



SAGE Research Survey: Results Report

July 2016

Amanda Lau; Lucie Richards



Policy Wise
for Children & Families

1. Introduction

The Secondary Analysis to Generate Evidence (SAGE) Researcher Survey 2016 was sent out to the research community in April 2016 to assess, from the perspective of a secondary data user (hereafter referred to as an “accessor”) as well as of a primary data producer (hereafter referred to as a “depositor”), what factors lead to successful sharing and re-use of data. These factors are all, to some extent, features and services that SAGE aims to build, and we hoped that feedback on the relative importance of each feature or service would enable us to prioritize our service development.

Approximately 950 members of the research community were sent an invitation to complete the SAGE Researcher Survey. We received 82 responses to the survey by late May 2016, 8 of which were partial responses. In all cases the partial responses were nearly complete, and so were kept in this descriptive analysis. In three cases, researchers responded more than once, but with different (complete) responses.

Please note these results should be interpreted with caution due to the small sample size and self-selection of respondents.

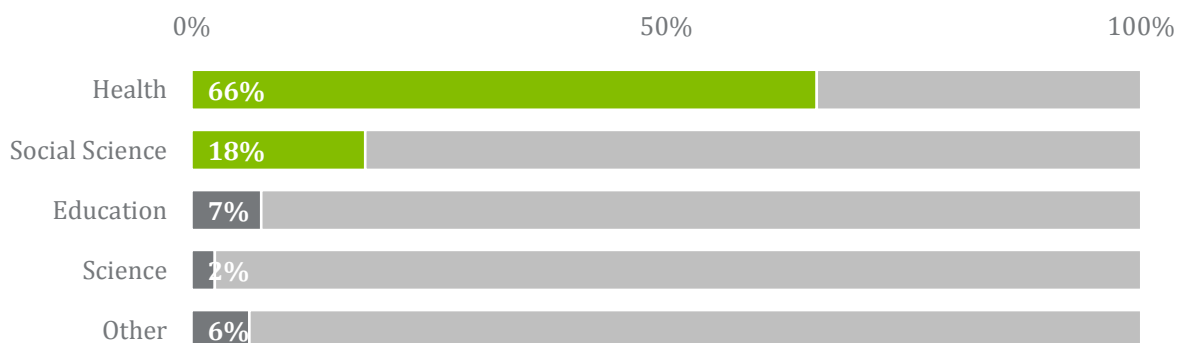
2. Results

2.1. Field of Research

The responding researchers work, for the most part, in health (i.e. mental health, nursing, health promotion, and clinical research) and social science (i.e. psychology, sociology, and kinesiology) fields. To a lesser degree, researchers that work in education or science fields have research that intersected with either health or social sciences.

What is your current field of research?

Percentages of researchers, categorized by field

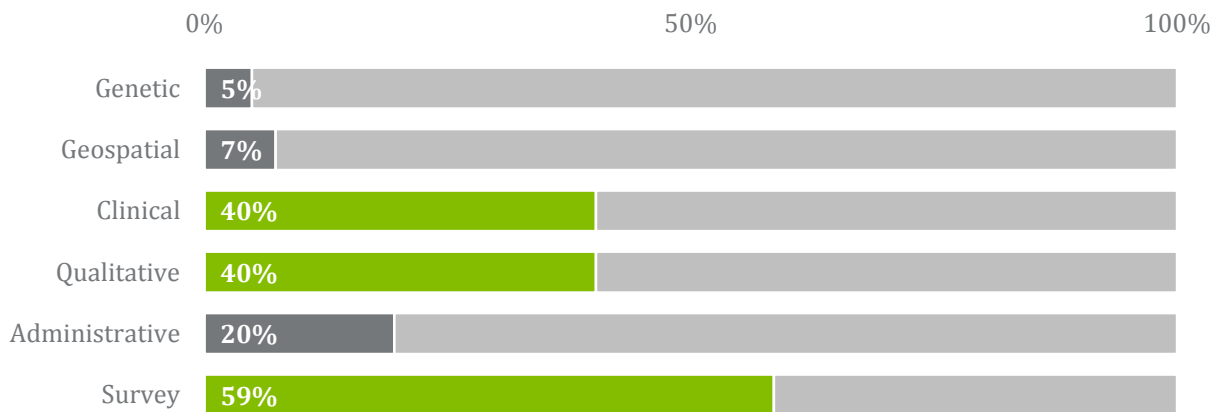


2.2. Data currently being used

Responding researchers use a variety of types of data, although survey, clinical and qualitative (eg. Interviews, focus groups) data make up the majority of listed data types. A moderately high number of respondents also use administrative data.

What types of data do you currently use?

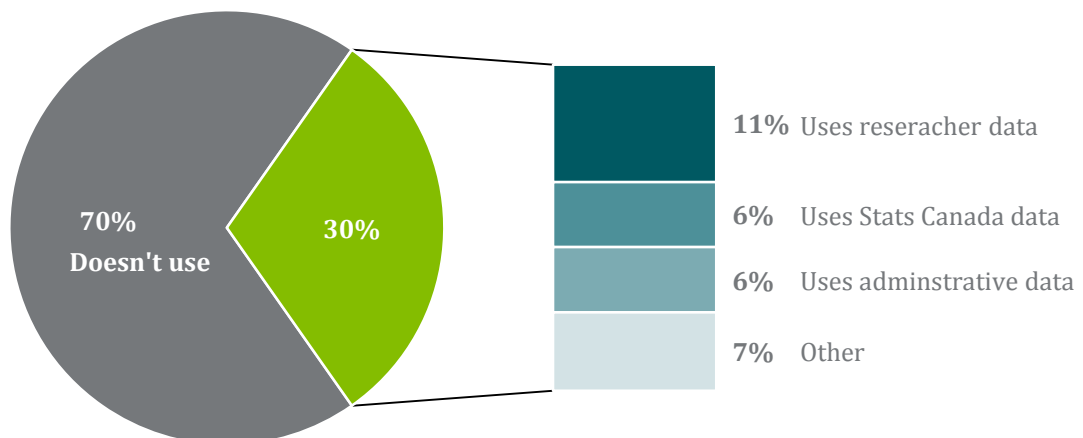
Selective list of research data currently used



2.3. Secondary Data Use

Only 30% of responding researchers indicate currently using secondary data for their research. Of these, approximately one third of researchers work with other researchers to access their data.

Do you currently use secondary data?



For those that did use secondary data, the majority were quantitative researchers followed by mixed methods (i.e. quantitative and qualitative) researchers. None of the qualitative researcher

surveyed used secondary data for their research. However, these results should be interpreted with caution due to the small sample size.

Secondary Use by Research Method

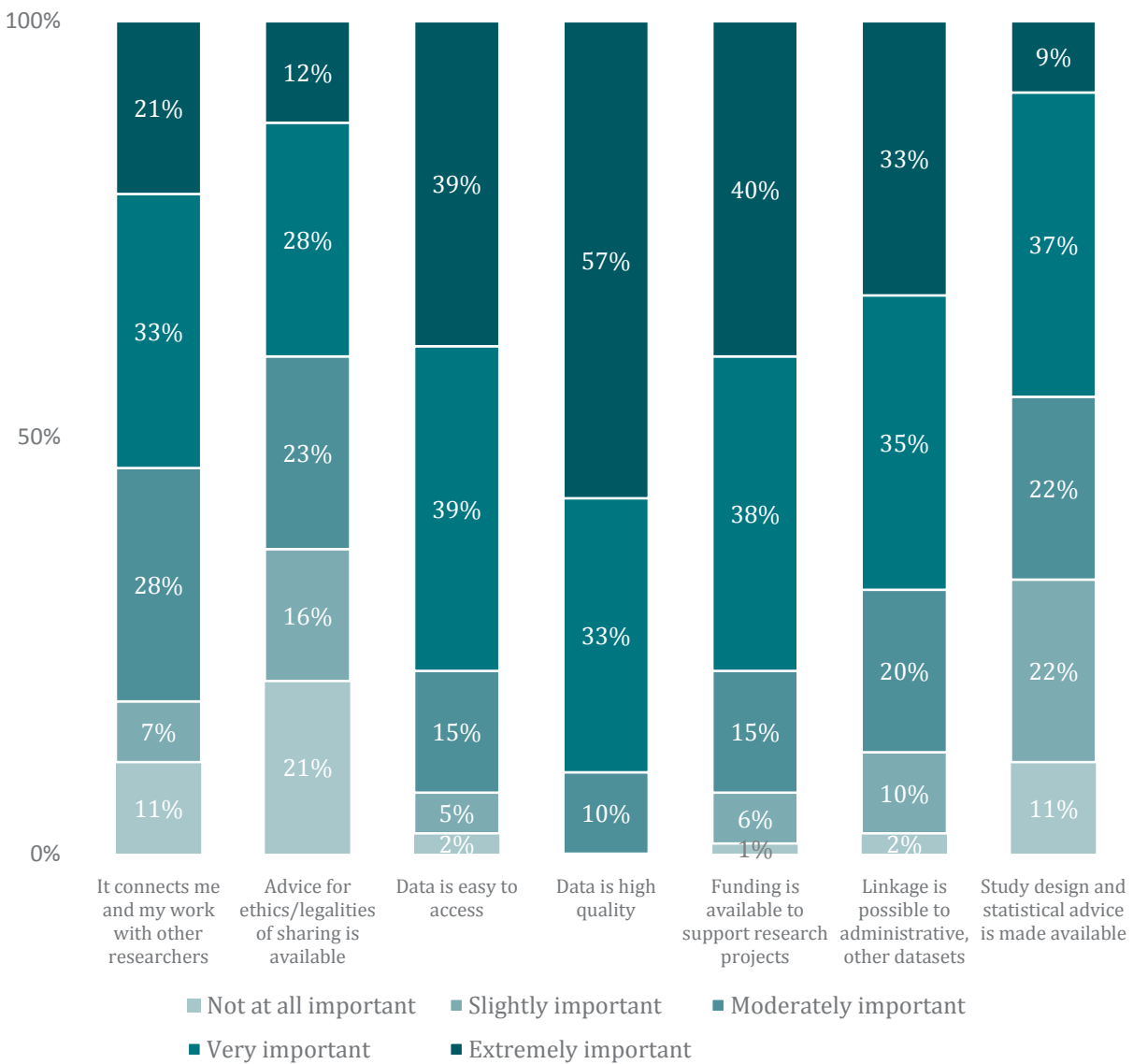
Percentages of researcher, categorized by method



Encouragement for Secondary Data Use

By contrast, respondents were more clearly in agreement on factors that would encourage them to use secondary data at SAGE. By far the most important of these is the presence of high quality secondary data assets. Closely following data quality was ease of access, funding availability and the ability to link to administrative and other datasets. Researchers on several occasions took the opportunity to comment in the “other section” about the importance of quality (or relevance) of data and ease of access as particularly important. In contrast to data management services, individual respondents tended to rate factors much less consistently.

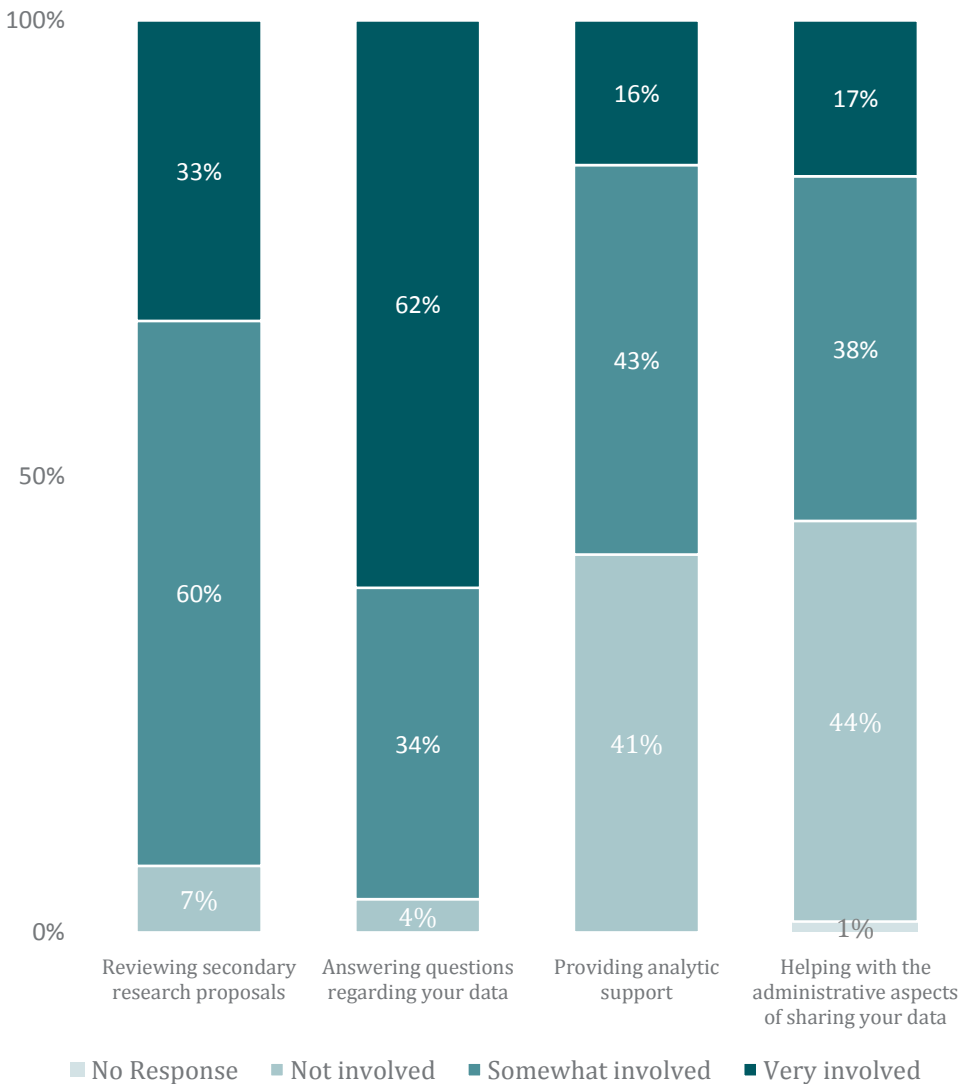
What would encourage you to conduct secondary analysis on data assets at SAGE?



2.5. Depositor's involvement in Secondary Data Use

Respondents asked to imagine themselves as depositors were much more interested in being involved with answering questions about their data and reviewing secondary research proposals than in providing analytic support or dealing with the administrative aspects of sharing data. Nevertheless, there were a fairly high number of researchers who wants at least some level of involvement in all aspects of sharing.

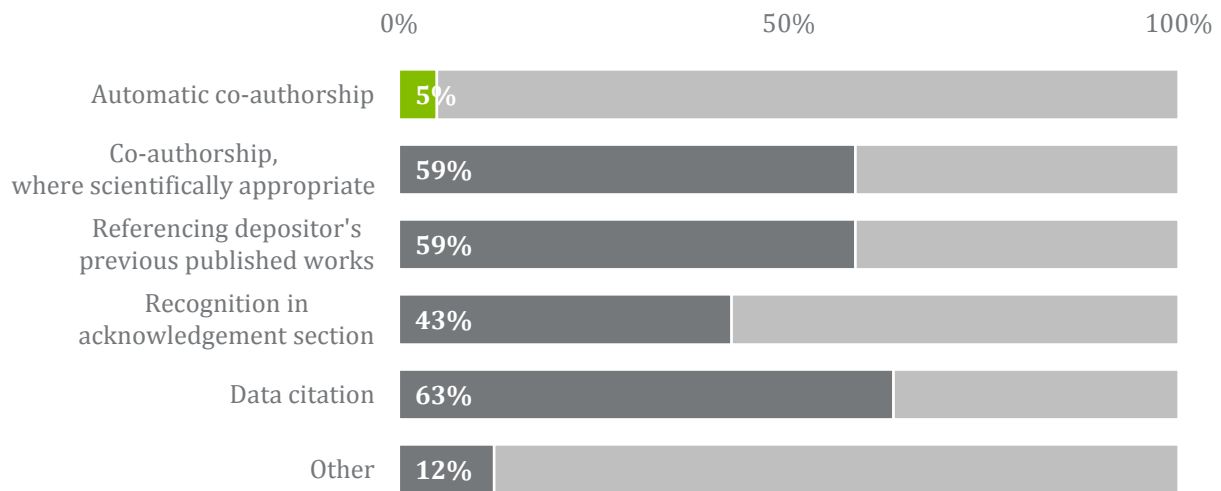
If you were to deposit data with the CDCA, how involved would you want to be in...



2.6. Depositor Recognition

Responding researchers agree with most forms of recognition for depositors, save automatic co-authorship on secondary research. There is also somewhat less agreement that data producers should be formally acknowledged in the publication.

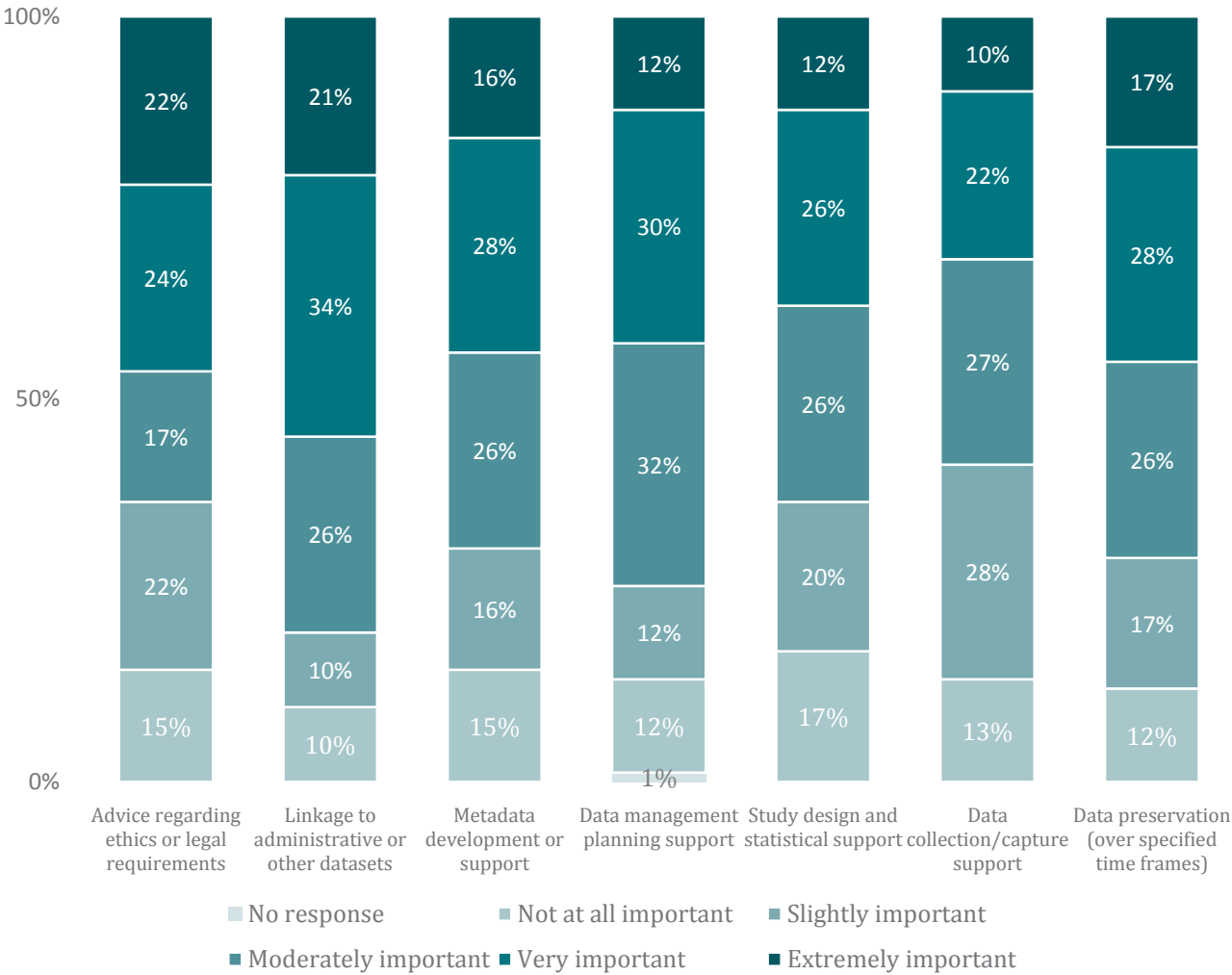
How should the depositor be recognized within the secondary research publication?



2.7. Data Management Needs

For the most part, respondents find all aspects of data management services at least moderately important. The ability to link to administrative and other datasets was somewhat more often rated as very or extremely important, while data collection support was somewhat more often rated as not at all or slightly important. It is worth noting, however, that individual respondents tended to have overarching attitudes towards data management services, either highly valuing such services or not at all.

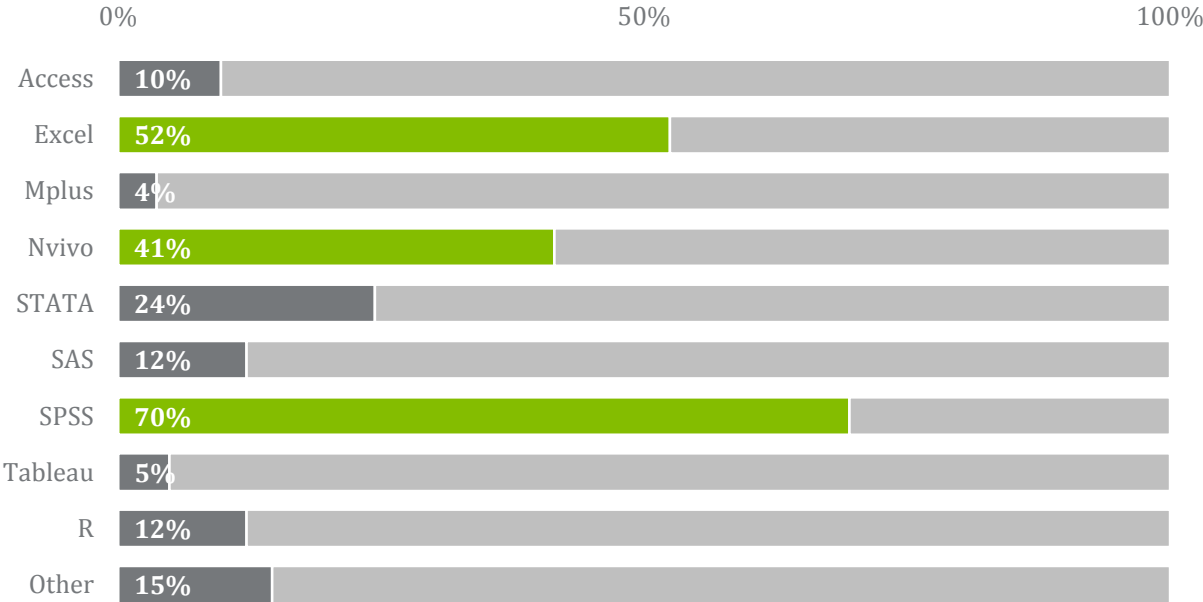
What data management services would meet an unmet need for your research?



2.8. Data-related software

Responding researchers also overwhelmingly use SPSS and NVivo for data manipulation, with Excel also ranked very highly. The design of the study left it unclear whether Excel is used primarily as a data input, visualization or analysis tool for these researchers, but it was listed alone in only 4 cases, which might indicate it is an all-purpose tool rather than used for data analysis.

Which data-related software do you use?



2.9. General Comments

Finally, of note: there were a few comments by researchers regarding the special nature of qualitative data, in particular data collected and shaped both by participants and researchers (community based participatory research, research with aboriginal communities). Because of the unclear or joint ownership of such data, sharing in these cases would require SAGE to develop a relationship not only with the researcher but the community as well, as well as special processes for access. In addition, certain types of “deep” qualitative data are not appropriate for secondary use, as by design such use would take it out of context.