

# Data curation practices amongst WDCs

Paul Laughton  
Department of Information and Knowledge Management  
University of Johannesburg  
[pau11@uj.ac.za](mailto:pau11@uj.ac.za)

CODATA 2010  
25<sup>th</sup> October 2010  
Stellenbosch



## 1. Introduction

- Data curation, according to Lord and McDonald (2003:12) can be defined as the “activity of managing and promoting the use of data from its point of creation, to ensure it is fit for discovery and reuse”
- A data curator’s role as a data professional includes: management of data, adding of value to data, data sharing and data preservation for later use (Rusbridge, 2008)
- Data curation remains a contested debate as how best to go about and ensure the survival of data
- Multiple approaches to effectively curating data
- Constant innovation and research is necessary for the survival of valuable research data
- OAIS standard seems to be the most accepted and publicised



## 2. Background to research

- The WDC was the target population for the study
- Since research commenced the WDC has been incorporated into the WDS along with FAGS
- Seemed necessary to determine common practices amongst WDC members to ensure new members conduct similar practices
- Appropriate for Africa's proposed WDC for Biodiversity and Human Health
- Aim of the research highlight data curation practices amongst WDC
- The approach is not to create a right or wrong scenario for data curation
- OAIS framework was used as a platform for analysis



### 3. Research problem

- The preliminary investigate and literature review led me to the creation of my research problem:

***To what extent is it possible to develop a framework for the curation of data in World Data Centres (WDCs)?***

- The findings will be used to contribute to the development of a framework that may highlight practices of other data centres who are members of the WDC and newly formed WDS



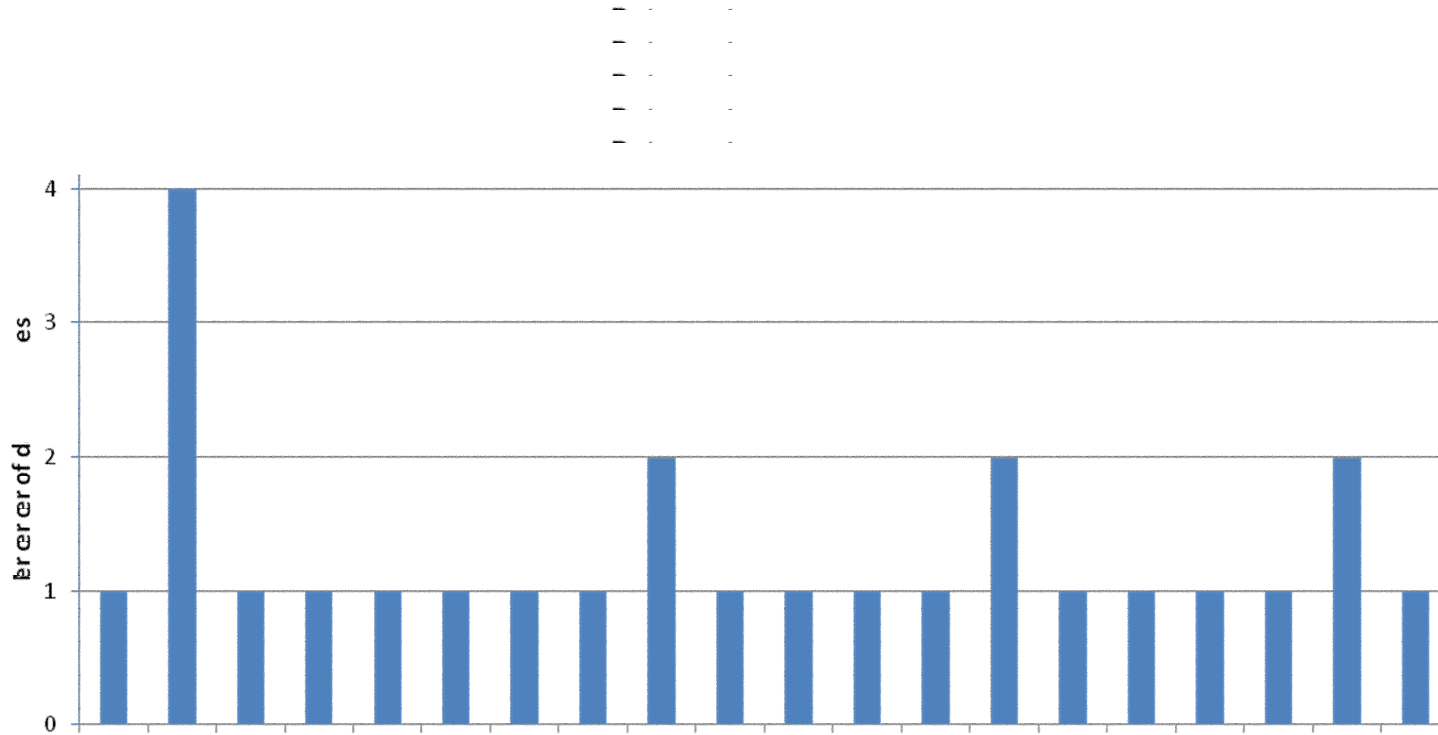
## 4. Research approach

- Mixed methods approach
- Combination of qualitative and quantitative data collection
- Each of these methods has their advantages and drawbacks
- The first round of data collection involved an online questionnaire (quantitative) which was distributed amongst the target population (WDCs)
- The software used to conduct the online questionnaires was Qualtrics ([www.qualtrics.com](http://www.qualtrics.com))
- Frequency analysis was conducted using SPSS
- A follow up of case studies needs to be conducted (qualitative)
- Cases have been selected based on the findings from the online questionnaires



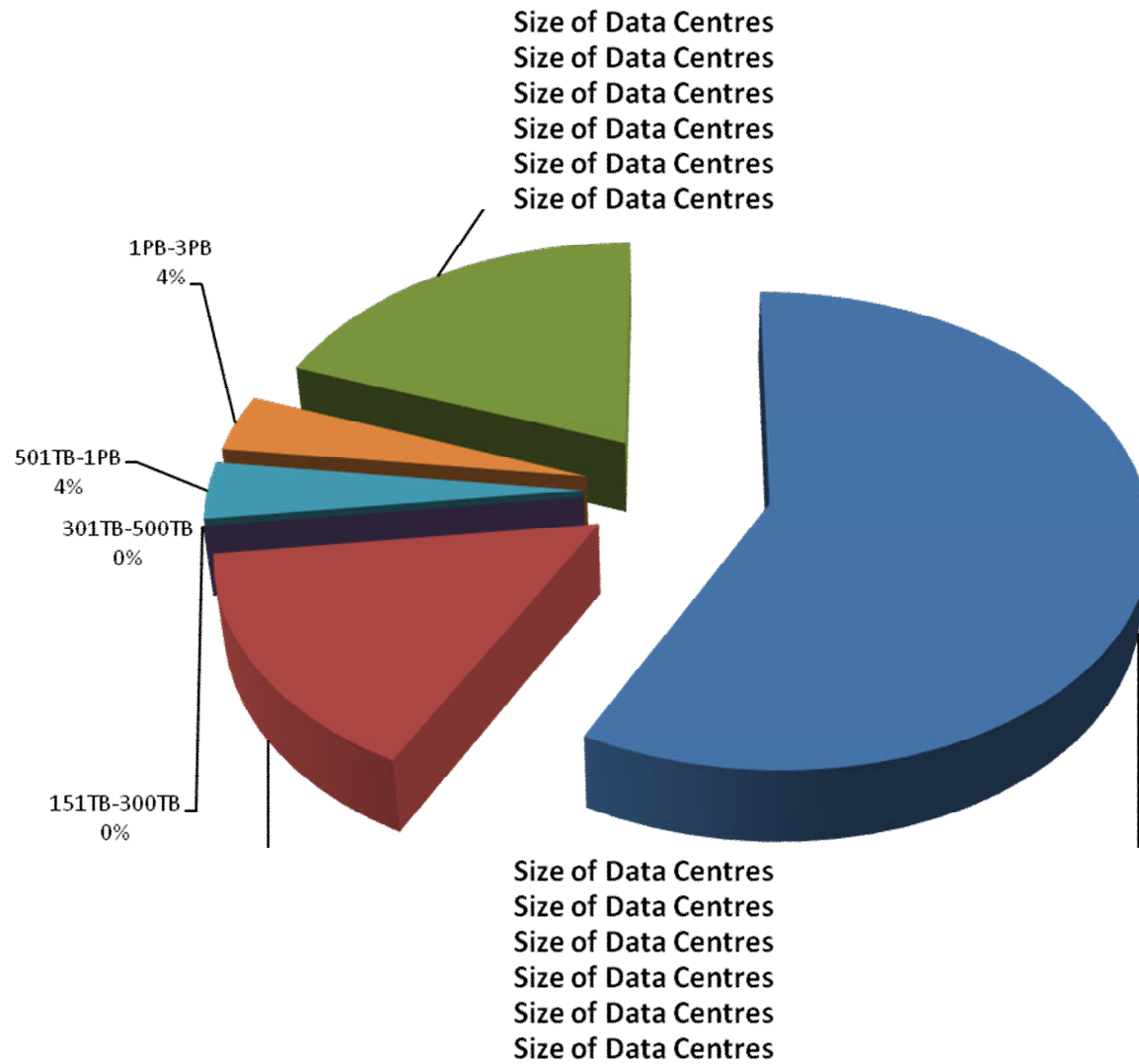
## 5. Preliminary Results

Responses

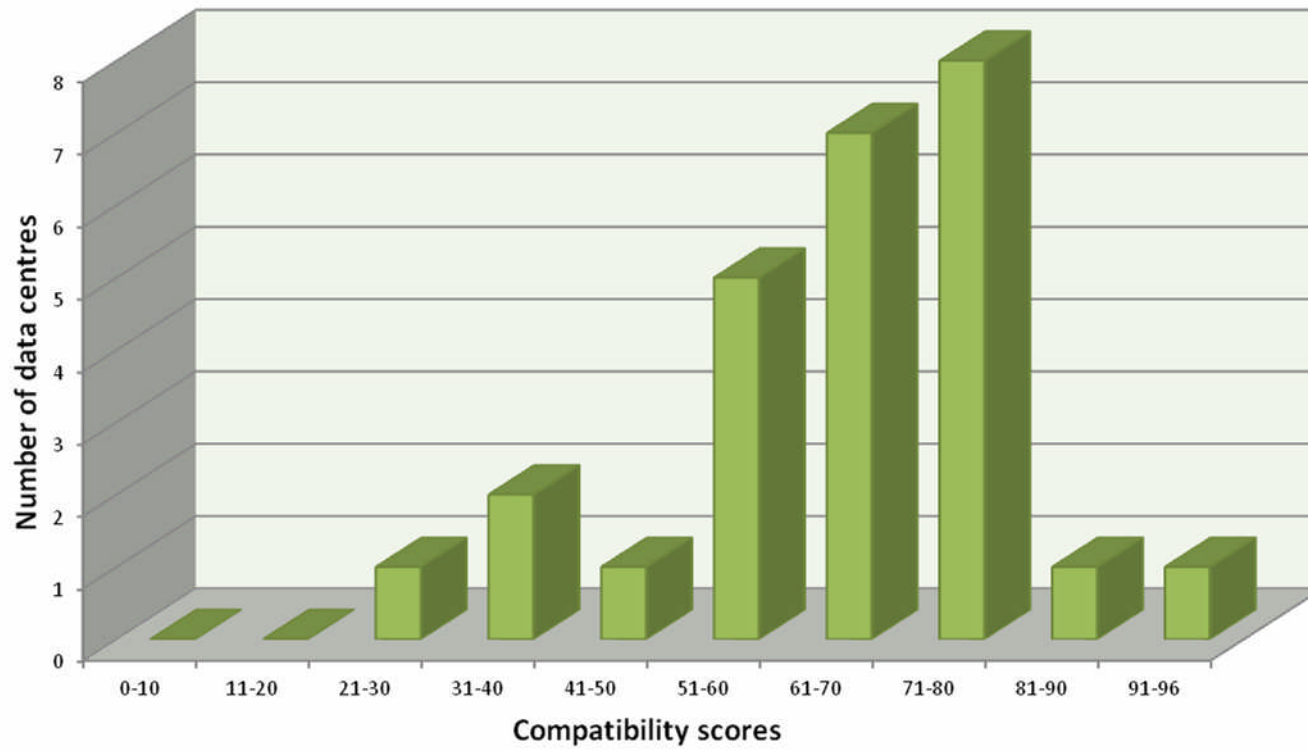




## 5. Preliminary Results



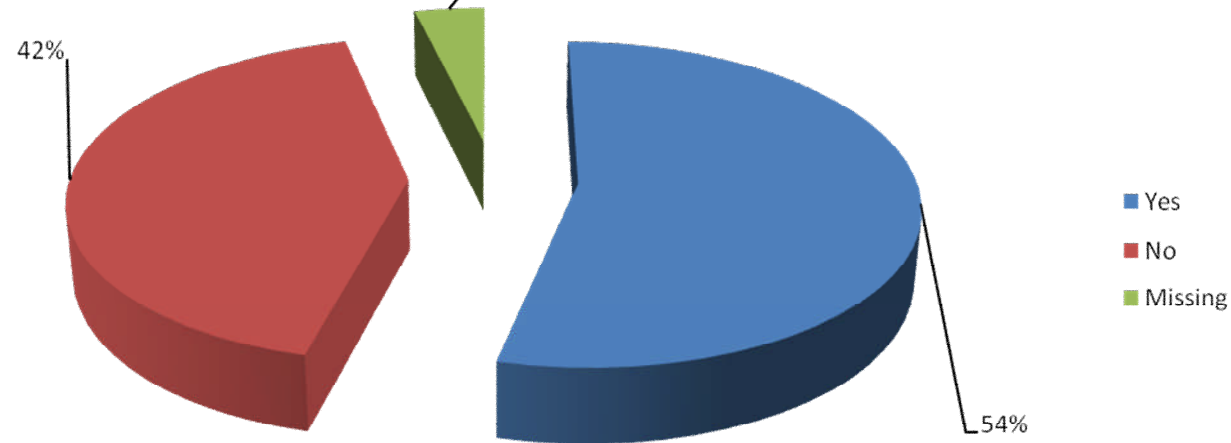
## 5. Preliminary Results





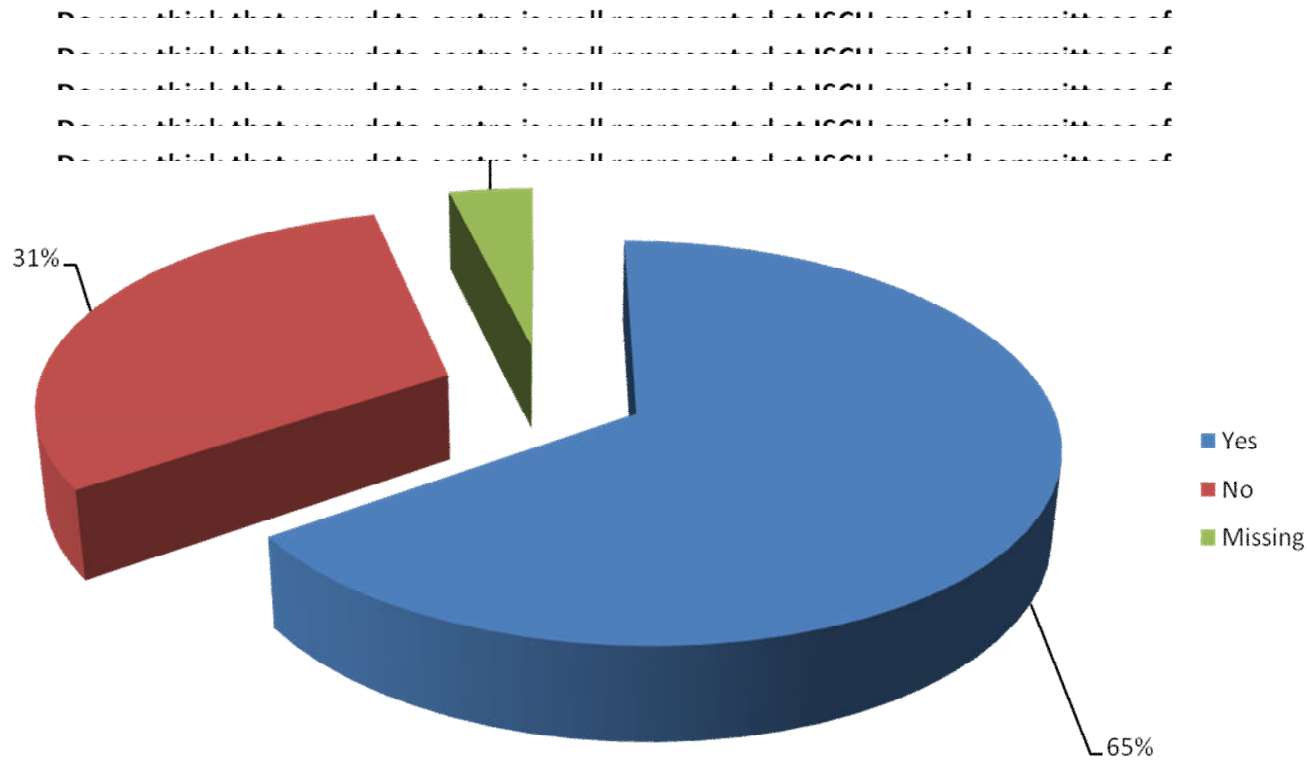
## 5. Preliminary Results

Do you feel your data centre gets enough support from ICSU and other  
Do you feel your data centre gets enough support from ICSU and other  
Do you feel your data centre gets enough support from ICSU and other  
Do you feel your data centre gets enough support from ICSU and other  
Do you feel your data centre gets enough support from ICSU and other



Do you feel your data centre gets enough support from ICSU and other  
Do you feel your data centre gets enough support from ICSU and other  
Do you feel your data centre gets enough support from ICSU and other

## 5. Preliminary Results



Do you think that your data center is well represented by ICGU essential committee of

Do you think that your data center is well represented by ICGU essential committee of

Do you think that your data center is well represented by ICGU essential committee of

## 6. Limitations of research

- Research is based on the former WDC (due to start date of research), results would differ if opened to the WDS
- Can be difficult to successfully implement a mixed method methodology
- Difficult to score data centres for OAI compatibility through a series of questions



## 7. Future research plans

- Next step would be to roll out the case studies on the identified data centres
- Study can be conducted over the wider WDS community
- OAIS compatibility test developed can be refined and tested on a more diverse selection of data centres



# Q & A

