

# Researchers' Perspective on the Publication of Research Data: Semi-structured Interviews from China

Interview: os\_007

|    |   |
|----|---|
| 1  | <b>Interviewer:</b> So first of all, thank you very much. That you are willing to do this interview with me. It is a pleasure to meet you. Erm... |
| 2  | <b>Researcher:</b> Nice to meet you.  |
| 3  | <b>Translator:</b> (unintelligible).  |
| 4  | <b>I:</b> So, the first question is// Please, in the beginning introduce yourself. Where you are coming from? And what your research is about.    |
| 5  | <b>T:</b> (unintelligible).   |
| 6  | <b>R:</b> Okay. I am speaking in Chinese. So//  |
| 7  | <b>I:</b> If somebody translates it's fine ((laugh))  |
| 8  | <b>R:</b> [NAME] is translator for me. Okay?  |
| 9  | <b>I:</b> Okay.   |
| 10 | <b>R:</b> (incomprehensible – Chinese)  |
| 11 | <b>T:</b> So, you first question is our introduction, is that correct?  |
| 12 | <b>I:</b> Introduction, exactly. With the name, your affiliate//affiliation. And what your research is about.                                     |
| 13 | <b>R:</b> (incomprehensible – Chinese)  |
| 14 | <b>T:</b> So, our centre is trying to using information technology. Trying to improve teaching and learning.                                      |
| 15 | <b>I:</b> Mhm ((affirmative sound)).  |
| 16 | <b>R:</b> (incomprehensible – Chinese)  |
| 17 | <b>T:</b> And, so. We currently use big data try to improve our... erm, our teaching and learning in education field.                             |
| 18 | <b>R:</b> (incomprehensible – Chinese)  |



|    |  |
|----|--|
| 19 | <b>T:</b> And so, our//our m//main research action is educational Big Data.  |
| 20 | <b>R:</b> (incomprehensible – Chinese)   |
| 21 | <b>T:</b> Yeah, we are, so... we are also very interested in the easier what you said about the... data, data publication and open, try to use a... and to make the data Open Access to other researchers around the world.  |
| 22 | <b>R:</b> (incomprehensible – Chinese)   |
| 23 | <b>T:</b> So currently in China, we launched a research group, that outlooks on the education and data collection and the, and the easier (unintelligible) data. And also, about the data quality. Try to get very high-quality data.  |
| 24 | <b>I:</b> Thank you very much. And how long have you been working in science already?  |
| 25 | <b>R:</b> (incomprehensible – Chinese)   |
| 26 | <b>T:</b> Er, I think more than ten years.   |
| 27 | <b>I:</b> Mhm ((affirmative sound)). Okay, and what research data exactly are you working with?  |
| 28 | <b>R:</b> (incomprehensible – Chinese)   |
| 29 | <b>T:</b> Have you got the, the Excel files?   |
| 30 | <b>I:</b> Yes.   |
| 31 | <b>T:</b> I think I send you by e-mail.  |
| 32 | <b>I:</b> Yes.   |
| 33 | <b>T:</b> So, yes, in that file we have our daily (unintelligible) among data that we are using currently.   |
| 34 | <b>I:</b> But I cannot. Erm. I am sorry, 'cause I got the table but I cannot really work with it as I need the interviews where I evaluate it. So, it was very nice for me to see it, but I really need this... exactly this information now in the interview as well. So... You can talk about this data, it is perfectly fine. I just need to have it recorded, that's the thing. I am sorry if I am asking the same question. |
| 35 | <b>R:</b> (incomprehensible – Chinese)   |
| 36 | <b>T:</b> Erm, so currently the data we are using all included three different parts.  |
| 37 | <b>R:</b> (incomprehensible – Chinese)   |
| 38 | <b>T:</b> And so, the first you have all the data is about the management data. So, we try to using this data to improve our educational experience.   |



|    |  |
|----|--|
| 39 | <b>R:</b> (incomprehensible – Chinese)   |
| 40 | <b>T:</b> So, this field is not just a interesting issue to research at university. But also, very interesting to the government (unintelligible)                            |
| 41 | <b>R:</b> (incomprehensible – Chinese)   |
| 42 | <b>T:</b> So, the second kind of data is the learning data.  |
| 43 | <b>R:</b> (incomprehensible – Chinese)   |
| 44 | <b>T:</b> So, the first type of this data is... we try to apply to learning the data.  |
| 45 | <b>R:</b> (incomprehensible – Chinese)   |
| 46 | <b>T:</b> So, and, so... (unintelligible) of learning an//ana//ana//analytics. So, we try to include in teaching.  |
| 47 | <b>R:</b> And learning.  |
| 48 | <b>T:</b> And, and learning.   |
| 49 | <b>R:</b> Yeah.  |
| 50 | <b>T:</b> Yeah.  |
| 51 | <b>R:</b> (incomprehensible – Chinese)   |
| 52 | <b>T:</b> So, the second kind is... we try to apply, we try to apply the technology into education field. So, we do the empirical research.                                  |
| 53 | <b>R:</b> (incomprehensible – Chinese)   |
| 54 | <b>T:</b> And, so. We try to validate why the//the new technology can improve learning and teaching. In our education field. And so...                                       |
| 55 | <b>R:</b> (incomprehensible – Chinese)   |
| 56 | <b>T:</b> And also the fact of (unintelligible).   |
| 57 | <b>R:</b> (incomprehensible – Chinese)   |
| 58 | <b>T:</b> And, so when data set is from is our//our classroom. So, in China we cover with math//math classroom. So, we collect the data from classroom by different devices. |
| 59 | <b>R:</b> (incomprehensible – Chinese)   |
| 60 | <b>T:</b> And, so... and the secondary set is from a//is from an online learning platform. It's like MOOC, our another online platform.                                      |



|    |   |
|----|---|
| 61 | <b>R:</b> (incomprehensible – Chinese)  |
| 62 | <b>T:</b> And... The set data is from//is from erm... for example... our WIFI data on campus and also student marker data.  |
| 63 | <b>R:</b> (incomprehensible – Chinese)  |
| 64 | <b>T:</b> Eh, so also learn we//we//we//we also do some research with the special education.  |
| 65 | <b>R:</b> (incomprehensible – Chinese)  |
| 66 | ((interruption by another person))  |
| 67 | <b>I:</b> Okay. So, you have to repeat the last part. Sorry.  |
| 68 | <b>R:</b> (incomprehensible – Chinese)  |
| 69 | <b>T:</b> And, so there is a group in our centre, they work on the//on special education. So, in that group we try to use learning technology to improve the education and for the special case. Especially for the children with [a disease] |
| 70 | <b>R:</b> (incomprehensible – Chinese)  |
| 71 | <b>T:</b> So, it is very important that you apply the data, (unintelligible) data and also do some interventions for those special case.  |
| 72 | <b>R:</b> Okay. That's all.   |
| 73 | <b>I:</b> Okay, perfect. And have you already published your research data?   |
| 74 | <b>R:</b> (incomprehensible – Chinese)  |
| 75 | <b>T:</b> We are a bit to... we are planning to... we are very interested in the data publication. But for the three different types of data. I think we are facing different challenges.   |
| 76 | <b>R:</b> (incomprehensible – Chinese)  |
| 77 | <b>T:</b> And for example, for the first data, the management data. Eh... in China, we try to promote to the data publica//publication for this data. But the problem for this data is this kind of data is (unintelligible).                 |
| 78 | <b>R:</b> (incomprehensible – Chinese)  |
| 79 | <b>T:</b> And so, for this data. We think that it is a little easier to publish on for researching.   |
| 80 | <b>R:</b> (incomprehensible – Chinese)  |
| 81 | <b>T:</b> And so, for the second type of data, our learning data, is facing two different problem.  |



|     |  |
|-----|--|
| 82  | <b>R:</b> (incomprehensible – Chinese)   |
| 83  | <b>T:</b> And... so we call the//this dat//this very spectrum of data is from our... our educational scenario. So it is... The data sets... is a bit internalized and (unintelligible).  |
| 84  | <b>R:</b> (incomprehensible – Chinese)   |
| 85  | <b>T:</b> And so, this kind of data is not very rare of (unintelligible).  |
| 86  | <b>R:</b> (incomprehensible – Chinese)   |
| 87  | <b>T:</b> And, so we think, a lot of work needs to be done before this kind of data can be published.  |
| 88  | <b>R:</b> (incomprehensible – Chinese)   |
| 89  | <b>T:</b> And so, the second problem to this type of data is ex-coalition. Also copyright.   |
| 90  | <b>R:</b> (incomprehensible – Chinese)   |
| 91  | <b>T:</b> And then for our learning behaviour data may relate to the privacy protection. For both teacher and the student.   |
| 92  | <b>R:</b> (incomprehensible – Chinese)   |
| 93  | <b>T:</b> And also, for example students learning behaviour data. Purely their own (unintelligible) by different//different organization.  |
| 94  | <b>R:</b> (incomprehensible – Chinese)   |
| 95  | <b>T:</b> We know, in Europe, especially from Germany.   |
| 96  | <b>R:</b> (incomprehensible – Chinese)   |
| 97  | <b>T:</b> We know Germany had a petition to the privacy property.  |
| 98  | <b>R:</b> (incomprehensible – Chinese)   |
| 99  | <b>T:</b> Currently, yes is in China we also petition to the privacy protection.   |
| 100 | <b>R:</b> (incomprehensible – Chinese)   |
| 101 | <b>T:</b> And, yeah so this... so this area of (unintelligible) data, it's difficult to publish this kind of data.   |
| 102 | <b>R:</b> (incomprehensible – Chinese)   |
| 103 | <b>T:</b> And so, to publish this learning behaviour data we see the true problem, is to be, is to be (unintelligible). So, the one simple, we need to do the same (unintelligible). To protect the pri... the privacy. So, the second is we need to get some//some standard and strage//strategy. |



|     |  |
|-----|--|
| 104 | <b>R:</b> (incomprehensible – Chinese)   |
| 105 | <b>T:</b> Erm, and generally speaking, we are very interested in the idea about publish this kind of data.   |
| 106 | <b>R:</b> (incomprehensible – Chinese)   |
| 107 | <b>T:</b> If we can promote the data publication in this field, we are pretty sure that research in this field can be... can be improved.  |
| 108 | <b>R:</b> (incomprehensible – Chinese)   |
| 109 | <b>T:</b> And so we think data in China, data in our land can have advantages.   |
| 110 | <b>R:</b> (incomprehensible – Chinese)   |
| 111 | <b>T:</b> So... we have a... you and other Germany researchers that we can help us to publish our data. Our data already gave us some very good ideas or status about how to do it correctly and properly.   |
| 112 | <b>R:</b> Thank you.   |
| 113 | <b>T:</b> Thank you.   |
| 114 | <b>I:</b> Thank you. Erm... I have few more questions. Erm... So, you said you were working or dealing with personal data as well. Do you know how you can make this te//type of data publicly available to others? Like what you need to do to make this... possible? |
| 115 | <b>T:</b> So when you say the personal data, do you mean is like your name, your age? Your//   |
| 116 | <b>I:</b> Yes.   |
| 117 | <b>T:</b> Gender?  |
| 118 | <b>R:</b> (incomprehensible – Chinese)   |
| 119 | <b>T:</b> Erm, so data publication is new of, is new of happening new here in China. So and, so and... to the best of our knowledge, so, in education field in China we have not have published such data currently.   |
| 120 | <b>R:</b> (incomprehensible – Chinese)   |
| 121 | <b>T:</b> Currently in China the open access data from education field it's more profound of managing data, rather than just the personalised data. (unintelligible) the personal information is marketed (unintelligible) through the open access data act.           |
| 122 | <b>R:</b> (incomprehensible – Chinese)   |
| 123 | <b>T:</b> So, in our own research and researching, so, we dealing with like a ID encryption, so we   |



|     |   |
|-----|---|
|     | translate to a different name//of students name into some digital number or... or it is just to try protect their privacy.  |
| 124 | <b>R:</b> (incomprehensible – Chinese)  |
| 125 | <b>T:</b> So, we think in the future if we were to publish, this kind of data, erm, if we are going to... collaborate with//with... German researcher. So, think of thing//thing//things (unintelligible) can be used. So (unintelligible) like//like//like of what we said previously. We don't like ID encryption. I think wasting of this kind solve this problem. |
| 126 | <b>R:</b> (incomprehensible – Chinese)  |
| 127 | <b>T:</b> If we are going to publish to... this data, I think we need to ask them to get some permission from our university. Also, we need to have some professional (unintelligible) or a better way to do it.  |
| 128 | <b>R:</b> (incomprehensible – Chinese)  |
| 129 | <b>T:</b> If the data pub//pub//publication is just//just for research use, and it//so it will be a little bit easier. So...  |
| 130 | <b>I:</b> ((affirmative sound))   |
| 131 | <b>R:</b> Okay.   |
| 132 | <b>T:</b> Maybe you can just...   |
| 133 | <b>R:</b> (incomprehensible – Chinese)  |
| 134 | <b>T:</b> Maybe you can send us an agreement for the (unintelligible) then we can share the data.   |
| 135 | <b>R:</b> (incomprehensible – Chinese)  |
| 136 | <b>T:</b> Yeah, because I, we think the public//publication is different with the data//data sharing. Publication for... our research data, if we get to do collaboration with the... with foreign research group (unintelligible), it would be much easier.  |
| 137 | <b>R:</b> (incomprehensible – Chinese)  |
| 138 | <b>T:</b> It is actually, it is related to the learn//learning//learner.  |
| 139 | <b>I:</b> ((affirmative sound)). Hm. Just one more question. Erm... Is the research data more or less published, in your discipline of course, in other countries than in China? More or less already heard it, but wanted to hear your opinion on it. And if you have any examples? Or do you think//what do you think the reason is. Alright?                       |
| 140 | <b>R:</b> (incomprehensible – Chinese)  |



|     |   |
|-----|---|
| 141 | <b>T:</b> In China previously we have//we have... some open access data set in the//in the education field. It's for some research fields.  |
| 142 | <b>R:</b> (incomprehensible – Chinese)  |
| 143 | <b>T:</b> So, for example, in China we have some in our... our data sets for example, (unintelligible), acoustic data set.  |
| 144 | <b>R:</b> (incomprehensible – Chinese)  |
| 145 | <b>T:</b> And in our education field, so, you and we can also use these common data set.  |
| 146 | <b>R:</b> (incomprehensible – Chinese)  |
| 147 | <b>T:</b> In China we have a (unintelligible) to publish (unintelligible) for data.   |
| 148 | <b>R:</b> (incomprehensible – Chinese)  |
| 149 | <b>T:</b> So, better in China we think, so the open access data for education field doesn't matter as much as the common data set.  |
| 150 | <b>R:</b> (incomprehensible – Chinese)  |
| 151 | <b>T:</b> So, the first reason//reason for this//for this... situation, we think is because of educational field. So most, so most research is in some special field. So, we do have a common data set or a common research interest to (unintelligible) a common data set like a metadata set. |
| 152 | <b>R:</b> (incomprehensible – Chinese)  |
| 153 | <b>T:</b> The second, the second reason (unintelligible) called education field. So, a lot of data are (unintelligible) through the learner. So, we are very careful to protect the privacy for those learners.   |
| 154 | <b>R:</b> (incomprehensible – Chinese)  |
| 155 | <b>T:</b> So therefore, erm... over, we have a lot of data//data from scientific field. But it is a little difficult for us to publish the raw data.  |
| 156 | <b>R:</b> Yes.  |
| 157 | <b>I:</b> Good. Thank you very much, that was my last question, so thanks a lot. It was very interesting  |
| 158 | <b>R:</b> Thank you. Thank you very much.   |
| 159 | <b>I:</b> Thank you very much.  |

