Differences in data-sharing attitudes and behaviours, extended version to African data curators and data management experts

Flavio Bonifacio[1] and Winny Nekesa Akullo[2]

Abstract

This article reports the results of a survey conducted between 16th November and 8th December 2020 among African data curators and data experts about different aspects of data sharing. The sample of respondents has been extracted from participants to the 1st IASSIST Africa Regional Workshop held on 11th -13th January 2021, Kampala, UGANDA and other data experts and practitioners. First, we recall the main results of a previous article published by IQ[3] about the same argument in order to introduce the new survey. After that we analyse the new findings comparing them with the previous results, splitting the samples between Africans and not Africans.

The idea of a new survey

The idea came from a previous survey conducted at the end of 2017 and published in the article "Differences in Data-Sharing Attitudes and Behaviours", Vol. 42 No. 3 (2018): IASSIST Quarterly. The questions regarded different aspects of data sharing: tools used in building metadata, problems encountered in order to share the data, the propensity to share the data, and the satisfaction obtained from different working tasks. This article extends the survey to African data practitioners. The occasion has been the 1st IASSIST Africa Regional Workshop, 11th -13th January 2021, Kampala, UGANDA. Winny Nekesa helped me to reach the attendees and other interested persons by email, sending them the link to the questionnaire. Thirty-one people took part in the survey.

Above all, this conclusion of the previous survey attracted my attention: "we have to broaden the scope of service promotion, moving from 'developed countries to 'developing countries', where data curation is less practised, to younger people, involving women in greater responsibility and more remunerative roles. How to do that is a matter that goes beyond the scope of this article, but one suggestion is to transform the self-referential meetings into open symposiums, for example moving from the usual locations (for IASSISTers USA, Canada, North Europe) to, perhaps, less easy locations, such as Southern Europe, Africa or Asia, and to less easy environments, outside the University, in an open public and private space."

We think that the IASSIST Africa Regional Workshop that took place in Uganda may be seen as a step moving in this direction.

2021 Survey, goal and general description

The goal of the work already reported in the previous survey "is to analyse attitudes toward data sharing in order to find strategies to expand data sharing and to show the best paths to gain new 'Premium Followers', as we call the best performers in Data Sharing (below). This pragmatic point of view comes from empirical evidence that is compliant with other more detailed analysis coming from wider surveys. We confirm what other sources say: 'Results show that researchers in different regions have different perceptions about data and different data behaviours'[4]. Although in literature there are many other well documented perspectives from which to examine Data Sharing, covering a range of aspects[5] from the political to the technical, we think that one of the most pressing issues is how to expand best practices in sharing data." In the actual survey we add the description of perceptions about data and data behaviours among African Data experts to the findings of the 2017 survey. The 2017 survey showed the existing differences between "Italians" as representative of a community where Data Sharing is less practised, and other more attentive communities.
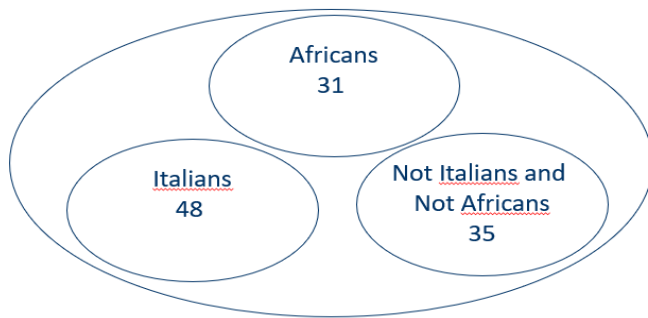
We start with a short description of background variables. Contrasting with the non-Africans sample, we note that Africans work more often in the public administration, are slightly more male than female, are younger (more than one half less than 40), more frequently they have a master degree and studied more in the scientific field while not Africans studied more in the social sciences field. Considering the use of metadata standards, an important indicator of data familiarity, Africans use them less than not-Africans. If we split the not-African between Italians and not-Italians, the Africans place themselves in the middle. Except for gender and educational qualifications, all other relations are significant. In the following figures one asterisk (*) means significance level at alpha level of 10%, two asterisks (**) means significance level at alpha level of 5%. Three asterisks (***) means significance level at alpha level of 1%.

Fig.1 – Significance (**)

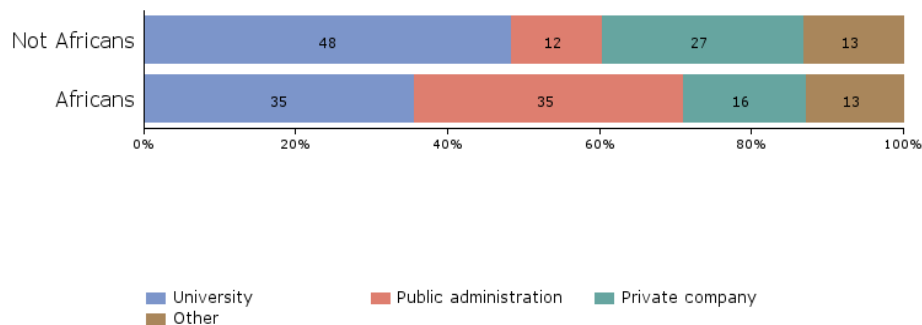**Which one of the following best describes your primary work sector?**
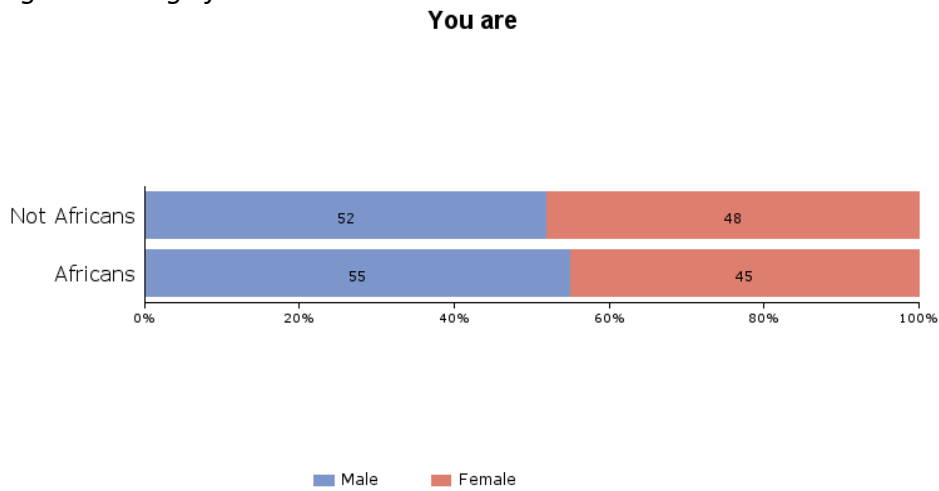
Fig.2 – Not significant

You are

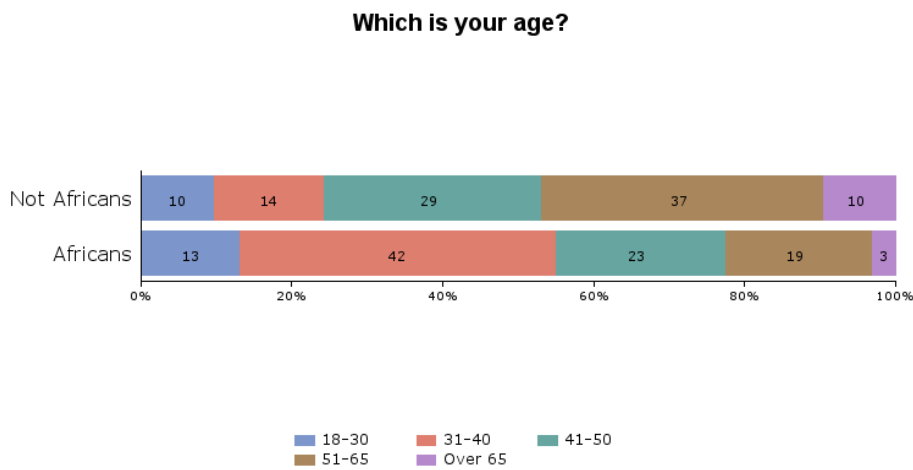|		
Not Africans	52	48
Africans	55	45

0% 20% 40% 60% 80% 100%

■ Male ■ Female

Fig. 3 – Significance (**)

Which is your age?

Not Africans	10	14	29	37	10
Africans	13	42	23	19	3

0% 20% 40% 60% 80% 100%

■ 18-30 ■ 31-40 ■ 41-50
■ 51-65 ■ Over 65

Fig. 4 – Not significant

What is your educational qualification?

Not Africans	2	41	37	19
Africans		19	68	13

0% 20% 40% 60% 80% 100%

■ High schools ■ Graduation ■ Master
■ PhD

Fig. 5 – Significance (***)

Which field of study?

Not Africans: Humanities 16, Scientific 34, Social 51
Africans: Humanities 26, Scientific 58, Social 16

Legend: ■ Humanities ■ Scientific ■ Social (sociology, psychology, economics)

Fig. 6 – Significance (*)

Use of metadata standard?

Not Africans: Yes 63, No 37
Africans: Yes 45, No 55

Legend: ■ Yes ■ No

Fig. 7 – Significance (***)

Use of metadata standard?

Not Italians and not Africans: Yes 74, No 26
Italians: Yes 29, No 71
Africans: Yes 45, No 55

Legend: ■ Yes ■ No

Data sharing attitudes and behaviours

Most of the questions have been extracted from IMDSM2017 (see the above quoted 2017 survey report, *Differences in Data-Sharing…*)[6] and from the survey Changes in Data Sharing and Data Reuse Practices and Perceptions (cf. *Differences in Data-Sharing…)* among Scientists Worldwide and grouped into three main conceptual frameworks:

1. items influencing data sharing propensity (6 items)
2. items influencing work satisfaction (7 items)
3. problems emerging in data sharing (21 items)

We report the results for each framework starting from a synthetic overview of the respondent answers frequencies for the African sample, ordered in descending order of agreement (agree somewhat, agree strongly are jointly considered as a single answer for ordering the frequencies). Then we analyse in deeper detail only the distributions of the items showing greater, significant differences between Africans and not Africans. Finally, we will show a synthetic index for each framework[7], summarizing all the information in a single measure.

Items influencing data sharing propensity (6 items)

In the following figure we report the items used to measure the sharing propensity. The most appreciated item refers to *the use of other researcher datasets*, while one of the less appreciated ones refers *to placing all "my" data into a central repository.*

Fig. 8 – The following statements relate to sharing scientific data. Tell us how much you agree with each statement…
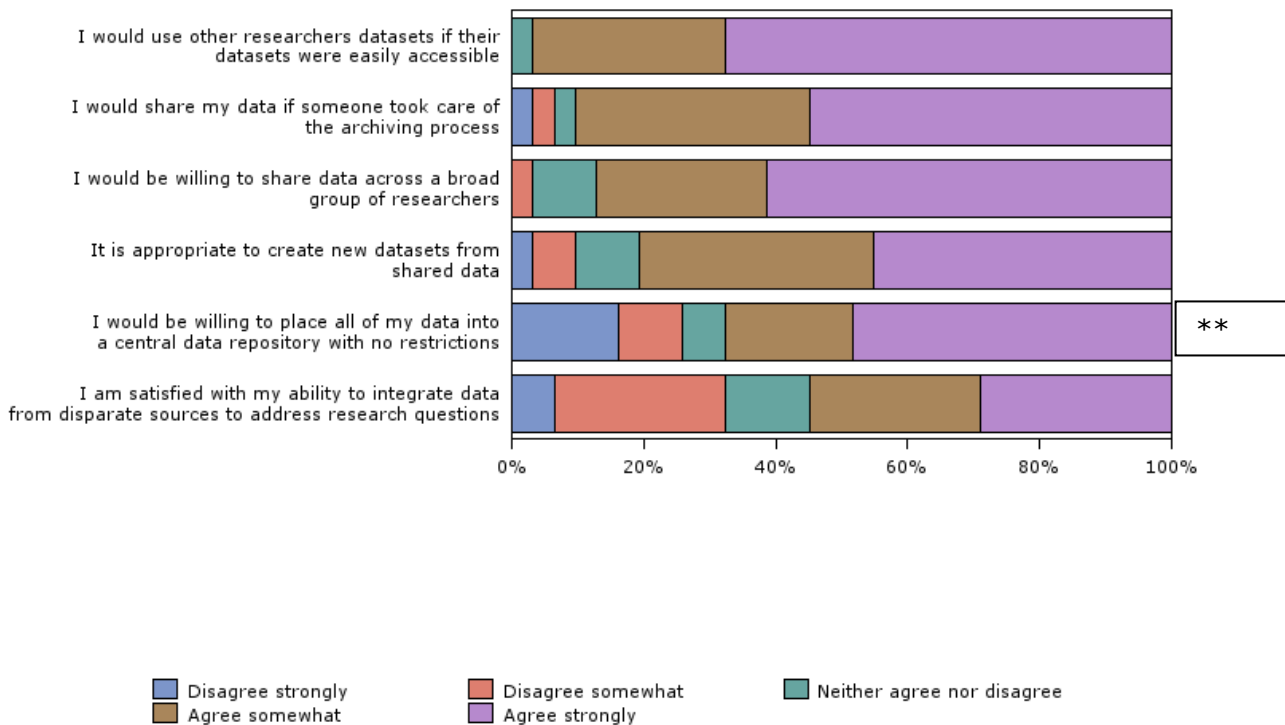
Details

The only distribution showing a significant difference between Africans and Not Africans (α=5%) relates the opinion to place data into a central repository. Looking at the distributions we observe a greater concentration of Africans, both among the most disagreeing and the most agreeing, with respect to not Africans. Interesting is the fact that about one half of Africans agree strongly with placing their own data into a central repository.
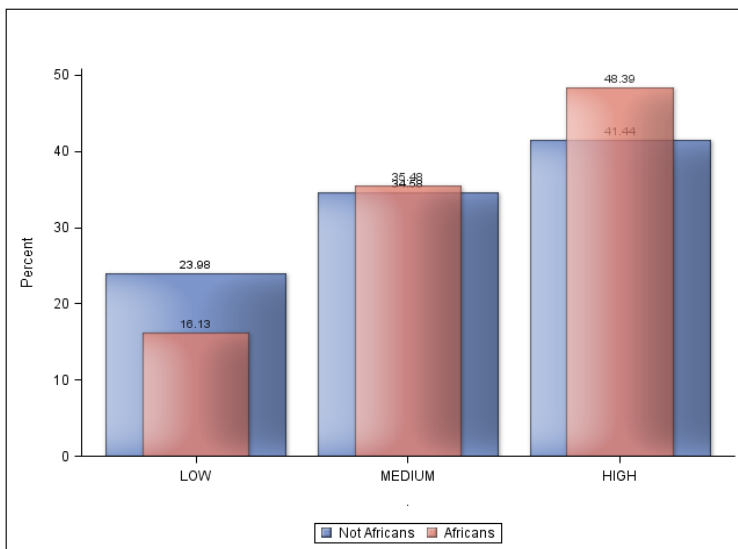
Fig. 9 - Significance (**)
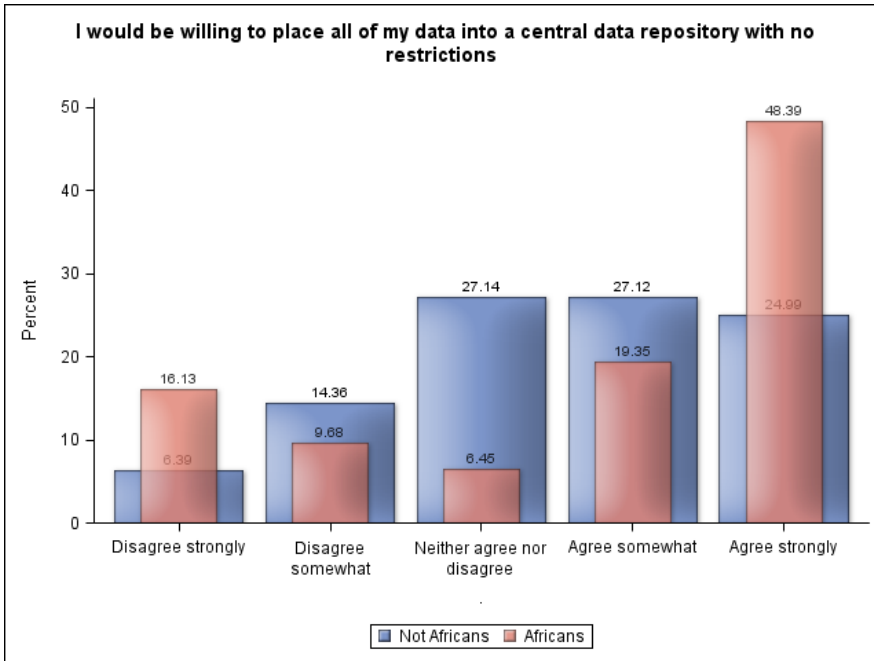
Sharing propensity index

The index summarizing all the items (*Data Sharing Propensity Index*) does not show significant differences. Therefore, we can say that from the *Data Sharing Propensity* point of view, *Africans and not Africans show similar behaviour.*

Fig. 10 – Sharing propensity index

Items influencing work satisfaction (7 items)

In this case four Items referring to the work satisfaction show a significant relation with the origin of the respondents. One of them *(tools for preparing my documentation)* is at 1% α level. Also significant are processes for storing, searching and collecting data.

Looking at the order of items the most satisfying is the *process for cataloguing-describing data* while *the process for analysing data* is the least satisfying.

Fig. 11 – The following statements relate to how you collect and use research data. Tell us how much you agree with the following ways to complete this sentence: I AM SATISFIED WITH THE…
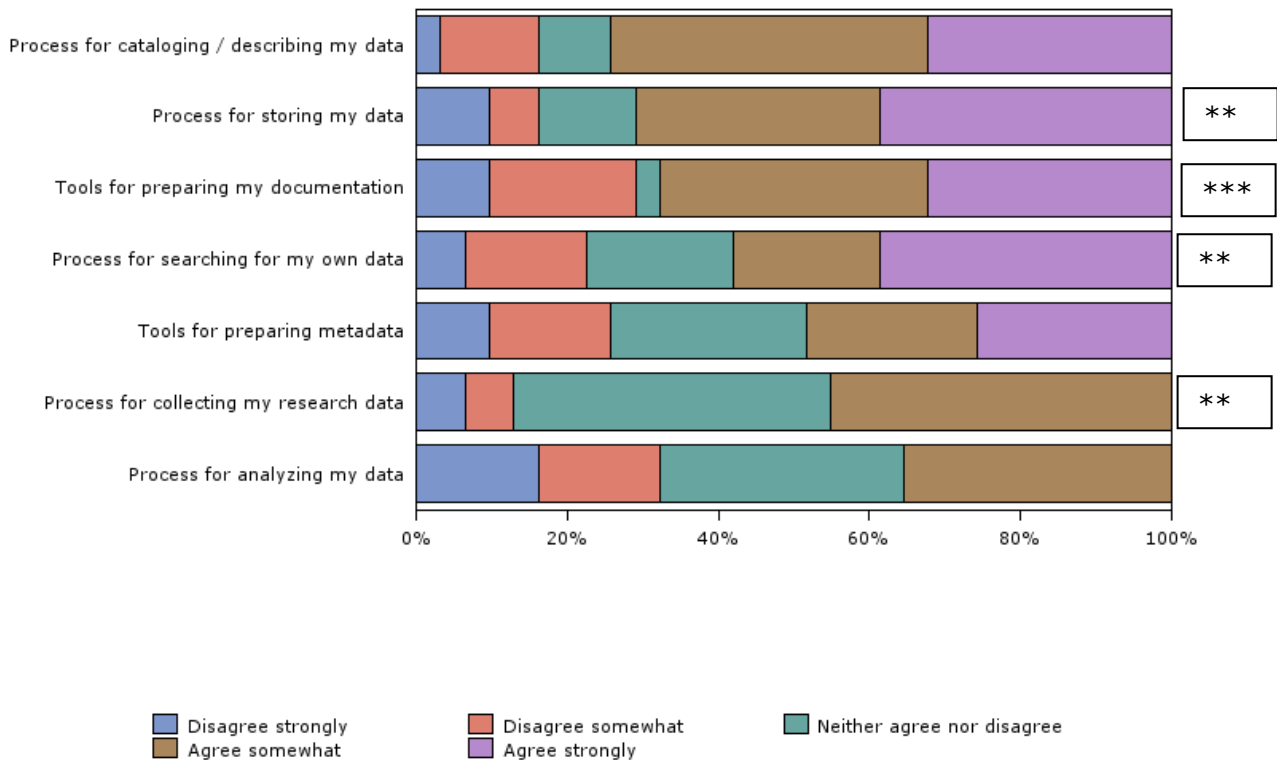
Details

Interesting is the fact that for almost all the items Africans show stronger Agreement.
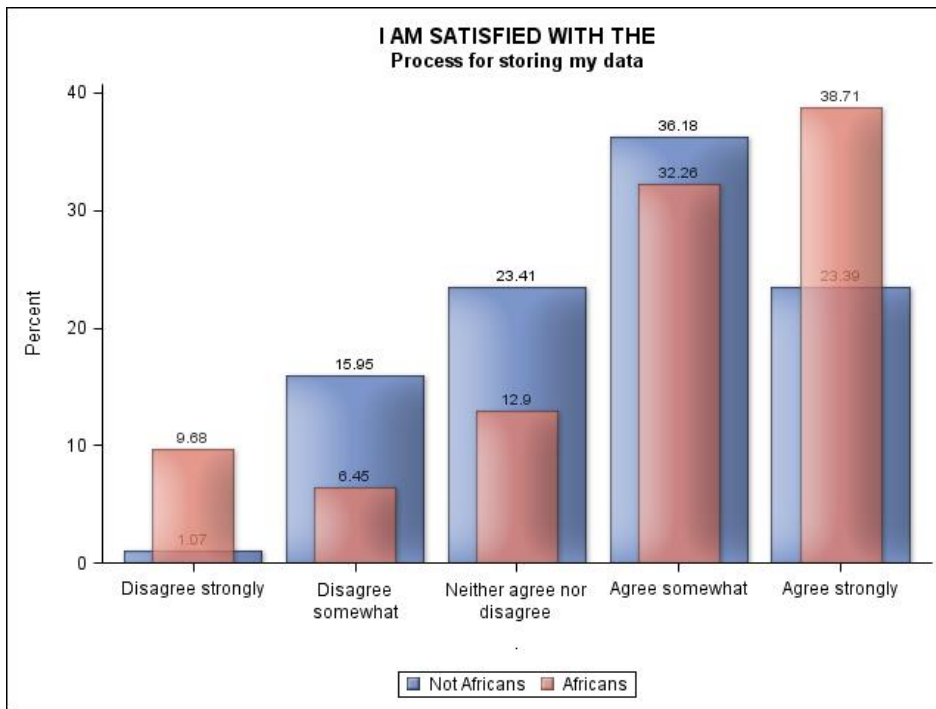
Fig. 12 - Significant (**)
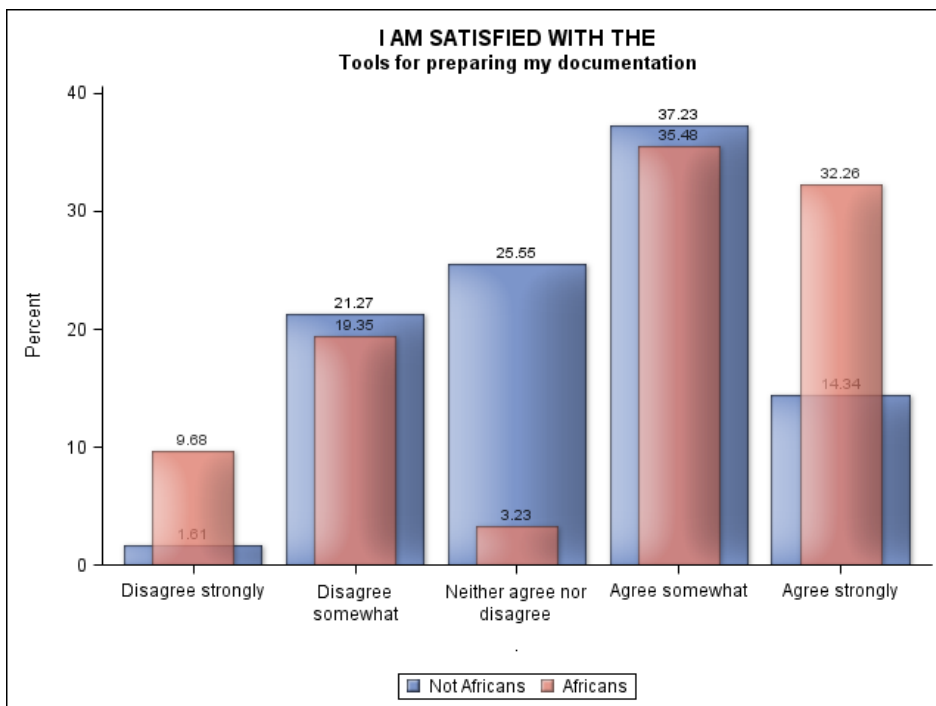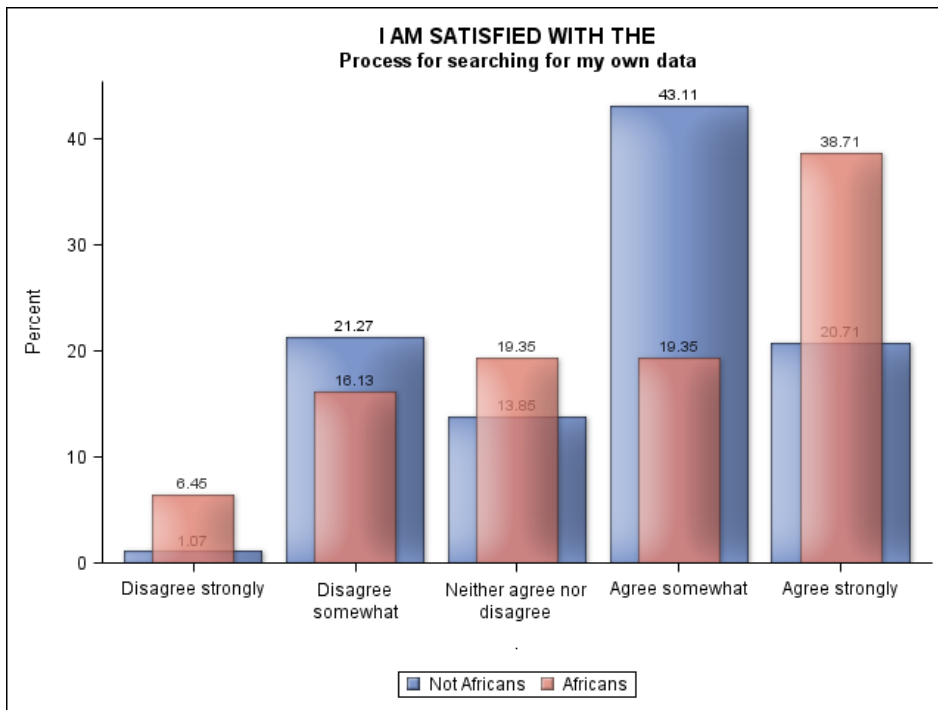

Fig. 13 - Significance (***)

 Bonifacio, Flavio; Akullo, Winny Nekesa (2021) Differences in data-sharing attitudes and behaviours, extended version to African data curators and data management experts, IASSIST Quarterly 45(3-4), pp. 1-17. DOI: https://doi.org/10.29173/iq993

Fig. 14 – Significance (**)

I AM SATISFIED WITH THE
Process for searching for my own data

Fig.- 15 Significance (**)

I AM SATISFIED WITH THE
Process for collecting my research data
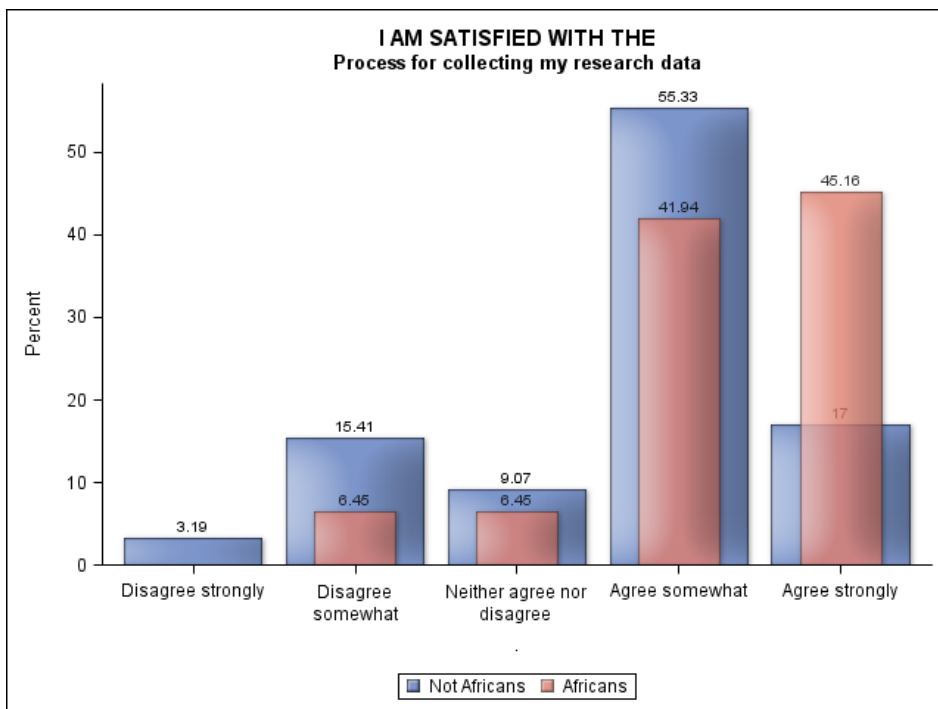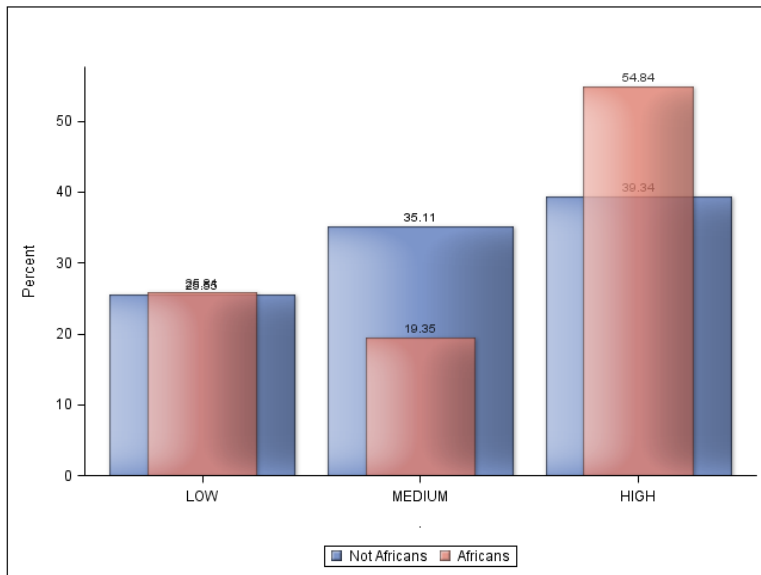
Work satisfaction index

This causes a higher concentration on higher scores of Satisfaction Index, although not yet significant.

Fig. 16 - Work satisfaction index



Problems emerging in data sharing (21 items)

In this context most of the item distribution differences between Africans and not Africans are significant. There are five items that reach three stars of significance, α=1%: *I would lose control of the data, My data change too quickly, My data are old - they don't answer to the questions researchers ask today, I spent a lot of money on this research and it is not economically convenient to share it, It could affect negatively my career.* Noteworthy is the fact that at least three items refer to personal views which directly contrast the Data Sharing positive vision.

Referring to the order of items in the table, on the top we find problems concerning data propriety, the shortage of funds, privacy considerations; on the bottom the aspects purporting personal interests instead of more general views.

Fig. 17 – How much do you agree with these statements about the reasons that prevent data sharing? I would share the data but…
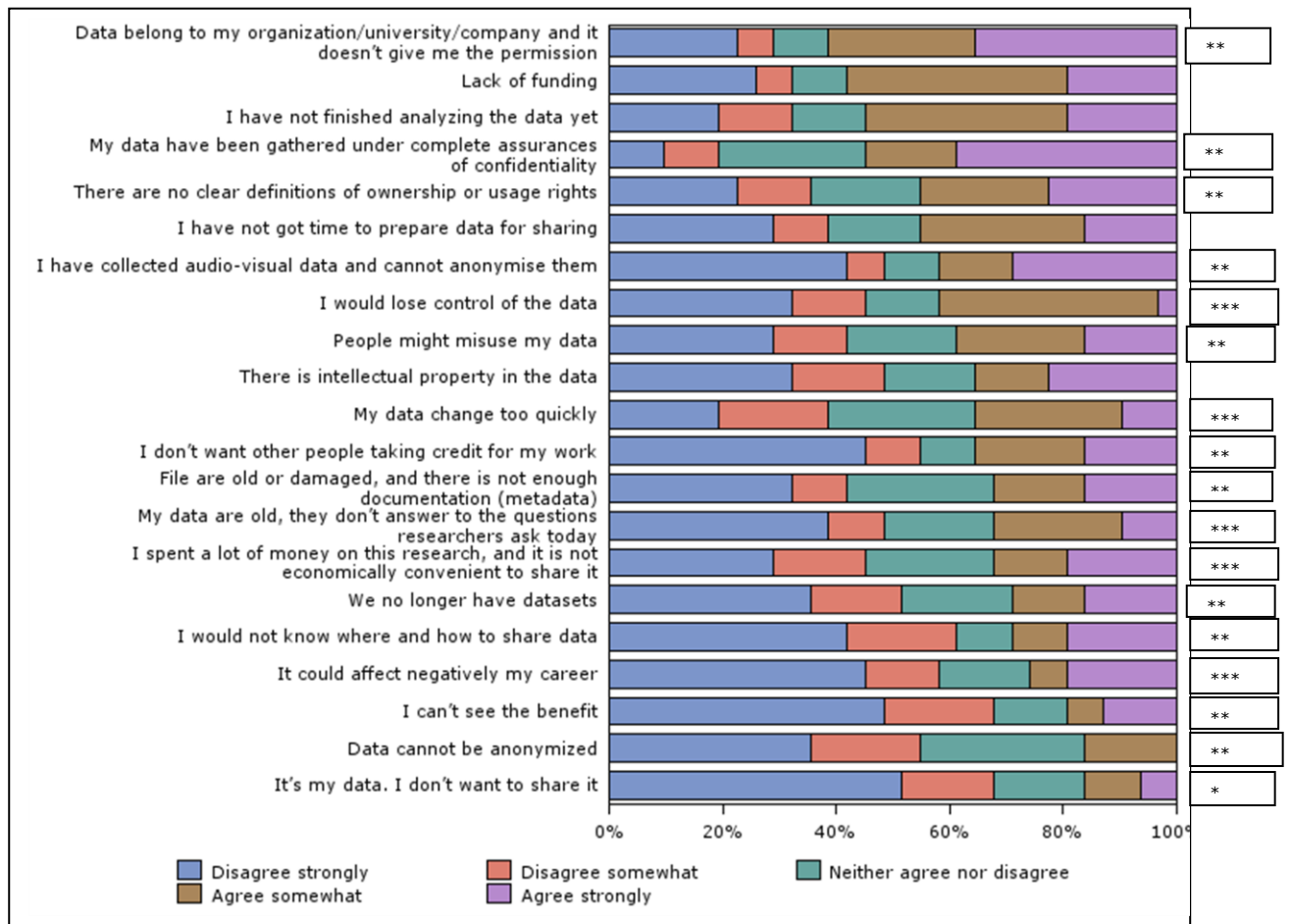
Details

In all distributions we note that Africans seem to have more problems in sharing data than not Africans: they fear more to lose control of the data, they think they have old or not affordable data, or think it is not convenient to share their data in terms of economic or personal values (career), as we can see in the following more significant tables. Also, in almost all other not reported less significant item tables this is true.
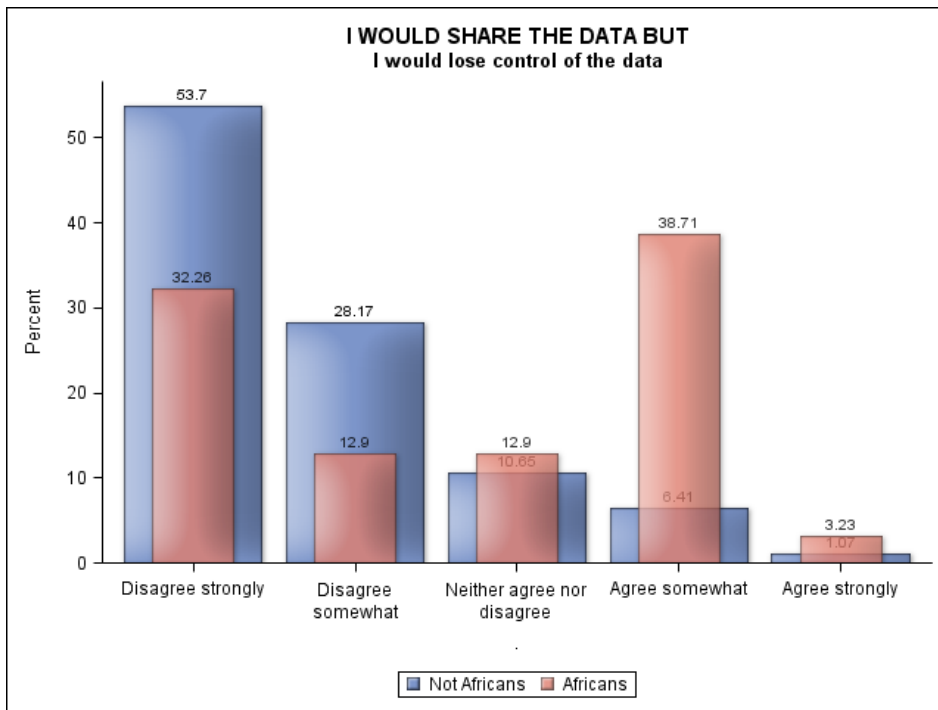
 Bonifacio, Flavio; Akullo, Winny Nekesa (2021) Differences in data-sharing attitudes and behaviours, extended version to African data curators and data management experts, IASSIST Quarterly 45(3-4), pp. 1-17. DOI: https://doi.org/10.29173/iq993

Fig. 18 – Significance (***)
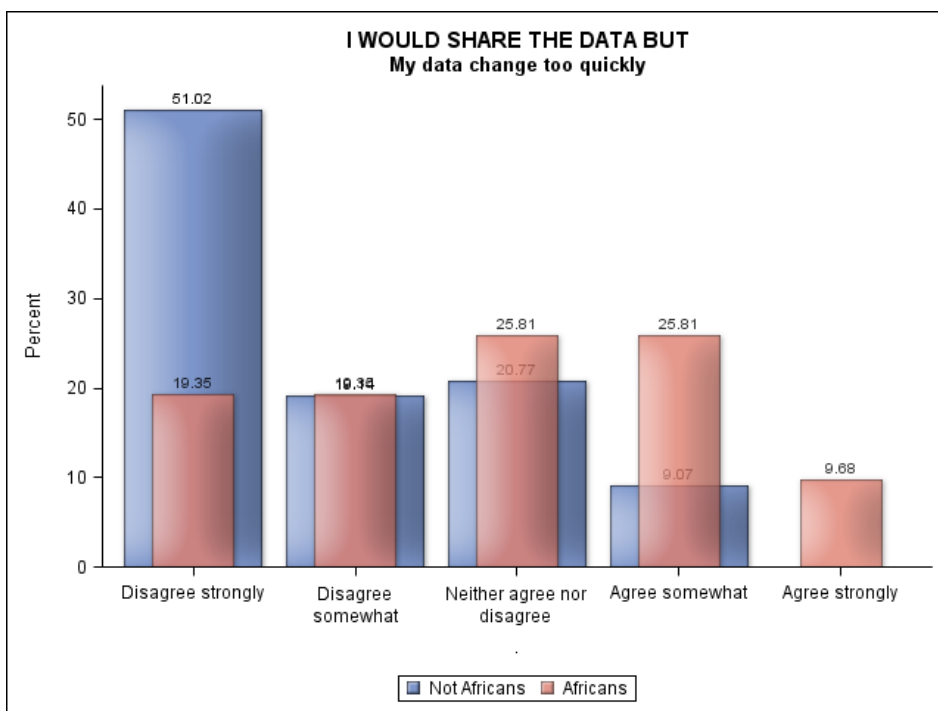

Fig – 19 Significance (***)

Fig – 20 Significance (***)
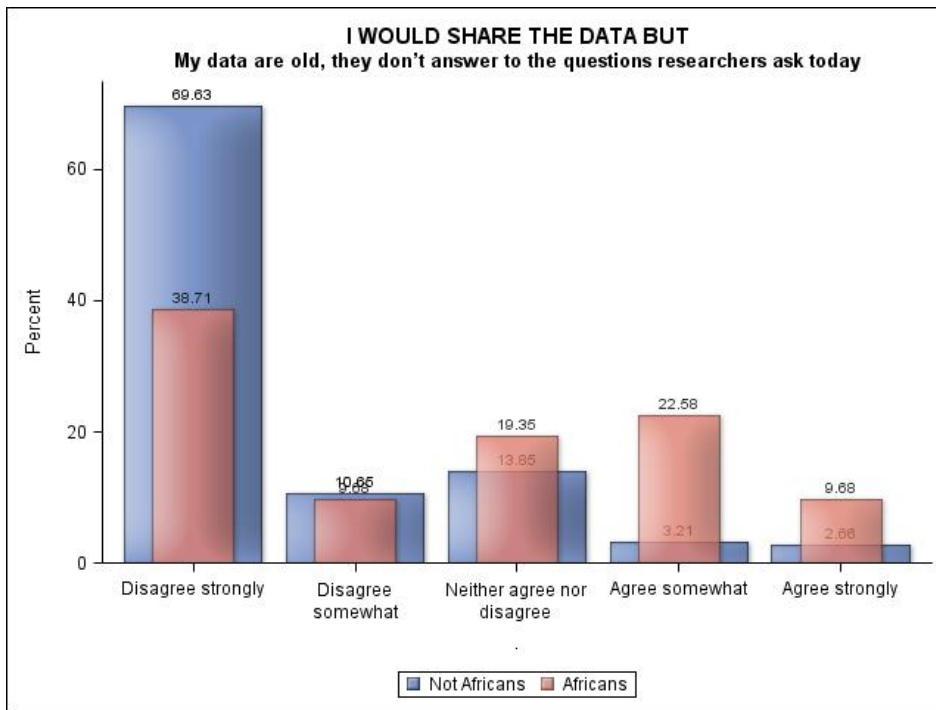
Fig – 21 Significance (***)

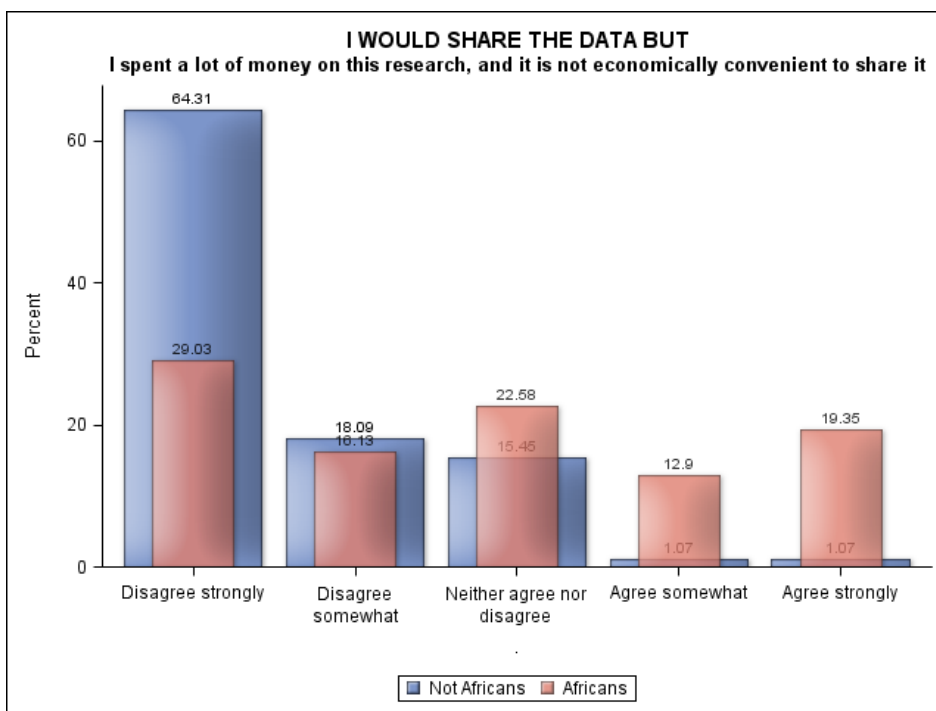*Fig – 22 Significance (***)*

Fig. 23 - Sharing Problems Importance Index Significance (**)
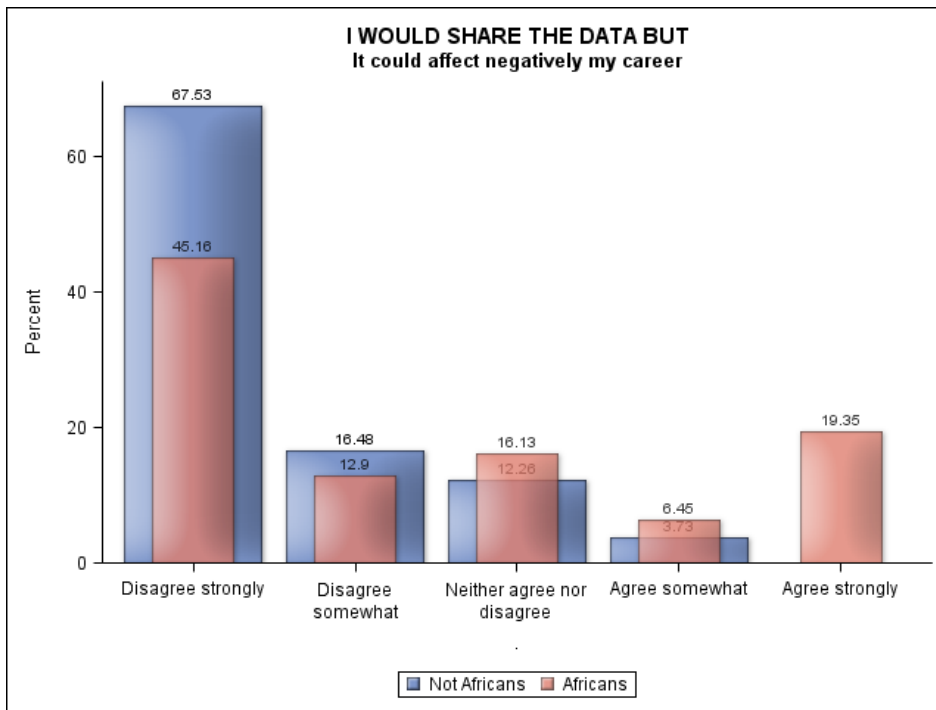
Sharing Problems Importance index

As we did before, we summarize the answer to the items in one single index, the Sharing *Problems Importance Index.* In this case also the synthetic index shows a significative difference (α=5%) between Africans and not Africans: Africans often recognise a high Problems Importance evaluation.

Fig. 23 - Sharing Problems Importance Index Significance (**)
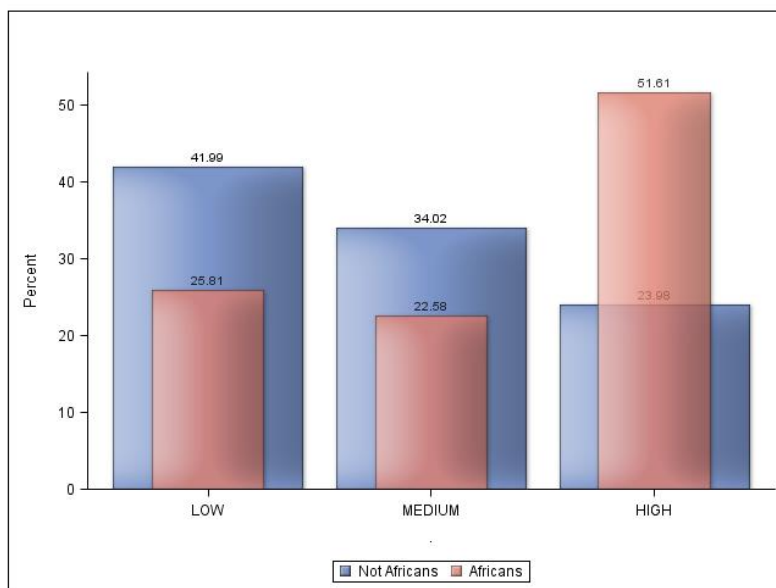
The model

Coding V3 (not Italy or Africa) as 1, otherwise coding V3 as 0, coding V4 (Africa) as 1, otherwise coding V4 as 0, coding V2 as 1 when the respondent works at University or in the Public Sector, otherwise coding V2 as 0, and taking Data Sharing Propensity Index as dependent Variable (V1) we obtain the model schema of the model reported below (direct and indirect effects). It seems to be the most appropriate and not reducible model. All the direct and indirect effects are significant at alpha=5%. The indirect effect of the Country of Origin (V3, V4) influences the Sharing Propensity via the Work Sector, which is positive. The total effect of V3 and V4 (Country) over V1 (Data Sharing Propensity) is given by 0.27+(0.25*0.27)+0.27+(0.38*0.27) that equals 0.71. In other words, in this sample Africans more often work at university or in the public sector than Not Africans and Not Italians that work more at University too (Not Africans and Not Italians means Northern Europa, Canada, USA, etc., respondents): they both add to their own higher sharing propensity also the fact that they work at University (or in the public sector). We indeed know that working at a university has its own positive effect on data sharing propensity too, as the Model shows.

Fig. 24 – Path model with sharing propensity as dependent variable (V1) and Work Sector (V2) and Country of Origin (V3, V4) as independent variables
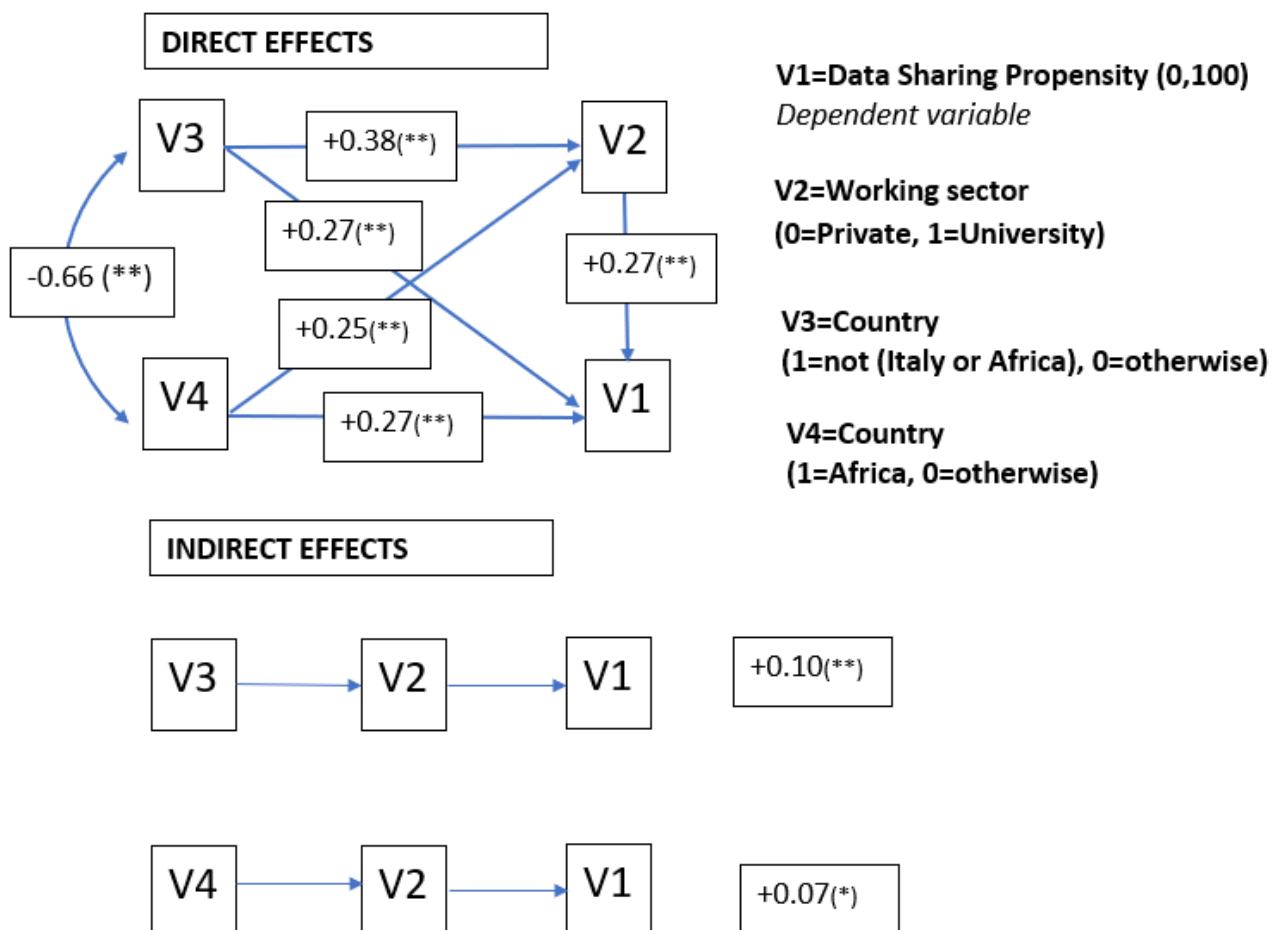
Conclusions

Africans experts are more likely to be male, are younger, with more scientific study, working more in university or public administration. They appear to be at the same time more attentive to problems that may arise in the data sharing activity and more satisfied. They seem also to have a higher propensity to share data.

In general, the distributions on the attitude items have not such a different shape if compared with the other "not African" experts. A warning comes from the work: Africans seem recognise more problems in sharing data, as said above. An effort may be necessary to make the issues of data sharing more attractive. Maybe it is only a feeling of mine but I observed more enthusiasm in the African sample. This will help in order to reach the goal.

Acknowledgements

References

Bonifacio, Flavio (2017) 'Working Across Boundaries – Public and Private Domains', Part 3 – A follow-up Survey, (Available at http://doi.org/10.5281/zenodo.1120237)

Doorn, Peter and Tjalsma, Heiko (2007) 'Introduction: archiving research data', Springer Science+Business Media B.V.

Hatcher, Larry (1994) 'SAS System Factor Analysis and Structural equation modelling', SAS Institute

Horton, Laurence (2016) 'LSE Research Data Management Data Sharing Objections FAQs and Noughts and Crosses Game', (Available at http://doi.org/10.5281/zenodo.61978)

Kim, Youngseek and M. Stanton, Jeffrey (2012) 'Institutional and Individual Influences on Scientists' Data Sharing Practices', Journal of Computational Science Education, Volume 3, Issue 1

Noble, Susan; Russel, Celia and Wiseman, Richard (2012) 'Mind the Gap: Global Data Sharing', IASSIST Quarterly, Vol. 35, n° 3

Qualtrics (2010) 'The 1936 Election – A Polling Catastrophe', Available at (https://www.qualtrics.com/blog/the-1936-election-a-polling-catastrophe)

Rasmussen, Karsten Boye (2014) 'Social Science Metadata and the Foundations of the DDI', IASSIST Quarterly, Vol. 37, n° 1

Ribeiro, Cristina and Matos Fernandes, Maria Eugenia (2012) 'Data Curation at U. Porto: Identifying current practices across disciplinary domains', IASSIST Quarterly, Vol. 35, n° 4

Tenopir, Carol; D. Dalton, Elizabeth; Allard, Suzie; Frame, Mike; Pjesivac, Ivanka; Birch, Ben; Pollock, Danielle and Dorsett, Kristina (2015) 'Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide', (Available at https://doi.org/10.1371/journal.pone.0134826)

Yang, Meng-Li (2013) 'Strategies of Promoting the Use of Survey Research Data Archive', IASSIST Quarterly, Vol. 36, n° 1

Endnotes

[1]Flavio Bonifacio, Metis Ricerche srl, Via Camerana 6, I-10128 Torino, Italy. Contact e-mail: flavio.bonifacio@metis-ricerche.it.

[2] Winny Nekesa Akullo, Public Procurement and Disposal of Public Assets Authority, Kampala, Uganda. Contact e-mail: winny.nekesa@yahoo.com

[3] Bonifacio, Flavio, Differences in Data-Sharing Attitudes and Behaviours, IASSIST Quarterly, Vol. 42 No. 3 (2018)

[4]Tenopir, Carol; D. Dalton, Elizabeth; Allard, Suzie; Frame, Mike; Pjesivac, Ivanka; Birch, Ben, et al. (2015)

[5]Kim, Youngseek and M. Stanton, Jeffrey (2012), Doorn, Peter and Tjalsma, Heiko, (2007), Noble, Susan; Russel, Celia and Wiseman, Richard, (2012), Ribeiro, Cristina ad Matos Fernandes, Maria Eugenia, (2012) Yang, Meng-Li, (2013), Rasmussen, Karsten Boye, (2014)

[6] We refer to the Annex 2: IMDSM2017, IASSIST Members Data Sharing Mail, FALL 2017. The mails reported suggestions about how to ask questions

[7] For the method used building the indexes (PCA Analysis) and for questions concerning the sample see the quoted Bonifacio, Flavio, Differences in…. While the subsample for Italians and (Not Africans and not Italians) have been weighted, the subsample of Africans is not weighted.