

Summary of responses to GBIF Consultation Licensing of data within the Global Biodiversity Information Facility (GBIF)

Overview

This document presents a summary of responses to a consultation carried out by GBIF Secretariat between April and June 2014 on proposed changes to the current system of licensing data within the global network. It documents the range of views received on this important issue from GBIF Participant countries and associates, other organizations and individuals. The report does not prejudge any decision on future licensing arrangements, but offers a transparent analysis of the breadth of opinion expressed in the thoughtful responses received through this open consultation. The Secretariat expresses its thanks to all respondents who put considerable effort into their submissions. The analysis will inform a proposal from the Secretariat to the 21st GBIF Governing Board meeting in New Delhi, India in September 2014, including possible alternative options for the way forward.

Background

Based on the outcomes of comments received to a first consultation on data licensing in August 2013, the GBIF Secretariat organised a second round, *Licensing of data within GBIF*, running from 14 April to 14 June 2014 to request feedback on a proposal with the following elements:

- 1) All data publishers through GBIF would be asked not to claim rights when publishing their data (through a Creative Commons CCO license or equivalent);
- 2) All data publishers and users would agree to follow 'publishing and usage norms', to be drawn up collaboratively with the GBIF community, for example following standard guidelines for data citation and defining what constitutes commercial use; and
- 3) The GBIF data publishing process would be modified to allow data publishers to flag datasets they wish to restrict from commercial use, as defined in the norms.

The consultation paper was made available in English and Portuguese, and the response form in English.

During the consultation period, 49 responses were received. Six respondents asked for their comments to remain anonymous. The remaining 43 responses were as follows:

- Voting Participant Countries (14): Finland (FI), Benin (BJ), Republic of Korea (KR), Germany (DE), Norway (NO), Ireland (IE), Argentina (AR), Sweden (SE), Colombia (CO), South Africa (ZA), Madagascar (MG), Netherlands (NL), Mexico (MX), United Kingdom (GB)
- Associate Participant Countries (3): Austria (AT), Brazil (BR), Israel (IL)
- Associate Participant Organizations (1): Society for the Management of Electronic Biodiversity Data (SMEBD)
- Other institutions or individuals (25): Patricia Mergen/Royal Museum for Central Africa, Robert S.
 Capers/GS Torrey Herbarium Univ. of Connecticut, Andy Bentley/Univ. of Kansas, Nicolas Noé, Gregor Hagedorn/Museum für Naturkunde Berlin, Lyubomir Penev/Pensoft, Jean Woods/Delaware Museum

of Natural History, Quentin Groom, Charles Roper/Sussex Biodiversity Record Centre, Peter Desmet, Dimitri Brosens, Shannon Asencio, Thomas Leya/ Fraunhofer-Institute for Biomedical Engineering IBMT, CCCryo, Ian Egerton, Roman David/IMBE: Mediterranean Institute of Biodiversity and marine and terrestrial Ecology, Station Marine d'Endoume, Marseille, Charlie Barnes, Tim Corner, David Baxter, Tom Hunt, Aaike De Wever/BioFresh, Rachel Stroud/NBN, Teresa Frost, Alan Paton & Cecilia Buffery/Kew Gardens, Steve Whitbread/NFBR, Paula Lightfoot

The consultation included four questions:

- 1. The approach outlined would significantly clarify licensing and aims to promote fair use and citation via a set of community norms. However, data publishers would no longer be in a position to claim copyright over data. Do you support GBIF taking this position?
- 2. Do you believe that there are additional factors which should be considered at this time?
- 3. Do you foresee any substantial risk arising from this approach, in particular regarding the willingness of data holders to continue publishing through GBIF?
- 4. Are you interested in contributing to collaborative documentation on this topic?

In practice, the topics raised by these questions are interconnected and some aspects of responses to each question also relate to the other questions. This summary attempts to integrate such comments regardless of how responses were structured.

Responses to the consultation document

Summary

The responses span a wide range of divergent views. Responses in favour of the proposal include comments that "the use of a CCO license would result in a more robust system which is well understood by both data publishers and data users" Lightfoot, and that "it should be communicated in the documentation (...) that this approach is [the] most honest and legally valid one GBIF can take, while still catering to the wishes of the GBIF community." Desmet . Several responses state support for a simplified licensing model in general, while not supporting exclusive adoption of CCO.

However, a number of submissions raised fears that data may be lost from publishers who cannot or do not want to waive their rights, due to a number of factors including:

- concerns around commercial use
- concerns about lack of attribution
- national differences in legislation ZA, AR; and
- different regional approaches^{BJ}.

Some respondents expected loss of data not even to be noticeable ("If some withdraw [their data] it may not even be noticed." but others consider the loss to be potentially significant. The responses identified a number of key requirements for implementing a CCO or equivalent licensing policy, including:

- the need to implement supporting mechanisms for data citation, attribution and tracking;
 and
- clear communication, addressing the legal basis, the reasoning behind the agreed solution, and the implementation plan Noé, Groom.

Question 1: Support for the proposed changes to licensing of data within GBIF

Category	VP	APC	APO	Other	Anon.	Total
Support suggested changes Finland, Norway, Ireland, Sweden, Colombia, Netherlands, Austria, Israel, SMEBD; Capers, Noé, Hagedorn, Wood, Groom, Desmet, Leya, David, Baxter, De Wever, Lightfoot	6	2	1	11	3	23
Support suggested changes, subject to appropriate implementation South Korea, Germany, Argentina, Madagascar; Penev, Brosens, Asencio, Corner, Stroud	4	-	,	5	-	9
Do not support suggested changes, unless alternative options considered South Africa, Mexico, United Kingdom; Mergen, Roper, Egerton, Barnes, Bentley, Paton/Buffery, Hunt, Frost, Whitbread	3	-	-	9	-	12
Do not support changes Benin, Brazil	1	1	-	-	1	3
No statement	-	-	-	-	2	2

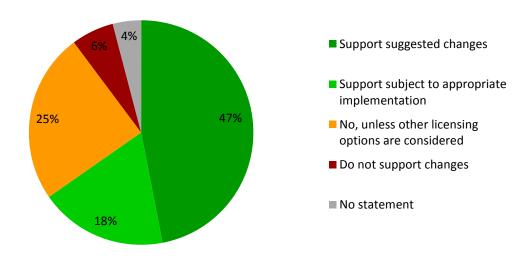


Figure 1 - Breakdown of responses to question 1 - ALL responses

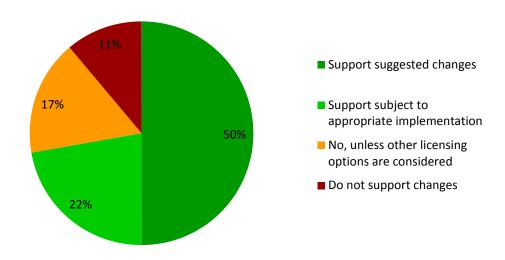


Figure 1 - Breakdown of responses to question 1 - Participant responses only

Question 2: Additional factors that need to be considered

- Communities and business models: at least in the UK, data collection and collation is often self-funded, partially through at-cost, tailored data services. Data publishers may themselves not be data owners, but aggregators, not able to make or overrule licensing decisions, and working with licensing schemes that resulted from community agreements Barnes, Hunt, Frost. Licensing decisions need to be made by the original data owners MG. Any solution on the side of GBIF should not compromise the integrity of such systems. A fallback option may be the publication of lower-resolution data Egerton, Barnes, Corner, Hunt, Stroud, anon. Consider the needs of the private sector and associations (e.g. BirdWatch), NGOs, individuals, and others not directly involved in the general open data trend AT.
- **Specification and norms:** Community norms will require a wide discussion^{Fl}. Consider the Canadensys norms as a starting point^{Baxter}. It is important to ensure integrity, traceability and attribution of data where possible, rather than to waive the need^{Roper}.
- Explore other options: Suggest reinforcing the use of CC0 to promote to data sharing^{CO}, but not to the exclusion of other licensing options; rather, provide a wider set of options and require publishers to select one of them in their data rights management^{Roper, Whitbread}. Consider CC 4.0¹, and especially CC-BY 4.0 (*sui generis* database rights), and at least document the reason for rejecting the option, if applicable^{Baxter, Paton&Buffery}. As attribution is key, CC-BY would be the minimum requirement for UK datasets^{Stroud, GB}; CC-BY-NC would be in line with the UK recorder network model^{Hunt}. Explore other options in consultation with CC², ODI³, OKFN^{4Roper}. There needs to be community involvement in the decision^{Whitbread}. Subsistence of copyright is not the only intellectual property right to consider^{Paton&Buffery}. If GBIF does not follow the approach outlined in the consultation call, then at least restrict the set of possible terms and conditions to make working with data downloads manageable^{Asencio}.
- Implementation plan: Before licenses can be changed, the appropriate supporting mechanisms need to be in place (citation, attribution, and use and re-use tracking mechanisms, supported by DOIs)^{AR, SE, ZA, Stroud, anon}. A clear implementation plan is needed, ensuring the right sequence of implementation, norms discussion and agreement, and evaluation of resource implications on the side of publishing networks for negotiating and implementing any changes. Investment in better attribution mechanisms is more important than the introduction of the proposed changes^{ZA}. The proposed timeframe of 6 month appears too short to several respondents^{Stroud, Lightfoot}.
- **Implementation**: the handling of licenses for already existing datasets in the network needs to be explicitly specified and communicated beforehand (possibilities: full migration, withdrawal of datasets, staying with earlier licenses etc, especially in cases of disagreement with the new model) Noé, De Wever, CO, anon.

¹ http://blog.okfn.org/2012/08/15/cc-license-version-4-0-helping-meet-the-needs-of-open-data-publishers-and-users

² http://creativecommons.org

³ http://theodi.org

⁴ https://okfn.org

- Other data aspects: apart from dataset citations, individual records need support for citation, referencing and linkage Hagedorn. Different models apply to different data, e.g. CC-BY-4.0 for taxonomic work Brosens. Also consider standard licenses for metadata (context: data papers) Multimedia objects need to be considered separately DE, anon. For observations, copyright regulations are insufficiently examined Barnes. Sensitive data should undergo automated tests to prevent accidental publishing I. The applicability to new types of data needs to be considered (multimedia, cultural heritage, genetic information) SE.
- **Non-commercial**: any restrictions on commercial use should be time-limited to avoid establishing permanent limitations^{Groom}. Offering the option to indicate non-commercial limitations is important^{KR, IE, AT, Stroud}. Equally important is the definition of what is meant by commercial use, which is controversial and difficult^{GB}. Careful discussion is needed considering whether we do indeed want to exclude particular categories of users and uses, as open data should support all^{NL}.
- **Enforcement:** Currently, the NBN citation guidelines are often not heeded, and this is likely to be an issues with new norms. Licences are often hard to police; education, automated support and a level of enforcement are all needed to make this happen Barnes, Egerton. It is important to move towards a culture where users abide by the common code of conduct SE.
- **Applicability**: copyrights and intrinsic rights differ between countries^{anon, anon}. In Argentina, licensing aspects for mandatory publishing of publicly funded data are under discussion^{AR}. In Brazil, the proposed changes are not compatible with existing laws on data publishing^{BR}.
- General: more discussion is needed with the Nodes^{MX}.

Question 3: Substantial risks arising from this approach

- Loss of buy-in from the publisher community, if a) the necessary support mechanisms (attribution, usage reporting) are not in place and properly introduced, b) the norms are not agreed and rolled out, and/or c) the concerns around commercial use have not been sufficiently addressed before the licensing change comes into effect^{AR, ZA, anon, Mergen, Woods, Penev}.
- Data loss, resulting from a), institutions or individuals withdrawing their data in disagreement with the new rules, b) institutions, individuals or networks who are not in a position to waive rights, and/or c) community dynamics, following the lead of an institution or group withdrawing data^{various(App.2)}.

Question 4: Contribution to collaborative documentation

Category	VP	APC	APO	Other	Anon.	Total
Interested in collaborative documentation on this topic SMEBD, Finland, Germany, Norway, Ireland, Argentina, Sweden, Colombia, South Africa, Brazil, Israel, Netherlands, UK; Mergen, Bentley, Noé, Hagedorn, Penev, Roper, Desmet, Brosens, David, Barnes, Baxter, De Wever, Stroud, Paton & Buffery	10	2	1	14	4	31
Interested in collaborative documentation, but ideally through suitable (national) agencies Egerton, Corner, Hunt, Frost, Whitbread, Lightfoot	-	-	-	6	-	6
May be interested in collaboration, if applicable	-	-	1	2	-	2
No, not at this time, not applicable, or not necessary Korea, Austria, Madagascar, Mexico; Capers, Wood, Asencio, Leya	3	1	-	4	2	10

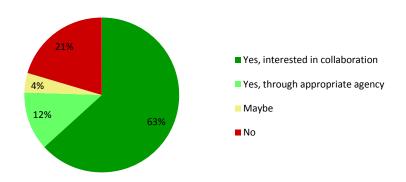


Figure 3 - Breakdown of responses to question 4 - ALL responses

Appendix 1 – Comments received in response to question 1 and 2 (General approval of the proposal, and additional factors to be considered)

This appendix brings together comments from responses underlying the summary issues regarding questions 1 and 2 identified in the main document.

- In general, respondents approve of the concept of free sharing and re-use this goes both for
 responses endorsing the full approach, as for others having some reservations concerning detail and
 implementation
 - o "Most have no problem with free sharing, re-use." (Mergen)
 - "Data publishers do not now have copyright authority over their data, so we would be losing nothing. Free use of our data is the best policy." (Capers)
 - o "I agree not to use CC licenses. It is better to simply state the terms, e.g. All data published through GBIF is free to use but the DATASET must be attributed (not just GBIF) in the following way: (...). However, I am not sure that the original data providers loose copyright by licensing GBIF to publish the data. Several journals now let the copyright of the CONTENT of papers reset with the authors but the paper in its published form rests with the Publisher. I suggest this is a optimal approach for GBIF." (SMEBD)
 - o "The Museum für Naturkunde, Berlin, Germany endorses the approach. It will contribute to theusability of GBIF as a source of research data, without undesired legal obstructions, while at the same time highlighting the scientific requirements of citation and attribution. The approach matches well with the Bouchout Declaration." (Hagedorn)
 - o "(...) this approach will clarify the licensing status of data available through GBIF and is the first step towards developing a set of community norms. As factual (e.g. DarwinCore) data is generally not subject to copyright this approach should not materially affect the rights of data publishers." (Woods)
 - o "(...) this is the only fair, sustainable and workable position. I imagine many people will wonder why licences equivalent to CC-BY could not be used, but in meta-analysis this could lead to the requirement to acknowledge many thousands of data providers. Furthermore, it is not usually the data providers that actually collected the data, more often than not they are just middle-men. It is better that the data is free to use and that scientific norms for data citation and acknowledgement are used." (Groom)
 - "We support the outlined approach and do not see the commercial use [of data] as a problem. The CCO license model can be used as a waiver (not an actual legal license) with data." (FI)
 - o "It should be communicated in the documentation (proposed change 1) that this approach is most honest and legally valid one GBIF can take, while still catering to the wishes of the GBIF community." (Desmet)
 - o "Yes, that is probably the only way to make this work. It will remove the threshold of making use of GBIF data for everyone who is not familiar with legal data issues, licencing etc and is possibly afraid to receive a legal claim because of fraudulent data use. It will also guarantee that end users do not have to worry about legal issues when they forget to attribute someone or something or when this is simply not possible." (Brosens)
 - o "Having open access to biodiversity data should be a top priority for GBIF." (Asencio)
 - o "In principle, the German GBIF Nodes are in favour of implementing data access free of obstacles. The arguments in favour of releasing our (textual) specimen and observation data under CCO as opposed to putting the data under a CC by (or similar) license may be summarised as follows:
 - Legal applicability: CC licenses are based on copyright. It is unclear to what extent NH metadata is copyrightable at all (Patterson et al. 20145), and, if they were, it is unclear who the actual holder of the copyright is.
 - Willingness to enforce: While there are numerous examples (e.g. in the context of large-scale analysis of GBIF data) where individual right statements were not followed, we are not aware of any institution that tried to enforce attribution for collection metadata in court.
 - Comparison to other forms of publishing: there is no legal requirement to cite sources from literature.
 However, citation of sources (i.e. attribution) is morally enforced because it is part of the recognised good practice in science.
 - Obstacle to research: especially in scenarios that would lead to attribution stacking (a situation where a
 very large number of rights holders should be cited, .e.g for an analysis across data sets), a legal
 requirement for attribution may lead to exclusion of data from analysis.
 - Part of the mission: natural history museums and herbaria are knowledge institutions, traditionally
 making their collections accessible to researchers and the public why taking a different stance on the

⁵ Patterson et al. 2014. Scientific names of organisms: attribution, rights, and licensing. BMC Res Notes. 7:79. doi: 10.1186/1756-0500-7-79

data once it is in electronic form? As stated by Creative Commons6: "Much of the potential value of data is to society at large — more data has the potential to facilitate enhanced scientific collaboration and reproducibility, more efficient markets, increased government and corporate transparency, and overall to speed discovery and understanding of solutions to planetary and societal needs." [http://wiki.creativecommons.org/Data].

■ The reasons to avoid the "non-commercial" clause in licensing (or, as suggested, inventing our own legal definition of "non-commercial" are outlined in the article by Hagedorn & al. (2011)7, who come to the conclusion that "NC licenses … provide a broad protection for the copyright owner, but strongly limit the potential for re-use, collaboration and sharing in ways unexpected by many users". In addition to clarifying the problems with the NC clause, the article provides background on CC licensing in general.

The German GBIF-Nodes are in favour of working on restricting the number of legal options under which GBIF data are published. Also, we suggest to implement a filter for searches on GBIF data that includes licensing criteria." (Germany)

- " (...) Ethic and responsibility of scientists to the rest of the society will no longer be forget for careers aspects only.
 (...) formation of new researchers, careers evolutions and recognizing must be indexed on outreach and capacity of communicate, share datas and results. Non open reviews must be excluding of the evaluations of research!" (David) " Such [copyright] claims would be void in all but a few exceptional cases, and this way both providers and end users are made aware of this. (...) GBIF would be freed from non-binding agreements as to what provided data may be used for and by whom, while setting new standards that are more in line with the legal reality and GBIF's objectives. " (Norway)
- o "[Ireland] recognises the strategic importance for GBIF data to be as free from restrictions as possible, in order for it to achieve its potential in providing a global biodiversity data service to international, regional and national partners and initiatives. It should also help to clarify the objective for mobilising data through the GBIF data portal. In this regard, a Creative Commons license is identified as the preferred option. However, Ireland recognises the value in having the ability to flag some datasets as restricted from commercial use." (Ireland)
- o "A good portion of biodiversity data (i.e. facts and measurements) are not copyrightable in some jurisdictions, so many data publishers are already not in a good position to be claiming copyright over data." (Baxter)
- "GBIF-Sweden together with its host + funding agency supports the proposed way forward. Clarity in this field is most important" (Sweden)
- o "I agree that it would be much more productive to invest in the development of a common set of "community norms" and clear recommendations to ensure proper attribution of datasets. (...) For new datasets, I believe this creates the needed clarity, and avoids the legal grey-zone as described in the consultation text" (De Wever)
- o "The National Biodiversity Network (NBN) supports GBIF's direction of clarifying and simplifying data publishing licenses to encourage use of data. A robust system, that is well understood by data publishers and data users, based around some form of Creative Commons has many advantages. The NBN believes that a move to more openness would be beneficial for biodiversity in the UK (and globally) in terms of research, conservation and education. Furthermore, a Creative Commons license is consistent with the INSPIRE Directive. However,(...)" (Stroud)
- o "We believe it is quite important to reinforce the strategy for promoting data sharing and using the CCO license among the publishers (and potential publishers)" (Colombia)
- "South Africa recognises the importance of free and open access to data, but does not support the position to waive copyright." (South Africa)
- o "The institutional policy of [our university] does not allow us licensing. Moreover, data from the collections deposited in the [physical collection] are public and available to the entire scientific community, both on site or on line, observing the rules of no-commercial use and citation existing today." (anonymous)
- o "[We support the proposal] if it does not hinder the sovereignty of the original data publisher and their right to be quoted in the citation." (Madagascar)
- "As the document clearly states, the developments around science data and data in general will move in a direction
 where it will be impossible to cling on to your copyrights (big data, individual records through web services, etc.). So
 yes, we need to establish a clear common GBIF statement on this and a kind of CCO status seems the most
 appropriate." (Netherlands)
- o "Clarifying licences issue is an important thing and we support the idea of CC0 with norms" (anonymous)
- o "I agree that the use of a CCO license would result in a more robust system which is well understood by both data publishers and data users. While I am perfectly aware that this will not be palatable to all data publishers, and may result in withdrawal of some data from GBIF, I believe it is the only honest, transparent and legally valid option.

 GBIF has a responsibility to provide leadership on the issue of open data access and to provide a system that

condition-implications-for-the-re-use-of-biodiversity-information

⁶ http://wiki.creativecommons.org/Data

Hagedorn, G. et al. 2011. Creative Commons licenses and the non-commercial condition: Implications for the re-use of biodiversity information. ZooKeys 150: 127–149. http://www.pensoft.net/journals/zookeys/article/2189/creative-commons-licenses-and-the-non-commercial-

facilitates fair and responsible data use with a firm legal basis. Many data publishers already strongly advocate open access to data for commercial and research purposes. GBIF standards should reflect the real legal situation regarding copyright and factual DwC-Archive data and the community's aspiration for open access, while acknowledging that some data publishers are not yet ready to be as open as they would like, and engaging with them to help them reach that position.

GBIF's proposal to enable datasets to be flagged as restricted from commercial use demonstrates their awareness of an important barrier to open data access affecting some data providers in some countries. Not all UK data suppliers will want to apply this restriction, but for those who do, it is important that the flagging system is in place, and that its legal and practical implications are fully understood, before the CCO license is introduced. From the point of view of data users, it is vital that this flag operates as a machine-readable filter to exclude datasets from downloads for commercial purposes.

I applaud the success of GBIF in increasing the use of biodiversity data for research. I understand that adoption of a CCO waiver does not remove the moral obligation to acknowledge data sources and I welcome GBIF's proposal to establish a citation model built upon the Digital Object Identifier (DOI) system." (Lightfoot)

- o "In [our country], the Open Access Policy was enforced since the beginning of May. All public funded scientific outputs (publications and data) from now on need to be deposited in open access repositories. The [environment ministry] also is publishing its geographic information system ([URL]) in open access platforms." (anonymous)
- o "The UK supports the position of clarifying and simplifying the licensing to encourage as great a level and range of use of the data as possible." (UK)

Where the answer to question 1 (support for the approach) is negative, the following additional arguments are supplied:

- "Open data sharing should be encouraged, but CCO should not be enforced." (Mergen)
- o "The approach outlined does not seem to have considered CC 4.0, which has been developed to tackle issues and improve clarity surrounding publishing of data. In particular, CC 4.0 covers sui generis database rights which apply throughout the EU. Summary of CC 4.0 here: http://blog.okfn.org/2012/08/15/cc-license-version-4-0-helping-meet-the-needs-of-open-data-publishers-and-users/. I do not oppose the option (and indeed encouragement) of CCO waiver and community norms, but it should come as part of a range of licensing options so as to enable publishing where rights cannot be waived." (Roper)
- o "(...) if we succeed up to now to have some data from publishers in Benin it is mainly because those who accepted to provide their data feel that they remain data owners and still have rights on their data. It is difficult at least in some African countries to handle this issue now. I advise GBIF to postpone that debate because it is quite sensible and can have inconvenient result on data mobilization" (Benin)
- o "Whilst we support the objective of making licensing simpler, it has been shown within the flow of data already in place in the UK, that licences are often hard to police when misuse is evident. Due to this situation and including the NBN Gateway as an example portal, DBRC and many other data providers are rightly cautious of sharing biodiversity data openly where sufficient assurances cannot currently be given. As such we cannot support the position of a copyright waiver in all cases. This is likely to be counter productive and actually reduce the amount of data available because unlike many other countries frameworks, the UK has a system of data collation, management and dissemination only partially assisted by public money. This need to principally self-fund is not ideal, but remains fact." (Egerton)
- o "We cannot currently support the proposed position as it is in direct contradiction to our agreements with our own data providers. We do not believe we are able to waive rights to data that is supplied to us by third parties. In order to retain the quality and accuracy of the data, the data supplier or designated custodian of the data needs to retain rights to exercise maintenance and editing of the data. Current standard guidelines for data citation for data published on the NBN Gateway is not being followed and we wouldn't expect 'publishing and usage norms' to be followed either. Rather than removing the problem entirely, more effort should be put into education and enforcement of current terms and conditions for use of data." (Barnes)
- "Whilst we support the objective of making licensing simpler, which is beneficial to both data users and data providers, in short, no, we don't support the position of a copyright waiver in all cases. This is likely to be counter productive and actually reduce the amount of data available. It needs to be understood that organisations such as records centres collect data recorded by a variety of different people, and it is these recorders who own the copyright of their records, and therefore only them who can waive it." (Hunt)
- "With a strong long-standing tradition of skilled amateur involvement in natural history, many of the biological records in the UK are not collected by professional taxonomists and surveyors, but by unpaid knowledgeable enthusiasts. By taking into account their needs and motivations the UK has an outstanding resource of detailed information on species distributions available to all. Although the data is not published, much is shared through the UK's national portal, the National Biodiversity Network Gateway (http://data.nbn.org.uk) with terms and conditions attached, including the possibility of adding individual provider's own terms and conditions, but no data licence is currently applied. User resignation is required so that data providers can see who is using the data and why and this helps motivate further sharing of data by public institutions such as Tullie House Museum and funding for support and data aggregators such as the Cumbria Biodiversity Data Centre. --- We recognise a multitude of data terms and

conditions can be confusing or burdensome for the end user, and so we do support the idea of introducing clear data licensing on GBIF based on the Creative Commons model, but we do not support a blanket CCO approach." (Frost)

- "The way the GBIF proposes using the CCO waiver with restriction is contradictory. If a dataset is published using CCO, how can someone flag it as non-commercial? It does not seem to be possible. As with the other CC licenses, CCO is not accepted in all jurisdictions, according to the Open Source Initiative (OSI), which does not recommend using it.(...) Publishers might be discouraged to improve the quality of data sets if they have waived their rights on them." (Brazil)
- " (...) Whereas the Forum [National Forum for Biological Recording] strongly supports the ambition of simplifying access to and use of data (as exemplified by the Bouchout Declaration), there are several options as to how this might be achieved. Whilst each of these would potentially result in some greater or lesser degree of less welcome consequences, GBIF's proposal, the rapid, untested and mandatory imposition of a Creative Commons C-0 waiver of rights as the only option, would be certain to maximise the negative. Following the initial 2013 consultation, it would have been appropriate to have conducted a detailed appraisal of the licensing options and the possible combination of means by which those able to do so could follow the CC-0 route and by which those constrained from doing so might agree to a brief suite of optional terms and conditions. Instead, the suggestion made by some of the respondents to that consultation appears to have persuaded GBIF to the view that any licences wouldn't be enforceable so, regardless of any consideration for those willing to provide access to data but unable to waive rights to it, CC-0 should be the only option. Waving the banner for Open Data, however welcome this might be where practicable, seems to have resulted in the very real difficulties this will pose for some being overlooked. In this regard, the present consultation is something of a disappointment' It is worth commenting that in relation to the number of Data Publishers that GBIF proudly proclaims (609 according to the banner today, yet 777 according to the number of data search entries), the 32 respondents to the initial survey may well not be representative of the views of its Data Publishers as a whole. Secondly, organisations such as the UK's National Biodiversity Network (seemingly misnamed as the National Biodiversity Information Network in relation to endorsements of Data Publishers from the UK) are aggregate Data Publishers providing access to Data Sets supplied by a multiplicity of diverse organisations and individuals who are the actual data owners or (e.g. local record centres) possibly intermediaries themselves, with no title to the data they are able to provide in accordance with agreements to which they are bound." (Whitbread)
- "The UK does not support the adoption of CCO (where there is no obligation to cite the sources of the data) and would rather see the adoption of CC-BY (or even CC-BY-SA). Given the efforts that many data collators have put into creating and maintaining data collations within the UK and elsewhere, the UK considers that it a reasonable stipulation that there be an obligation on the users of the data to cite the sources. The UK accepts that even if this is stipulated it will not always occur but believe that this is not a reason to wave the obligation. The adoption of CC-BY at least ensures that there is some recourse open to the providers which can be instigated in significant cases. It should be noted that CC-BY is very much in line with the UK Governments Open Government License (which stipulates that the source should be acknowledged).
 - Therefore, overall the UK supports a single and open licensing model for GBIF but one **which requires the attribution of the sources**. We believe this is a reasonable stipulation (particularly with current technical developments) and is likely to lead to a greater volume of higher quality data being made available." (UK)
- Some communities and aggregators have established models that need to be taken into account –
 A number of respondents raised the point that in their communities, data exchange agreements
 existed between data contributors and publishing agent that could not be easily overruled. These also
 touch on the business model of data services.
 - "Devon Biodiversity Records Centre (DBRC) works in partnership with local and national data providers to mobilise species, habitat and sites data, and acts as a cost effective and quality controlled mechanism through which data flows. A high proportion of DBRC data is also self-generated through proactive survey and recording schemes & projects, therefore the integrity & sustainability of the funding model for this work must be protected to support continued collation of new data in the county." (...) The resolution of data provided to users by DBRC is normally tailored to recognise the actual need of the organisation/individual, and whether its use is commercial or not for profit. The principle variance is resolution at which it is provided depending on the above. Potentially an agreed resolution at which data could be released could be used. One which did not compromise the integrity of our current system. (Egerton)
 - "We cannot currently support the proposed position as it is in direct contradiction to our agreements with our own data providers. We do not believe we are able to waive rights to data that is supplied to us by third parties."
 (Barnes)
 - o "The NBN Gateway currently provides two levels of access to data 'public' and 'enhanced'. Our data providers are happy for a certain level of information to be displayed at 'public' level (e.g. 10km resolution) whilst restricting the use of further data to vetted users. GBIF should consider whether a two tier system such as that provided by the NBN Gateway would be suitable" (Barnes)

- o "By waiving all rights to data, there may be a business risk to the current operations of the GLNP. Income from commercial use of data is used to ensure datasets are kept up-to-date and of high quality. If this income stream is lost, the GLNP may not be able to continue to provide up-to-date and high quality data." (Barnes)
- o "BRERC is a not for profit partnership and only receives funding sufficient to cover approximately 60% of our core funds, the other 40% is made up from contracted work and our data search enquiry service. Subject to sensitive data constraints, anyone can ask BRERC for BRERC held data at full resolution for a charge that covers the costs of that service. Without the money from the search service BRERC would not be able to continue to provide data to the NBN or GBIF. The reduction of BRERC data to 1km resolution on the Gateway and GBIF is to ensure we are able to continue to maintain our search service. Unless extra compensatory funds were made available and if the data were made freely available at full resolution then commercial and other users would no longer pay for our services which in turn would mean we could not continue to collate, manage and disseminate the existing and future data. This would mean much less data being made available to anyone. To ensure stability and continuity BRERC wish to see appropriate safeguards introduced/maintained to ensure data we provide to the NBN (and hence GBIF) at full resolution is not made freely available. We are happy to provide public access at 1km resolution." (Corner)
- " Any approach to the release of data that is not owned or collected by the party that wishes to release, publish, etc. can only occur in consultation with the data custodian/owner. What steps are being made operational that ensures that the integrity relationships that [aggregators] have with their data suppliers/contributors/recorders are not being jeopardised? [Our institution] is keen that data is made available and utilised for conservation, etc. purposes. What we cannot afford is to side-line our core function. Any solution must ensure that the expectations of the local recording community are addressed and met." (anonymous)
- o "It is necessary to have in mind that providers are part of either the public or private sector. A general CCO licence for data shared through the GBIF network would not touch the public sector providers, since it is the way to go forward envisaged by governments throughout Europe cf. OpenData declaration by the UNESCO 2004. But providers from the private sector e.g. associations (BirdWatch) or small environmental agencies and NGOs or even individuals contributing to GBIF may decline their support and withdraw "their" data from the network." (Austria)
- "The United Kingdom has a long tradition of biodiversity data being collected by volunteers and not for profit organisations. This can be supported by investment from the commercial sector, if free access to data is restricted in some way for commercial users. In order to preserve this valuable investment, ALERC cannot support a simple blanket waiver. More details on data exchange in the UK are set out by the NBN Trust's seven data exchange principles." (Hunt)
- o "Tullie House Museum would not be in a position to continue to upload data to GBIF in the case of the records held by Cumbria Biodiversity Data Centre if a CCO licence was applied. This is because the funding for continued existence of the Centre means that commercial use of online data is not possible at time but more importantly because the Centre does not own the data and cannot waive the rights of the providers. There is a clear difference between these data holdings, and the data the museum holds about its own collections where the museum would be able to choose whether to waive rights, but would still have concerns about commercial use. With a choice of attribution licences including no commercial use options, the Museum would be able to choose the appropriate licence of its own data, and publish derived summary data from the collated records data (for example presence at 10km resolution by decade)." (Frost)

Community Norms

- "It is very confusing that the norm in science and popular media would be to refer to GBIF as the Publisher. Authors of books and papers do not call themselves the Publisher unless they self-publish. I agree assigning DOI is a high priority." (SMEBD)
- o "There should be wide discussion on the suggested community norms (...) [for the] publishing of sensitive data with CCO. For example flagging of local endangered species and expressing that details (including more accurate locality information and/or more precise coordinates) can be requested from the source for research. This is especially important if we wish to publish collection data from developing countries as part of data repatriation. (...)." (Finland)
- o "By introducing a mechanism for expressing "levels of publicity" we could enhance the willingness of participants to mobilize more of their data to be as public as possible. By defining standards for sensitivity and especially for expressing whether more/what-kind-of details can be acquired under stricter terms, we can assure data publishers that the risks involved to sensitive data are manageable. This would make it easier to involve the governmental sector to also share (not only use) the data through GBIF." (Finland)
- o "The continued use of a Data Use Agreement will provide a hindrance to reuse of published data and at best remain in a legal grey area. We recommend to seek alternatives including community norms for citation and to remove the Data Use Agreement restrictions. It is important that both the legal and practical implications of requesting not to use data in certain contexts (such as commercial, however this is to be defined) rather than prohibiting it are clear to all. This is one of the main challenges for the documentation and code of conduct." (Norway)
- "Ireland would be supportive of the approach proposed by GBIF to clarify licencing issues while also promoting fair use and citation of data via a set of community norms." (Ireland)
- "Provided the data we submit to the NBN (and therefore to the GBIF) is safeguarded by the terms and conditions regarding resolution and commercial uses then BRERC is not unduly concerned at the proposal." (Corner)

- o "I think that the Canadensys norms are a good launching point for the discussion." (Baxter)
- o "Another norm for which I will advocate is to encourage users interested in raw biodiversity data to retrieve it from the official source (i.e. GBIF or a more local portal) whenever possible so that they are always working with the data providers' most up-to-date work." (Baxter)
- o "We believe that it is essential for users to formally oblige to abide by a common code of conduct accessing data both through international and national/thematic GBIF-portals." (Sweden)
- "We welcome the initiative to set up clear community norms, which can be endorsed by the wider GBIF network"
 (De Wever)
- **Explore other licensing options** a number of responses suggest to explore other options, ranging from dropping a licensing scheme altogether and concentrating on community norms and best practices to a number of other licensing schemes
 - o "As CC licences do not seem appropriate for data, why not drop them completely and work with Ask proper citations, Flag if have special clauses like Noncommercial use or better only for "non-profit uses" which seems a better definition of what is meant; Agree that if shared to GBIF data has to be openly accessible and remain as such also when third parties use them; In some rare cases add that no derivative work is allowed; Give contacts for feed-back; Keep the notification to the original provider on usage of the provided data " (Mergen)
 - " CC 4.0 should be fully considered. Publishers should certainly be invited to use CCO, but it shouldn't be to the exclusion of other licenses. I see it as better to publish under a license than not at all, even if that license imposes extra burden on the user. I would encourage GBIF to consult with CC, the ODI and OKFN to gain a deeper understanding of options." (Roper)
 - o "If GBIF decides not to adopt the proposed changes, there should be a restriction on the terms and conditions that data publishers are able to attach to their datasets. Allowing providers to set their own widely varying restrictions makes it impossible to determine the rights for individual records within a large download. This problem is one reason I support the proposed changes." (Asencio)
 - "The issue of copyright of data hasn't been fully explored whilst the presence of a specimen labelled 'xyz' in a
 museum may well be a 'fact', where no specimen or supporting evidence exists for the record (e.g. field
 observations) there is originality in the selection and arrangement of data and substantial investment in obtaining,
 verifying and presenting the information." (Barnes)
 - "I think that sui generis database rights, as addressed in CC-BY 4.0 should be fully considered in this process.
 Although I believe that the CCO waiver is the most appropriate choice for GBIF biodiversity data, the reasons for rejecting a sui generis option should be properly documented. " (Baxter)
 - "We also believe that use of Creative Commons licences should be considered. CC BY-NC is a license that allows for data to be used freely, so long as it is not for commercial purposes. This is in line with United Kingdom data sharing systems which appear to be beneficial to the majority of UK data suppliers and users." (Hunt)
 - "(...) given the considerable efforts that many data collectors, curators and publishers have put into creating and
 maintaining data collections the NBN considers that there should be an obligation on data users to cite all sources.
 Therefore, a Creative Commons Attribution license (CC-BY) is a preferred solution allowing re-distribution and re-use
 of a licensed work on the condition that the creator is appropriately credited." (Stroud)
 - o "One additional factor to consider is that CC v4 licenses cover database rights so subsistence of copyright is not the only intellectual property right to consider under Q[uestion] 1." (Paton & Buffery)
 - o "In contrast to GBIF's proposal, NFBR suggests that, whilst enabling those who wish to adopt CC-0 to do so, GBIF needs to (a) provide a somewhat wider set of Creative Commons licensing options, and (b) require its Data Publishers to adopt a consistent approach to their Data Rights Management which will be compatible with this. This 'short suite' compromise would provide the required simplification and greater open data access whilst minimising the likely negative effects. (...) Creative Commons licensing appears to be well developed (by effective communities), offers a range of advantages including (for the EU region) being INSPIRE-compliant. Attribution licensing (CC-BY) which enables data dissemination and re-use with the proviso that the owner is credited, or (CC-BY NP) is not available for commercial use should be investigated further. Whereas this might take more time to grapple with than the blanket CC-0 proposal, it is likely to achieve the stated objectives as well or better whilst avoiding the drawbacks that have already been identified or seeming to have no regard to those members of the GBIF community that it would affect. Moreover, improving access without impacting of data flow would benefit research, planning, conservation, environmental management, mitigation and adaptation." (Whitbread)
- Other essential components that need to be considered in the implementation plan and implementation sequence — Before licenses can be changed, the appropriate supporting mechanisms need to be in place (citation-, attribution- and use and re-use tracking mechanisms, supported by DOIs)
 - "The main concern of the community according to past experiences is proper citation of the data source, owner, institution, involved individuals. There is an expectation of feed-back on re-use and possibilities to react in case of misuse. "(Mergen)

- "Acknowledging use of our data is always appreciated, although we understand it is difficult to enforce this."
 (Capers)
- o "As I see it the issue is less about copyright and more about providers being able to choose a CC-BY license to ensure attribution of data. The fact that this could be addressed in a norms document is not lost on me but I would prefer to see this explicitly associated with each record downloaded for my data in an attempt to try and enforce this even further as it is always a chore to try and track down usage of data through publication or Genbank submissions."

 (Bentley)
- "Citing data providers through data set citation created by GBIF is important. However, it is also important that individual records, as published by the original institution become dereferenceable, citable, and semantically linkable. The Linked Open Data (LOD) framework should be supported in GBIF, providing the URI for the physical specimens as an LOD data citation and as a clickable link in the user interface for humans." (Hagedorn)
- o "Yes [for support to the approach], given that GBIF elaborates efficient and working mechanisms for data provenance, attribution and citation. The community norms in data citation may not work, if a proper mechanism is lacking. Provenance, attribution, data citation and data usage mechanisms should be elaborated simultaneously with acceptance of the new policies. [I foresee substantial risk] if data attribution and citation mechanisms are not in place when the new policy will be launched." (Penev)
- o "I also look forward to a more robust citation model to ensure that publishers receive appropriate credit for the data they have collected, curated, and supplied." (Woods)
- "Of course GBIF should provide all the possible mechanisms for attributing, citing... the data in the most easy and efficient way." (Brosens)
- o "One of the barriers to data use is described as the onerous task of collating data provider acknowledgments. As the majority of our data providers are not obligated to provide data to the wider public and receive no remuneration as a result of providing the data; acknowledgement of use of their data is one of the few direct benefits of them providing it. These 'barriers' are easily overcome in such publications as Bird Atlas 2007–11 (BTO 2013) and the Atlas of Dragonflies in Britain and Ireland (CEH 2014) where each individual contributor is acknowledged. Building a system on GIBF which automatically assembles data provider information would not be a great technical challenge." (Barnes)
- o "As the proposal is stated, changing to CCO would have an immediate effect, while the complementary norms on fair use are only planned for the future. We see attribution as an essential part of data sharing and would prefer not to relax licensing without a compensatory action to promote attribution." (Argentina)
- o "The development of a working citation mechanism [is] necessary" (Sweden)
- "Finally, we fully support the proposal to establish a citation model built upon the Digital Object Identifier (DOI)
 system, and ask that this be in place before the agreed Creative Commons license or waiver is introduced." (Stroud)
- o "Another factor to consider is archiving of non-static datasets when attributing for research. Due to the nature of the data collected, and taxonomic advances, the data held is regularly revised. For research purposes, it is important that unique identifiers are used for datasets, so that if a dataset is subsequently revised and data points changed or removed it is clear this revised dataset is different from the original dataset used." (Frost)
- o "Issues of attribution and citation are key and needs to be dealt with alongside considerations of licensing. (...)

 Overall mechanisms such as persistent identifiers should be considered." (South Africa)
- "The benefit of reuse and increased citations for scientists are important since the currency of science is citation. I therefore suggest that the Norms are developed/rolled out with the Community first and discussions about proposed changes be done with them.
 - There are resource implications for this process if changes are recommended and stakeholders nationally don't agree to these changes. The resource requirements of convincing people to sign away copyright data could be best spent on developing alternative citation methods. Furthermore, the ceding of copyright can alienate the community leading to a decline in data sharing." (South Africa)
- "The three data holders did not bother about the licencing but are concerned with citation and sovereignty."
 (Madagascar)
- "The major concern of data providers is that they should be accredited for their data. We feel that GBIF should offer the option of CC BY licensing in addition to CCO." (Paton & Buffery)
- "We would suggest that (a) improving the frequency and quality of citation and (b) tracking data use which would help to highlight the value of biodiversity information (and GBIF) to different sectors and for different uses (as well as showing where it was not being employed) - should be seen as priorities for GBIF during this process."
 (Whitbread)
- o "To compensate for any doubts that will arise with data providers it will be needed to put more efforts on issues such as stable ID's for records and datasets." (Netherlands)
- o "Changes should be implemented only when a fully implemented citation system is in place to allow recognition to the data publishers." (anonymous)
- Commercial use even though non-commercial use cannot be policed, the possibility of flagging datasets intended only for non-commercial use is important to a number of publishers and users, though there are also arguments against.

- o "Some have issues with commercial use" (Mergen)
- o "You have outlined the difficulty in preventing for-profit use of public data, and we understand this. I can imagine cases where our data might be used in ways to which we would object, but it is impossible to prevent these uses in advance so there is no way for GBIF to reasonably restrict data use in a way that would avoid this potentially undesirable situation. We will just have to live with the occasionally uncomfortable uses in order to ensure that our data are freely available for the more frequent uses that prompted us all to digitize our collections in the first place." (Capers)
- o "If data use has restrictions then it should simply not be published. It is impractical to have some datasets with different conditions, such as no commercial use. Such restrictions are already ignored and will continue to be. Such restrictions are NEVER applied when publishing in scientific journals and books, so why should they be through GBIF? Also, contrary to what is implied, there are a range of interpretations of what commercial use means. IF a commercial organisation publishes a scientific paper can it not use some GBIF data without further permissions?" (SMEBD)
- o "While not legally binding I support the ability to flag data for non-commercial use." (Woods)
- o "Because copyright is not applicable, if any non-commercial restrictions are used there must be an embargo period.

 Otherwise licence agreements will continue forever." (Groom)
- o "However, data publishers should be able to set restrictions on the commercial use of their institution's data when deemed appropriate. The ability to flag datasets that are not to be used commercially must therefore be assured. While I have concerns about these restrictions on use having no legal basis, I do think the disadvantages pale in comparison to the consequences of limiting access to data, which is primarily in place to drive biodiversity research forward." (Asencio)
- o "(...) it should be essential to flag datasets when data publishers do wish to restrict from commercial use. In fact, Korea has those datasets currently." (Korea)
- o "The reasons to avoid the "non-commercial" clause in licensing (or, as suggested, inventing our own legal definition of "non-commercial" are outlined in the article by Hagedorn & al. (2011)8, who come to the conclusion that "NC licenses ... provide a broad protection for the copyright owner, but strongly limit the potential for re-use, collaboration and sharing in ways unexpected by many users". In addition to clarifying the problems with the NC clause, the article provides background on CC licensing in general." (Germany)
- o "Ireland recognises the value in having the ability to flag some datasets as restricted from commercial use." (Ireland)
- o "Our community is fond of the ideas of attribution and non-commercial use, and we have promoted GBIF for many years on that ground. A change to a more relaxed licensing should be done with time, and after those principles are taken care by other actions." (Argentina)
- "I also support the portion of the proposal regarding the ability for data publishers to indicate which datasets they
 wish to restrict from commercial use. This will be an important part of the discussion for many data providers."
 (Baxter)
- o "It's important that commercial use can be restricted by the data providers. Thus it is necessary to provide a kind of definition of commercial use." (Austria)
- o "We expose no datasets where limitations for commercial use are stated (rather all our providers have formally agreed upon free and open access to data a few potential providers have declined participation)." (Sweden)
- "[We] appreciate the efforts to work out what constitutes commercial use and what is permitted under commercial use and what's not. Providers to BioFresh have e.g. expressed their concerns that they would not want anyone to "sell" their data." (De Wever)
- o "(...) There are already many forms of commercial use of biodiversity data in the United Kingdom. The NBN will work to encourage all biodiversity data owners to move towards an operating model compatible with data made available under a CC-BY license. NBN expects biodiversity data users to ensure that they respect and acknowledge the proposed CC-BY licence and operate in a way that supports the sustainability of biodiversity data owners. The proposal to enable some datasets to be flagged as restricted from commercial use is seen by the NBN as critical. We feel that an open data model from which some suppliers can opt out will be more conducive to increasing data use than a model that restricts use of all data based on the wishes of the most conservative data publishers. The ability to flag a dataset as restricted from commercial use will enable those who do not need to comply with INSPIRE to opt out if they wish." (Stroud)
- "Commercialisation has been raised, but what is more important is that where data is used for commercial purposes, which can certainly benefit the green economy, science and society, that due credit and acknowledgement is provided." (South Africa)
- o "Restricting use in non commercial activity would comfort data publishers in some extent but we know we cannot have that in control, good to mention anyway." (Madagascar)
- o "The issue about flagging datasets that should be restricted from commercial use is something that should be very carefully discussed. We will create a division in the GBIF data, is that desirable? Do we want to tell the

⁸ Hagedorn, G. et al. 2011. Creative Commons licenses and the non-commercial condition: Implications for the re-use of biodiversity information. ZooKeys 150: 127–149. http://www.pensoft.net/journals/zookeys/article/2189/creative-commons-licenses-and-the-non-commercial-condition-implications-for-the-re-use-of-biodiversity-information

- pharmaceutical industry that they can not use all GBIF data? Open data are there to stimulate scientific, societal and economic developments, this includes commercial use of the data." (Netherlands)
- "Commercial use is controversial and difficult to define and there is a divergence of views within the data provider community in the UK. This area needs further consideration and the UK intends to undertake additional consultation here." (UK)
- Other data aspects licensing approaches and options may be different for other types of data than
 occurrence records, including multimedia content, taxonomical work, metadata etc, as well as
 sensitive data and data resolution
 - o "Concerning multimedia, the other machine readable CC license models should be supported by GBIF" (Finland)
 - There should be (...) further plans for technical solutions that enable safe "partial or coarse resolution" publishing of sensitive data with CCO. (...) There should be a mechanism for preventing "accidental" publishing of sensitive locality data from other countries without consulting the locals." (Finland)
 - o "At some point a standard license for the metadata has to be considered as well. Copyright does apply here (e.g. these metadata can form the basis of a data paper) so something like CC-BY makes sense, while on the other hand some parts (e.g. the dataset title, abstract, keywords or publisher) need to be reusable without burden (which is why e.g. Europeana releases metadata under CCO). Not sure if the discussion regarding licensing/norms for data is a good opportunity to tackle this issue or if it would complicate things." (Desmet)
 - "(...) there can be substantial difference in the "form & content" of the "possibly" published dataset. Using CCO for occurrence data seems like the most obvious thing to do, but a (CC BY-4.0) licence can work perfectly well for "taxonomical works" like Fauna Europaea, FADA and WORMS... I would leave the option for such taxonomical datasets to use a cc-By-4.0 licence (which is also an open licence)" (Brosens)
 - o "Access rights for multimedia objects should be discussed separately from textual datasets, because different legal conditions exist." (Germany)
 - o "Data submitted to GBIF from the UK NBN Gateway is provided at a spatial resolution that the original data provider is happy to have publically available. In many cases this is less than the greatest possible spatial resolution. It may therefore be possible to find a spatial resolution that suits most of GBIF's users as well as providing the confidence to data providers that will allow them to waive restrictions on the use of their data. This idea should be explored by GBIF, who should investigate what resolution their data users require, and whether or not data providers would be happy to relinquish rights to their data at this resolution." (Hunt)
 - "Subsequently we will have to reconsider whether conditions apply differently as new (types of) data (multimedia; cultural heritage with biodiversity contents; genetic information) are successively incorporated into the offerings of GBIF." (Sweden)
 - "Whilst The NBN agrees that an open data model is the way forward, it is important that sensitive information (e.g. exact locations of threatened or at risk species) be redacted or downgraded (in terms of data resolution). Likewise personal information about recorders should be withheld to comply with the UK Data Protection Act 1998" (Stroud)
 - o "Metadata and multimedia should be clearly separated from the primary data, and receive the same "copyright" attention in due time." (Netherlands)
 - o "Multimedia objects have different legal characteristics, it should be taken in account" (anonymous)
- Differences in legislation differences between countries may make a global approach challenging
 - o "One of he most important factors is the disparity in copyright and intrinsic rights between sovereign states. Disputes regularly arise concerning the uses of data in this regard." (anonymous)
 - "Argentina recently issued a law on the obligation of researchers and institutions to make accessible the primary scientific data produced with public funds. The law still has to be regulated, and that will define some operative aspects of access and licensing. The tendency is to follow international parameters." (Argentina)
 - "South Africa has addressed the issue of open access to data as far back as 2000, in the form of the Promotion of Access to Information Act (PAIA). This act elaborates that publicly funded institutions are legally bound to make their data/records accessible.
 - In 2010 the SANBI Biodiversity Information Policy Framework was developed, which strives to ensure easy access to information whilst simultaneously providing protection to sensitive data and maintaining intellectual property rights, and basically managing information in compliance with PAIA. The framework provides the norms and standards for managing biodiversity information.
 - Whilst PAIA gives effect to the constitutional right of access to information (record), this is not an absolute right, as PAIA balances the right to information against the right to refuse access to information, if granting the information would cause harm.
 - The law of each country determines who owns intellectual property rights where intellectual property is created. In South Africa, intellectual property rights include copyrights. IPRs refer to the rights a person may have in certain property or products they have created as a result of their intellectual activity (such as research results and records). The development of a data/record can be considered intellectual property, with a record defined as primary species occurrence data, legacy data, observational and multimedia records and includes without limitation, photographs, images, recordings, specimens, monographs and other documents. PAIA applies to requests for records of private bodies and public bodies. The term "record" refers to any recorded information in the possession of SANBI,

regardless of the form or medium of that information and whether or not it was created by SANBI. This places a huge obligation to disseminate data according to this policy framework to abide by the terms and conditions of release of the data providers.

Many other references including VertNet, Canadensis, and Hagedorn G. 2011, take a different approach to copyright and indicate that data does not contain copyright as it is in itself not a creative work. The South African BIMPF takes into account a number of legislative acts to develop a framework that can support open access but also protect the rights of the data provider. " (South Africa)

- "CCO is not accepted in all jurisdictions, according to the Open Source Initiative (OSI), which does not recommend
 using it. (...)This proposal is not compatible with the laws on publishing information of certain institutions and of the
 Brazilian government." (Brazil)
- o "An effort should be made to clearly explain to data publishers why they cannot claim to copyright rights. This should be done taking into account the different jurisdictional regions. The Secretariat could get assistance from the Participants to identify the situation in each region." (anonymous)

• Clear definitions and communication are very important

- "The text of GBIF deals with these issues but the way it is formulated is rather confusing and seems contradictory?

 Why put CCO by default and then use flags that are equal to CC-BY, CC-NC and such? Plus if a data is put in CCO is it still allowed to put other restrictions, I though normally not? So the approach of GBIF is on the right track, but the way it is implemented may need to be clarified." (Mergen)
- o "The concept rights are related to propriety and the exploitation. It's necessary more definition" (anonymous)
- "Communication on these topics is really hard. I think these risks are real and may have catastrophic consequences, so the communication around this proposal has to be of very, very high (crystal clear, very reactive, anticipating common concerns*, based on solid evidence and experience...). I think it's critical for GBIF to recognize the need for excellence at this level, and take all necessary measures: solid preparation, external review/ help, Canadensys way of communicating on this topic has been great and may be a good basis, IMHO." (Noé)
- "Communication on that subject is really important and it must be clearly explain but it is not easy. It is a real challenge." (anonymous)
- "An effort should be made to clearly explain to data publishers why they cannot claim to copyright rights."
 (anonymous)
- o "Data publishers is a generic term that includes many entities with different roles: institution, collector, curator, database curator, etc. Each one might have different perceptions on how the norms affect their work. The norms also might not affect all of them equally. This needs to be clearly identified in the documentation. It is also important to keep in mind that GBIF is a global organisation, and changes will affect different regions with different cultures. The willingness to open data is not the same everywhere. If we do not want to limit the growth of GBIF In the future, these differences need to be considered at this time." (anonymous)

Stakeholder involvement – both data publishers and data users and their needs have to be considered

- "Content mobilization, especially in mega-diverse countries are still a very critical activity to be undertaken by many Nodes. We need to keep focus on both ends of the spectrum –the data provider and the end user and try to ensure we meet the needs of both. The re-use is equally important, and through this process it would be good to bring the data provider and user closer together through better tracking and understanding of how the data is used." (South Africa)
- o "The plan only considers data users, but not requirements from data providers. Data use without giving credit to the data sources is a very serious issue for data providers (it is not only about commercial use). It is essential to give credit and recognize scientific merit, when data are used by other scientists." (Mexico)
- o "Although the current cultural change in the public administration of [our country] to open access to data, the actual actors (curators, database curators) need to be involved, so that they are motivated to adhere." (anonymous)

• Timescale and migration plan for existing datasets – several times, the question was raised how a change in licensing would be rolled out to existing datasets

- "How will we manage existing data? Data owners have accepted to share them under certain terms, so AFAIK we can't just re-license them without at least their written consent. And we'll need concrete plans for the data that will have to stay in the current terms: flag them and display them in a specific way on the data portal? Something else?" (Noé)
- o "Will existing datasets be allowed to remain on GBIF without a CCO waiver? Some may opt to leave an existing restricted dataset on GBIF if required to waive rights on update." (Baxter)
- "As proposed change 4 suggests that all datasets should be 'migrated' to CCO within 6 months, I would appreciate
 advise on how this can be achieved for datasets for which data holders already chose an alternative license. We
 would not want to withdraw datasets for which the data holder does not agree with this migration." (De Wever)
- o "The NBN Trust would prefer a 12 month time scale for implementation to enable the NBN to align their inhouse Terms and Conditions accordingly to align more closely with those of GBIF and to provide time for in country

- consultation and development work. Please also refer to answer to Q.3 below where we recommend greater levels of consultation by GBIF over the changes proposed." (Stroud)
- o "We wonder how would be the process for the current published data, and for data contacts that may not reply to this process." (Colombia)
- o "I wonder how if later on they decided to license their database?" (Madagascar)
- o "What about the data sets already connected to GBIF? It is possible that they won't agree with this decision. And how we will proceed to make the changes?" (anonymous)
- "The UK biological recording community is diverse and dispersed, comprising recorders from the voluntary, public, private and research sectors. Some data publishers would need to consult with recorders before agreeing to make the data available under a CCO waiver. It is therefore important to allow sufficient time for consultation before applying a CCO waiver to GBIF. I feel the proposed timeframe of 'within 6 months' is too short and should be within 12 months." (Lightfoot)

Appendix 2 - Comments received in response to question 3(Substantial risks)

This appendix brings together comments from responses underlying the summary issues regarding question 3 identified in the main document.

- Loss of buy-in from the publisher community, and data loss Many responses commented on the expected loss of publishers, datasets or data detail. This is often related to other aspects of e.g. attribution. The estimates of the impact cover a wide range, from negligible to substantial.
 - o "A quite high risk, that the community may get cross if especially citations is no kept clearly and it is asked to go to CCO for all content. (...) The risk is that the amount and details of data provided to GBIF may be reduced." (Mergen)
 - "I think there will be a small number of data publishers for whom this approach is not acceptable according their institution's policies." (Woods)
 - o "(...) some data holders will remove records and/or refuse to change licensing of shared data" (anonymous)
 - o "(...) information retention is still very strong in some of our countries. (...) We sincerely think that if this issue is raised, it will be negative on data mobilization in our country. It is also no doubt that Benin is not the only country suffering from this. (Benin)
 - o "Many data publishers, particularly within the UK's NBN (which holds in excess of 95 million species occurrence records on the NBN Gateway alone, not including the data held by Local Record Centres which is not yet present on the NBNG) will not be able to waive rights as required by CCO. In effect, CCO limits the scope of GBIF. CCO also allows for users to extract, remix and republish data without any attribution and without any recourse whatsoever on the part of the original creator. In an age where data is increasingly extracted, remixed and republished, and in as-yet-unknown contexts, it is ever more important that integrity, traceability and attribution are maintained where possible, and that there can ultimately be recourse where agreements are broken." (Roper)
 - "There will always be risks associated with free access to data and its use. Some data publishers may be hesitant to
 publish data within this framework, but it is doubtful that many, if any, will refuse to publish data to GBIF, especially
 considering this is becoming the global standard." (Asencio)
 - o "From our experience with data providers and also particularly with the case of Europeana (the European Digital Library) who changed their contribution scheme to CCO for textual data, there certainly is the risk of data holders withdrawing from GBIF, or (as in the case of the OpenUp! project) to restrict data publishing to some core data items (e.g. excluding geographic information)." (Germany)
 - o "(...) biodiversity recorders in the UK are mainly volunteers and are likely not to submit data into the system if they think it will mean commercial users have free and unimpeded access to their data. That ownership and potentially the integrity of the record is diminished, or if they felt that this practice impacts negatively on their local LRC through which they are supported locally." (Egerton)
 - o "Certain researchers will say they risk to lost their freedom BUT it is the opposite who will append, because now, if you don't have money, you can not publish papers in good journals. Another important aspect: to be sure that your name will not be expose in a bad papers without your accord. But I think that the risk is minimum" (David)
 - o "There is certainly a risk in that some data providers will wish to retain copyright over data. However, it is anticipated that this would be only for a minority of data providers, and efforts should be made to engage with these providers to stress the research and conservation benefits of an open access model for data." (Ireland)
 - o "Since LERC was set up in 2007, it has spent, and continues to spend time consulting with data providers, both current and potential, on sharing data and discussing terms and conditions for it use. Those data sharing agreements (22) which are currently set up would have to be renegotiated to take into account the proposed licensing terms or data withheld entirely from GBIF. This could result in approximately 2 million records not being made available to GBIF. --- As at 5 June, the NBN Gateway, the UK Node of GBIF currently holds 96,679,457 records from 828 datasets (including 316,553 provided by the Greater Lincolnshire Nature Partnership, expected to increase to around 2

million in the near future). These have been shared with the NBN Gateway under a set of terms and conditions that are incompatible with the proposed licensing terms. If GBIF goes ahead with the proposed licensing terms, all data provided by the NBN Gateway would have to be withdrawn from GBIF and each data provider consulted. Given the number of data providers involved, this would take a considerable amount of time and may result in a portion of UK-wide data not being published through GBIF." (Barnes)

- o "Many recorders are already suspicious of the intent of the NBN, DEFRA and other government bodies. The suggestion of a removal of certain rights will undoubtedly move some to consider withdrawing data as well as decline to submit future records. Our experience is that often the most vocal tend to be committee members of various taxonomic recording groups. If they choose to withdraw they can influence many of their members to do likewise." (Corner)
- "Although some data holders will be apprehensive or uncomfortable about publishing under CCO, I think it is generally recognized that "open data is the future" and the cultural shift will continue steadily in that direction." (Baxter)
- o "[We see a substantial risk in] the level of access that is permitted and the applications to which the data are used." (anonymous)
- o "It may happen, that some of the Austrian data publishers will decide to withdraw their consent to access their data via GBIF if they have to abandon their own licensing details in the metadata they have provided for each data set (e.g. University of Vienna Institute for Botany Research Group for Plant Biogeography; Österreichische Mykologische Gesellschaft)" (Austria)
- o "There is a risk that fewer records are made available as a result of applying the CCO waiver. Biodiversity recorders in the UK, particularly voluntary ones, are likely not to mandate LRCs to submit data to GBIF if they think it will mean commercial users have free and unimpeded access to their data. However, this may be addressed by considering the alternatives outlined above." (Hunt)
- o "We are unable to see this happening with our present providers but on the other hand we already experience problems with restricted access at the source to certain information e.g. regarding protected species' whereabouts or even existence as specimens. Providers may want to withdraw their data unless the issue is properly addressed and problems solved. Users and providers are at risk of thinking differently on this topic." (Sweden)
- "The NBN believes that the proposed changes will be seen as a threat to some UK data providers and some providers will almost certainly respond by requesting the removal of their data from GBIF. It is currently impossible to quantify the risk to GBIF of the changes being proposed and the NBN believes greater consultation by GBIF is needed. To assist with this, the NBN is willing to undertake an in-country round of consultation over the next four months to more completely understand the implications from the perspective of UK data providers. While there are enormous benefits in open data, some data providers have reasons for retaining copyright and control for various reasons (including adhering to copyright held by original data recorder or collator, and for the purpose of monetising and publishing). Some data providers may also lower the resolution of the data that they provide to GBIF." (Stroud)
- o "We can not deny it would be a high probability that many publishers (at least in Colombia) will disagree with this approach, and they might [withdraw their data]" (Colombia)
- "The approach as described would be likely to disengage many UK institutions and lead to many datasets being removed from the GBIF portal, including our own datasets." (Frost)
- "The data provider communities support is critical. Alienating this community is going to be a significant risk."
 (South Africa)
- "Several publishers and institutions may be discouraged to publish data due to incompatibility between their internal data publishing policy and CCO." (Brazil)
- o "If the data holders do not fill secure with their sovereignty on the data they published and/or to be quoted in literatures there would be a high risk they were not willing to continue publishing their data." (Madagascar)
- "CC licences are irrevocable; if providers chose to release under CCO and decided they were not getting due recognition, the conditions under which the data is already licensed cannot be changed. Ability to use CC BY may reduce the impact of this risk and encourage more data provision." (Paton & Buffery)
- "There is a risk of less willingness of data holders to publish data through GBIF but this concerns the minority of data holders. We hope that the benefits for research and conservation of an open access model will help in convincing potential reluctant data holders. The ability to flag datasets that are restricted from commercial use is important for reducing this risk." (Israel)
- "The probability is that there would be withdrawal of existing data from GBIF and/or the Data Publishers providing records, e.g. the NBN (Gateway) and fewer records being provided in future. Any suggestion that 'there are so many records on the database that this wouldn't matter' would appear to miss the point and be dismissive of the concept of community that GBIF has worked hard to develop.
 - Pursuing the CC-0 plan as GBIF has outlined is likely to damage the relationship it has with some of its Data Publishers (and theirs with their own providers in turn). This would be a misstep and one that could be avoided by taking the time to pursue alternative courses and, involving the community in these.
 - Within the UK, some organisations, e.g. not for profit local record centres rely on the charges levied for commercial data searches to fund their operations including their involvement in and support of the local collection, validation,

verification and validation of data.

Enforcing Open Data at GBIF now will impact adversely on the quality and quantity of data which it receives in future - and on recording and biodiversity information supply in the UK." (Whitbread)

- o "Yes, [we foresee] some [issues]. But the sooner this issue is clarified the better. GBIF / the Nodes could anticipated to possible problems by considering methods / standards to blur data to a level that is acceptable for the data providers that have problems with a CCO status for the GBIF data" (Netherlands)
- "Citation is essential" (Mexico)
- o "There is a risk that data publishers didn't accept the changes and that we loose some of them." (anonymous)
- "If the CCO waiver is not applied, there is a risk that GBIF-mediated data could be used in a way that contravenes the wishes of some data publishers if users don't understand (or don't read) the Data Use Agreement, especially if the data publishers have included additional licensing details in the dataset metadata which are more restrictive than the standard GBIF terms.

If the CCO waiver *is not* applied, there is a risk that some potential users will simply decide not to use GBIF-mediated data for research or commercial purposes, if they feel that there is too much uncertainty over their permission to use the data.

If the CCO waiver *is not* applied, there is a risk that data publishers will share data under a false expectation that they or GBIF will be able to take legal action against 'misuse' of their data.

If the CCO waiver *is* applied, there is a risk that some datasets will be withdrawn from GBIF, but this risk should be mitigated to some extent by the ability to flag datasets as restricted from commercial use, and by the improved citation model. Both of these measures must be in place and publicised well before the CCO waiver is introduced, to enable data publishers to take an informed decision as to whether they meet their needs and will enable them to continue to publish their data under the CCO waiver.

The risk of data loss from UK publishers could be mitigated further by the fact that publishers can reduce the spatial resolution of taxon occurrence data using the NBN Data Access Controls if they wish. They may do this instead of, or as well as, flagging the datasets as restricted from commercial use.

I believe that some loss of public access to data via GBIF is an acceptable trade-off for a system that is clear and legally defensible." (Lightfoot)

- o "The UK is worried about ensuring the value added to many of the data collations by the database owners (so called sui generis database rights recognised within the EU) is acknowledged. This includes the time to collate but more critically to improve the quality of the collation. The UK believes that failing to acknowledge those carrying out this function would lead to a significant disincentive to provide these quality assured sources which could be detrimental to GBIF overall (lower volume of data, possible generally lower quality). Hence the recommendation to adopt a CC-BY license. The UK judges that these changes represent a reasonable risk to GBIF in terms of providers withdrawing their data and would like to see more evidence of the likely impacts ahead of a final decision. In particular the UK is aware of concerns within its own data providers regarding the use of the data for commercial purposes and the UK intends to explore the potential impacts with its own providers. We would anticipate a final evidence backed decision being made at the next Governing Board." (UK)
- No substantial risk foreseen a number of responses to question 3 remark that no substantial risks are foreseen to arise from this approach. Where comments are present, they include
 - "No, I believe most data publishers will understand the impossibility of limiting data use in such a way that would prevent an occasional undesirable commercial use. Accepting these occasional uses are the inevitable consequence of making our data freely available for all the good uses to which they can be put." (Capers)
 - " If some data sources did not want their data publicly available (=published) then they would not have provided it to GBIF. There is a huge amount of data in GBIF. If some withdraw it may not even be noticed." (SMEBD)
 - "No, providers are already aware that people are ignoring the licence agreements. Often licenses are so
 contradictory that they are impossible to follow. This change will allow the data to be used freely and citations will
 result from that. People who are not already providing data are unlikely to change their position, whatever, system
 was used." (Groom)
 - "While there might be an initial backlash by some, I think most publishers will appreciate the effort GBIF and others have put into these recommendations and the tools to simplify giving and getting credit." (Desmet)
 - "No, open data is the future. But, like I said, for non-occurrence data (cc.by 4.0) seems like an option. And might be more acceptable for some dataholders." (Brosens)
 - "I suppose those data holders holding data which are sensitive will withhold them anyway. The opportunity to flag data sets as confidential or to withhold them from commercial use is acceptable way." (Leya)
 - o "Data will be retracted and withheld, but this would be done on the basis of a more accurate understanding of the current legal situation rather than an actual change. GBIF would be freed from non-binding agreements as to what provided data may be used for and by whom, while setting new standards that are more in line with the legal reality and GBIF's objectives. We hope and expect that it will not be long until these standards will become an undisputed norm, with negligible effects on the willingness to publish through GBIF in the long run." (Norway)