



# Marine Microbial Biodiversity, Bioinformatics & Biotechnology



Grant agreement n°287589

Acronym: Micro B3

Start date of project: 01/01/2012, funded for 48 months

## Deliverable 8.4

### Report on the IP model agreements for pre-competitive access to microbial genomic research databases

Version: Final

Expected Submission Date: 31.12.2013

Actual submission Date: 20.02.2014

Lead Party for Deliverable: UCL, Belgium, Partner #23

Mail: [tom.dedeurwaerdere@uclouvain.be](mailto:tom.dedeurwaerdere@uclouvain.be)

Tel.: +32 10 478 882

#### Dissemination level:

Public (PU)	X
Restricted to other programme participants (including the Commission Services) (PP)	
Restricted to a group specified by the consortium (including the Commission Services) (RE)	
Confidential, only for members of the consortium (including the Commission Services) (CO)	



The Micro B3 project is funded from the European Union's Seventh Framework Programme ( Joint Call OCEAN.2011-2: Marine microbial diversity – new insights into marine ecosystems functioning and its biotechnological potential) under the grant agreement no 287589. The Micro B3 project is solely responsible for this publication. It does not represent the opinion of the EU. The EU is not responsible for any use that might be made of data appearing herein.





Version	Author	Date	Status	Reviewer
V.1	Arul George Scaria, Arianna Broggiato and Tom Dedeurwaerdere	01.02.2014	Draft	Tom Dedeurwaerdere [Contributions from WP8 members through several WP8 Skype meetings and emails]
V.2	Arul George Scaria, Arianna Broggiato and Tom Dedeurwaerdere	20.02.2014	Final	

## Summary

The primary objective of Deliverable 8.4 is to report on the IP model agreements for pre-competitive access to microbial genomic research databases, based on the responses collected from different stakeholders. The deliverable provides novel and surprising insights from diverse stakeholders in the field, particularly on the current data exchange practices within the genomic research community and the different influencing aspects of genomic data sharing. Some of the important findings include –

- i. High prevalence of data sharing among the genomic research community;
- ii. Importance attached to attribution by the genomic research community;
- iii. Perceptions regarding the benefits received from data sharing;
- iv. Motivations to share data; and
- v. Factors dissuading researchers from sharing data.

Those insights are highly important not only in the optimisation of the IP model agreements in this area, but also in the formulation of genomic data sharing policies at the national, regional and global levels.



## Contents

Introduction .....	4
Methodology .....	6
Major findings from the study.....	8
3.1 Insights on current data sharing practices .....	8
3.2 Effectiveness of the existing database infrastructures in data sharing .....	10
3.3 Perceptions of the benefits received from data sharing.....	12
3.4 Importance of providing attribution .....	14
3.5 Importance of metadata .....	15
3.6 Policies and regulations on data sharing.....	16
3.7 Motivations to contribute data .....	20
3.8 Factors that dissuade researchers from sharing data .....	22
3.9 Factors that can create more willingness to share data .....	24
3.10 Some of the specific implications of the findings for the model agreements .....	25
Conclusion and future directions.....	27
Appendix 1: Questionnaire used for the semi-structured interviews.....	30
Appendix 2: Detailed descriptive statistics.....	43



## Introduction

One of the important contributions made by the Micro B3 project through Deliverable 8.2, submitted last year, was the detailed illustration of the legal and policy framework governing the area of large scale genomic data sharing. Deliverable 8.2 had also provided some model contractual clauses for access to large scale genomic databases. Those model contractual clauses were not only fulfilling the necessary legal and policy conditions outlined in the deliverable, but was also combining the best licensing approaches existing in different organisations.

As one may recall from Deliverable 8.2, the three core aspects of the model contractual clauses developed under the Micro B3 Project for pre-competitive access to microbial genomic research databases were ensuring - 1. appropriate rules for attribution; 2. data interoperability and 3. automated data integration. As mentioned in the conclusion and future directions section of Deliverable 8.2 and also the description of Deliverable 8.4 in the Micro B3 project proposal, the next logical step in the work was to seek the views of different stakeholders on the model IP model agreements for pre-competitive access to microbial genomic research databases, particularly through semi-structured interviews. Such inputs from stakeholders are highly important to get a better picture of the actual data related transactions within the community and also to identify the conditions generally associated with those transactions. Only through such a participatory research approach, one can bring forward the diverging perspectives of various stakeholders on different aspects of data sharing and thereby optimise the model contracts in the area.



As envisaged, we have undertaken a detailed study to explore the perspectives of the diverse stakeholders in the area of genomic data sharing. The innovative participatory research approach taken by the Micro B3 project has been highly helpful in bringing forward the voices of the different stakeholders in this area and also to get a better understanding of the actual data related transactions within the genomic research community. Instead of asking their direct opinions on the model agreements, what we have attempted is to seek their opinion on the core aspects included in the model agreement. This not only helped in avoiding biased responses of the subjects, but also helped in identifying areas where we could further optimise the IP model agreements. We firmly hope that this exploratory approach will also help in guiding the evolution of data policies at the organisational level as well as at national, regional and international levels.

This deliverable report will highlight in detail the major outcomes of the responses collected from different stakeholders in the area on all the important issues covered in our study. This report is divided into four different sections. This brief introduction is followed by section 2, which explains in detail the methodology used for collecting the perspectives of the stakeholders. As this was an exploratory study in this field, replications of this study are highly necessary and the detailed discussion on the methodology will enable easier replications of this study by other researchers. Section 3 provides the major findings from the study and they are provided under different sub-headings for easier navigation through the results. Section 4 highlights the major conclusions from the report and also outlines some of the limitations of the study and possible future steps. Appendix 1 provides the



questionnaire used for the study and this not only provides more transparency for this study, but will also enable easier replications of this exploratory research work by other researchers. Appendix 2 provides detailed descriptive statistics in graphical formats.

## Methodology

As one may notice from the Micro B3 project proposal and the Description of Work (DoW) document, the methodology proposed for collecting the perspectives of the different stakeholders was to conduct semi-structured interviews with the stakeholders. The semi-structured interview methodology was suggested in those documents to make sure that important and contextual insights of diverse stakeholders in this area could be captured in the most optimal way.

We began by drafting a semi-structured questionnaire with this objective, based on extensive literature review and discussions with different experts working in the area of genomics as well as questionnaire designing. The first draft of the questionnaire was completed in April 2013. Special care was taken during the questionnaire drafting to avoid social desirability bias in responses, as studies like the present one on sharing data can be highly susceptible to biased responses from the respondents. To get a more open and unbiased approach to the questions, we explicitly assured the participants that “[a]ny personal information collected during the interview will be used only for data treatment purposes and will not be published or disseminated, except in an aggregated form”.<sup>1</sup> The questionnaire was subjected to intense revisions based on discussions with experts

---

<sup>1</sup> See the covering letter of the questionnaire provided in Appendix 1. For reasons of protecting the confidentiality of respondents, we have also avoided the names of the interviewees in this report and we have only referred to them with the date of their interviews. We would like to thank all the respondents for sharing their valuable time with us for this study.



including Prof. Matthew Bietz (University of California, Irvine).<sup>2</sup> After several rounds of revisions, the final version of the questionnaire was ready by November 2013 and we used the surveymonkey platform for collecting the responses.<sup>3</sup>

The responses from the stakeholders were collected during the month of December 2013. As genomic research is a comparatively new and highly specialised area, we had to adopt a non-random sampling approach in the study to reach the target respondents. The sampling approach taken was purposive sampling, which broadly refers to a non-random sampling approach wherein the researcher chooses the appropriate respondents based on the requirements and characteristics of the study. Purposive sampling approach is particularly recommended in cases wherein there are only very limited number of people who can provide valuable information on the subject matter of research. As the aim of our study was to get the perspectives of the stake holders from the genomic research community, there was a very specific need to ensure that responses were from respondents who have expertise in the area of genomic research. The purposive sampling approach was a highly useful tool in this regard.

With the purposive sampling approach, we were able to collect responses from a total of 21 people, from 19 different organisations, located in 11 different countries. Responses from 8 of them were collected through Skype/ telephone interviews and the responses from the remaining respondents were collected through an online survey instrument created in this regard, due to their time constraints and preference to answer the questions according to their time convenience. In order to ensure that we did not lose any valuable contextual

---

<sup>2</sup> Discussion with Prof. Matthew Beitz on 26 April 2013 (Washington D.C.).

<sup>3</sup> For the copy of the final version of the questionnaire used in the study, see Appendix 1.



information even when respondents were responding through the online questionnaire, we provided them an option to directly email their additional remarks on the questions to one of the authors of this study.

For reasons of protecting the confidentiality guarantee given to the respondents, we have avoided the names of the interviewees inside this report and we have only referred to them with the date of their interviews. We remain thankful to them for the valuable time shared by them for this study.

## Major findings from the study

### 3.1 Insights on current data sharing practices

One of the most important aspects observed from our study is the high prevalence of data sharing among the genomic research community. As high as 94% of the respondents are found to be contributing data at least once a year to Genbank or other similar public online genomic sequence databases.<sup>4</sup> Similarly, 94% of the respondents are also found to be sharing data with colleagues at least once a year.<sup>5</sup> An interesting factor highlighted during the interviews was that most journals in this field do not insist on submitting data to the publishers as part of manuscript submissions and they generally only mandate submission of data to a public depository.<sup>6</sup> This might have also been contributing to the high prevalence of data contribution to public depositories in the field.

---

<sup>4</sup> N=18. See also question 6 in Appendix 1.

<sup>5</sup> N=18. See also question 6 in Appendix 1.

<sup>6</sup> Interviews dated 9 December 2013 and 17 December 2013.



Another important aspect explored was the time point at which researchers shared their data. With regard to sharing of data to GENBANK and similar public online databases, it is found that 82% of the respondents are submitting data at the time of publication of a paper based on that data.<sup>7</sup> On the other hand, when it comes to sharing of data with colleagues, 69% of the respondents are found to be sharing data as soon as they generate the data.<sup>8</sup> Interestingly, interviews with industry representatives indicated that they generally share data only at the time of submission of a related patent application and they share data with colleagues only when the collaboration requires it.<sup>9</sup>

We also tried to explore the type of data being sharing by the researchers.<sup>10</sup> Interestingly, while sharing data with online public depositories like Genbank, 56% of the respondents shared the core data set plus some additional data such as related publications, geographical origin and identity of the data contributors.<sup>11</sup> Only 25% of the respondents are seen sharing the complete original datasets, which might allow other researchers to do deeper analyses of the data.<sup>12</sup> 19% of the respondents shared only the minimal data set required to understand their research/the minimal data set mandated by the data repository.<sup>13</sup>

On the other hand, while sharing data with colleagues, 69% of the respondents shared the complete original dataset and only 8% of the respondents restricted sharing to the minimal

---

<sup>7</sup> N=17. See also question 7 in Appendix 1.

<sup>8</sup> N=16. See also question 7 in Appendix 1.

<sup>9</sup> Interview dated 13 December 2013.

<sup>10</sup> See question 8 in Appendix 1.

<sup>11</sup> N= 16.

<sup>12</sup> N=16.

<sup>13</sup> Ibid.



data set required to understand their research.<sup>14</sup> But some interviewees further clarified that the quantum of the data shared depends on the type of project and the concerned colleague.

It is also important to mention here that most of the inter-personal data sharing within our sample (which has only researchers from Europe and North America) are found to be happening between researchers located within Europe or North America.<sup>15</sup> Among those who shared data with colleagues at least occasionally, we tried to explore the geographical location of those colleagues. While all those researchers in our sample had the experience of sharing data with colleagues located in Europe, 95% also had the experience of sharing data with colleagues located in North America.<sup>16</sup> But only 16% of them had the experience of sharing data with colleagues located in Africa or Middle East and only 21% had the experience of sharing data with colleagues located in Asia or South America.<sup>17</sup> These figures indicate a sharp contrast in the geographical reach of inter-personal data sharing among researchers.

### 3.2 Effectiveness of the existing database infrastructures in data sharing

The opinions of the stakeholders regarding the effectiveness of the existing public online databases was a particularly important dimension for our study, as the information about the actual working of the existing public online databases is highly required for the designing of optimal model contracts in this area. This sub-section highlights in detail the

---

<sup>14</sup> N=13.

<sup>15</sup> See question 10 in Appendix 1.

<sup>16</sup> N=19.

<sup>17</sup> Ibid.



opinions shared by different stake holders regarding the effectiveness of the existing public online databases.<sup>18</sup>

Two important factors explored in this regard were the opinions of the respondents regarding the costs as well as level of efforts involved in contributing to the existing databases in the field. Interestingly, 61% of the respondents are found to be of the view that the way the databases are currently organized allow them to contribute data at low costs.<sup>19</sup> Similarly, 61% of the respondents also shared the opinion that the way the databases are currently organized allow her or him to contribute data with reasonable efforts.<sup>20</sup> However some interviewees also expressed the opinion that there could be improvements in the current databases to reduce the costs as well as efforts required for submission of data.<sup>21</sup> Some interviewees also further pointed out that the actual cost with regard to uploading of data is not the mere act of uploading data, but the cost involved in hiring a qualified person to upload data.<sup>22</sup>

Interestingly, only 50% of the respondents are found to be of the view that the quality of the databases in the field is good enough for their research.<sup>23</sup> This indicates the scope and need for further improving the databases in the field. Similarly, only 50% of the respondents are of found to be agreeing to the view that there are software tools that can increase the

---

<sup>18</sup> See also question 11 in the questionnaire provided in Appendix 1.

<sup>19</sup> N=18

<sup>20</sup> N=18

<sup>21</sup> Interview dated 16 December 2013.

<sup>22</sup> Interview dated 17 December 2013.

<sup>23</sup> N=18



value of databases in their field of research and that they are easily accessible to them.<sup>24</sup>

Some of the interviewees clarified the situation by mentioning that “[t]here are software tools available for increasing the value of databases in my field of research. But it is doubtful whether they are accessible to all. The problem is the usability of those tools.”<sup>25</sup> Some other interviewees pointed out that “[t]here are software tools scattered around, but they are not so easily accessible.”<sup>26</sup> Some of the interviewees also pointed out that data integration is still missing and there is further scope for tools for more automation and more integration.<sup>27</sup> They also added that they expect the Micro B3 project to contribute primarily on improving these aspects.<sup>28</sup>

In a rather provocative way, the study had also asked the respondents how far they agree to the statement “It is not worthwhile to invest time in contributing to the databases in my field”. Interestingly, as high as 83% of the respondents are seen disagreeing or strongly disagreeing to that statement.<sup>29</sup> This in turn indicates the strong preference of the stakeholders in the area to contribute to the databases in the field, irrespective of the bottlenecks they might be facing in this regard.

### 3.3 Perceptions of the benefits received from data sharing

What are the different benefits that are received by different stakeholders when they share their data in a public database or when they share data with their colleagues? This was one

---

<sup>24</sup> N=18

<sup>25</sup> Interview dated 9 December 2013.

<sup>26</sup> Interview dated 17 December 2013.

<sup>27</sup> Interview dated 16 December 2013.

<sup>28</sup> Ibid.

<sup>29</sup> N=18



of the important aspects explored in our study.<sup>30</sup> It is highly important to know the benefits currently received by different stakeholders, as any model contracts that do not support the expected benefits of stakeholders are bound to have lesser acceptance and lesser compliance within the community.

The most frequently received benefit from contributing data to a public database is found to be the new research insights gained by using the software tools of such databases. A total of 79% of the respondents admits to have gained this benefit occasionally or regularly.<sup>31</sup> Similarly, 68% of the respondents mentioned that they occasionally or regularly gained new personal contacts from among the people who accessed the database to which they contributed.<sup>32</sup> 58% also received occasionally or regularly additional information from other data contributors on the data item that she or he contributed (for example, by way of curation of the data entry, additional complementary data, etc.).<sup>33</sup> Interestingly, 58% of the respondents also admit to have received attribution in publications based on the data submitted by them.<sup>34</sup> Finally, 50% of the respondents also had benefited occasionally or regularly from the quality improvement of their data (for example, through highlighting of errors in the data).<sup>35</sup>

The benefits received when sharing data with colleagues present a comparatively different picture. The most important benefit claimed to have received by most respondents are new

---

<sup>30</sup> See questions 13 and 14 in Appendix 1.

<sup>31</sup> N=19

<sup>32</sup> N=19

<sup>33</sup> N=19

<sup>34</sup> N=19

<sup>35</sup> N=18



research collaborations based on the shared data. 95% of the respondents received such new collaborations occasionally and 5% of the respondents received such new collaborations regularly.<sup>36</sup> The second most cited benefit is new insights from the colleagues on the provided data (for example, by showing new research or complementary data related to the shared data). 79% received such new insights occasionally and 11% received such new insights regularly.<sup>37</sup> Finally, 68% of the respondents also claim to have received the benefit of quality improvement of their data, regularly or occasionally.<sup>38</sup>

### 3.4 Importance of providing attribution

One of the most important insights we received from our study is the importance attached to attribution rights by the genomic research community. The researchers were not only seen expecting attribution when someone uses their data, but they also claim to provide attribution to the original data contributors whenever they are using data from public databases.

For example, 53% of the respondents are found to be of the view that proper attribution to the creators of a dataset, in any publications based on that data, is an important factor that can induce more data sharing.<sup>39</sup> Interestingly, when using data from public databases, nearly all of the respondents provided attribution to the original data contributors, though the type of attribution provided and the type of data use wherein attribution is provided

---

<sup>36</sup> N=19

<sup>37</sup> N=19

<sup>38</sup> N=19

<sup>39</sup> N=19



differed among the respondents.<sup>40</sup> For example, 47% of the respondents provided attribution to the original contributors when they reproduced an exact copy of data in a research publication, conference presentation or personal website and 33% provided attribution when the data was directly relevant for generating the research results presented in a research publication or conference presentation.<sup>41</sup> 20% provided attribution even when it is part of the general information search that contributed to reaching the results presented in a publication or conference presentation.<sup>42</sup>

Majority of the respondents (56%) are also found to be of the view that current database infrastructure allows improving their own visibility upon contribution, as there is a clear way to link the contributed data to their identity.<sup>43</sup> One of the private sector representatives even provided an interesting example wherein the data submitted by him as part of a patent application was subsequently accessed by many researchers and he was receiving appropriate attribution in publications, though he himself had never published any article based on those data.<sup>44</sup> All these aspects clearly illustrate the emergence of attribution to data contributors as a strong norm within the genomic research community.

### 3.5 Importance of metadata

Yet another important question we tried to explore in the study was the importance of sharing raw data with colleagues and also the importance of having proper metadata. 84% of the respondents are seen agreeing or strongly agreeing to the view that sharing raw data

---

<sup>40</sup> See, also, question 24 in the questionnaire provided in Appendix 1.

<sup>41</sup> N=15

<sup>42</sup> Ibid.

<sup>43</sup> N=18

<sup>44</sup> Interview dated 13 December 2013.



with colleagues can help to gain new insights and new perspectives in their own research.<sup>45</sup>

Interestingly, only 37% of the respondents are found to be agreeing or strongly agreeing to the view that *[s]haring raw data with colleagues is of limited importance, as such data will not have much use without extensive additional background information.*<sup>46</sup> We tried to explore during the interviews the possible reasoning behind this approach and many interviewees are found to be of the view that raw data can also be helpful at times in the genomics field.<sup>47</sup> However, they also clarified that this depends completely on the type of data and the same cannot be true for all types of data.<sup>48</sup>

### 3.6 Policies and regulations on data sharing

Another aspect of specific interest for our study was the policies and regulations that currently guided data sharing among different stakeholders. One of the important questions we asked in this regard was who was having ownership rights on the data produced from their research (for example, who can decide on matters relating to contribution of their data to public repositories or colleagues).<sup>49</sup> Interestingly, 47% of the respondents responded that they are having sufficient ownership rights on the data produced from their research and they don't have to seek any permission for sharing data produced by them.<sup>50</sup> On the other hand, 53% of the respondents responded that their employer has ownership or co-ownership rights on data produced from their research and so they have to ask their

---

<sup>45</sup> N=19

<sup>46</sup> N=19. See also question 12 in the questionnaire provided in Appendix 1.

<sup>47</sup> Interviews dated 9 December 2013 and 13 December 2013.

<sup>48</sup> Ibid.

<sup>49</sup> See question 17 in Appendix 1.

<sup>50</sup> N=17.



employer before sharing the data.<sup>51</sup> But one of the interesting things observed from the interviews in this regard is that a considerable section of the respondents are not so confident about the actual legal position within their organisation in this regard and some also suggested that clear official data policies at institutional level in this regard would make things far more easier for them as researchers.

During the interviews it is also noticed that some of the interviewees consider the data ownership by individual scientists as not necessarily a good thing.<sup>52</sup> For example, one researcher pointed out that she has ownership rights over all the data she produces, as her institute never made her sign on any document that shifts ownership of data produced by her during her official duties as a researcher.<sup>53</sup> But she thinks that such individual ownership at the scientist level (as opposed to ownership of data by the institute) is not good, as she cannot demand sharing of data produced by even colleagues at the same institute.<sup>54</sup> She also pointed out that her institute does not have a data policy. According to her, for the progress of science and for the benefit of all researchers, it is important to have institutional level data policies and if there are clear data related guidelines at the institutional level, it would be easier to make people comply with them and everyone would benefit.<sup>55</sup> She also cited an interesting experience from one of her past projects.<sup>56</sup> In one of the Italian projects she collaborated earlier, there was a requirement in the proposal that all the data had to be shared by the participants. But once the project was completed, no one asked for her data.

---

<sup>51</sup> N=17

<sup>52</sup> Interview dated 12 December 2013.

<sup>53</sup> Ibid.

<sup>54</sup> Ibid.

<sup>55</sup> Ibid.

<sup>56</sup> Ibid.



She had to remind them about the data policy and submit her data. According to her, such situations have to change for the benefit of scientific progress!

We had also asked the respondents who determines within their organisation how the data produced by the research activities of the organisation should be made available to the public.<sup>57</sup> Interestingly 79% of the respondents mentioned that it is a case by case decision by individuals.<sup>58</sup> But during the interviews, some of the respondents further clarified that in cases wherein external funding is involved, the rules of the funding agencies are also followed.<sup>59</sup> During the interviews, it was also pointed out that organisations like CNRS has data policies which are obligatory in nature and they also have service departments within the organisation to manage all data related issues with external parties.<sup>60</sup>

Interestingly, some of the interviewees expressed the view that there are comparatively advanced data sharing requirements at the European level now. They cited as an example the requirements in the European Commission grant applications to include a data sharing plan as part of the proposal.<sup>61</sup> They expressed the view that similar efforts must be taken in this regard at the national levels also.<sup>62</sup>

We had also tried to explore the restrictions generally imposed on the dissemination of the data shared by the respondents. We specifically asked the respondents whether they

---

<sup>57</sup> See question 20 in Appendix 1.

<sup>58</sup> N=19

<sup>59</sup> Interview dated 9 December 2013.

<sup>60</sup> Interview dated 9 December 2013.

<sup>61</sup> Interview dated 12 December 2013.

<sup>62</sup> Ibid.



generally insisted on not further disseminating data without their permission, when they provide data to colleagues.<sup>63</sup> Only 11% of the respondents are seen not imposing any such restrictions and 32% of the respondents impose restrictions against further dissemination of the shared data.<sup>64</sup> More strikingly, 47% of the respondents answered that it depends on the type of data they are sharing. During the interviews, some of the respondents further clarified this by saying that they generally insist on not disseminating data further when they have not yet made any publications based on that data.<sup>65</sup> In other words, publication of research results based on the data can play a crucial role in determining the conditions imposed with regard to further dissemination of that data. 11% of the respondents also mentioned that it depends on the type of colleague with whom they are sharing the data.<sup>66</sup> During the interviews some of the respondents also added that sometimes data are provided only for a specific time period and for a specific purpose.<sup>67</sup>

To better understand the data usage practices in the field, we had also asked the respondents at what point they generally sought permission from the original data contributors, while using data from public databases.<sup>68</sup> Interestingly, 37.5% of the respondents mentioned that they never seek permissions from the original data contributors while using data from public databases.<sup>69</sup> But 50% of the respondents mentioned that they will seek permission when they reproduce an exact copy of the data in

---

<sup>63</sup> See question 22 in Appendix 1.

<sup>64</sup> N=19.

<sup>65</sup> Interview dated 9 December 2013.

<sup>66</sup> N=19.

<sup>67</sup> Interview dated 12 December 2013.

<sup>68</sup> See question 23 in Appendix 1.

<sup>69</sup> N=16.



a research publication, conference presentation or personal website and 12.5% of the respondents mentioned that they will seek permission when the data was directly relevant for generating the research results presented in a research publication or conference presentation.<sup>70</sup>

### 3.7 Motivations to contribute data

As part of this study, we had also tried to explore the motivations for contributing data to public databases and for sharing data with colleagues. Such an inquiry was highly important to ensure that the IP model contracts evolved under the project are in tune with those motivations.

The most important motivation seen in the decision to contribute data to public online databases is the strong belief that data sharing is necessary for the progress of science and research and it is her or his duty as a scientist to contribute data to public online databases. As high as 89% are seen considering it as an important factor that motivates them to share data to public online databases.<sup>71</sup> The second most cited factor is the belief that contributing to public online databases helps in better dissemination of their own research works and 79% of the respondents consider this as an important factor.<sup>72</sup> Finally the belief that contributing to public online databases helps in building new research collaborations is also an important motivating factor and 58% of the respondents consider it as an important factor.<sup>73</sup>

---

<sup>70</sup> Ibid.

<sup>71</sup> N=19

<sup>72</sup> N=19

<sup>73</sup> N=19



Interestingly, when we look at the most relevant factors in the decision to provide data to a colleague who requests data from their research, it could be seen that the possibility of building research collaborations receives the highest importance with 84% of the respondents considering it as an important factor.<sup>74</sup> The second most frequently cited factor, with 79% of the respondents sharing the view, is the belief that it is their duty as a scientist to provide data to colleagues who request data.<sup>75</sup> Finally, 63% of the respondents are also seen thinking that providing data to colleagues creates a social network which increases their own possibilities to ask for data from others when they need some data.<sup>76</sup>

We had also tried to explore the extent of influence of the major data sharing initiatives in the field.<sup>77</sup> Interestingly, majority of the respondents are found to be not aware of the major data sharing initiatives that have happened in the field. For example, 56% of the respondents are seen not aware of the Bermuda principles on release of human genomic sequence data.<sup>78</sup> Similarly, 68% of the respondents didn't have any information on the Fort Lauderdale principles on rapid prepublication release and 79% of the respondents are seen not aware of the Toronto statement on rapid prepublication data release.<sup>79</sup> This result is particularly interesting, as their high motivations to share data appears to have been not arising from the awareness of such data sharing initiatives in their field.

---

<sup>74</sup> N=19

<sup>75</sup> N=19

<sup>76</sup> N=19

<sup>77</sup> See question 26 in the questionnaire provided in Appendix 1

<sup>78</sup> N=19

<sup>79</sup> N=19



### 3.8 Factors that dissuade researchers from sharing data

Similar to the need for identifying the motivations for sharing data, it is also highly important to get a better picture of the factors that dissuade researchers from sharing their data in public databases and sharing data with their colleagues. Hence we also tried to explore the factors that dissuaded researchers from sharing data to a public database and also the factors that dissuaded them from sharing data with colleagues.

Interestingly, the factor which most number of respondents cited as the most important factor in dissuading them from contributing their data to a public database is the belief that as the contributor of data, they must be able to complete all research on that data and they can contribute data to a public database only when all research and publications based on those data are completed.<sup>80</sup> Lack of time to do the required quality management of data before uploading is also considered by many respondents as one of the important factors in dissuading them from contributing data. But it is important to mention here that during the interviews, some of the respondents strongly disagreed to this view and opined that “[I]ack of time cannot be a reason for dissuading the scientists from contributing data.”<sup>81</sup> Some of the respondents further added that researchers will always be able to find time for data contribution.<sup>82</sup>

---

<sup>80</sup> n=8

<sup>81</sup> Interview dated 11 December 2013.

<sup>82</sup> Interview dated 16 December 2013.



Similarly many respondents also cited commercial value of data for their organisation as one of the important factors that dissuade them from sharing data to a public database. The commercial value aspect of data is particularly emphasised by the private sector representatives as the most important factor that dissuades them from sharing data in a public database.<sup>83</sup> The commercial value of data is also seen highlighted by some of the scientists working in public funded research institutions and according to them, even though exceptional, when collaborating in projects with private firms, they are generally obliged not to reveal data outside and in those cases this becomes a significant factor.<sup>84</sup> During the interviews, some respondents also mentioned that lack of proper attribution could be an additional factor dissuading sharing of data.<sup>85</sup>

Similarly, it is important to see the factors that dissuade people from sharing data with their colleagues. Two factors are considered by equal number of respondents as the most important factor that dissuades them from sharing data to colleagues. The first one is the lack of time to do the required quality management of data before providing it to someone else.<sup>86</sup> An equal number of respondents are also seen considering that as the contributor of data, they should be the first one to do research on that data and that they can provide data to someone else only when all their research and publications based on those data are

---

<sup>83</sup> Interview dated 13 December 2013.

<sup>84</sup> Interview dated 16 December 2013.

<sup>85</sup> Interview dated 9 December 2013.

<sup>86</sup> n=5. But some of the interviewees also mentioned during the interviews that lack of time cannot be a reason for dissuading a scientist from providing data to a colleague who requests data. Interview dated 11 December 2013.



completed.<sup>87</sup> Commercial value of the data is cited by respondents from private sector as the important factor that dissuades them from sharing data with colleagues.<sup>88</sup>

### 3.9 Factors that can create more willingness to share data

In order to better understand how we can further augment data sharing and also fine tune the model agreements in this regard, we had also asked the respondents which factors they consider as the most relevant factors for creating more willingness among researchers to provide data to other researchers and to public repositories.<sup>89</sup>

The most frequently cited factor, with 58% of the respondents, is ensuring legal certainty on data ownership and sharing issues.<sup>90</sup> Some of the interviewees further clarified this by adding that legal certainty will help to create more trust.<sup>91</sup> One of the respondents added that “[d]ata sharing is all about creating trust. Anything that can facilitate in creating trust will certainly induce more data sharing.”<sup>92</sup> The second most frequently cited factor, with 53% of the respondents supporting, is ensuring proper attribution to the creators of the dataset, in any publications based on the data.<sup>93</sup> Finally, meetings with other researchers at global and regional conferences is also considered as an important factor that can motivate data sharing and 47% of the respondents are found to be considering this as an important

---

<sup>87</sup> n=5

<sup>88</sup> Interview dated 13 December 2013.

<sup>89</sup> See question 29 in Appendix 1.

<sup>90</sup> N=19

<sup>91</sup> Interview dated 16 December 2013.

<sup>92</sup> Ibid.

<sup>93</sup> N=19



factor.<sup>94</sup> It is also important to mention here that some interviewees also expressed the opinion that “[o]nly when people are forced to engage in data sharing, they will share data”.<sup>95</sup>

### 3.10 Some of the specific implications of the findings for the model agreements

All most all the findings we have outlined in the previous sub-sections have major implications for the genomic research community and the model contracts which are evolving in this area. However, we would like to specifically outline in the below table five most important findings from this study, which have significant implications on the model contracts developed under the Micro B3 project.

	<b>Some of the important findings</b>	<b>Specific implications for the model contracts developed under the Micro B3 project</b>
a.	High prevalence of data sharing within the genomic research community	Both the model ABS Agreement (developed under Deliverable 8.1) and the model data license agreement (developed under Deliverable 8.2) were drafted with the objective of facilitating better pre-competitive access to microbial research data
b.	Biggest motivation for contribution of data to public databases is the belief that data sharing is necessary for the progress of science and research	

<sup>94</sup> N=19

<sup>95</sup> Interview dated 12 December 2013.



		<p>generated in marine bio-prospecting missions. See, for example, Art. 6 of the Model ABS Agreement. The model contractual clauses are not only in consonance with the existing data sharing practices within the community, but also in tune with the motivations of the genomic research community to contribute data.</p>
<p>c.</p>	<p>Overcome the limitations of existing database infrastructures in data sharing, particularly with regard to data integration</p>	<p>Micro B3 Information System, which is one of the important expected outcomes of the Micro B3 project, will help to overcome the existing bottle necks in data integration. The Micro B3 model ABS Agreement has also been particularly designed to support the integration of data, particularly the data generated from samples collected as part of the Ocean Sampling Day project.</p>
<p>d.</p>	<p>Importance of attribution for data contributors</p>	<p>Micro B3 model ABS Agreement and the model data licensing provisions</p>



		provided in Deliverable 8.2 give at most importance to attribution rights of data contributors. See for example, Art.3 of the model data license agreement provided in Deliverable 8.2.
e.	Importance of legal certainty in data related transactions	The model contractual provisions developed under the Micro B3 project will create more legal certainty for all the stakeholders involved.

## Conclusion and future directions

As one may notice from the discussions in the previous section, the participatory research approach taken in the study has enabled to identify many important and interesting perspectives from diverse stakeholders in the field. They not only provide highly valuable insights for the optimisation of the model agreements developed as part of the Micro B3 project, but will also help in guiding the evolution of data policies nationally, regionally and globally.

As one could notice from the responses of different stake holders, the genomic research community has been highly active in data sharing. Even the private sector representatives are seen actively contributing to public databases, even if they are doing it as part of their efforts for IP protection in other areas. It is highly important that any model agreements



developed in this area should be supporting the desire of the community to engage in data sharing.

As rightly pointed out by some of the respondents in the study, legal certainty on data ownership and sharing issues can play a major role in increasing data contribution from researchers.<sup>96</sup> The model IP agreements developed as part of Deliverable 8.1 and Deliverable 8.2 in Micro B3 project are certainly useful legal tools in this direction and we hope that the genomic research community will be actively using such model agreements to create more legal certainty on data related transactions. Similarly, as we could notice from the discussions in the previous section, the genomic research community attaches high importance for proper attribution to the creators of datasets during any use of the data and we hope that the model IP agreements developed as part of the Micro B3 project can contribute positively in this regard. Finally, it is also important to mention here that the IP model agreements developed as part of the Micro B3 project are in tune with the general motivations of the researchers in sharing data, particularly because they are built on the premise that data sharing is necessary for the progress of science and research.

It is also important to mention here some of the possible limitations of our study. The first one is the reliance on non-random sampling to reach the target audience. Second one is the limited sample size in our study. However, these are acceptable limitations in view of the fact that this was an exploratory study in the field. We strongly hope that we will be able to do a wider survey in future, based on the learnings from this study, and thereby reach far more researchers in this field. We also encourage other researchers to use the

---

<sup>96</sup> N=19



questionnaire provided in the Appendix of this study and try to replicate the study. Only such replications using different sampling approaches and methodologies can help us to learn more about this emerging and unexplored field.

We would like to end the report with two interesting comments, from two of our interviewees, that very well characterise the current situation and future direction of data sharing in this field -

“Even within my own lab, it is often a struggle to convince the younger generation that when they share data they will be adding to the total value of available data. Far more needs to be done for bringing this message to the wider scientific community.”<sup>97</sup>

“We must recognise that it is an emerging world and data sharing is an emerging cultural thing. We must try to promote this culture.”<sup>98</sup>

We truly hope that this report will help not only in the more optimal use of the model IP agreements that can promote data sharing among researchers, but also help in the designing of better data sharing policies at national, regional and international levels.

---

<sup>97</sup> Interview dated 17 December 2013.

<sup>98</sup> Interview dated 16 December 2013.



## Appendix 1: Questionnaire used for the semi-structured interviews

## Survey on Data Sharing Practices

Within the EU FP7 project Micro B3 (biodiversity, bioinformatics, biotechnology), WP8 is dealing with intellectual property management issues. One of the important things envisaged within WP8 is evolving model contracts for pre-competitive access to large scale microbial genomic databases, which will be proposed by MICRO B3 as a project result, but might also have a broader impact on EU data policy.

As an important participant in the Micro B3 project, we would like to hear your views on data sharing and we are conducting this short semi-structured interview (around 20 minutes) for this purpose. Your views will play a significant role in the evolution of data policy and model contracts within the project and we aim to discuss the resulting data policy aspects in the upcoming Micro B3 General Assembly meeting in April 2014.

Any personal information collected during the interview will be used only for data treatment purposes and will not be published or disseminated, except in an aggregated form.

We thank you for your time and consideration and welcome you to the questionnaire.

best regards,

Tom Dedeurwaerdere  
Professor at the Université catholique de Louvain  
Director of the BIOGOV Unit  
Centre for Philosophy of Law (CPDR)  
Université catholique de Louvain

## Personal and institutional profile

[Note: All the information collected will be used only for data analysis purposes and will not be published or disseminated, except in an aggregated form.]

### 1. Name

### 2. Organisation

### 3. In which country is the organisation located

Country:

### 4. Which of the following best describes your organisation (may select more than one option)

- University
- Public funded research institution
- Commercial research institution
- Private company

Other (please specify)

### 5. Which of the following best describes your profession (may select more than one option)

- Academician (duties are mainly related to teaching/training at an educational institution)
- Researcher (duties are mainly related to research activities)
- Database manager
- Data curator

Other (please specify)

## Data contribution practices

### 6. Which of the following best describes the frequency of your contribution of data to the following:

	never	once in a year	twice in a year	once in a month	several times each month
Genbank or similar public online genomic sequence databases	<input type="radio"/>				
Proteomic sequence databases	<input type="radio"/>				
Website of the publisher of journals, as part of your manuscript submissions	<input type="radio"/>				
Colleagues	<input type="radio"/>				

### 7. In general, at what point do you submit data to the following:

	Immediately after the generation of data	At the time of deposit of the associated biological material in a public collection	At the time of publication of a paper based on that data	After the publication of all papers based on that data	Never
Genbank or other genomic sequence databases	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Proteomic sequence databases	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Website of publishers (of journals)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Colleagues	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### 8. In general, what type of data do you submit to the following:

	The minimal data set required to understand my research/the minimal data set required by the data repository	The core data set plus some additional data such as related publications, geographical origin and identity of the data contributors	The complete original dataset so as to allow other researchers to do further analysis of the data
Genbank and other genomic sequence databases	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Proteomic sequence databases	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Website of the publisher	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Colleagues	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**9. How often do you contribute to the databases which are geographically located in the below regions**

	never	occasionally	regularly
Africa	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Asia	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Europe	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Middle East	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
North America	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Oceania (comprising Australia and proximate islands)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
South America	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**10. If you share data with colleagues at least occasionally, could you please indicate which of the following regions represent the geographical location of such colleagues (you may choose more than one option)**

- Africa
- Asia
- Europe
- Middle East
- North America
- Oceania (comprising Australia and proximate islands)
- South America

## Effectiveness of existing databases

### 11. Could you please express your extent of agreement with the below statements with regard to the public online databases (databases which provide free online access) in your field

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
The quality of the databases in my field are good enough for my research	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The databases in my field are not well developed to make data contributions to them	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The way the databases are currently organized allow me to contribute data at low cost	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The way the databases are currently organized allow me to contribute data with reasonable effort	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is not worthwhile to invest time in contributing to the databases in my field	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The current database infrastructure allows to improve my visibility upon contribution, as there is a clear way to link the contributed data to my identity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
There are software tools that can increase the value of databases in my field of research for me and they are easily accessible to me	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### 12. Could you please express your extent of agreement with the following statements

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
Sharing raw data with colleagues is of limited importance, as such data will not have much use without extensive additional background information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sharing raw data with colleagues can help to gain new insights and new perspectives in my own research	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## Benefits from sharing

### 13. How frequently have you received the following benefits from contributing data to a public database

	never	occasionally	regularly
A new personal contact from someone who accessed the database to which I contributed	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Additional information from other data contributors to the data item that I contributed (For example, by way of curation of the data entry, additional complementary data, etc.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
New research insights by using the software tools of the database	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Attribution in a publication based on my data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quality improvement of my data (for example, highlighting of errors on my data)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Others (please specify)

### 14. How frequently have you received the following benefits when you provided data to colleagues:

	never	occasionally	regularly
New insights from those colleagues on the provided data (for example, by showing new research or complementary data related to my data)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quality improvement of my data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A new research collaboration based on that data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## Factors that dissuade sharing of data

**15. How important are the following factors in dissuading you from contributing data to a public database, even when similar kind of data are regularly uploaded by your colleagues or other scientists in your field to such databases (Rank 1 indicates the highest importance, Rank 4 indicates the lowest importance)**

Lack of time to upload data to the database

Lack of time to do the required quality management of data before uploading it

As the contributor of data, I must complete all research on that data and I can contribute data to a public database only when all research and publications based on those data are completed

Data has commercial value for my organisation (for example, it serves the organisation to sell services relating to data or data analysis expertise)

**16. In general, how important are the following factors in dissuading you from providing data to a colleague who requests your data (Rank 1 indicates the highest importance, Rank 3 indicates the lowest importance)**

Lack of time to do the required quality management of data before providing it to anyone else

As the contributor of data, I should be the first one to do research on that data and I can provide data to someone else only when all my research and publications based on those data are completed

Data has commercial value for my organisation (for example, it serves the organisation to sell services relating to data or data analysis expertise)

## Policies and regulations on data sharing

### 17. Who has ownership rights on the data produced from your research (for example, who can decide on matters relating to contribution of your data to public repositories or colleagues) :

- I have sufficient ownership rights on the data produced from my research and I don't have to seek permission for sharing data
- My employer has ownership or co-ownership rights on data produced from my research and so I have to ask my employer before sharing data

Others (please specify)

### 18. In general, when you produce and share data from your research projects, how far the official policies from following sources have to be followed:

	Obligatory	Optional	Not applicable	I don't know
Data policy of your organisation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Data policies prescribed by the government	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Data policy of the project funder	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### 19. In general, when you provide data to a publisher (as part of a publication), how far the following official policies apply:

	Obligatory	Optional	Not applicable	I don't know
Data policy of your organisation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Data policy of publishers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Data policies prescribed by the government	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Data policy of the project funder	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### 20. In general, who determines within your organisation how the data produced by the research activities of the organisation should be made available to the public:

- Case by case decisions by individuals
- Centralized decision by (a) manager(s) of your organisation
- Rules imposed by the government
- Rules imposed by the research funders

**21. In general, how frequently are the following categories of people consulted when your organisation formulates the general guidelines for sharing data generated from your organisation**

	Never	Occasionally	Always
Scientists	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Private sector companies	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
National/regional/local public authorities (for example science ministries)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Publishers of scientific journals in your area	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**22. When providing data to a colleague, do you generally insist on not further disseminating data without your permission**

- No
- Yes
- Depends on the type of colleague with whom I am sharing
- Depends on the type of data I am sharing

## Data use practices

### 23. When you use data from public databases, at what point do you generally seek permission from the original data contributors:

- when you reproduce an exact copy of the data in a research publication, conference presentation or personal website
- when you do not reproduce an exact copy of the data, but when the data was directly relevant for generating the research results presented in a research publication or conference presentation
- when you do not reproduce an exact copy of the data, but when it is part of the general information search that contributed to reaching the results presented in a publication or conference presentation
- Never

Others (please specify)

### 24. If you are using data from public databases, when do you generally provide attribution to the original data contributors:

- when you reproduce an exact copy of data in a research publication, conference presentation or personal website
- when you do not reproduce an exact copy of the data, but when the data was directly relevant for generating the research results presented in a research publication or conference presentation
- when you do not reproduce an exact copy of the data, but when it is part of the general information search that contributed to reaching the results presented in a publication or conference presentation
- Never

Others (please specify)

## 25. To what extent does your research group participate in the activities of the following international research organisations/initiatives/science federations

	No participation	Passive participation (for example, receiving emails from them)	Active participation (for example, attending their conferences, meetings, etc.)	Organisational role (for example, being in the governing body)	I don't know
Global Genome Initiative (GGI)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Global Genome Biodiversity Network (GGBN)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Genomic Standards Consortium (GSC)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Human Genome Organisation (HUGO)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ERA-Net Plant Genomics	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The genome consortium for active teaching	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Other major international research organisations/initiatives/science federations, where your research group has active participation or organisational role:

## 26. Are you aware of any of the following initiatives:

	Yes I have heard about it, but I am not aware of the details	Yes I have heard about it and I am aware of the details	No, I do not have any information about this initiative.
Bermuda principles on release of human genomic sequence data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fort Lauderdale principles on rapid prepublication release	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Toronto statement on rapid prepublication data release	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## Motivations to contribute data

### 27. Which of the following factors are the most relevant in your decision to contribute data to public online databases (You may select up to 3 relevant factors)

- Data sharing is necessary for the progress of science and research and it is my duty as a scientist to contribute data to public online databases
- Contributing to public online databases helps in increasing my reputation within the scientific community
- Contributing to public online databases helps in building new research collaborations
- Contributing to public online databases helps in better disseminating my research work
- Contribution to public online databases is mandated by my funding agency/employer
- The type of license (for example, Creative Commons license) used in that website for providing data to users

Others (please specify)

### 28. Which of the following factors are the most relevant in your decision to provide data to a colleague who requests data from your research (You may select up to 3 relevant factors)

- It is my duty as a scientist to provide data to colleagues requesting data
- Providing data to colleagues helps in increasing my reputation within the scientific community
- Providing data to colleagues helps in building research collaborations
- Providing data to colleagues creates a social network which increases my own possibilities to ask for data from others when I need some data

Others (please specify)

### 29. Which of the following are the most relevant factors for creating more willingness among researchers to provide data to other researchers and to public repositories (you may select up to 3 factors)

- Meetings with other researchers at global and regional conferences
- Legal certainty on data ownership and sharing issues
- Proper attribution to the creators of the dataset, in any publications based on that data
- Mandatory directions from the funding agencies with regard to sharing of data
- Mandatory directions from the employers with regard to sharing of data
- Better funding of the existing research infrastructure

Others (please specify)

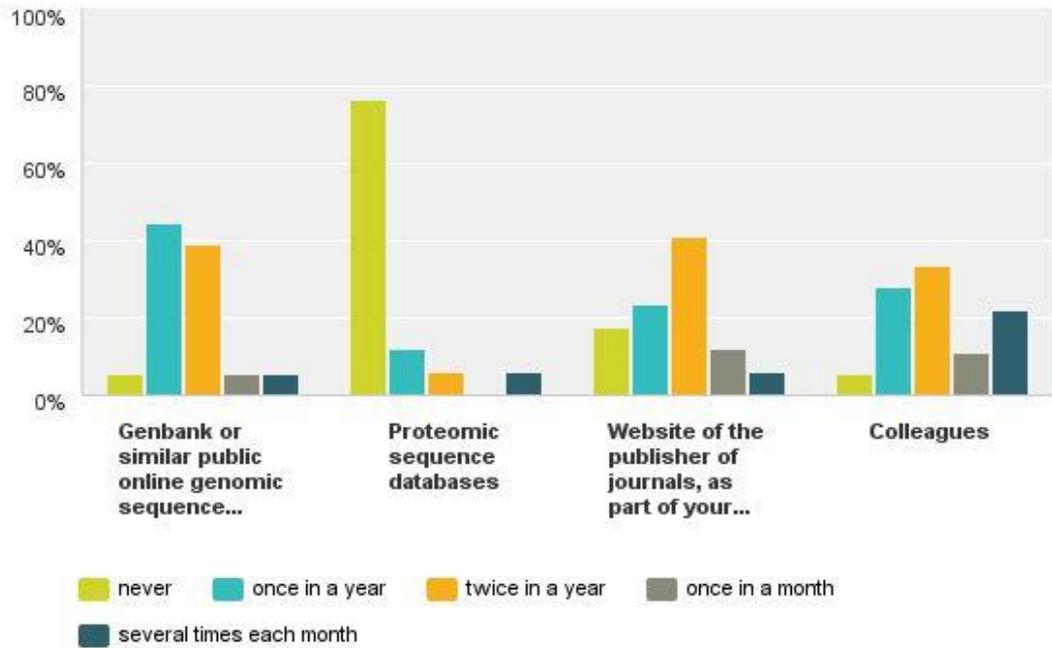
## **Appendix 2: Detailed Descriptive Statistics**

## General Remarks:

We collected responses from 21 scientists working in 19 different organisations that manage/produce large scale genomic data sets in the field of life science research. Due to the anonymity assurance we have provided to the participants in the study, we have presented in this deliverable only the aggregate data. The organisations to which the participants belong are: Aix Marseille University, AWI, Bangor University, CNRS - Station biologique de Roscoff, Ecole normale supérieure and CNRS, ICES, ICM-CSIC, IMBBC-HCMR, Institute for Coastal Marine Environment, Institute of Marine Biology, Biotechnology and Aquaculture HCMR, Marine Biological Association, MARIS, Matis ldt., Max Planck Institute for Marine Microbiology, Marine Biological Laboratory, PharmaMar, Prokazyne ehf, Stazione Zoologica Anton Dohrn, and Tokyo Institute of Technology.

### Q6 Which of the following best describes the frequency of your contribution of data to the following:

Answered: 19 Skipped: 2

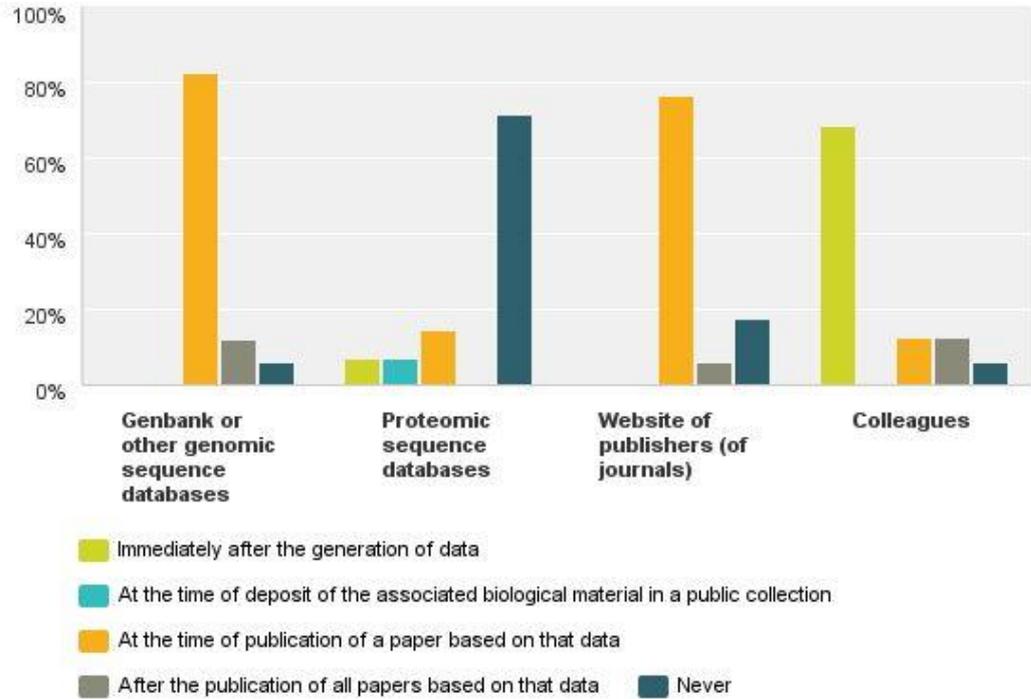


#### Full text of the categories:

- Genbank or similar public online genomic sequence databases
- Proteomic sequence databases
- Website of the publisher of journals, as part of your manuscript submissions
- Colleagues

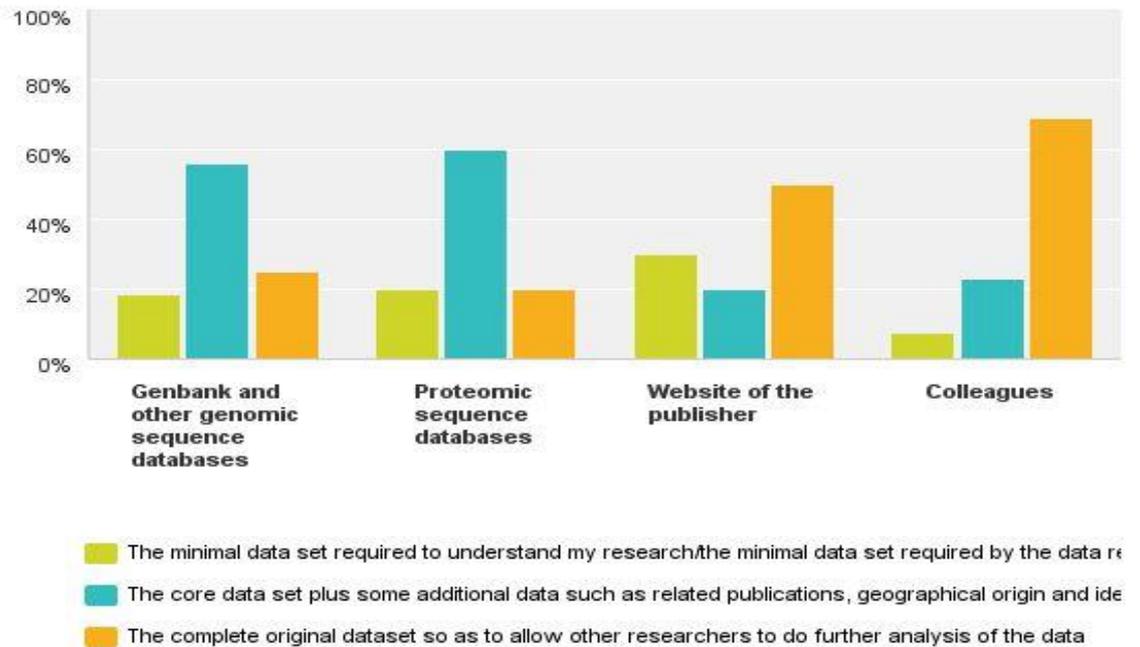
### Q7 In general, at what point do you submit data to the following:

Answered: 19 Skipped: 2



## Q8 In general, what type of data do you submit to the following:

Answered: 17 Skipped: 4

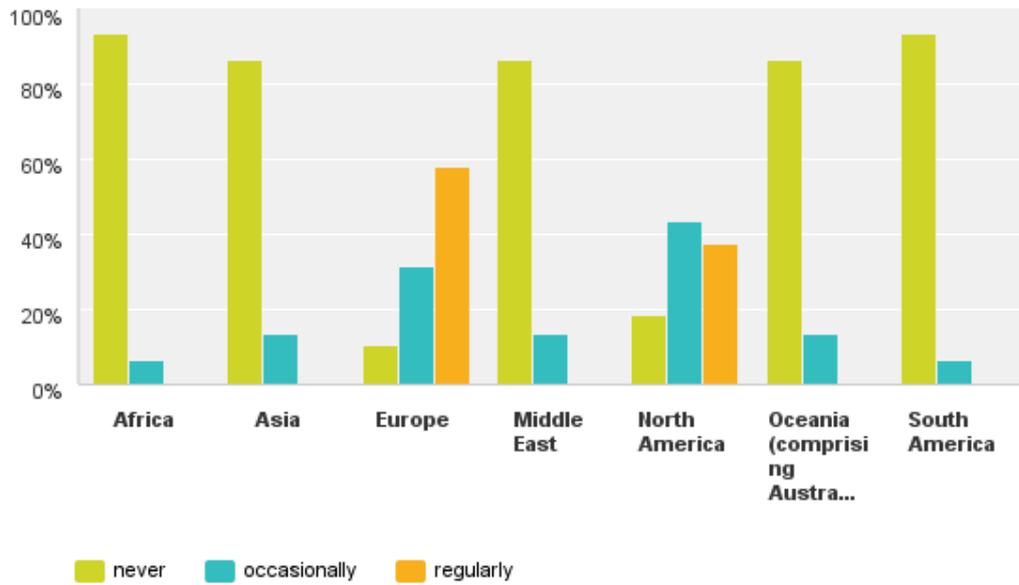


### Full text of the response options:

- The minimal data set required to understand my research/the minimal data set required by the data repository
- The core data set plus some additional data such as related publications, geographical origin and identity of the data contributors
- The complete original dataset so as to allow other researchers to do further analysis of the data

### Q9 How often do you contribute to the databases which are geographically located in the below regions

Answered: 19 Skipped: 2

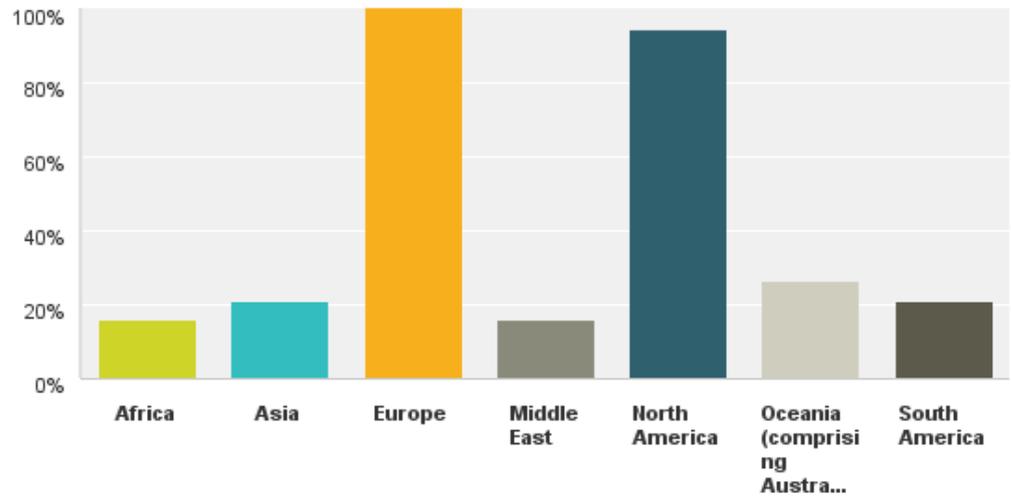


Full text of the categories:

- Africa
- Asia
- Europe
- Middle East
- North America
- Oceania (comprising Australia and proximate islands)
- South America

**Q10 If you share data with colleagues at least occasionally, could you please indicate which of the following regions represent the geographical location of such colleagues (you may choose more than one option)**

Answered: 19 Skipped: 2

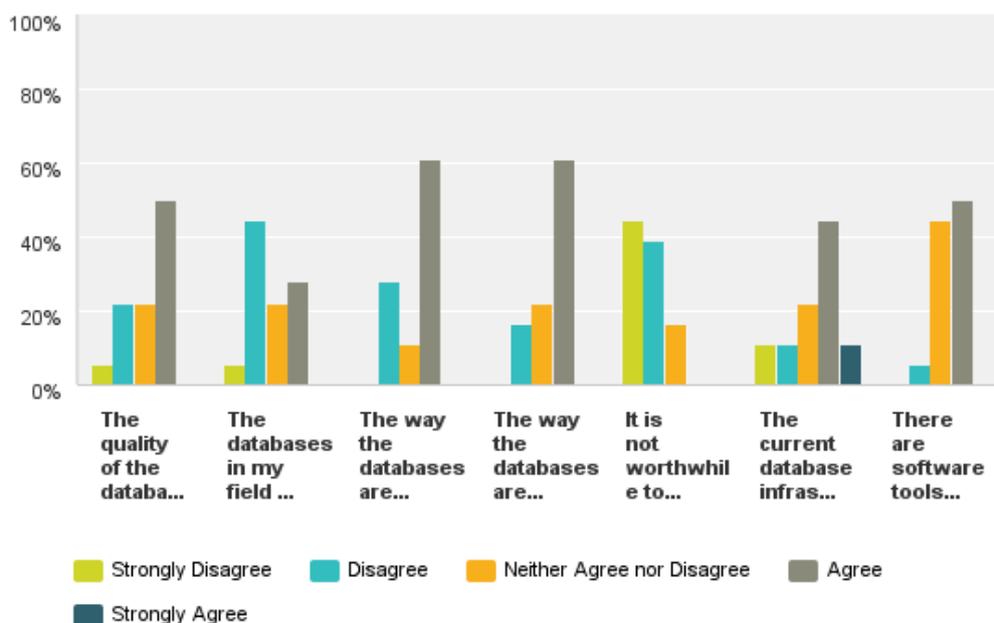


Full text of the categories:

- Africa
- Asia
- Europe
- Middle East
- North America
- Oceania (comprising Australia and proximate islands)
- South America

**Q11 Could you please express your extent of agreement with the below statements with regard to the public online databases (databases which provide free online access) in your field**

Answered: 18 Skipped: 3

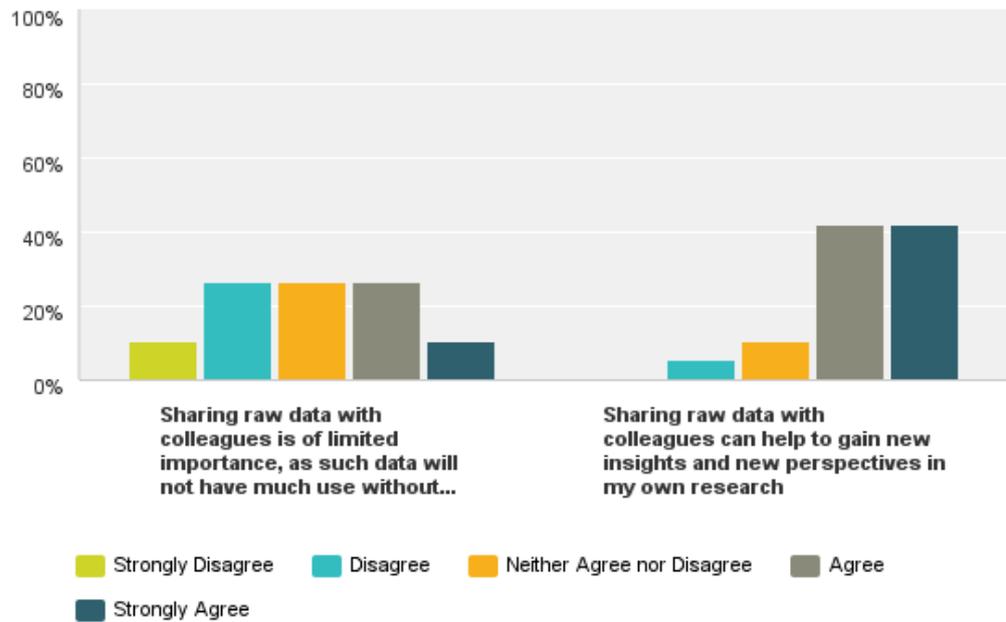


**Full text of the categories:**

- The quality of the databases in my field are good enough for my research
- The databases in my field are not well developed to make data contributions to them
- The way the databases are currently organized allow me to contribute data at low cost
- The way the databases are currently organized allow me to contribute data with reasonable effort
- It is not worthwhile to invest time in contributing to the databases in my field
- The current database infrastructure allows to improve my visibility upon contribution, as there is a clear way to link the contributed data to my identity
- There are software tools that can increase the value of databases in my field of research for me and they are easily accessible to me

## Q12 Could you please express your extent of agreement with the following statements

Answered: 19 Skipped: 2

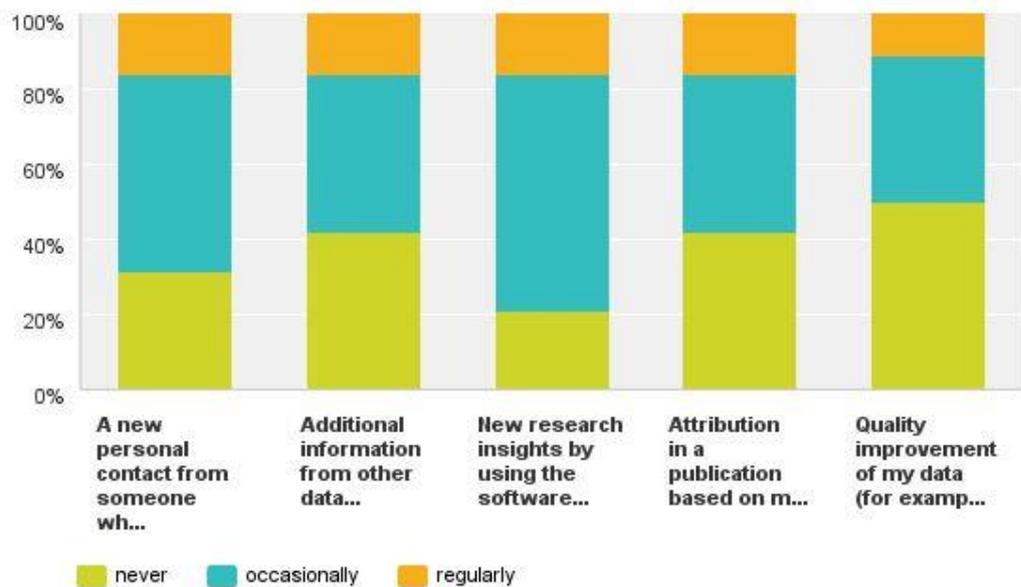


### Full text of the categories:

- Sharing raw data with colleagues is of limited importance, as such data will not have much use without extensive additional background information
- Sharing raw data with colleagues can help to gain new insights and new perspectives in my own research

### Q13 How frequently have you received the following benefits from contributing data to a public database

Answered: 19 Skipped: 2

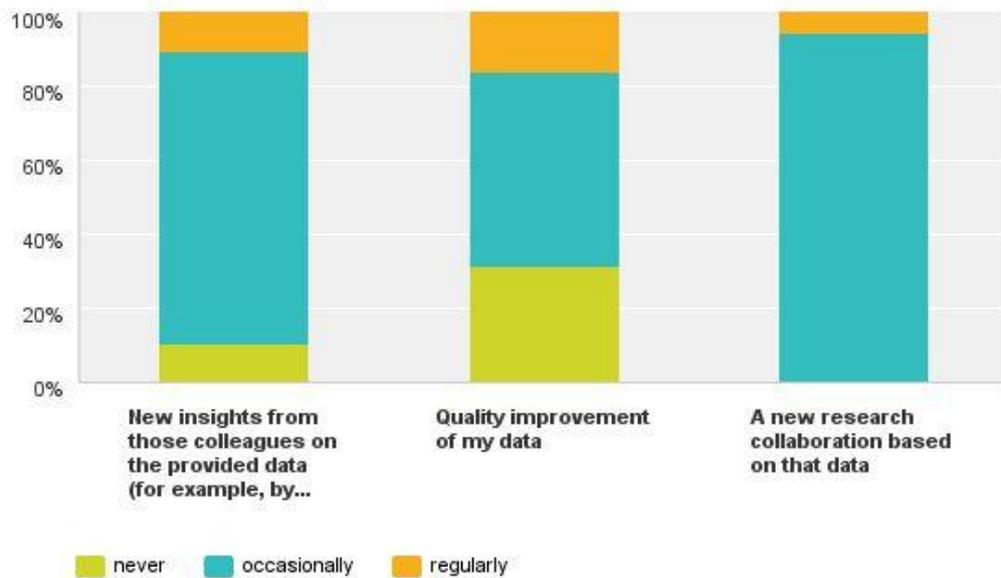


#### Full text of the categories:

- A new personal contact from someone who accessed the database to which I contributed
- Additional information from other data contributors to the data item that I contributed (For example, by way of curation of the data entry, additional complementary data, etc.)
- New research insights by using the software tools of the database
- Attribution in a publication based on my data
- Quality improvement of my data (for example, highlighting of errors on my data)
- Others (please specify)

### Q14 How frequently have you received the following benefits when you provided data to colleagues:

Answered: 19 Skipped: 2

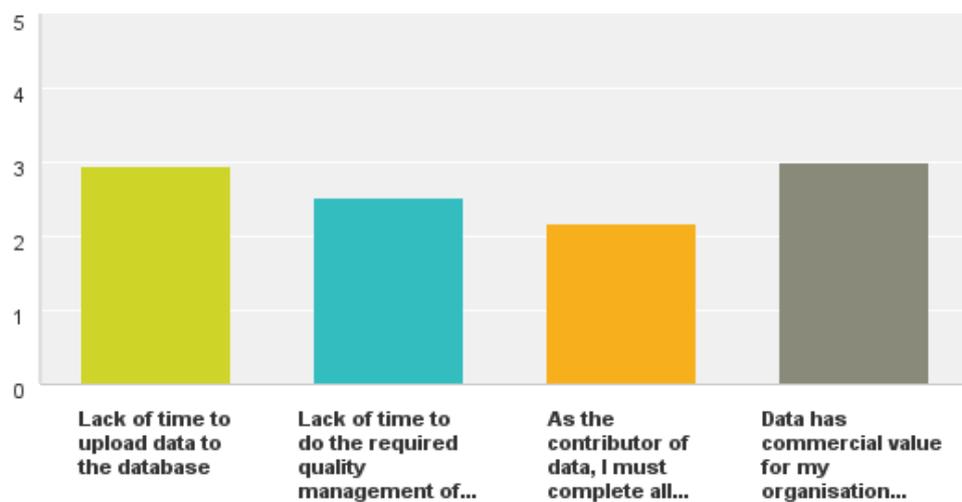


#### Full text of the categories:

- New insights from those colleagues on the provided data (for example, by showing new research or complementary data related to my data)
- Quality improvement of my data
- A new research collaboration based on that data

**Q15 How important are the following factors in dissuading you from contributing data to a public database, even when similar kind of data are regularly uploaded by your colleagues or other scientists in your field to such databases (Rank 1 indicates the highest importance, Rank 4 indicates the lowest importance)**

Answered: 19 Skipped: 2

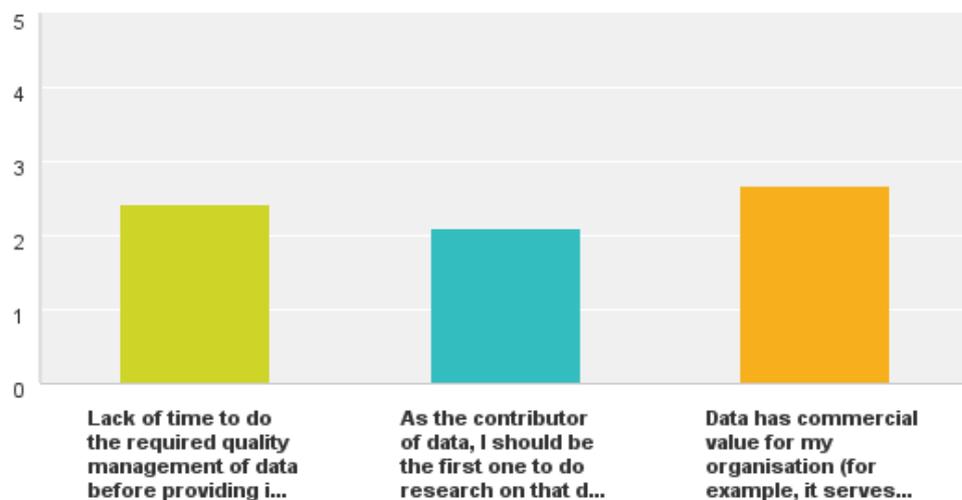


**Full text of the categories:**

- Lack of time to upload data to the database
- Lack of time to do the required quality management of data before uploading it
- As the contributor of data, I must complete all research on that data and I can contribute data to a public database only when all research and publications based on those data are completed
- Data has commercial value for my organisation (for example, it serves the organisation to sell services relating to data or data analysis expertise)

**Q16 In general, how important are the following factors in dissuading you from providing data to a colleague who requests your data (Rank 1 indicates the highest importance, Rank 3 indicates the lowest importance)**

Answered: 19 Skipped: 2

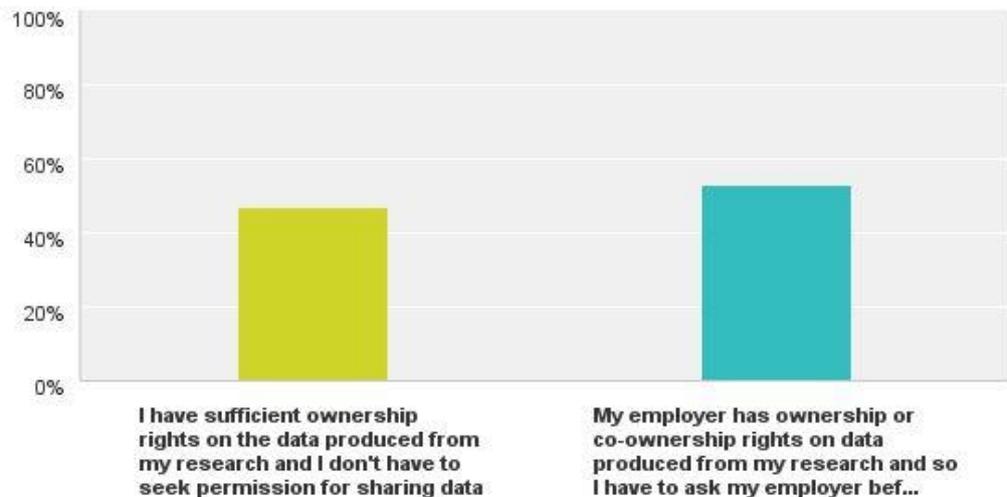


**Full text of the categories:**

- Lack of time to do the required quality management of data before providing it to anyone else
- As the contributor of data, I should be the first one to do research on that data and I can provide data to someone else only when all my research and publications based on those data are completed
- Data has commercial value for my organisation (for example, it serves the organisation to sell services relating to data or data analysis expertise)

**Q17 Who has ownership rights on the data produced from your research (for example, who can decide on matters relating to contribution of your data to public repositories or colleagues) :**

Answered: 17 Skipped: 4

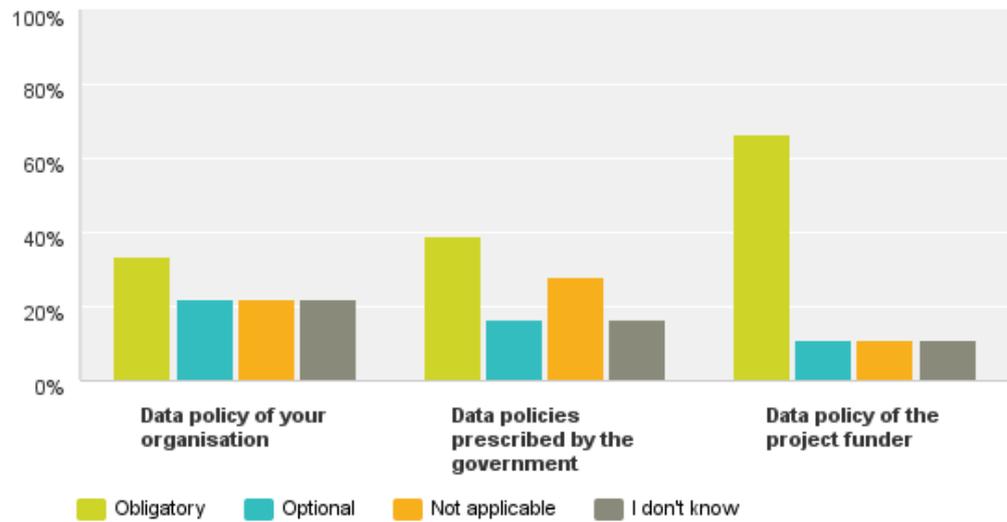


**Full text of the response options:**

- I have sufficient ownership rights on the data produced from my research and I don't have to seek permission for sharing data
- My employer has ownership or co-ownership rights on data produced from my research and so I have to ask my employer before sharing data
- Others (Please specify)

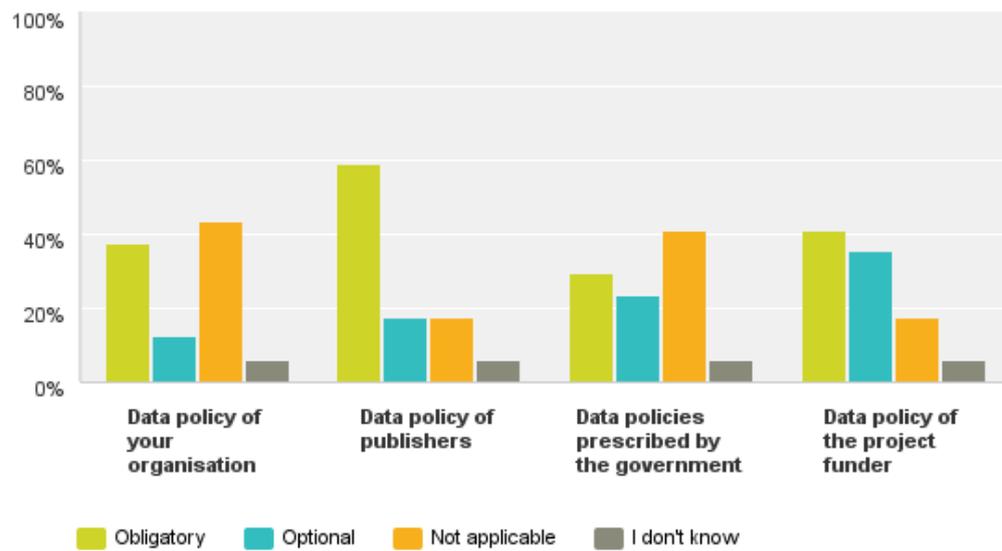
**Q18 In general, when you produce and share data from your research projects, how far the official policies from following sources have to be followed:**

Answered: 18 Skipped: 3



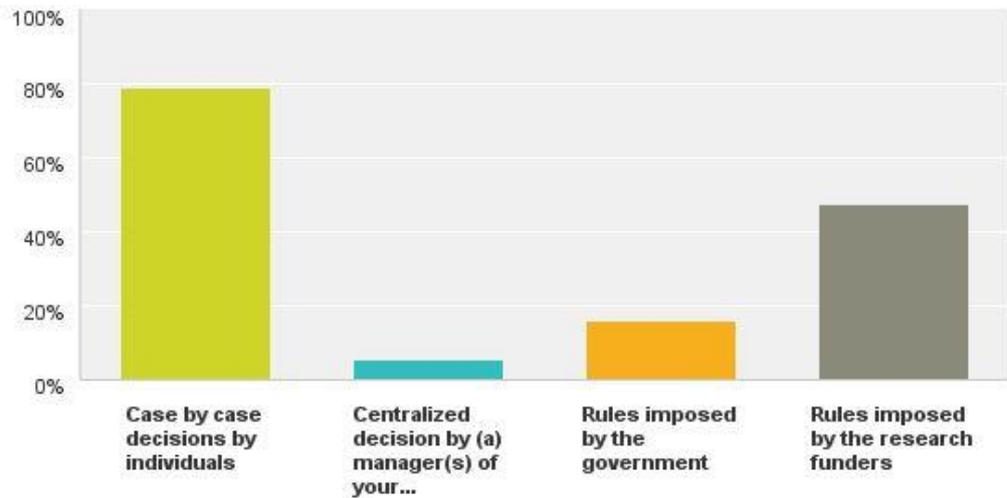
### Q19 In general, when you provide data to a publisher (as part of a publication), how far the following official policies apply:

Answered: 17 Skipped: 4



**Q20 In general, who determines within your organisation how the data produced by the research activities of the organisation should be made available to the public:**

Answered: 19 Skipped: 2

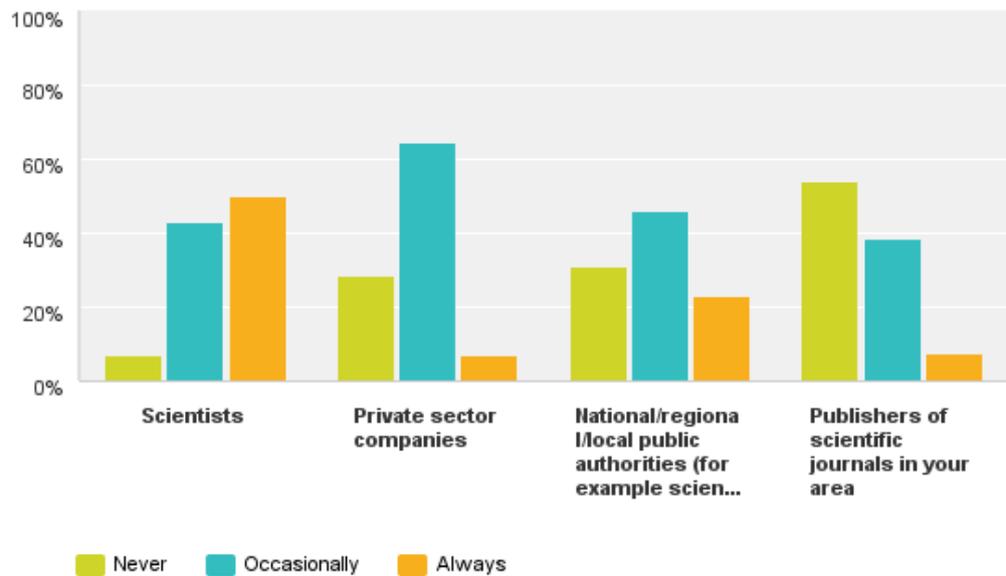


Full text of the response options:

- Case by case decisions by individuals
- Centralized decision by (a) manager(s) of your organisation
- Rules imposed by the government
- Rules imposed by the research funders

**Q21 In general, how frequently are the following categories of people consulted when your organisation formulates the general guidelines for sharing data generated from your organisation**

Answered: 14 Skipped: 7

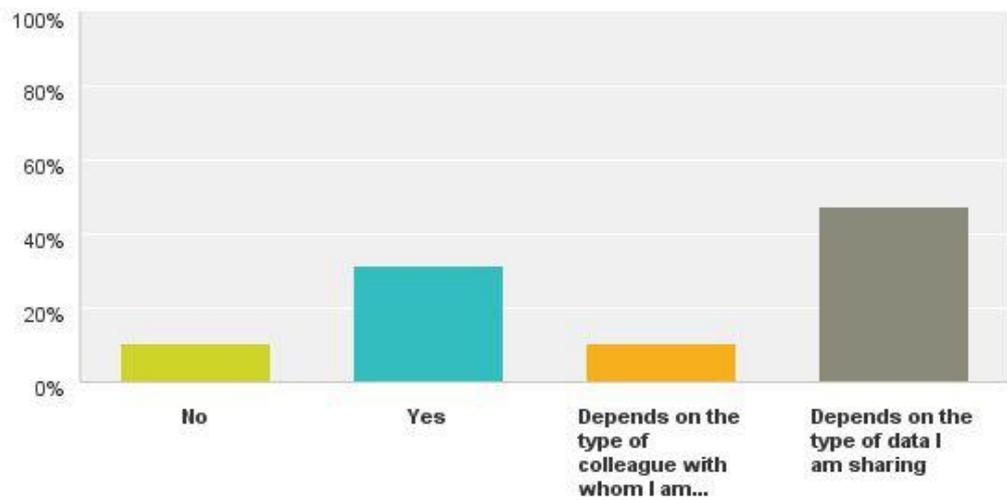


**Full text of the categories:**

- Scientists
- Private sector companies
- National/regional/local public authorities (for example science ministries)
- Publishers of scientific journals in your area

## Q22 When providing data to a colleague, do you generally insist on not further disseminating data without your permission

Answered: 19 Skipped: 2

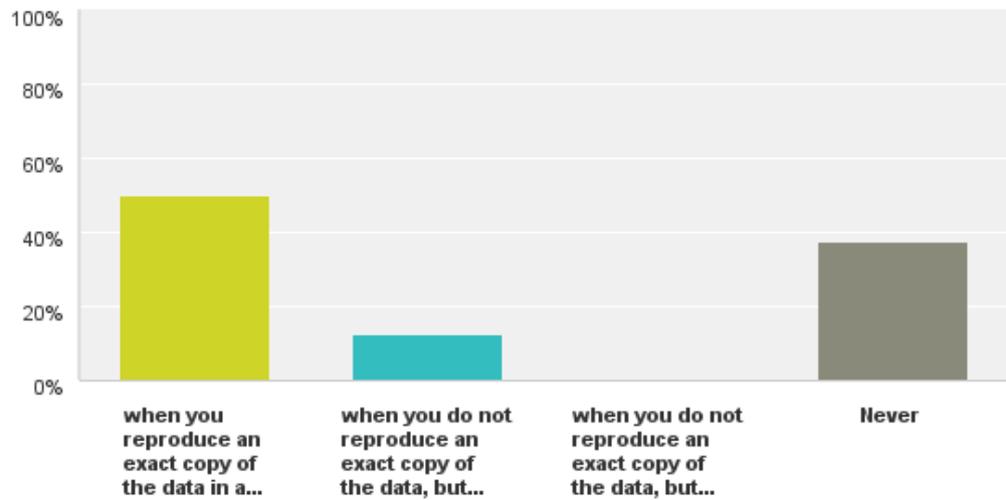


Full text of the response options:

- No
- Yes
- Depends on the type of colleague with whom I am sharing
- Depends on the type of data I am sharing

### Q23 When you use data from public databases, at what point do you generally seek permission from the original data contributors:

Answered: 16 Skipped: 5

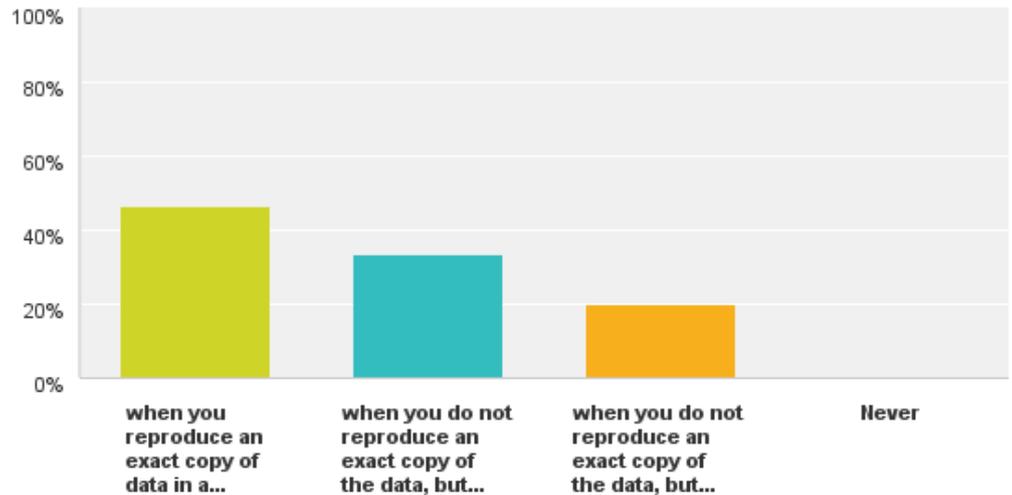


#### Full text of the response options:

- when you reproduce an exact copy of the data in a research publication, conference presentation or personal website
- when you do not reproduce an exact copy of the data, but when the data was directly relevant for generating the research results presented in a research publication or conference presentation
- when you do not reproduce an exact copy of the data, but when it is part of the general information search that contributed to reaching the results presented in a publication or conference presentation
- Never
- Others (please specify)

## Q24 If you are using data from public databases, when do you generally provide attribution to the original data contributors:

Answered: 15 Skipped: 6

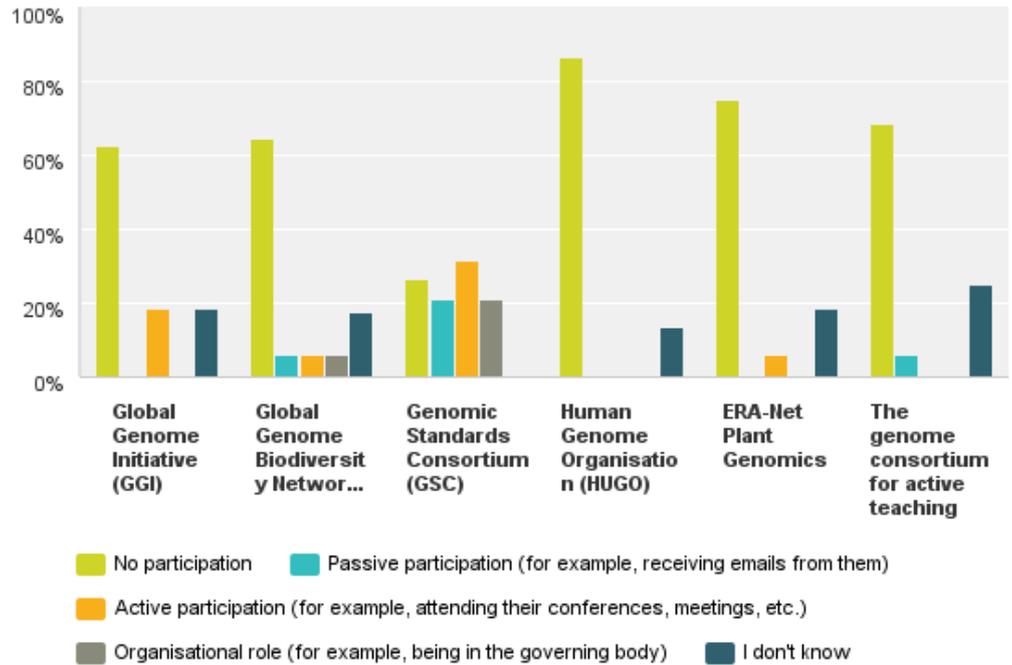


### Full text of the response options:

- when you reproduce an exact copy of data in a research publication, conference presentation or personal website
- when you do not reproduce an exact copy of the data, but when the data was directly relevant for generating the research results presented in a research publication or conference presentation
- when you do not reproduce an exact copy of the data, but when it is part of the general information search that contributed to reaching the results presented in a publication or conference presentation
- Never
- Others (please specify)

**Q25 To what extent does your research group participate in the activities of the following international research organisations/initiatives/science federations**

Answered: 19 Skipped: 2

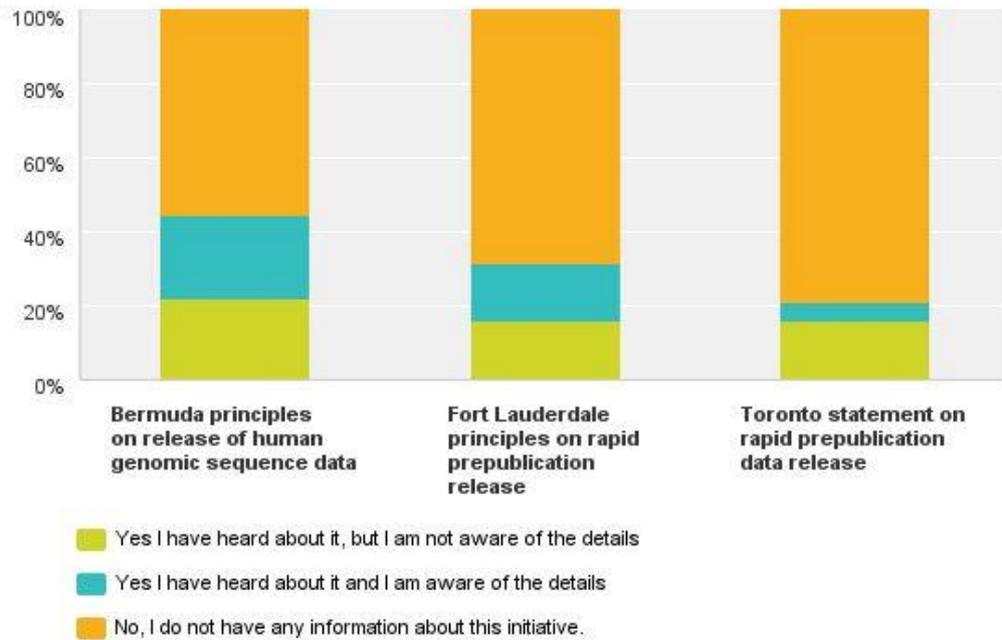


**Full text of the categories:**

- Global Genome Initiative (GGI)
- Global Genome Biodiversity Network (GGBN)
- Genomic Standards Consortium (GSC)
- Human Genome Organisation (HUGO)
- ERA-Net Plant Genomics
- The genome consortium for active teaching
- Other major international research organisations/initiatives/science federations, where your research group has active participation or organisational role:

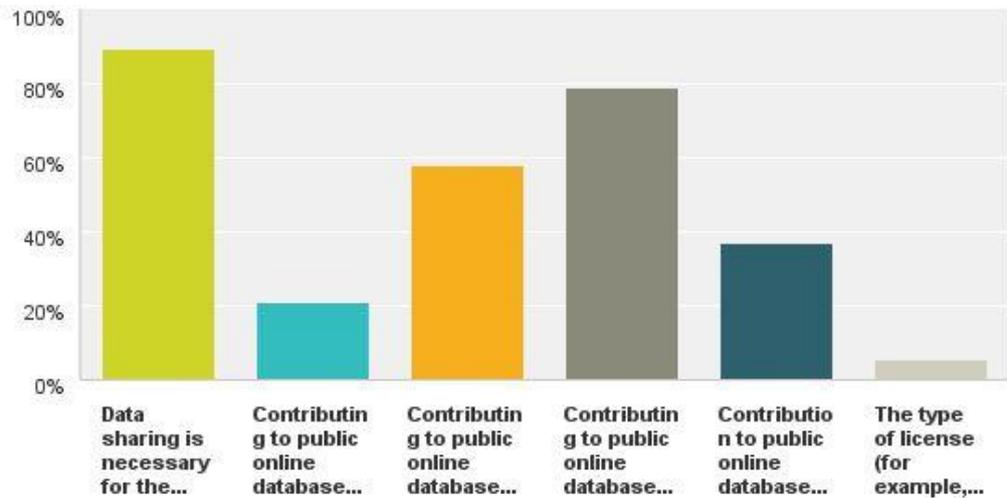
## Q26 Are you aware of any of the following initiatives:

Answered: 19 Skipped: 2



**Q27 Which of the following factors are the most relevant in your decision to contribute data to public online databases (You may select up to 3 relevant factors)**

Answered: 19 Skipped: 2

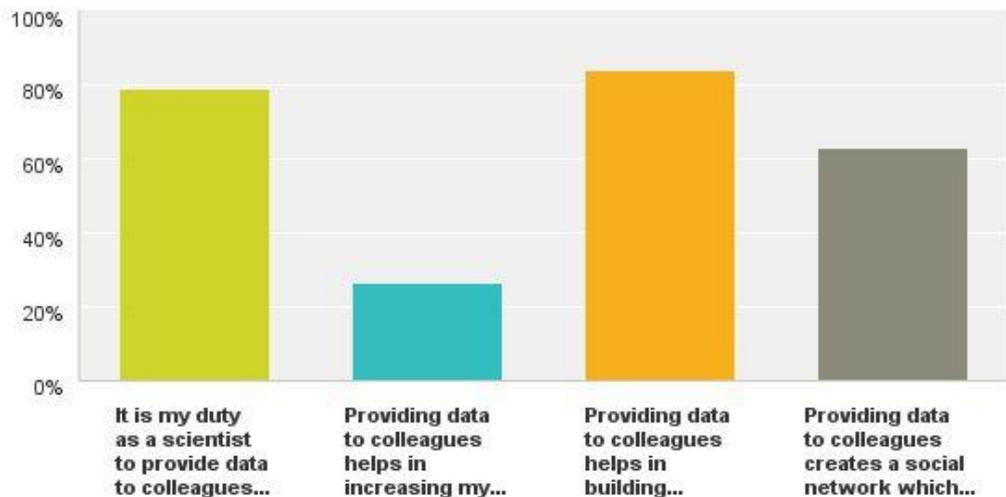


**Full text of the categories:**

- Data sharing is necessary for the progress of science and research and it is my duty as a scientist to contribute data to public online databases
- Contributing to public online databases helps in increasing my reputation within the scientific community
- Contributing to public online databases helps in building new research collaborations
- Contributing to public online databases helps in better disseminating my research work
- Contribution to public online databases is mandated by my funding agency/employer
- The type of license (for example, Creative Commons license) used in that website for providing data to users
- Others (please specify)

**Q28 Which of the following factors are the most relevant in your decision to provide data to a colleague who requests data from your research (You may select up to 3 relevant factors)**

Answered: 19 Skipped: 2

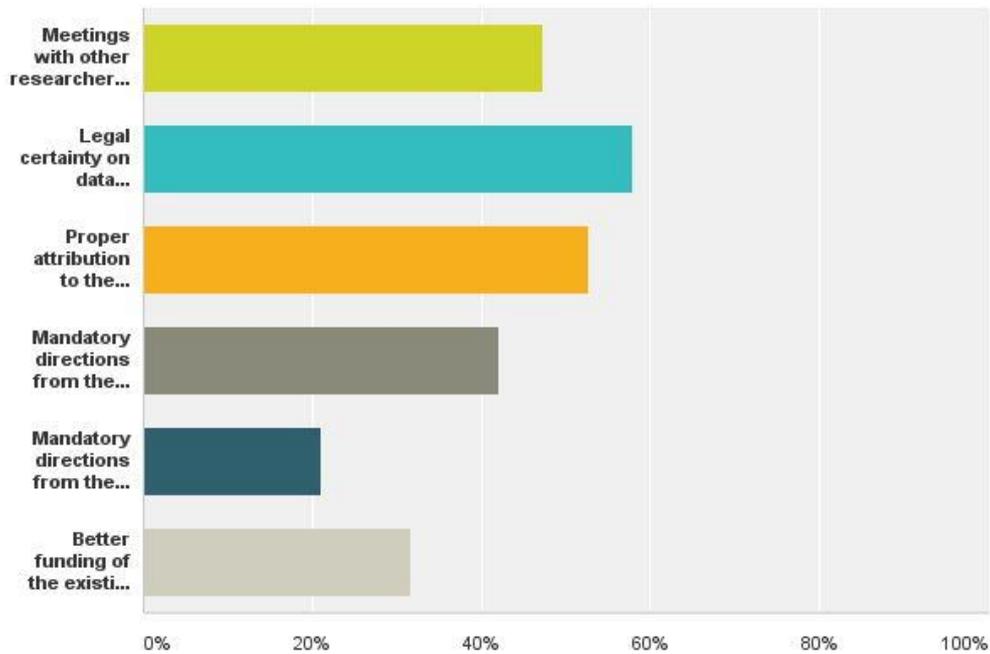


**Full text of the categories:**

- It is my duty as a scientist to provide data to colleagues requesting data
- Providing data to colleagues helps in increasing my reputation within the scientific community
- Providing data to colleagues helps in building research collaborations
- Providing data to colleagues creates a social network which increases my own possibilities to ask for data from others when I need some data
- Others (please specify)

**Q29 Which of the following are the most relevant factors for creating more willingness among researchers to provide data to other researchers and to public repositories (you may select up to 3 factors)**

Answered: 19 Skipped: 2



**Full text of the categories:**

- Meetings with other researchers at global and regional conferences
- Legal certainty on data ownership and sharing issues
- Proper attribution to the creators of the dataset, in any publications based on that data
- Mandatory directions from the funding agencies with regard to sharing of data
- Mandatory directions from the employers with regard to sharing of data
- Better funding of the existing research infrastructure
- Others (please specify)