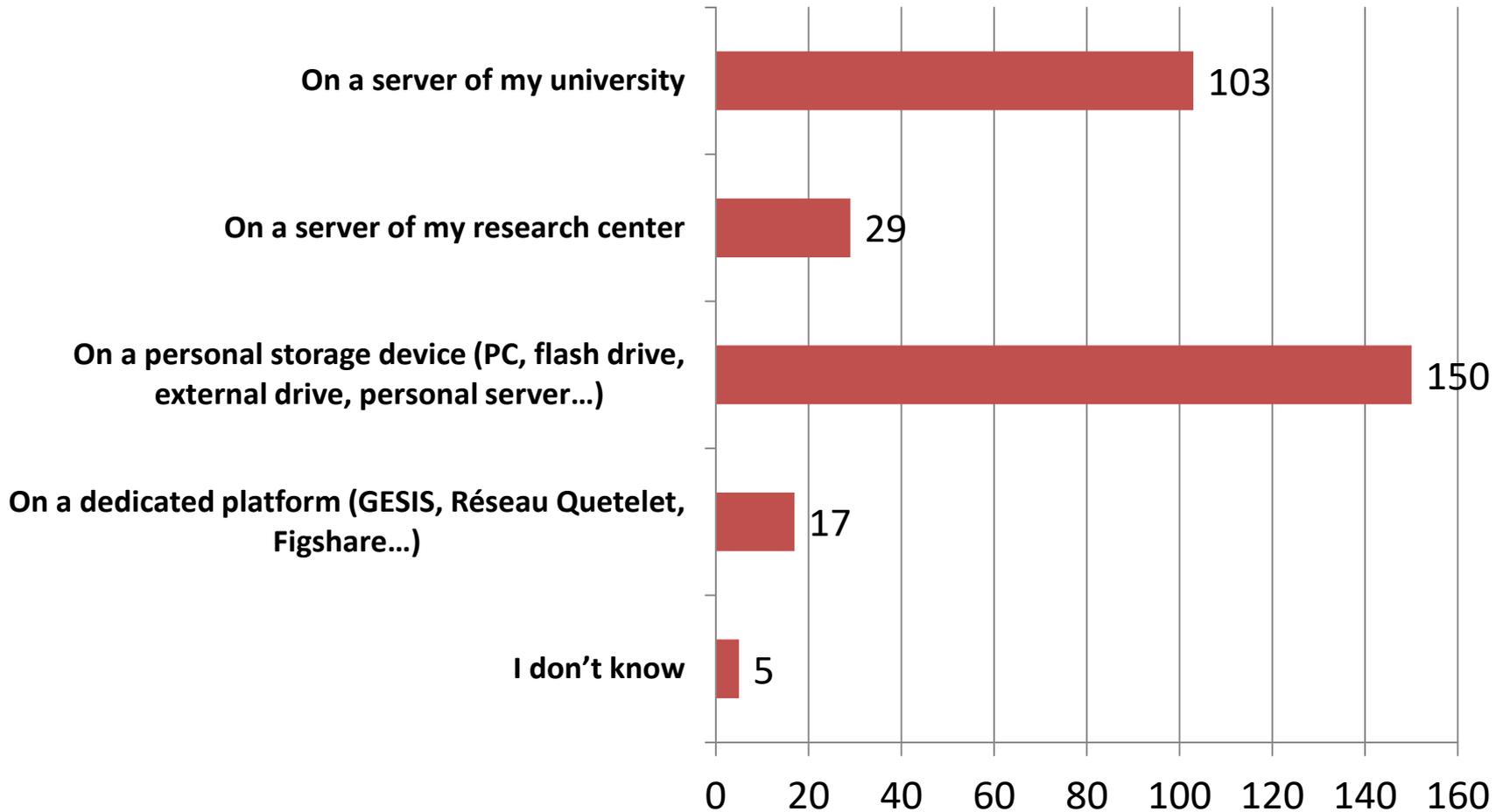
Q3. Do you archive your research data?



Total: 193

Skipped: 36

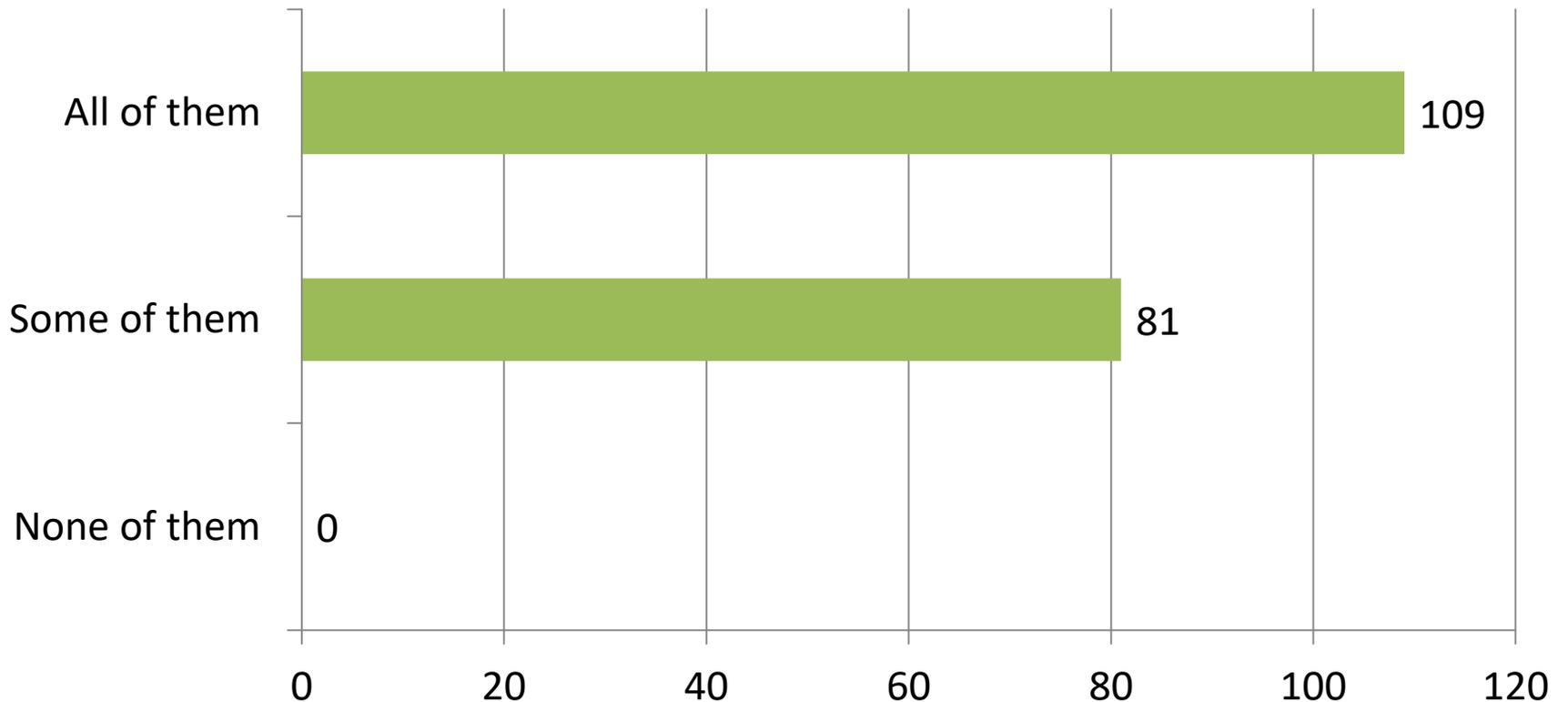
Q4. What type(s) of media is/are used in the course of your research for archiving your research data? (You may choose more than one answer.)



Total: 191

Skipped: 38

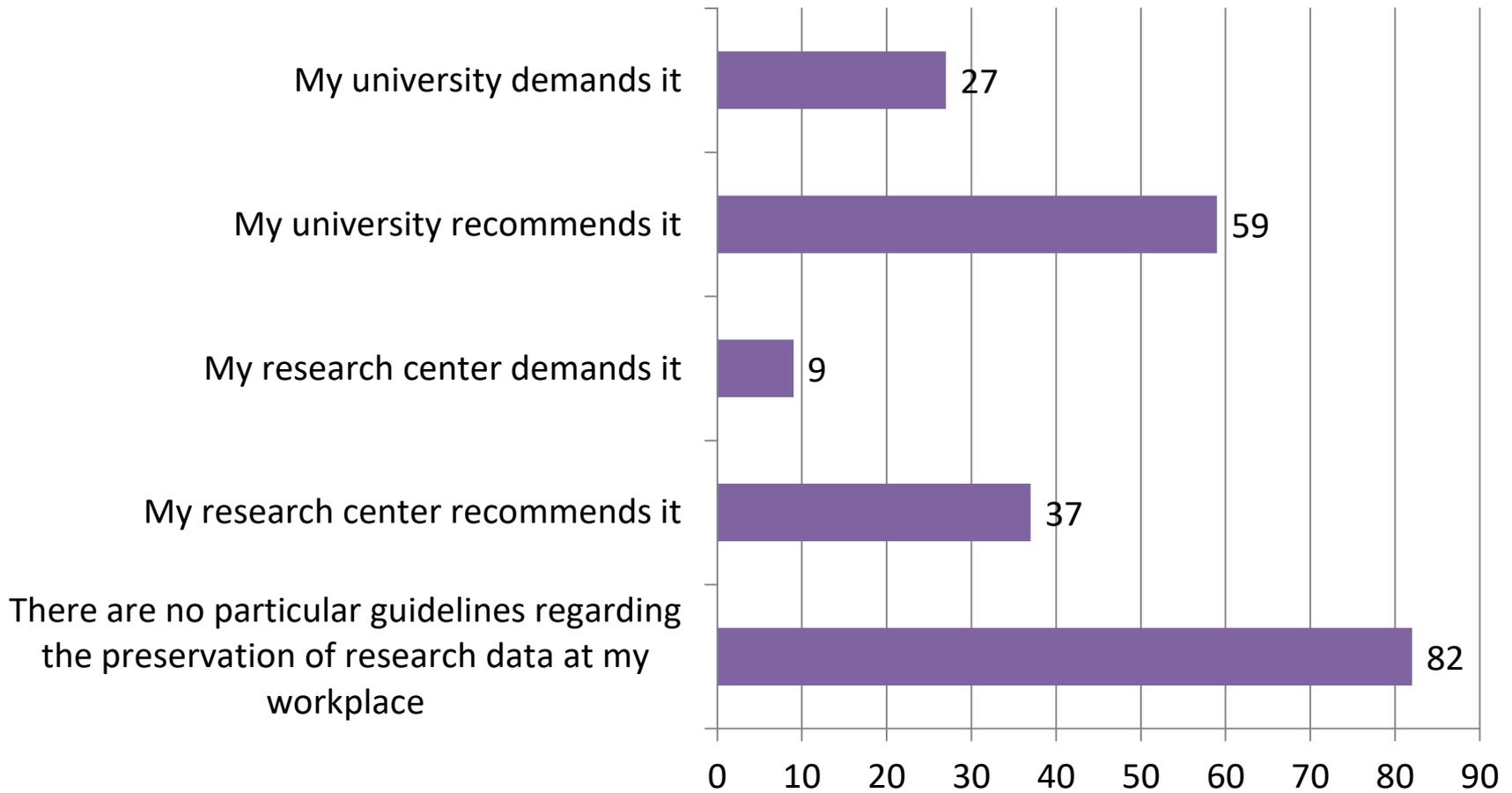
Q5. If you archive your research data yourself or if you partake in this operation, do you proceed accordingly for all of the research data that you handle?



Total: 190

Skipped: 39

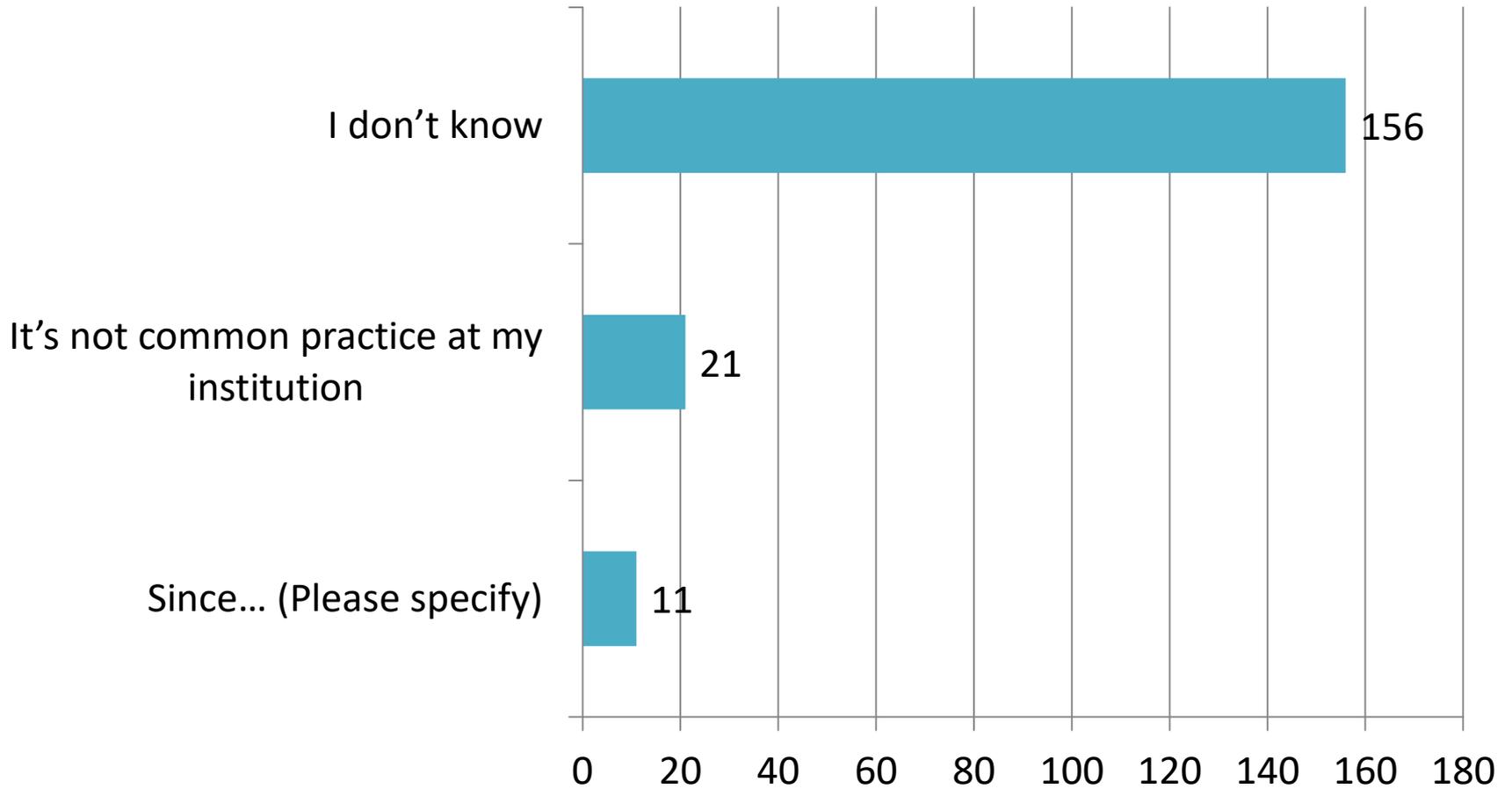
Q6. Is the preservation of the research data required from you by your institution? *(You may choose more than one answer.)*



Total: 184

Skipped: 45

Q7. Do you know when the preservation of research data became a practice in your institution?



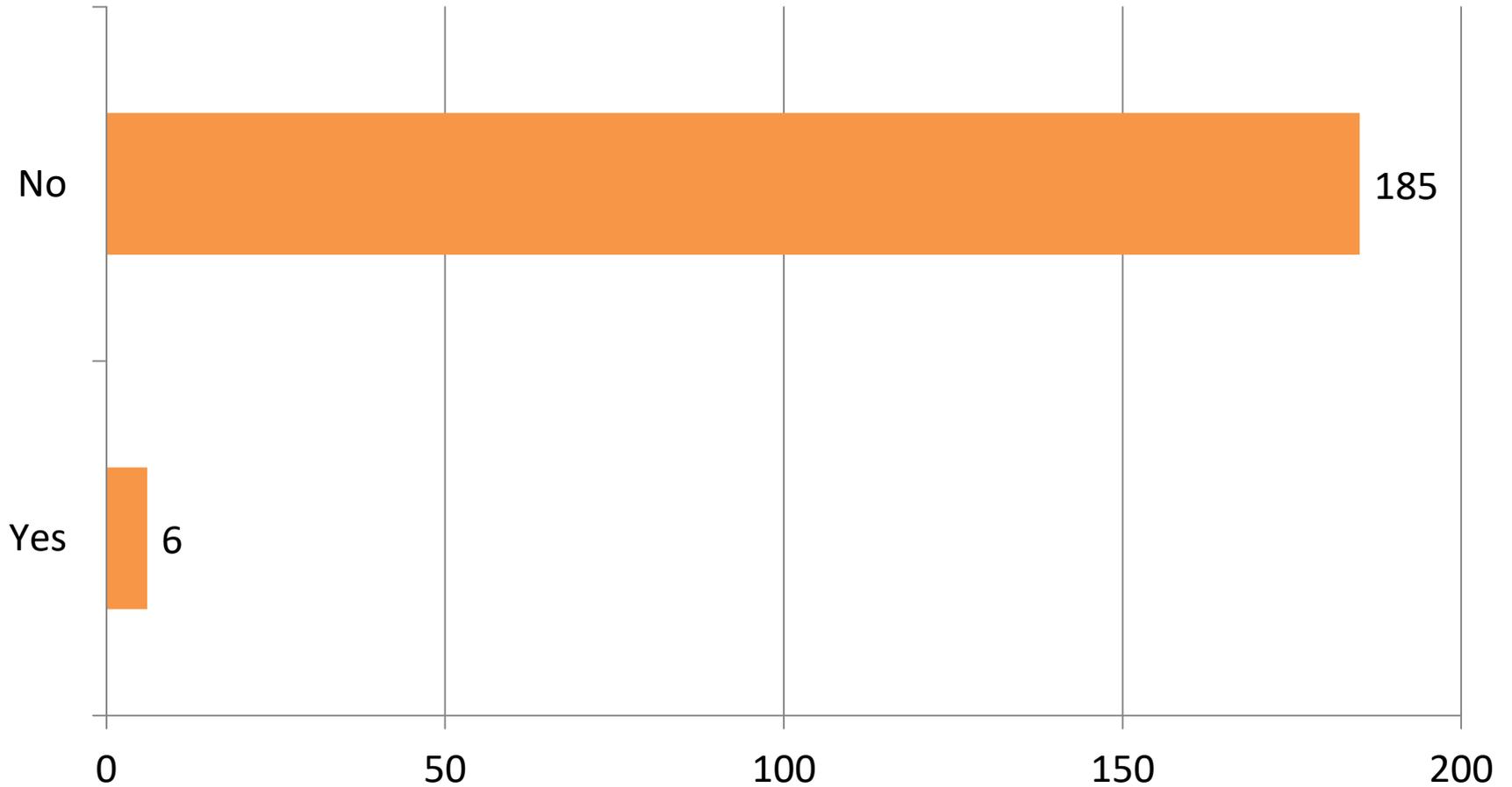
Total: 188

Skipped: 41

Q7. Do you know when the preservation of research data became a practice in your institution?

- The 11 respondents who replied “Since...” gave the following starting years: 2013, 2014, 2015, 2017.
- One respondent wrote: “[For] about 20 years.”
- Another said that a data preservation policy was being developed in their work environment though it was still a work in progress.

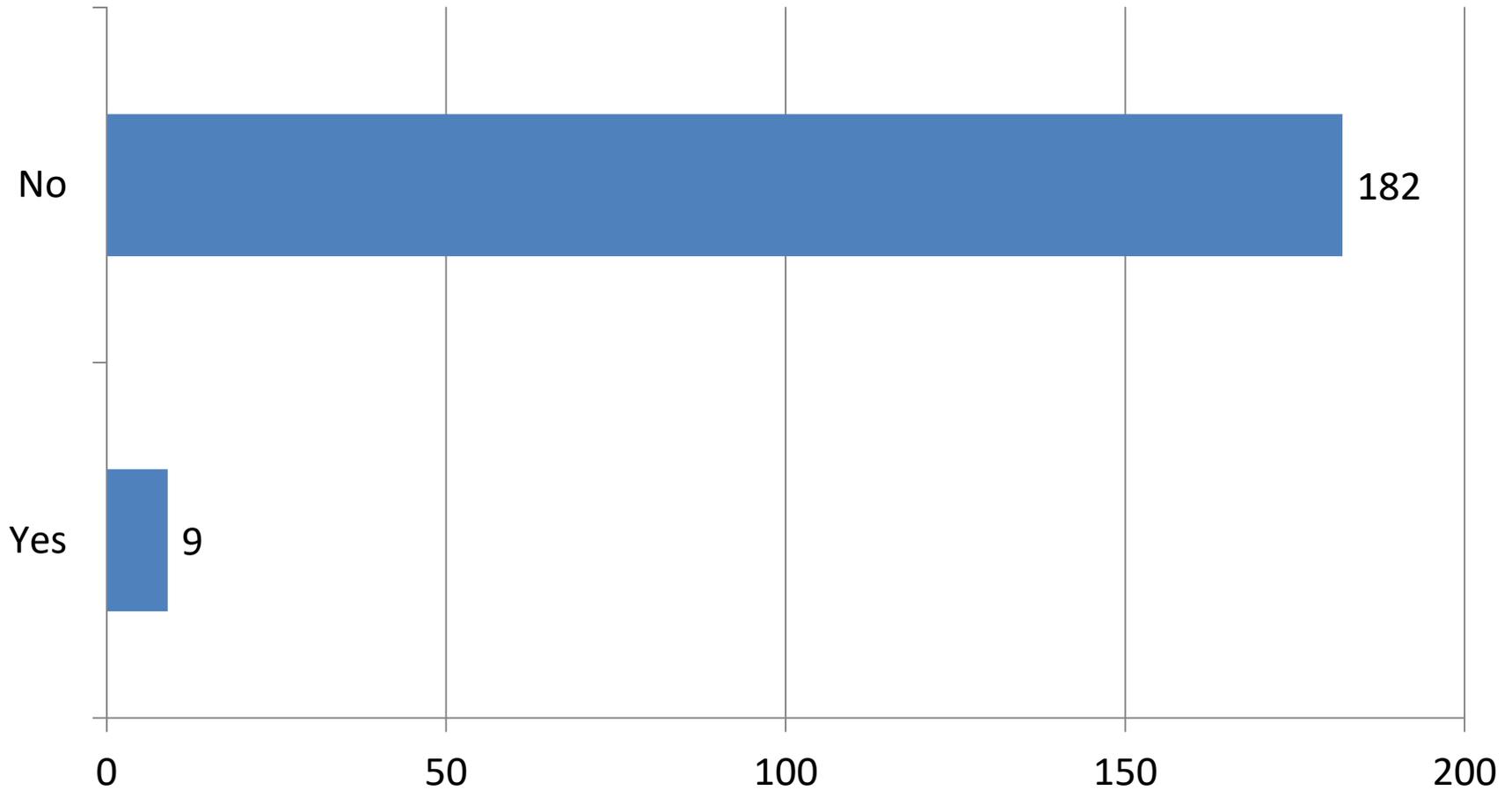
Q8. Do you work with your local academic librarian(s) for the preservation of your research data?



Total: 191

Skipped: 38

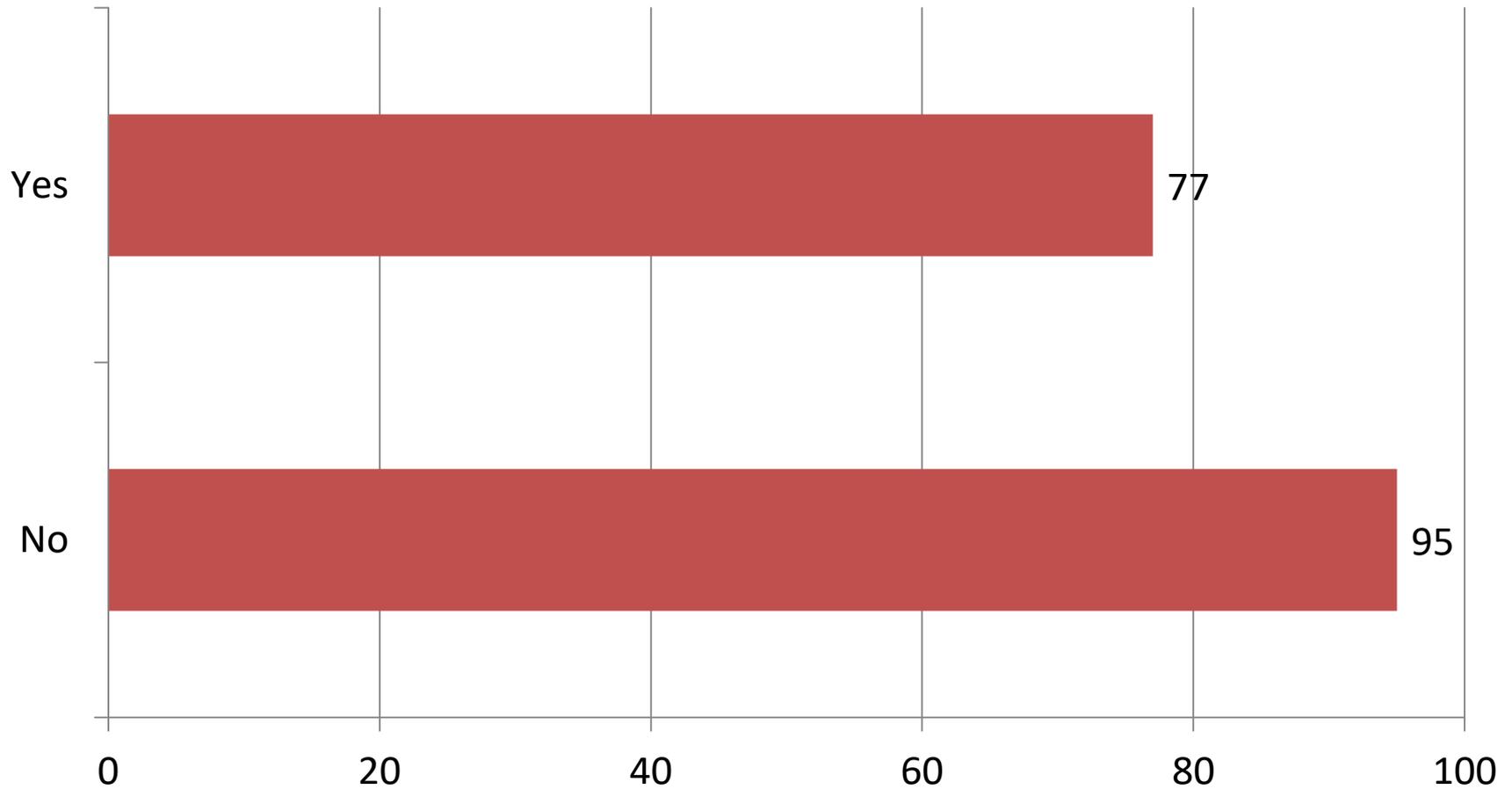
Q9. Do you work with your local archivist(s) for the preservation of your research data?



Total: 191

Skipped: 38

Q10. Have you ever heard about data management plans (DMPs)?



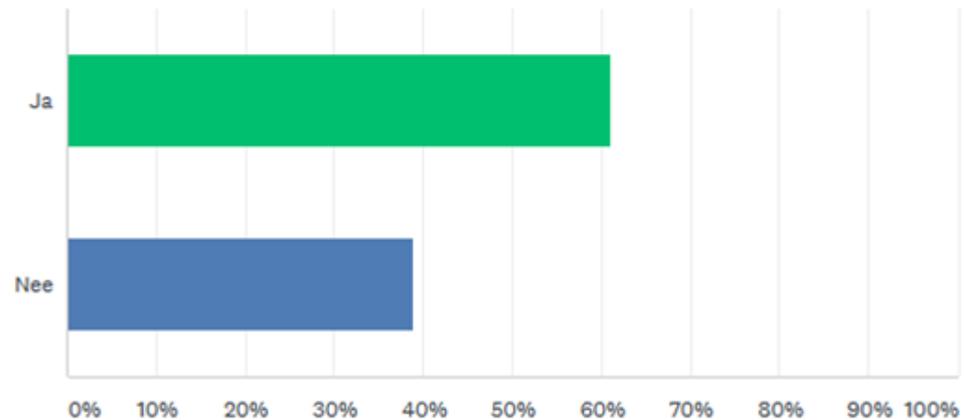
Total: 172

Skipped: 33

Q10. Have you ever heard about data management plans (DMPs)?
[Reactions for the questionnaire in Dutch]

Heeft u al gehoord over de data management plans (DMP)?

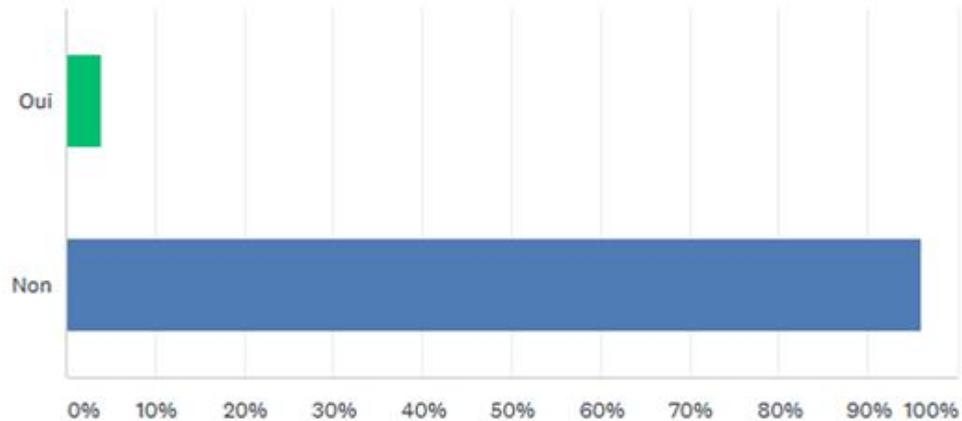
Answered: 113 Skipped: 20



Q10. Have you ever heard about data management plans (DMPs)?
[Reactions for the questionnaire in French]

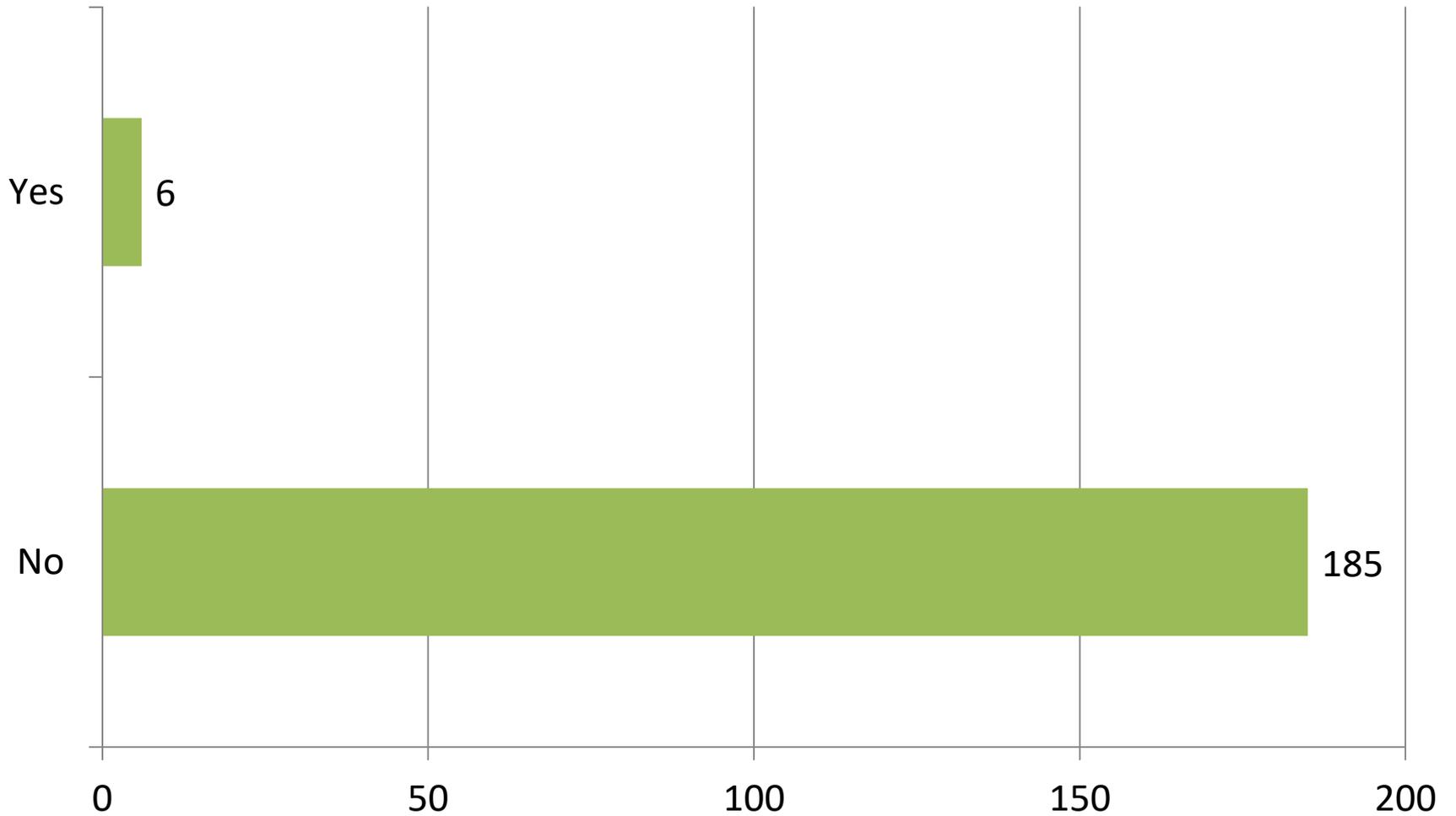
Avez-vous déjà entendu parler des data management plans (DMP) ?

Answered: 50 Skipped: 11



Documentation and metadata

Q11. Have you ever heard about the Data Documentation Initiative (DDI) standard?



Total: 191

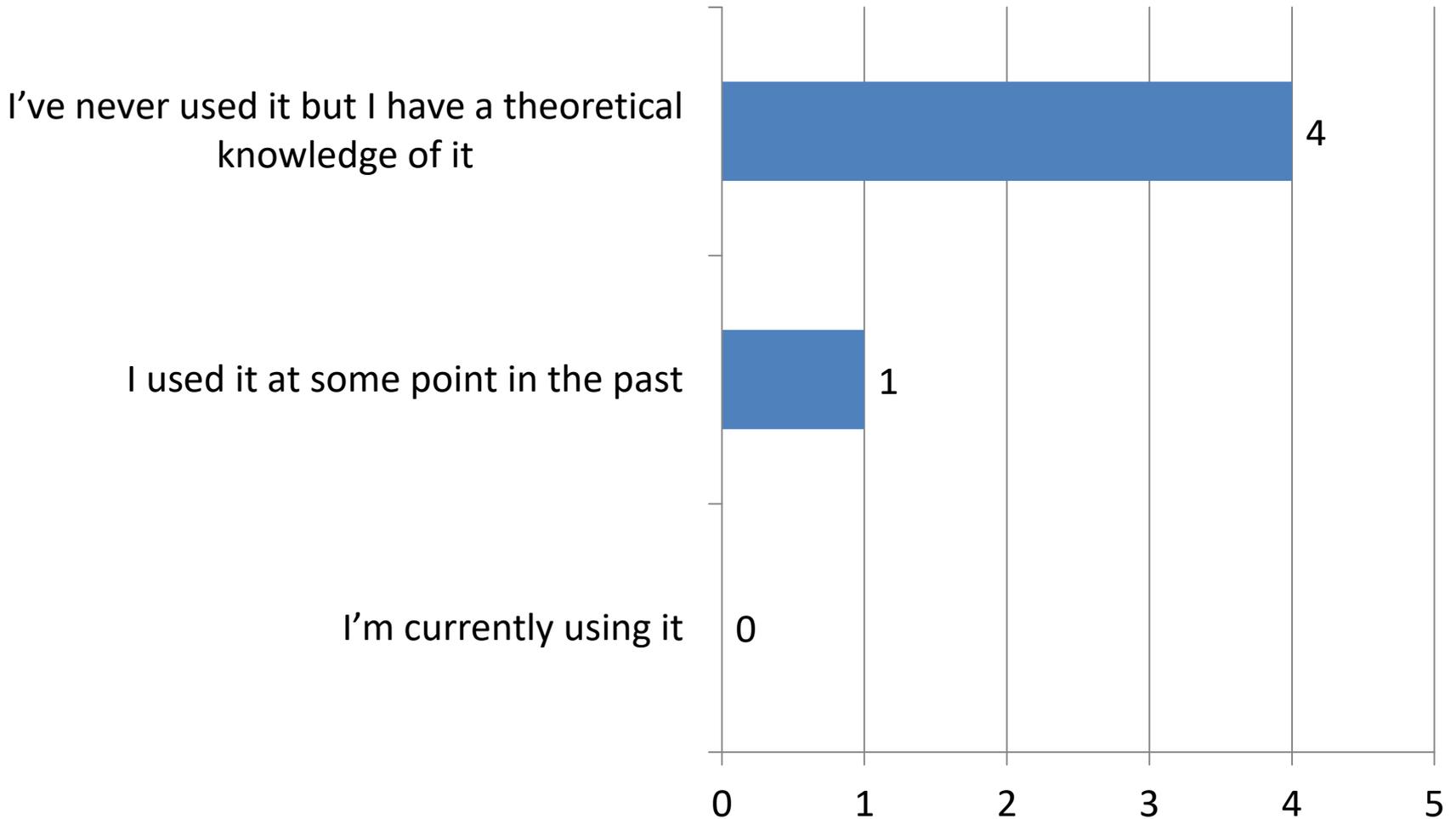
Skipped: 38

Q11. Have you ever heard about the Data Documentation Initiative (DDI) standard?

- Respondents who indicated that they had heard about DDI hail from the following institutions:
 - 2 – Université catholique de Louvain
 - 1 – Universiteit Antwerpen
 - 1 – Universiteit Gent
 - 1 – Vrije Universiteit Brussel
 - 1 – [undisclosed]

[If the respondent chose "Yes" for Q11; else → Q13]

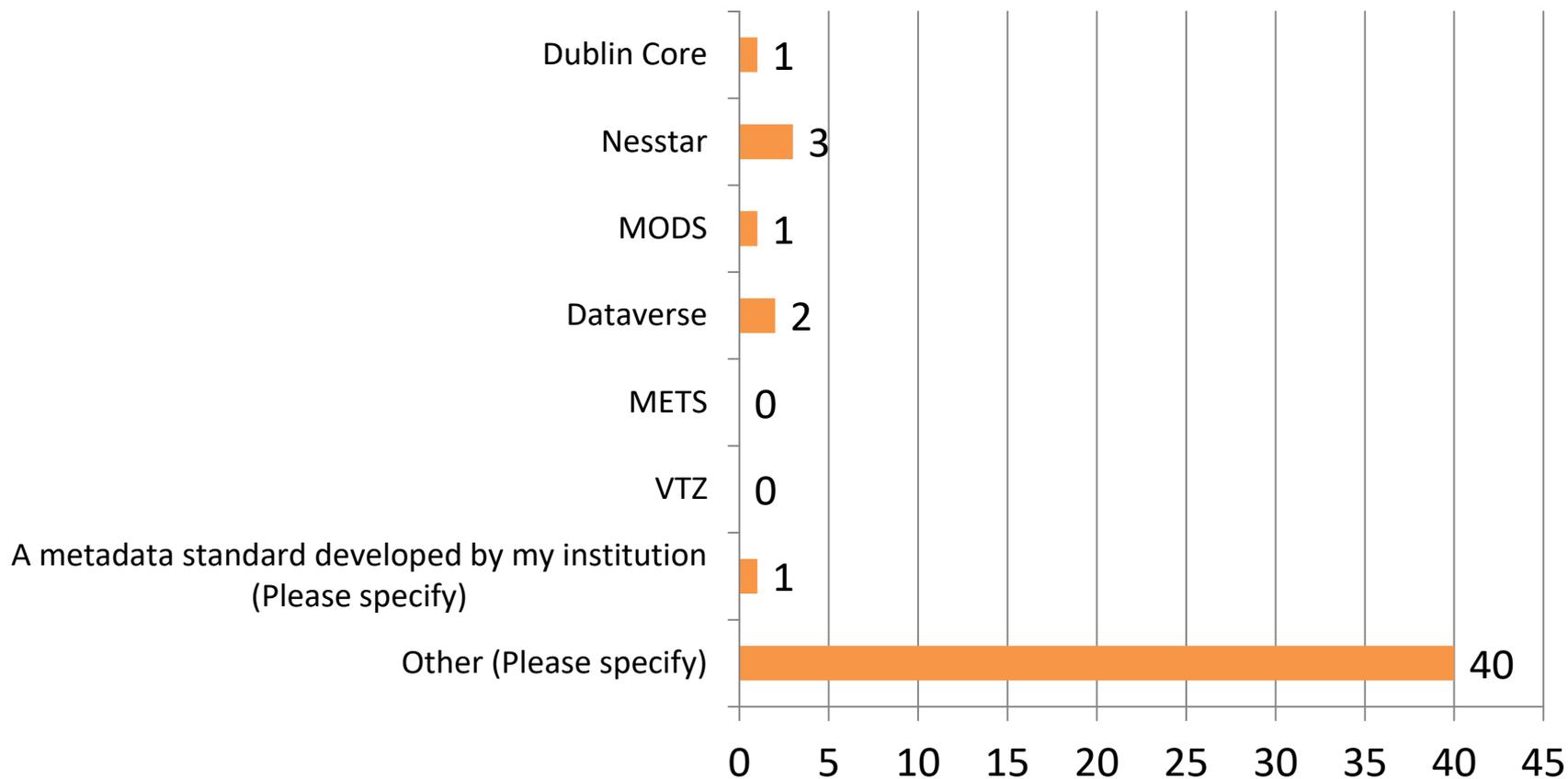
Q12. Are you currently using it, or did you use it at some point in the past?



Total: 5

Skipped: 1

Q13. What kind of metadata do you supplement your research data with and what aspects do they bear upon? (*More than one answer is possible. If you wish, you may send us an example of a metadata file that you created in the course of your research.*) (*You may choose more than one answer.*)



Total: 47

Skipped: 182

A word on DDI

We decided to ask our respondents whether they were familiar with Data Documentation Initiative, the metadata standard for social science datasets. The reason is that we did not know to what extent researchers themselves might know about it, or whether it was an element of the data archiving process that fell more to data technicians and archivists.

An overwhelming proportion of respondents chose “No” when asked whether they knew DDI, indicating that social scientists only play a minor role in the data archiving process.

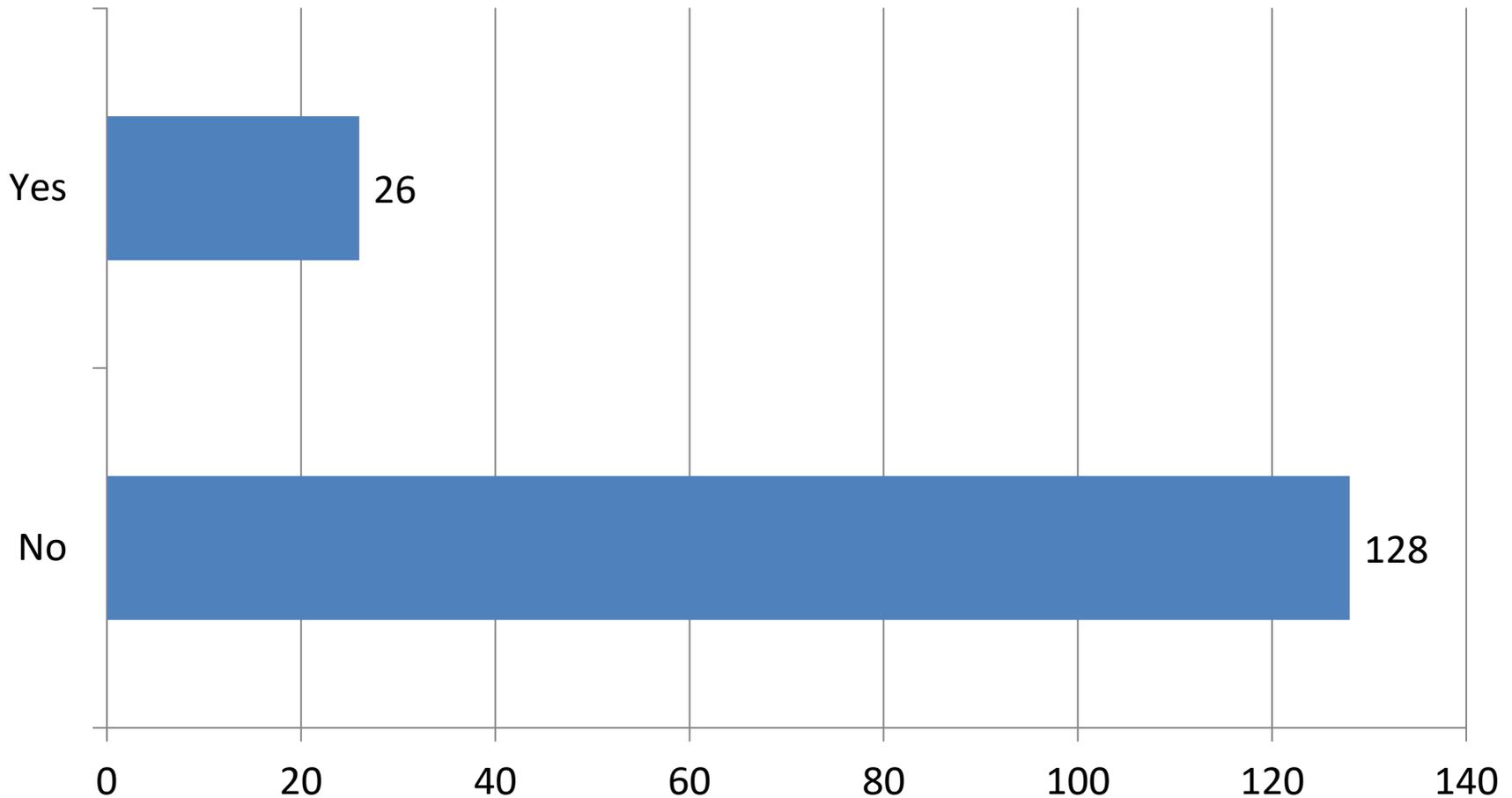
Hence, we decided that knowledge of DDI was not something that researchers should be burdened with. Rather, because DDI enables standardized data documentation and dissemination, it is imperative that researchers be given adequate tools to document their datasets and automatically generate DDI-compliant files so that information about datasets (i.e. metadata) can be easily circulated.

The IT personnel and the digital archivists of the future data archive will be responsible for DDI implementation and metadata management. Researchers will not need worry about this technical specification.

A few researchers also indicated (Q13) that they used local templates to document their datasets.

Re-use of research data

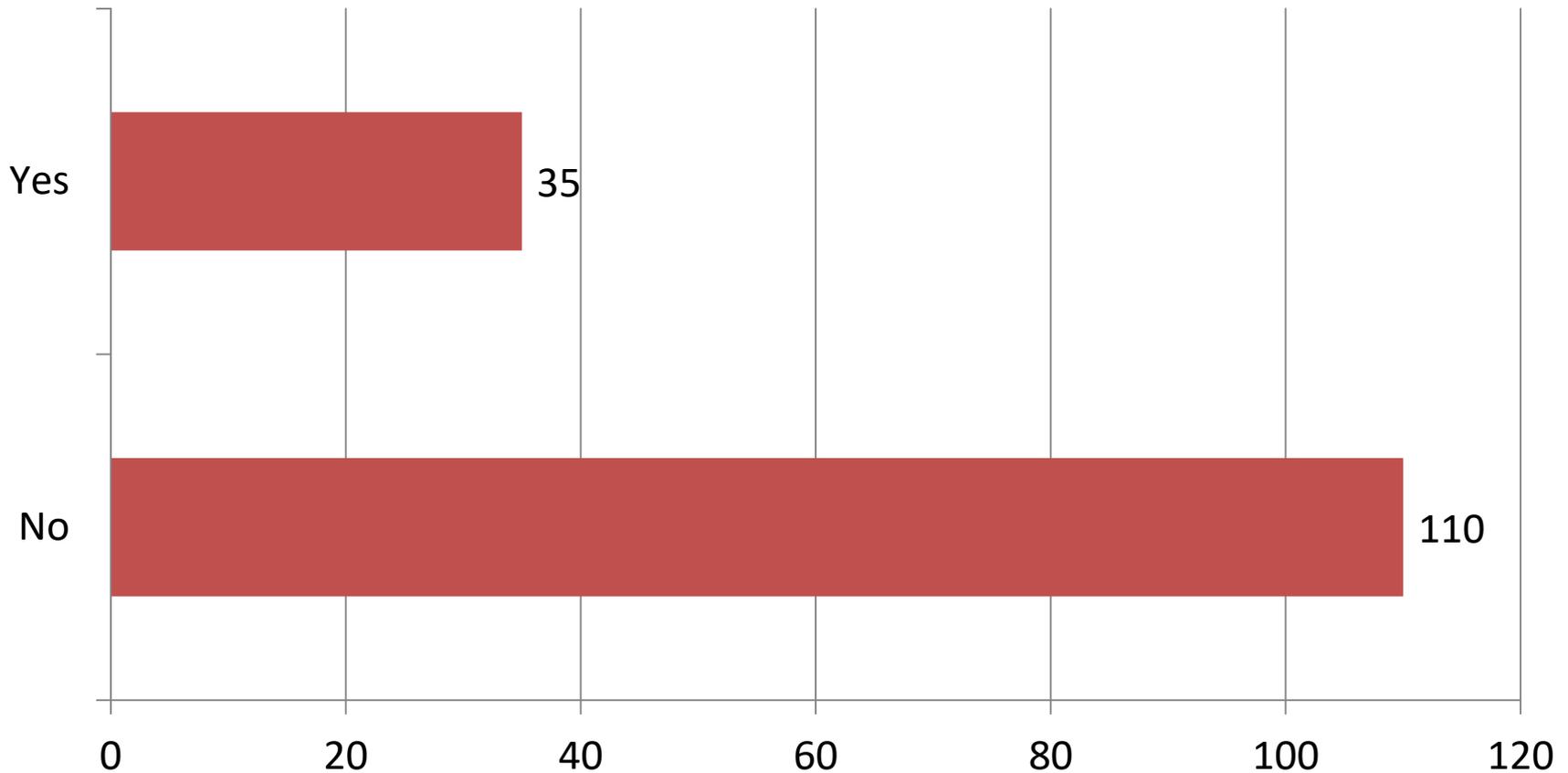
Q14. Do you have a good overview of the already existing, re-usable survey data produced in Belgium in your research area that you may request access to?



Total: 154

Skipped: 51

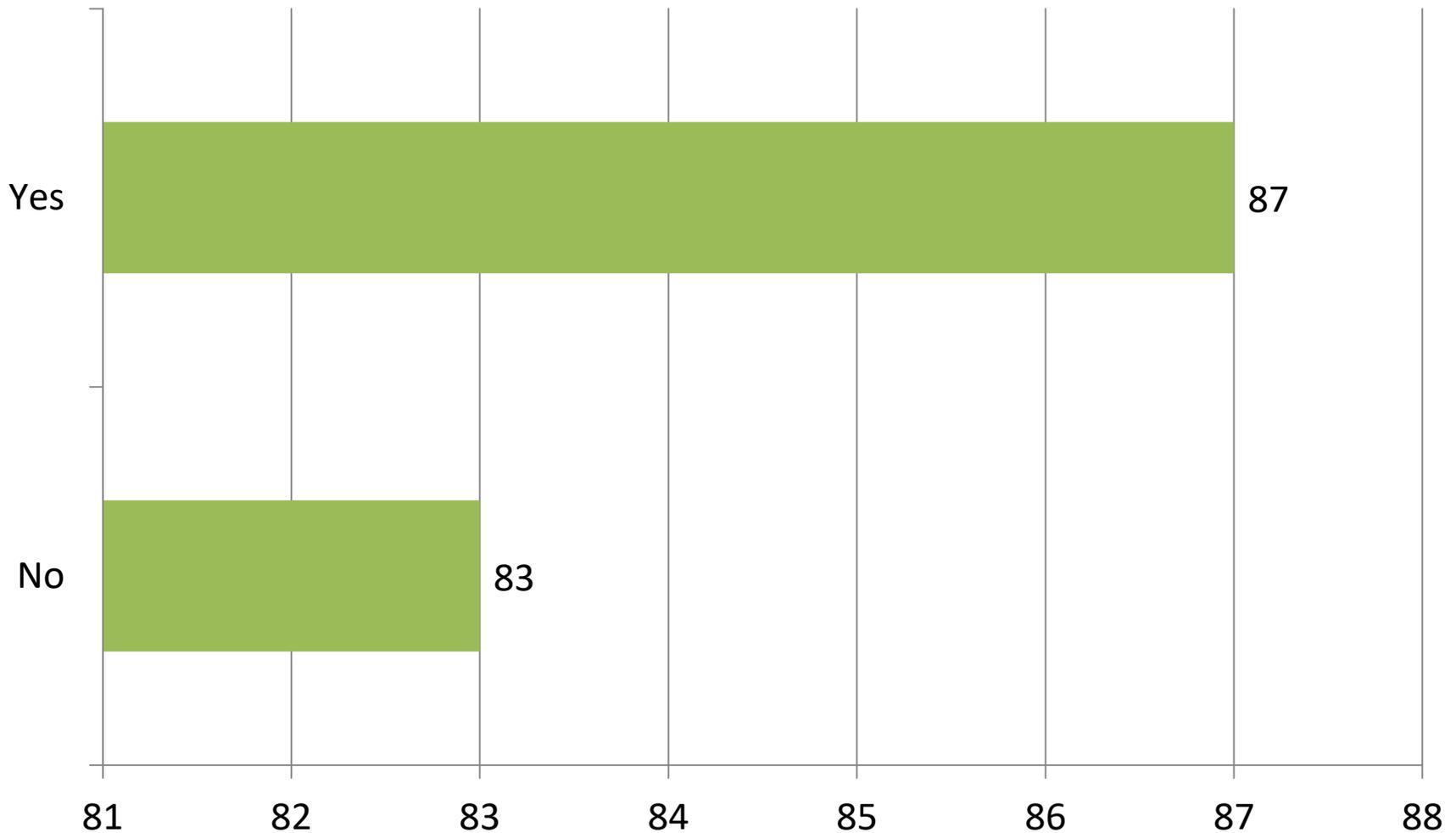
Q15. Do you have a good overview of the already existing, re-usable administrative data (e.g. data produced by government offices for statistics or civil registry data) produced in Belgium in your research area that you may request access to?



Total: 154

Skipped: 51

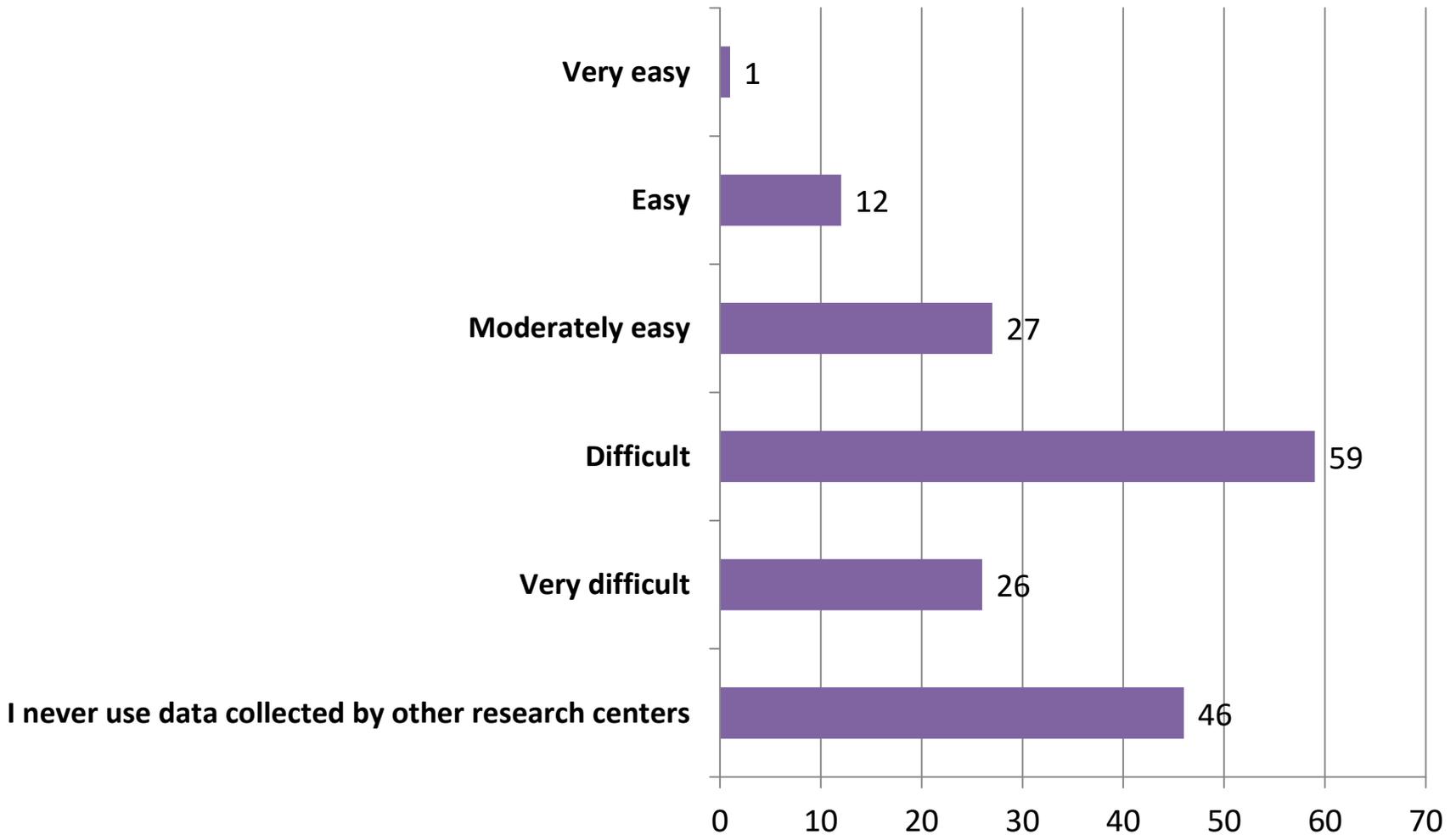
Q16. Do you often have occasion to re-use research data, whether published or unpublished, for your own research?



Total: 170

Skipped: 69

Q17. How easy is it for you to access and re-use published or unpublished data that do not originate from your own research institution?

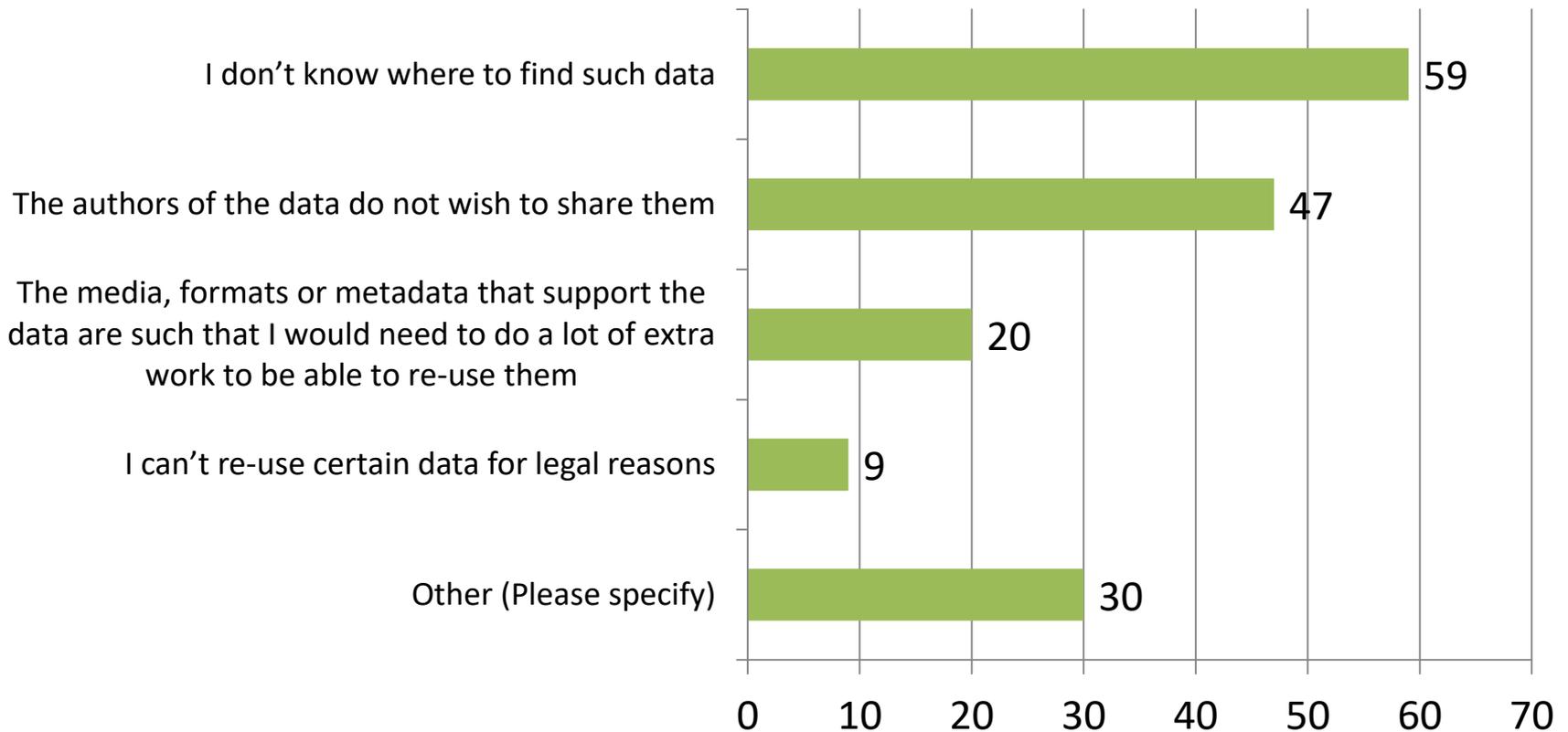


Total: 171

Skipped: 58

[If respondent chose “Moderately easy”, “Difficult” or “Very difficult” for Q17; else → Q19.]

Q18. What makes it difficult for you to access and re-use data produced by another research institution than yours? (You may choose more than one answer.)



Total: 93

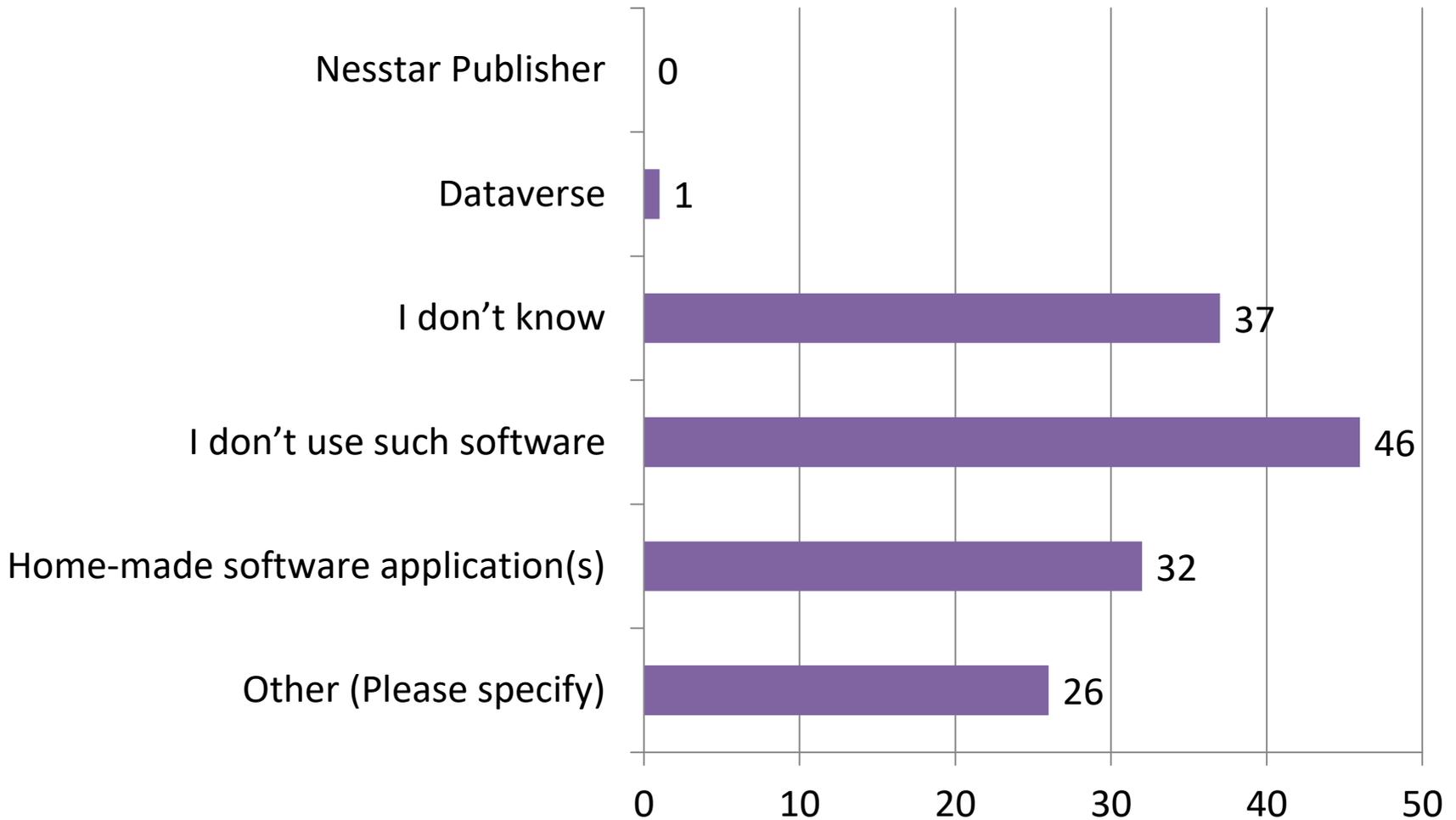
Skipped: 19

Q18. What makes it difficult for you to access and re-use data produced by another research institution than yours? (*You may choose more than one answer.*)

- Knowing which obstacles researchers face when trying to access research data for re-use was essential for us. The extra comments that several researchers wrote were also very valuable for our purposes.
- Because the upcoming Belgian data archive ought to work as a research support service, working on overcoming those obstacles will be one of its top priorities.
- Further reasons mentioned by some respondents include:
 - ethical concerns (consent of respondents is limited to the original study)
 - private data suppliers (not in favor of data re-use)
 - the online platform for data dissemination is no longer accessible
 - qualitative research data are especially difficult to re-use

Tools and software

Q19. What types of software do you rely on to document your research and create your datasets? (*You may choose more than one answer.*)



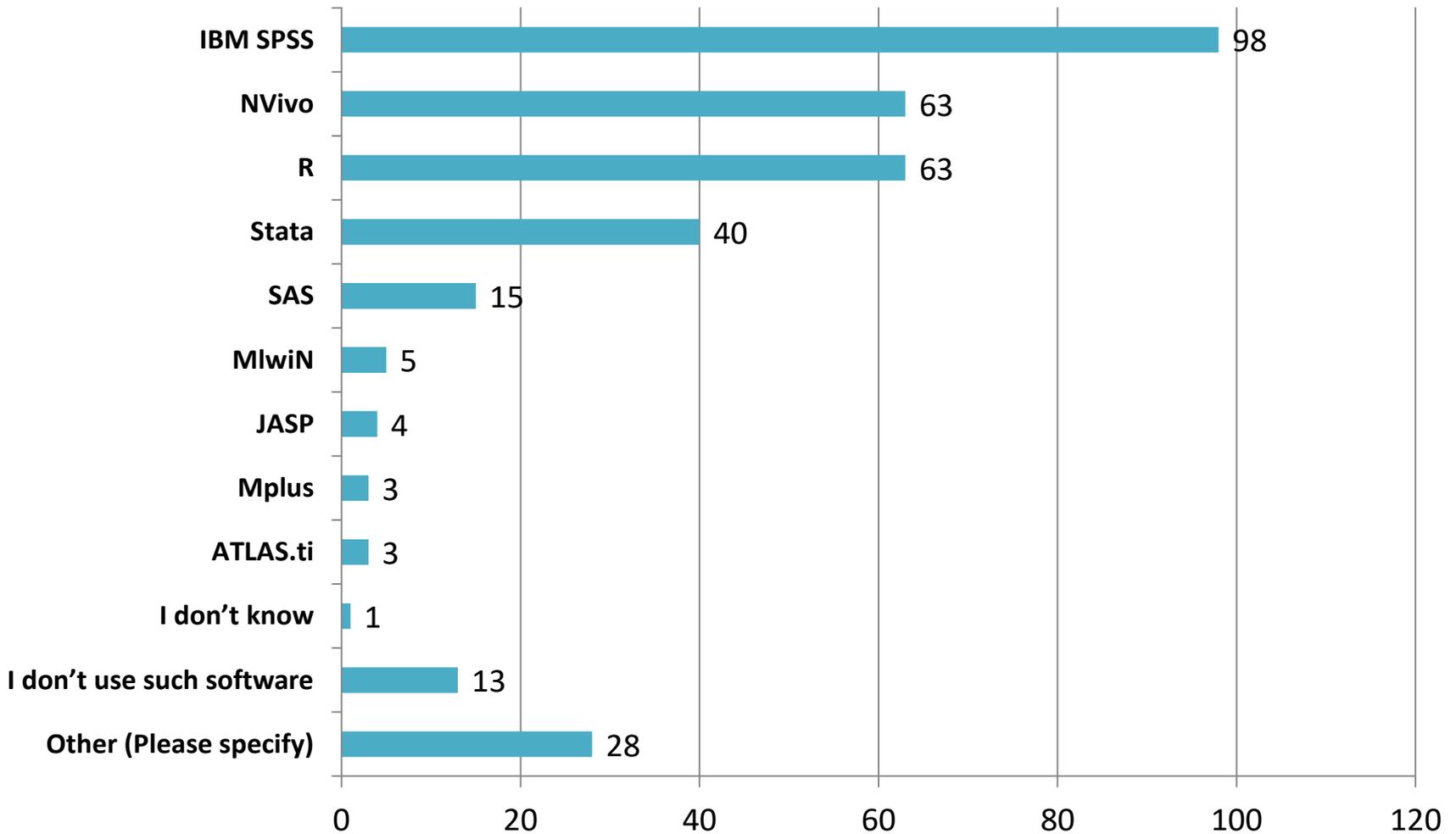
Total: 140

Skipped: 89

Q19. What types of software do you rely on to document your research and create your datasets? (*You may choose more than one answer.*)

- We wanted to know whether researchers in social sciences resort to specialized software to produce metadata about their datasets.
- Software like Nesstar can automatically produce DDI files. Regarding Nesstar however, it used to be a commercial product. Dataverse can also generate DDI, but researchers will likely need to be supported with materials such as user guides, tutorials, training, etc.
- As shown by the respondents' replies, researchers simply document their datasets with multi-purpose software programs that they also use for data processing and analysis, such as SPSS, NVivo, Stata, Microsoft Word, Microsoft Excel...

Q20. Do you use any of the following tools for data analysis? (*You may choose more than one answer.*)



Total: 168

Skipped: 61

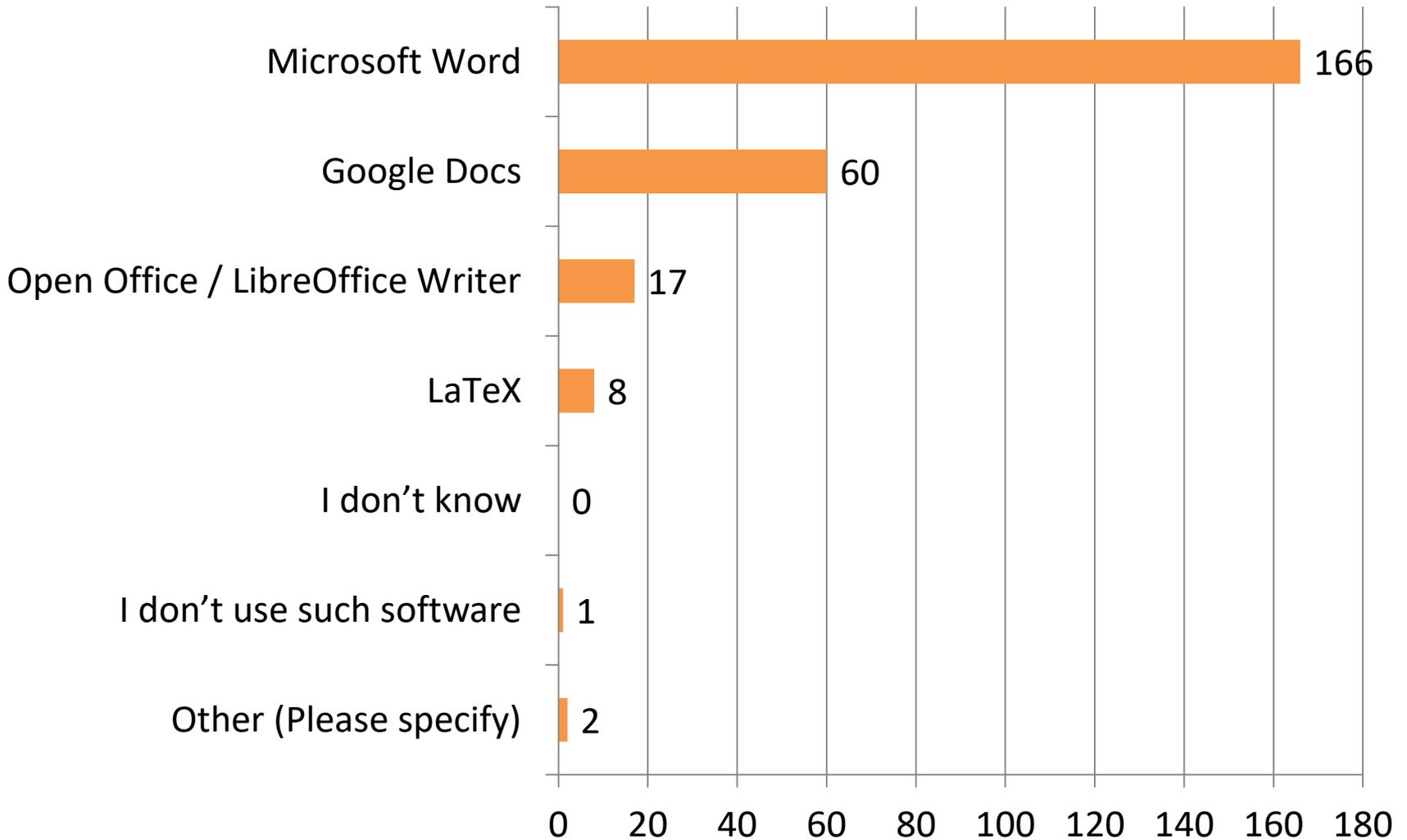
Q20. Do you use any of the following tools for data analysis? (*You may choose more than one answer.*)

Other programs mentioned:

AMOS	2
Excel	2
MaxQDA	2
ArcGIS	1
Brain Voyager	1
Evernote	1
Gephi	1
GNU PSPP	1
HLM	1

Jmp	1
MATLAB	1
Qualtrics	1
Sem	1
SPAD	1
Ucinet	1
VOSON	1
Word	1
WordStat QDA Miner	1

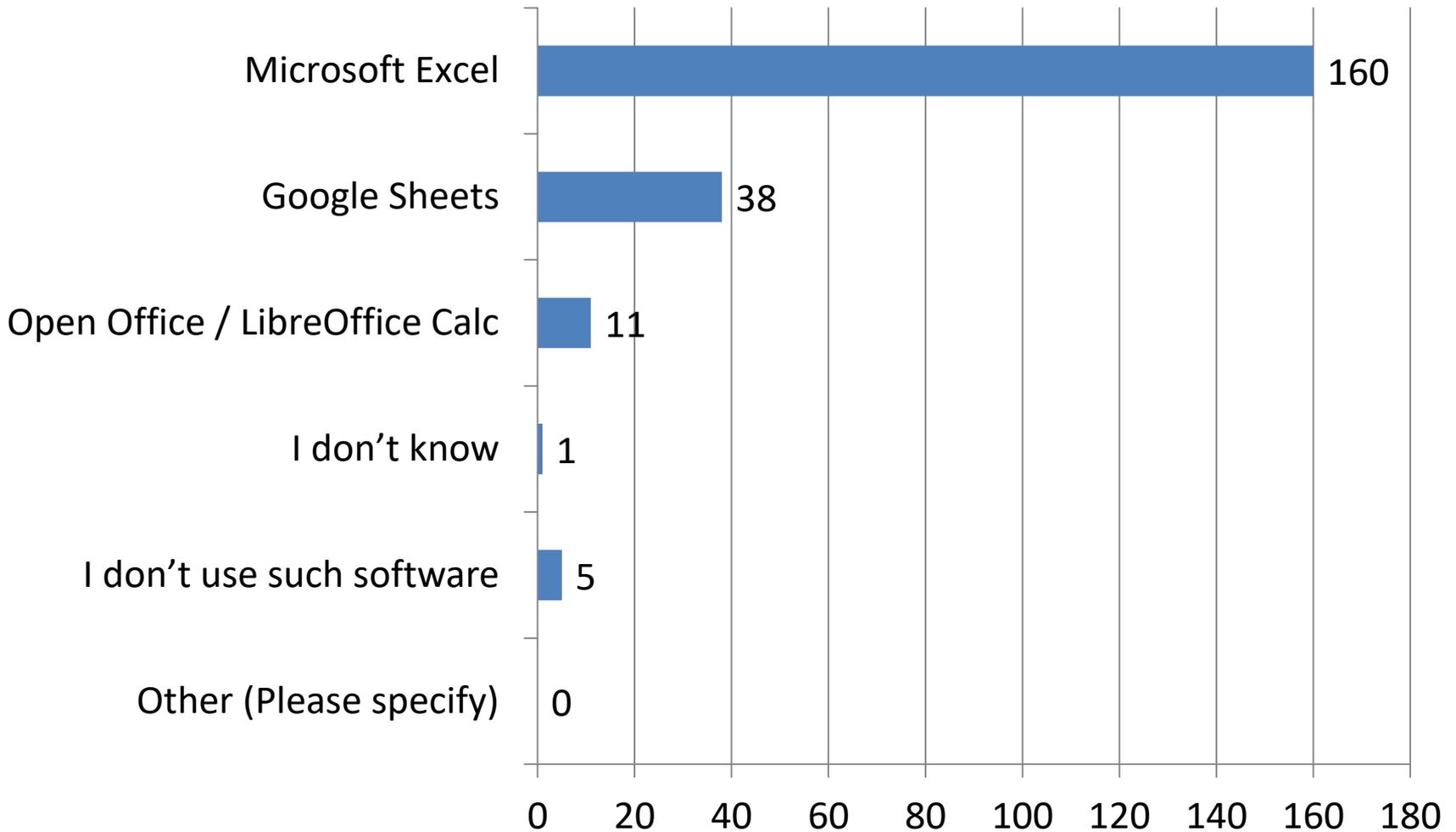
Q21. Do you use word processors? (*You may choose more than one answer.*)



Total: 168

Skipped: 61

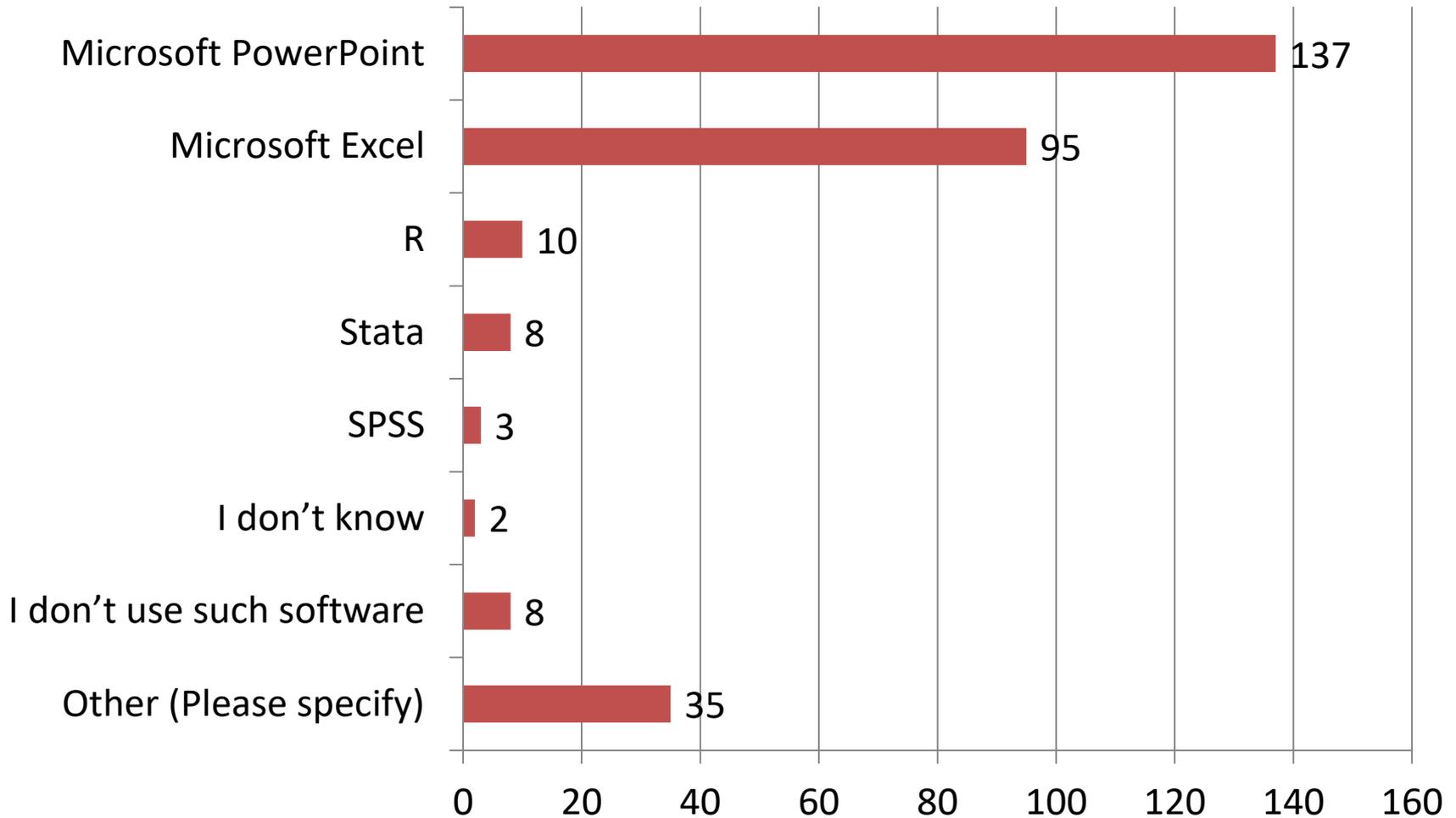
Q22. Do you use spreadsheet software? (You may choose more than one answer.)



Total: 166

Skipped: 63

Q23. Do you use programs for presentations and conceptual models? (*You may choose more than one answer.*)



Total: 166

Skipped: 63

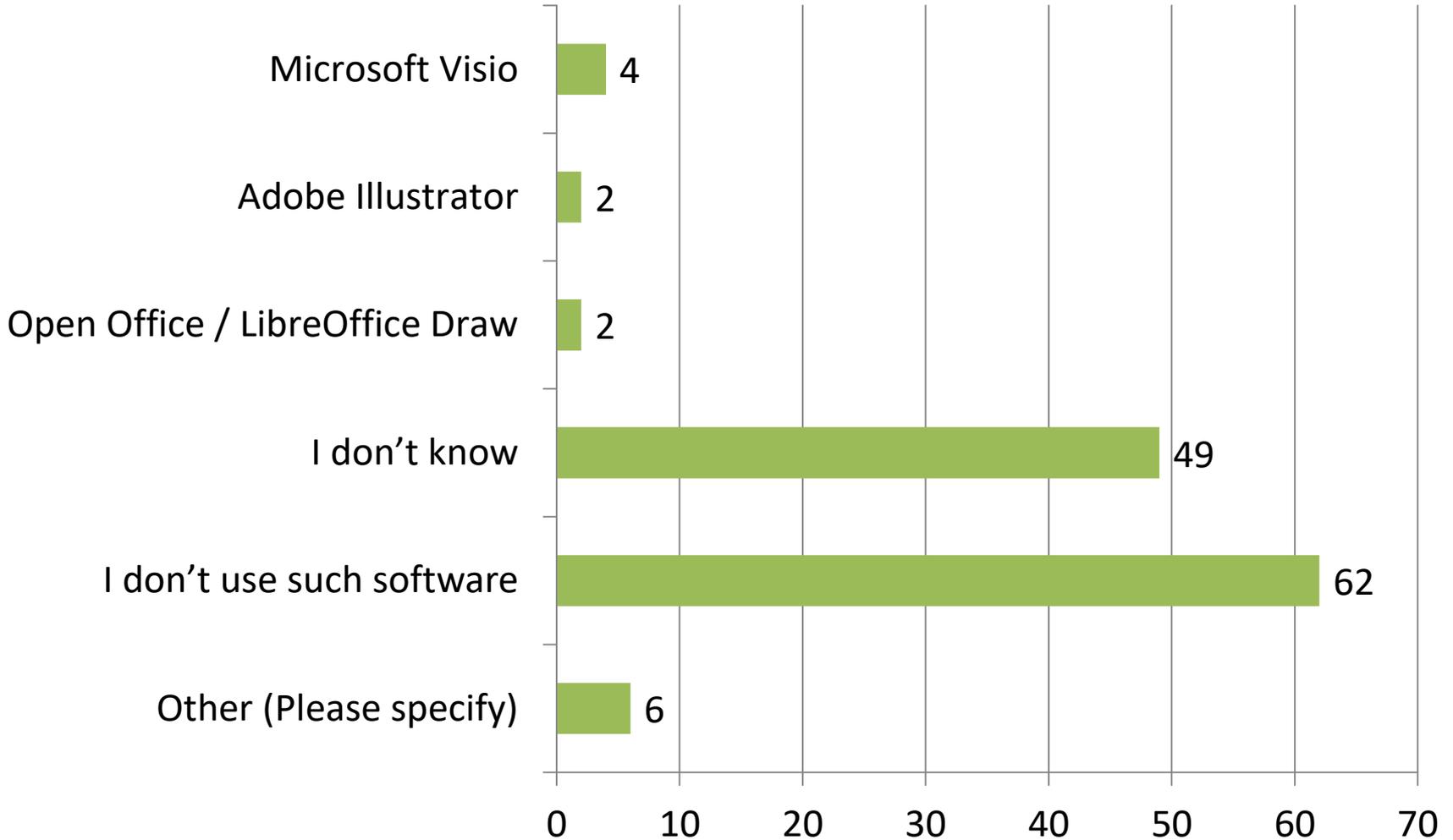
Q23. Do you use programs for presentation and conceptual models? (*You may choose more than one answer.*)

Other programs mentioned:

DataViz	2
Dia	2
Draw.io	2
Gephi	2
Keynote	2
Microsoft Visio	2
Prezi	2
XMind	2
yEd Graph Editor	2
ArcGIS	1
Canvas	1
Corel Draw	1
GIS Software	1

Googles Slides	1
Insight Maker	1
LucidChart	1
MATLAB	1
Microsoft Publisher	1
NVivo (Matrix Coding Queries)	1
Qualtrics	1
R (ggplot)	1
RStudio	1
Tableau	1
Ucinet	1
Visionary	1

Q24. Do you use software for making vector graphics? (*You may choose more than one answer.*)



Total: 126

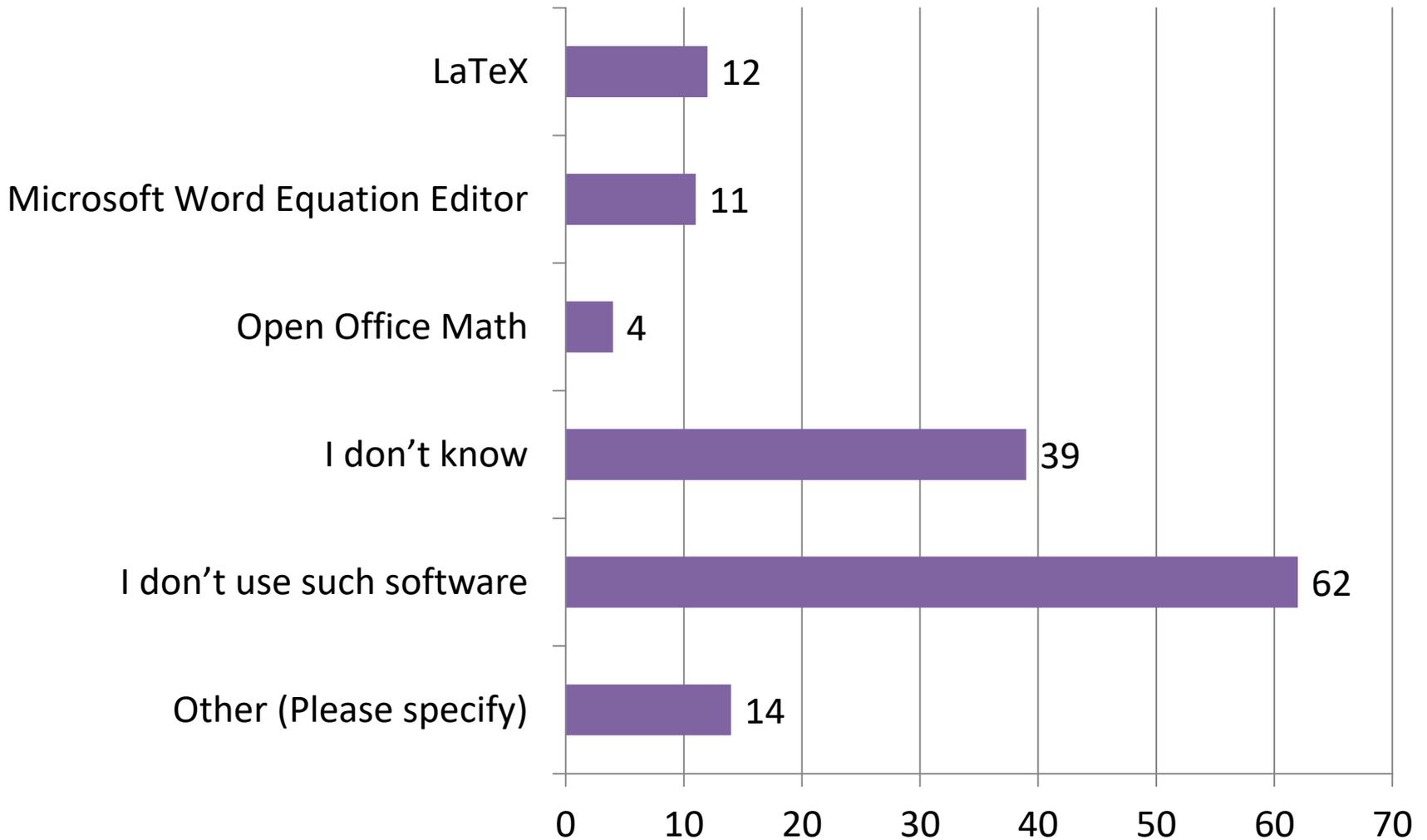
Skipped: 103

Q24. Do you use software for making vector graphics? (*You may choose more than one answer.*)

Other programs mentioned:

CMAPS Tool	1
Corel Draw	1
Dia	1
Draw.io	1
Inkscape	1
R	1

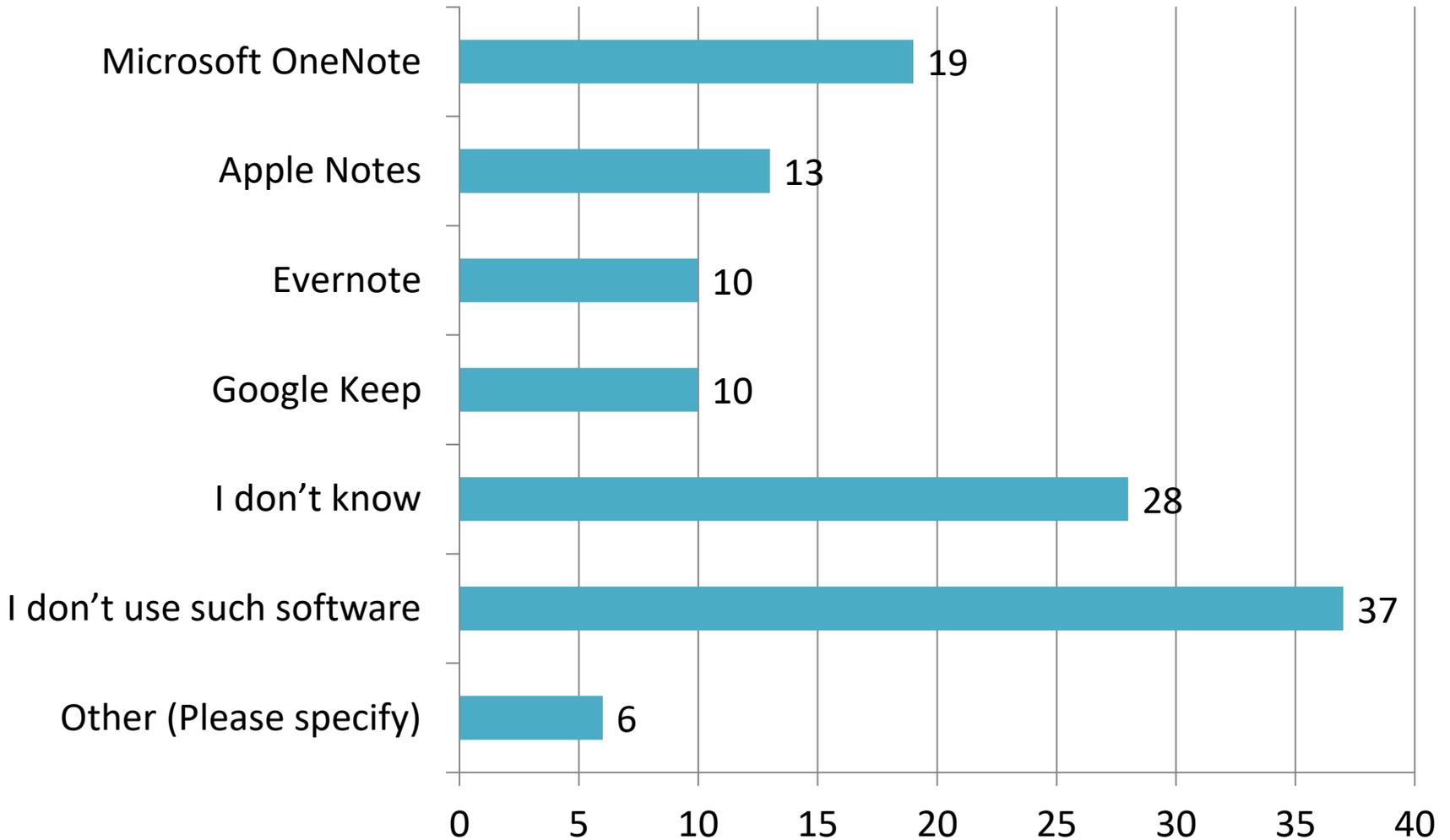
Q25. Do you use software for mathematical formulae? (*You may choose more than one answer.*)



Total: 129

Skipped: 100

Q26. Do you use note-taking programs? (You may choose more than one answer.)



Total: 131

Skipped: 98

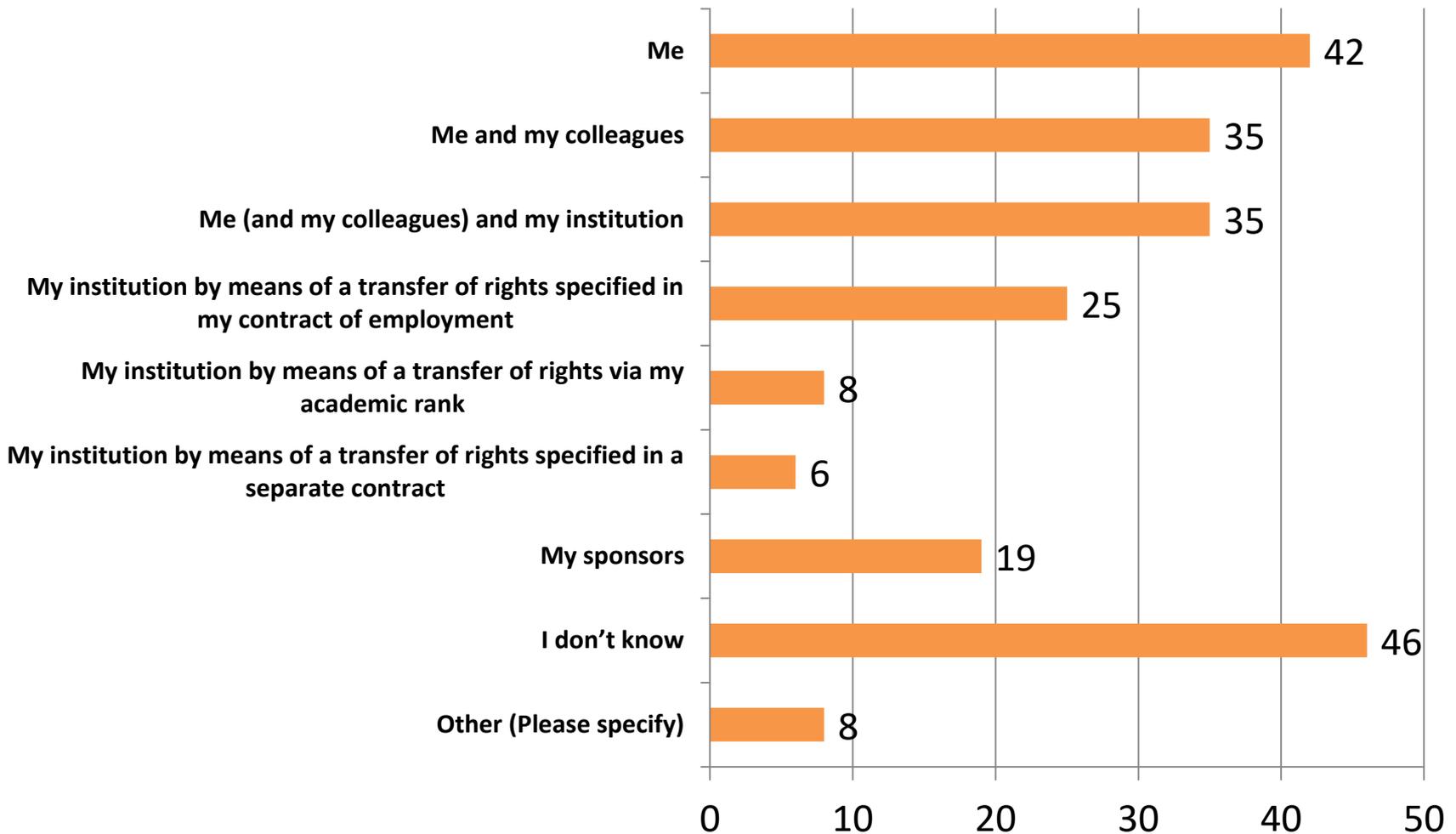
Q26. Do you use note-taking programs? (*You may choose more than one answer.*)

Other programs mentioned:

Bear	1
BoxNotes	1
DrawBoard	1
Microsoft Word	1
Notepad	1
Notes on Mac	1
StickyNotes	1
Thunderbird	1
Turtl	1
Zotero with zotfile extension	1

Legal aspects

Q27. Do you know who is the owner of the datasets that were put together during your research projects? (*You may choose more than one answer.*)



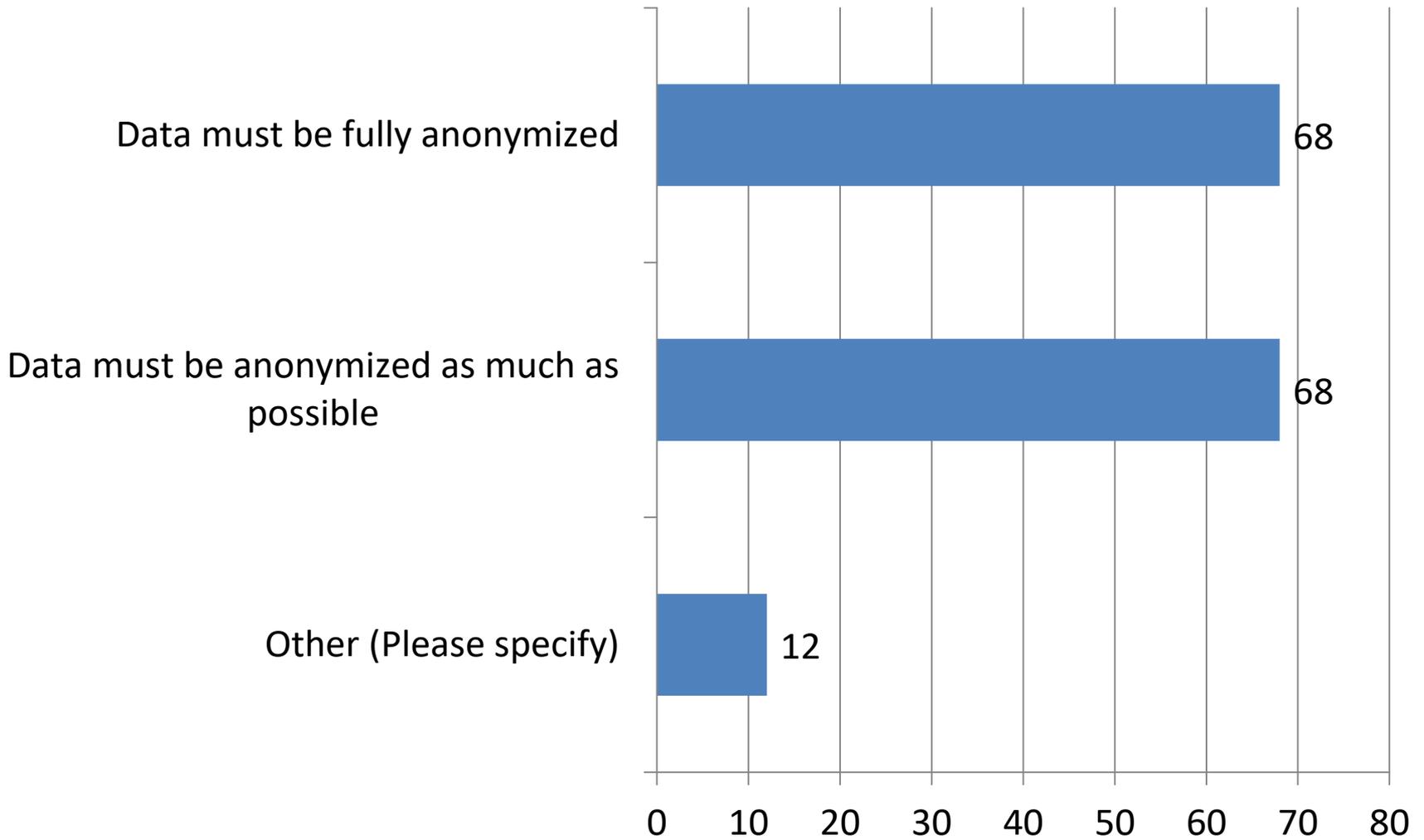
Total: 166

Skipped: 63

Q27. Do you know who is the owner of the datasets that were put together during your research projects? (*You may choose more than one answer.*)

- Some respondents mentioned that they use open data, therefore the question is not relevant to their situation.
- Others specified that data could belong to an external service provider, such as government agencies or newspaper archives.

Q28. To what extent are you required to anonymize the data that you collect?



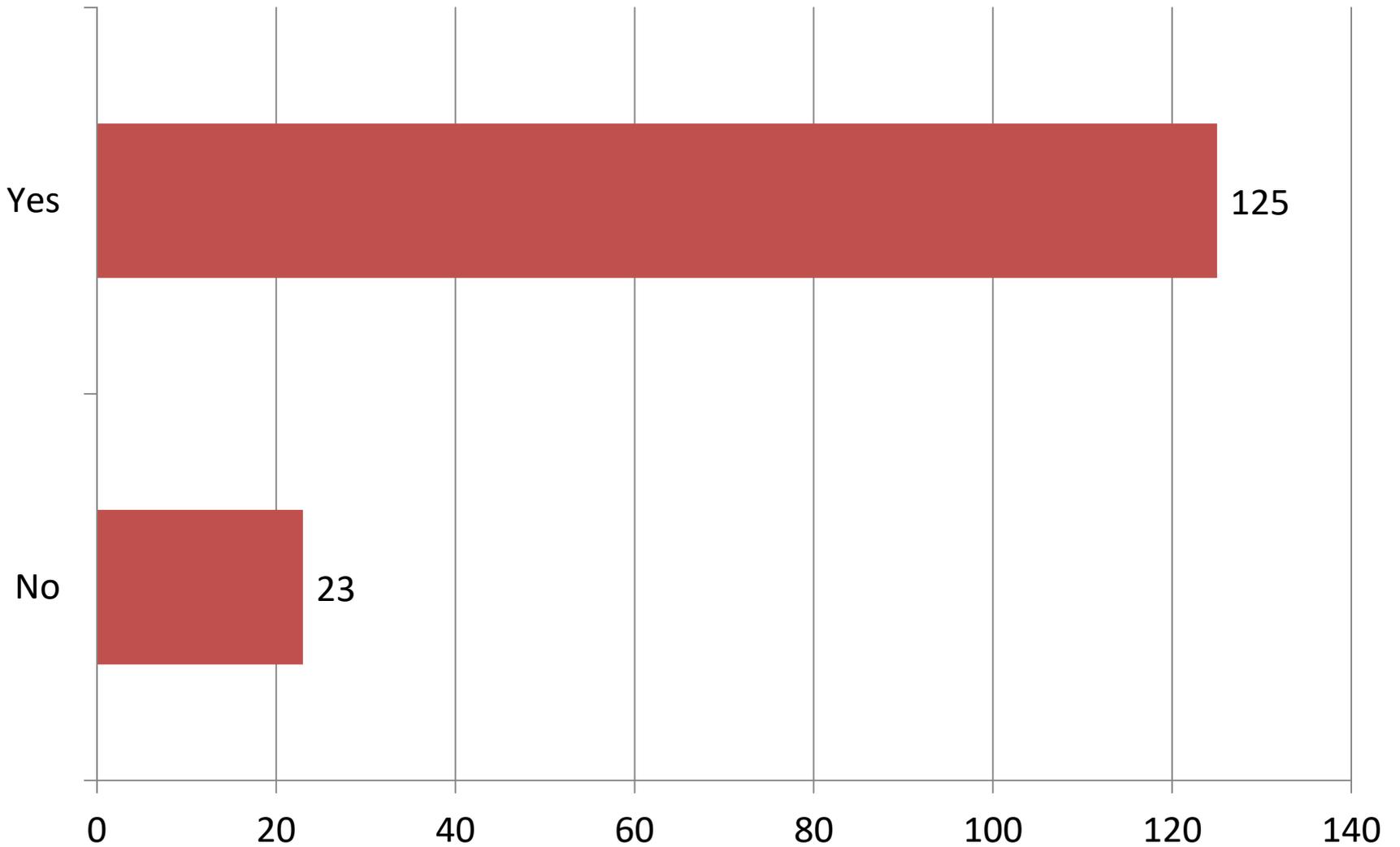
Total: 148

Skipped: 57

Q28. To what extent are you required to anonymize the data that you collect?

- Several respondents pointed out that, while some if not most of the data are anonymized, certain subsets are not.
- This is for instance relevant with oral history data, which, with the respondents' consent, are not anonymized for practical reasons.
- Another possible scenario is that data are anonymized when mentioned in a publication, but not otherwise.
- Finally, certain types of data need not be anonymized in the first place, e.g. data from works of fiction.

Q29. Do you anonymize your data by yourself?

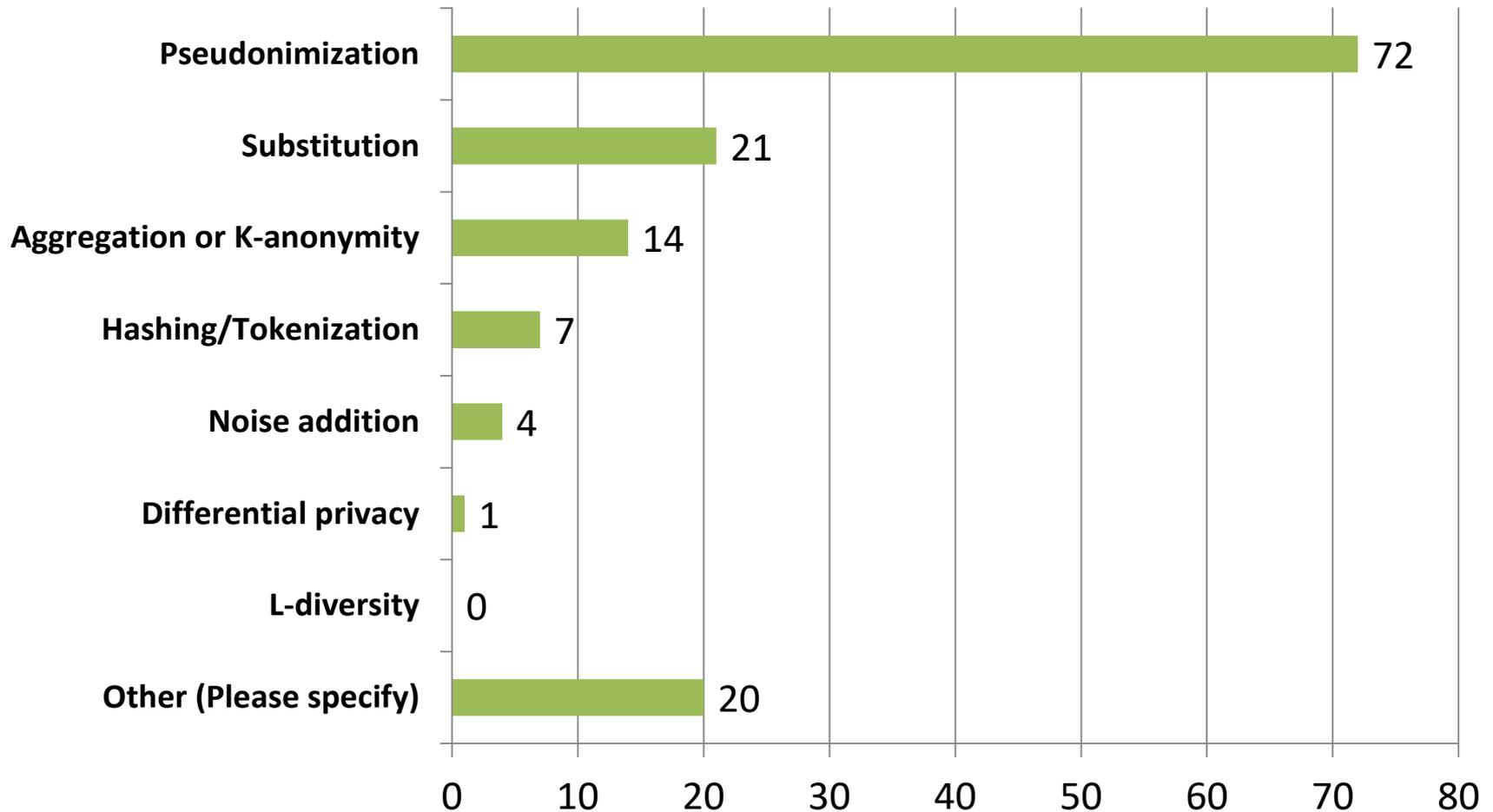


Total: 148

Skipped: 57

[If respondent chose "Yes" for Q29; else → Q31]

Q30. Which anonymization techniques do you use? (You may choose more than one answer.)



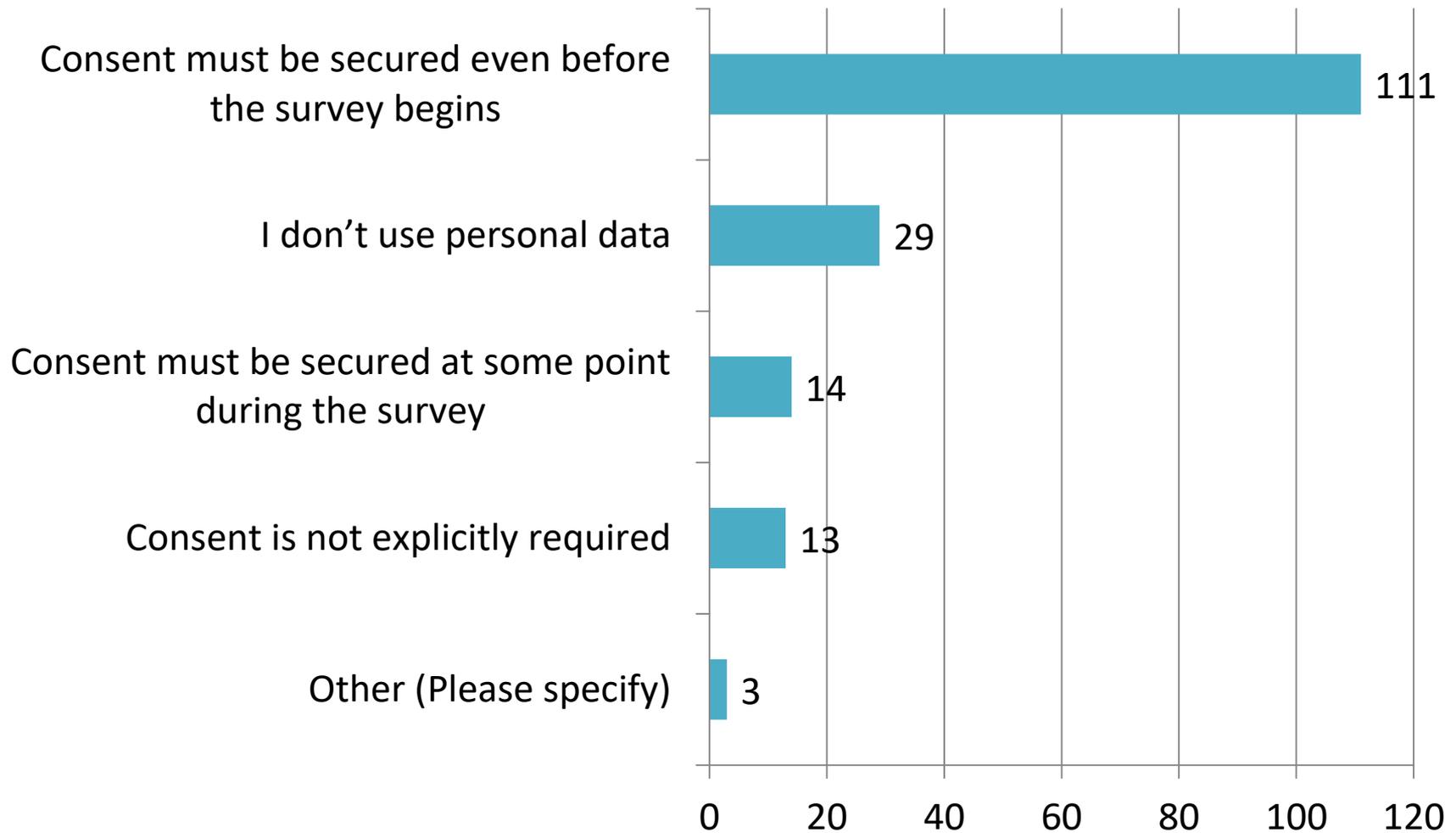
Total: 108

Skipped: 17

Q30. Which anonymization techniques do you use? (*You may choose more than one answer.*)

- Other techniques of anonymization mentioned by the respondents include:
 - deletion of nominal variables and any other personal information
 - coding of names and any other personal information
 - dissociation (*loskoppeling*) of originally related data elements
 - (not collecting personal data in the first place)

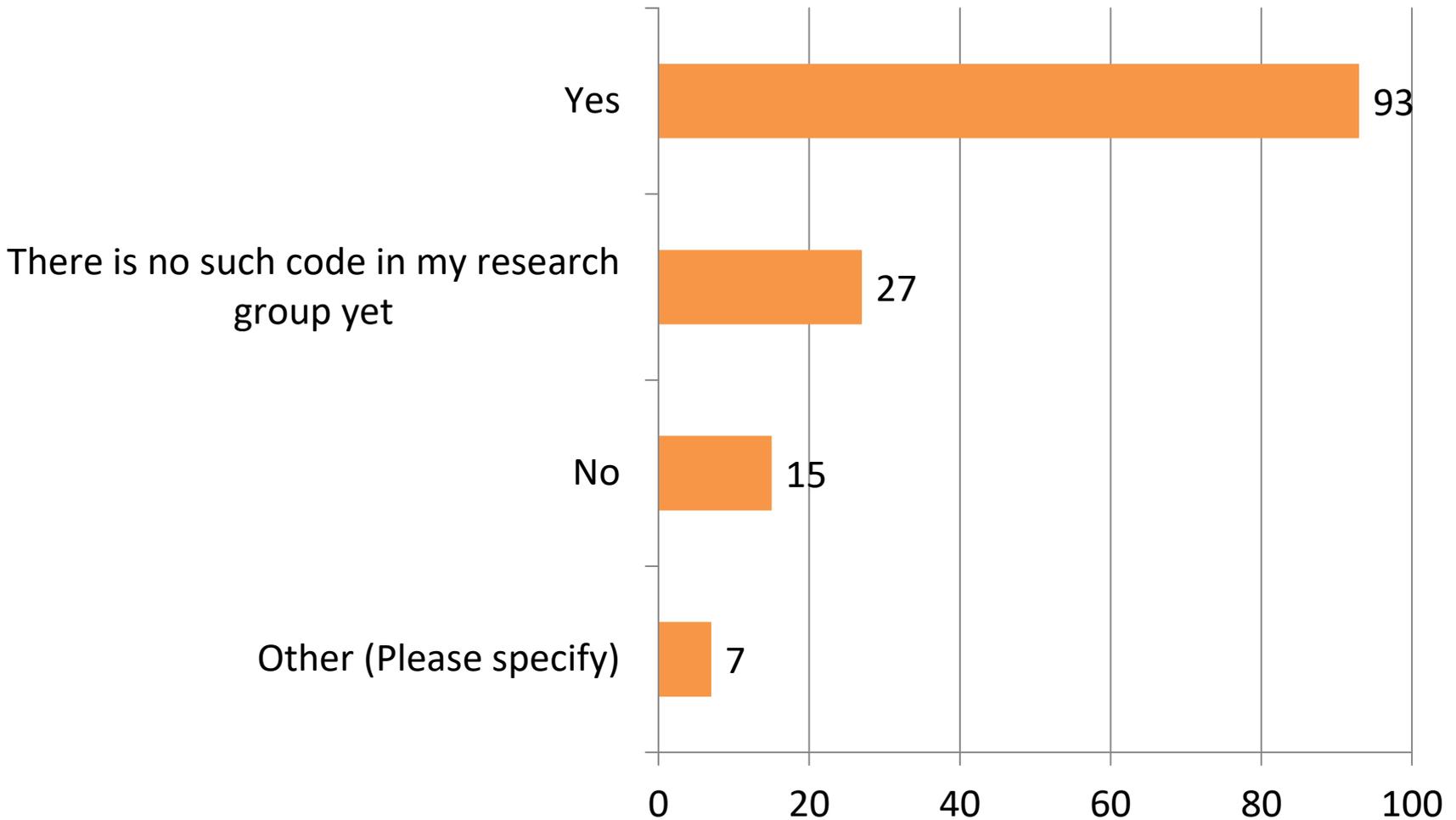
Q31. To what extent are you required to secure the consent of your respondents regarding the use of their personal data?



Total: 146

Skipped: 59

Q32. Are you required to follow codes of conduct that are specific to your research group for the handling of personal data?



Total: 142

Skipped: 63

Q32. Are you required to follow codes of conduct that are specific to your research group for the handling of personal data?

- Several respondents said that they did not know whether their research unit had such a code.
- Nevertheless, they follow certain general rules or guidelines as well as more specific codes, like the General Data Protection Regulation (GDPR).
- Some respondents remarked that their research unit's code of conduct was not always explicit or helpful.
- A few respondents mentioned that such a code was being developed in their research unit.

List of recipients

Recipients of the survey (1/7)

Katholieke Universiteit Leuven

- Behavioral Engineering Group
- Centre for Research on Peace and Development
- Centre for Sociological Research
- CMPO – Centrum voor Methodologie van het Pedagogisch Onderzoek
- CSCP – Centrum voor Sociale en Culturele Psychologie
- Institute for Anthropological Research in Africa
- Institute for Media Studies
- Interculturalism, Migration and Minorities Research Centre
- LICOS – Centre for Institutions and Economic Performance
- LINES – Leuven International and European Studies
- LRisk – Leuven Research Center for Insurance and Financial Risk Analysis
- OKP – Kwantitatieve Psychologie en Individuele Verschillen
- POOLL – Professionele Opleiding en Ontwikkeling en Levenslang Leren
- School for Mass Communication Research
- Steunpunt Werk
- VIVES – Vlaams Instituut voor Economie en Samenleving
- WOPP – Arbeids-, Organisatie en Personeelspsychologie

Recipients of the survey (2/7)

Université catholique de Louvain

- CIRFASE – Centre interdisciplinaire de Recherche sur les Familles et les Sexualités
- CIRTES – Centre Interdisciplinaire de Recherche Travail, État et Société
- DEMO – Centre de Recherche en Démographie (pretest phase)
- GIRSEF – Groupe interdisciplinaire de Recherche sur la Socialisation, l'Éducation et la Formation
- ISPOLE – Institut de sciences politiques Louvain-Europe
- LouRIM – Louvain Research Institute in Management and Organizations

Université de Liège

- CEDEM – Centre d'étude de l'ethnicité et des migrations
- CLEO – Centre d'étude de l'opinion
- CRIS – Centre de recherches et d'interventions sociologiques
- Pôle SuD – Pôle Liégeois d'Études sur les Sociétés urbaines en Développement
- Unité de recherche interfacultaire DIDACTIfen

Recipients of the survey (3/7)

Université de Mons

- humanOrg – Institut de recherche en Développement Humain, Sociétal et des Organisations
- Risques – Institut de recherche en Sciences et Management des Risques
- Santé – Institut de recherche en Sciences et Technologies de la Santé

Université de Namur

- Institut de recherche Transitions

Université libre de Bruxelles (1/2)

- CEB – Centre Émile Bernheim
- CEESE – Centre d'études économiques et sociales de l'environnement
- CeRePOI – Centre de Recherche en Psychologie des Organisations et des Institutions
- Cevipol – Centre d'étude de la vie politique
- CRePSI – Centre de Recherche en Psychologie sociale et interculturelle
- CRSE – Centre de Recherche en Sciences de l'Éducation
- Département d'Économie appliquée

Recipients of the survey (4/7)

Université libre de Bruxelles (2/2)

- Ecares – European Center for Advanced Research in Economics and Statistics
- Faculté de Philosophie et Sciences sociales
- GERME – Groupe de recherche sur les Relations ethniques, les Migrations et l'Égalité
- iCite – International Centre for Innovation, Technology and Education Studies
- IEE – Institut d'Études européennes
- IGEAT – Institut de Gestion de l'Environnement et d'Aménagement du Territoire
- Institut de Sociologie
- PsyTC – Centre de Recherche en Psychologie du Travail et de la Consommation

Université Saint-Louis – Bruxelles

- CASPER – Centre d'Anthropologie, Sociologie, Psychologie – Études et Recherches
- CEREC – Centre de Recherche en Économie
- CES – Centre d'études sociologiques
- CRESPO – Centre de recherche en science politique
- Observatoire du SIDA et des sexualités

Recipients of the survey (5/7)

Universiteit Antwerpen

- Antwerp Centre for Institutions and Multi-Level Politics
- Antwerp Centre of Evolutionary Demography
- Centre for Migration and Intercultural Studies
- Centre on Inequalities, Poverty, Social Exclusion and the City
- Centrum voor sociaal beleid Herman Deleeck
- Department of Economics
- Edubron
- FAMCARE – Family Dynamics and Care
- Institute of Development Policy
- M²P – Media, Movements and Politics
- Media, Policy & Culture
- PA&M – Public Administration and Management
- Research Group International Politics
- Society & Environment
- Urban Studies Institute

Recipients of the survey (6/7)

Universiteit Gent

- CEEM – Centre for Environmental Economics and Environmental Management
- CERISE – Centre for Russian, International Socio-Political and Economic Studies
- CESSMIR – Centre for the Social Study of Migration and Refugees
- End-of-Life Care Research Group
- Vakgroep Accountancy, Bedrijfsfinanciering en Fiscaliteit
- Vakgroep Bestuurskunde en Publiek Management
- Vakgroep Communicatiewetenschappen
- Vakgroep Data-Analyse
- Vakgroep Experimentele Psychologie
- Vakgroep Marketing, Innovatie en Organisatie
- Vakgroep Politieke Wetenschappen
- Vakgroep Sociologie

Recipients of the survey (7/7)

Universiteit Hasselt

- Expertise Centre for Digital Media
- IMOB – Transportation Research Institute

Vrije Universiteit Brussel

- BRIO – Brussels Informatie-, Documentatie- en Onderzoekscentrum
- Brussels Centre for Urban Studies
- CEMESO – Culture, Emancipation, Media and Society
- Cosmopolis – Centre for Urban Research
- Department of Applied Economics
- Department of Political Science
- Expertisecentrum Gender, Diversiteit en Intersectionaliteit
- ID – Interface Demography (pretest phase)
- Tempus Omnia Revelat