Summary of responses to GBIF Consultation
Endorsement and evaluation of datasets within the Global Biodiversity Information Facility

Overview

This document presents a summary of responses to a consultation carried out by GBIF Secretariat between April and June 2014 on possible changes to the current system of endorsing data publishers, and on proposals to help users evaluate datasets accessed through GBIF. 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 endorsement 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 background document and proposal from the Secretariat to the 21st GBIF Governing Board meeting (GB21) in New Delhi, India in September 2014, including possible alternative options for the way forward.

Background

The GBIF Secretariat organized a consultation, *Endorsement and evaluation of data sets within GBIF*, running from 16 April to 16 June 2014 to assess the opinions of its Participants and other stakeholders regarding 1) possible modifications to the processes used within the GBIF network to approve new data publishers and datasets and 2) to assist users in evaluating the relevance of different datasets and records for particular analyses. The consultation paper and response form were made available in English, French, Portuguese and Spanish. We thank the GBIF language networks for their work in developing these materials.

During the consultation period, 32 responses were received. Five respondents asked for their comments to remain anonymous. The remaining 27 responses were as follows:

- **Voting Participants (16):** Andorra, Argentina, Benin, Colombia, Costa Rica, Finland, Germany, Republic of Korea, Madagascar, Mexico, Netherlands, Norway, South Africa, Sweden, United Kingdom
- **Associate Participant Countries (3):** Central African Republic, Israel, Switzerland
- **Associate Participant Organisations (2):** Albertine Rift Conservation Society (ARCOS), Society for the Management of Electronic Biodiversity Data (SMEBD)
- **Other institutions or individuals (6):** Shannon Asencio/Canadian Museum of Nature, Aaike De Wever/BioFresh, Alan Paton/Royal Botanic Gardens Kew, J.H. Beach, Quentin Groom, David Romain/IMBE: Mediterranean Institute of Biodiversity and marine and terrestrial Ecology, Station Marine d'Endoume, Marseille

Responses

The consultation included three questions:

1. Do you have any comments on the proposed changes to the way that data publishers receive endorsements from GBIF Participants?

2. Do you have any comments on the proposal to associate all GBIF-mediated datasets with a range of fitness-for-use indicators?
3. Do you have suggestions for possible enhancements to these proposals?

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. No attempt is made in this document to discuss possible mechanisms to resolve problems identified or to reconcile conflicts implied between recommendations received. The background document accompanying the Secretariat proposal to the GBIF Governing Board (GB21) will consider these points.

Overall process

The topics addressed in this consultation are complex and have significant implications for how GBIF develops at both the global and national levels. Several responses (Argentina, Mexico, one anonymous response) emphasized the importance of engaging the Governing Board in any final decisions. Following further review and discussion, the Governing Board will be offered the opportunity at GB21 to vote on possible follow-on actions.

Question 1: Proposed changes to endorsement processes

The proposed changes include the following key elements:

1. Abandon the requirement for new data publishers to be approved by the relevant GBIF country participant or a relevant organisational participant.
2. Introduce tools and processes to allow any number of Participants to provide endorsements for each individual dataset – most data publishers offer multiple data sets, each of which could gain separate approval.
3. Enable users to separate between datasets which have received endorsements and those which have received no such endorsements ("unevaluated data").
4. Integrate endorsements with other measures of the apparent fitness-for-use of each dataset.

The consultation demonstrated significant difference of opinion on these suggestions. Below, we provide some detail on the views presented. However, the positions of the respondents can be categorised in a simplified form, as follows:

- **Support suggested changes** – responses agreeing with the proposed changes
- **Support suggested changes, subject to appropriate implementation** – responses agreeing that the proposed changes should be adopted, provided that the implementation includes certain features.
- **Support ONLY for non-Participant countries** – responses opposed to adopting the proposed changes in countries which are already GBIF Participants, but open to the possibility of adopting them elsewhere.
- **Do not support change** – responses which argued that the suggested changes would harm the development of GBIF
- **No opinion** – respondents indicated that they did not, as yet, have a view on this question

The following table and charts show a breakdown of the responses using these categories. Figure 1 presents the breakdown for all responses received. Figure 2 presents the breakdown only for GBIF Participants, i.e. the 21 responses received from Voting Participants (VP), Associate Participant Countries (APC) or Associate Participant Organizations (APO). Note that some Participants chose to respond anonymously and so do not appear in this second breakdown. As can be seen, none of these anonymous responses were in support of the proposed changes (except in one case for use in non-Participant countries).

Category	VP	APC	APO	Other	Anon.	Total
Support suggested changes Andorra, Benin, Israel, SMEBD, Asencio, David, De Wever, Groom, Paton	2	1	1	5	-	**9**
Support suggested changes, subject to appropriate implementation Finland, Ghana, Norway, Sweden, United Kingdom,	5	-	-	-	-	**5**
Support ONLY for non-Participant countries Argentina, Colombia, Costa Rica, Netherlands, Switzerland + 1	4	1	-	-	1	**6**
Do not support change Costa Rica, Germany, Korea, Madagascar, Mexico, South Africa, ARCOS, Beach + 4	5	-	1	1	4	**11**
No opinion Central African Republic	-	1	-	-	-	**1**

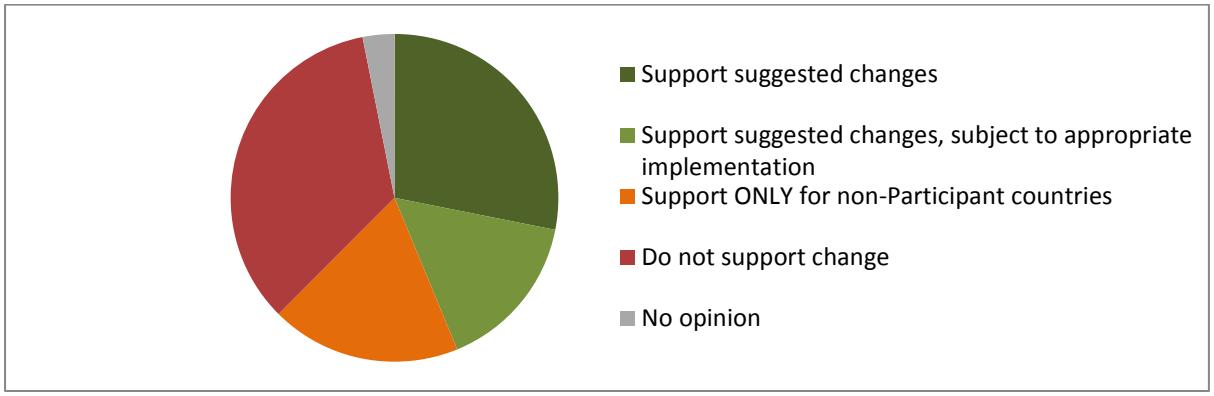

Figure 1 - Breakdown of responses to question 1 - ALL responses

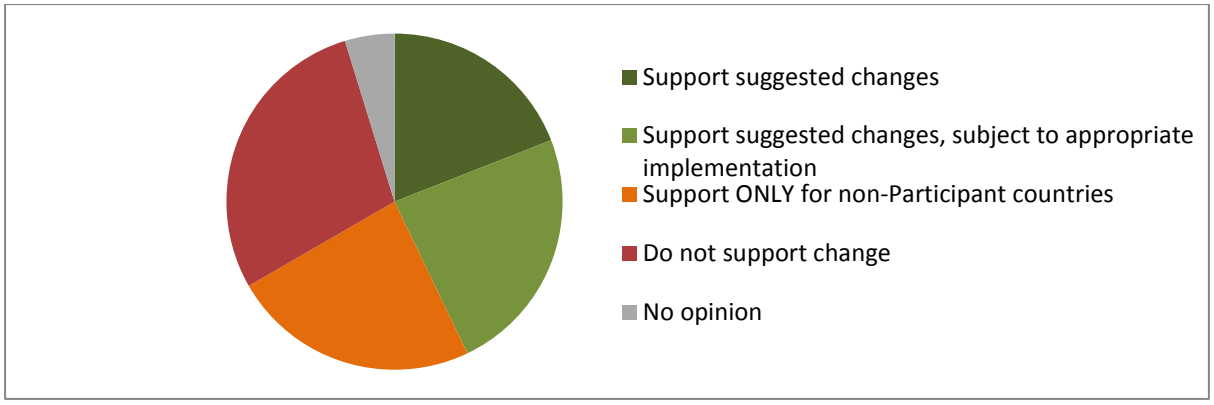

Figure 2 - Breakdown of responses to question 1 - Participant responses only

The following concerns were raised in regard to the proposed changes. See Appendix 1 for the detailed comments that underlie this summary.

- **The current endorsement process is critical for the strength and health of national Nodes –** Many responses expressed concern that changes to the endorsement process would undermine the role of existing GBIF Nodes and would reduce the focus in some countries on funding coordinated activity.
- **The current endorsement process adds significantly to the quality of data published** – A number of participants emphasized the role of well-functioning Nodes in assisting data publishers and reviewing

data to ensure high-quality data and clarify issues around ownership. One possible consequence of the proposed change would be to add significant quantities of data that have never been evaluated.

- **National language needs** – An associated benefit of the current endorsement process is that data publishers work with Nodes that share the same language and can therefore provide more appropriate support.
- **Confusion and lack of clarity** – Some respondents raised concerns that confusion may arise from disagreement about particular datasets and that it will be difficult to present this information clearly to users
- **Increases in workload** – The proposed changes effectively include a shift from endorsing only data publishers to endorsing individual datasets. Some respondents indicated a concern that this may introduce an unmanageable amount of additional work, particularly for the GBIF Secretariat.

Some respondents recognised that the proposed changes involved both 1) a shift from endorsing data publishers to endorsing datasets and 2) relaxation of the requirement that this endorsement must be provided by the GBIF Node for the publisher's country or the GBIF Node of a relevant organizational Participant. Several of these recommended modifications to the proposal:

- **Adopt the proposed changes, but ONLY for publishers in non-Participant countries** – Recognizing the importance of strengthening GBIF Nodes, and recognizing the value for GBIF of engaging with data holders outside current GBIF Participant countries, some respondents suggested using different processes for the two situations.
- **Retain the current model, but expand the role of other Participants in endorsing publishers without a relevant GBIF Node** – Several GBIF organizational Participants already endorse data publishers, and in some cases, GBIF country Nodes have hosted data for countries that are not yet members. These approaches could be expanded to enable endorsement of unaligned data publishers.
- **Introduce separate endorsement processes for data publishers and datasets** – There are advantages in maintaining the current system of endorsing publishers while introducing the proposed process for endorsing datasets to help evaluate fitness for use.

As indicated in the breakdown above, several responses supported, or partially supported, making the proposed changes but cautioned that certain features should be included in the implementation:

- **Relevant Nodes should be notified automatically when relevant data publishers or datasets are added** – Nodes should always at least be given the opportunity to make early contact with new data publishers or when an existing publisher adds new data.
- **Introduce minimal requirements that must be satisfied for any new dataset** – The addition of large numbers of new data sets may create excess workload, which might be relieved if data publishers were required to organize data only into larger data sets.

Question 2: Proposal to associate datasets with a range of fitness-for-use indicators

The proposed changes include the following key elements:

- As part of GBIF data indexing processes, automated checks will be performed for standards compliance and data consistency at the metadata, record and whole dataset levels.
- All relevant GBIF Participants will be encouraged to review each dataset and provide an evaluation of the dataset. Other data sets will be marked as "Unevaluated".
- Users of data will be encouraged to provide comments on individual datasets and records and to report on the usefulness and actual use of particular datasets in particular research activities and publications.

- GBIF portal dataset pages will report on all of these (automated checks, endorsements, user feedback) as separate aspects of fitness for use.
- Based on all available elements for evaluating fitness for use, datasets should be associated with one or many indicators to allow users to filter "Unevaluated" data or data with other perceived issues from views and downloads.

The consultation demonstrated strong support for this proposal, although many responses identified specific aspects that could cause problems and that need to be explored further. Below, we provide some detail on the views presented. However, the positions of the respondents can be categorized in a simplified form, as follows:

- **Support suggested changes** – responses agreeing with the proposed changes
- **Support suggested changes, subject to appropriate implementation** – responses agreeing that the proposed changes should be adopted, provided that the implementation includes certain features.
- **Support proposal, with some concerns** – responses support the recommendations, but indicate concern that some negative consequences may follow.
- **Do not support change** – responses which argued that the suggested changes would harm the development of GBIF
- **No opinion** – respondents indicated that they did not, as yet, have a view on this question

The following table and charts show a breakdown of the responses using these categories. Figure 1 presents the breakdown for all responses received. Figure 2 presents the breakdown only for GBIF Participants, i.e. the 21 responses received from Voting Participants (VP), Associate Participant Countries (APC) or Associate Participant Organizations (APO). Note that some Participants chose to respond anonymously and so do not appear in this second breakdown. As can be seen, all of these anonymous responses were in support of the proposed changes (except for one that expressed no opinion).

Category	VP	APC	APO	Other	Anon.	Total
Support suggested proposal Andorra, Argentina, Benin, Colombia, Costa Rica, Ghana, Mexico, Netherlands, Norway, South Africa, Israel, Switzerland, ARCOS, SMEBD, De Wever + 4	10	2	2	1	4	**19**
Support suggested proposal, subject to appropriate implementation Finland, Germany, Sweden, United Kingdom, Asencio, David, Paton	4	-	-	3	-	**7**
Support proposal, with some concerns Madagascar, Central African Republic, Groom	1	1	-	1	-	**3**
Do not support proposal Beach	-	-	-	1	-	**1**
No opinion South Korea + 1	1	-	-	-	1	**2**

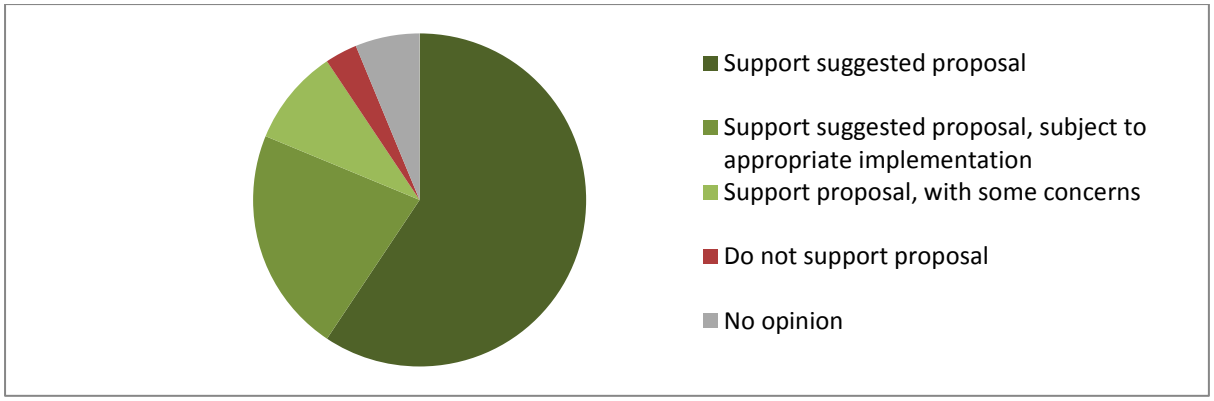

Figure 3 - Breakdown of responses to question 2 - ALL responses

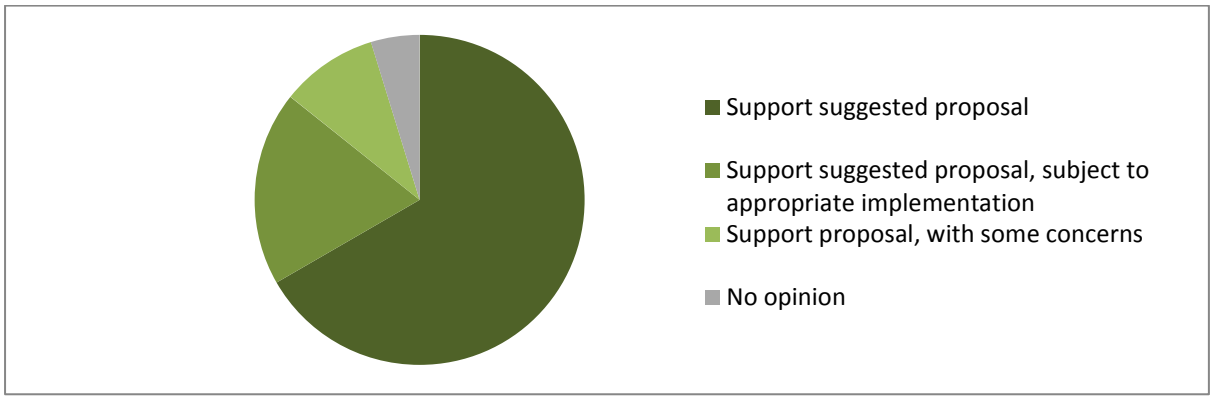

Figure 4 - Breakdown of responses to question 2 - Participant responses only

The following concern was raised in regard to the proposed changes. See Appendix 2 for the detailed comments that underlie this summary.

- **Indicating fitness for use for data sets may create divisions –** Some responses expressed concern that the proposed approach would lead to a perceived division between "good" and "bad" data publishers and might lead to conflict and withdrawal of data. This was considered a particular concern if subjective assessments are permitted.

As indicated in the breakdown above, several responses supported, or partially supported, making the proposed changes but cautioned that certain features should be included in the implementation:

- **Indicators need to be clear and objective** – Many responses supported the need for useful indicators of the fitness for use of different datasets, but emphasised the need for such indicators to be meaningful and helpful to users, and not problematic or contentious for publishers. In particular there was repeated emphasis on the need for these indicators to avoid subjective and controversial judgments (see also some of the comments above regarding the risk that indicators might cause divisions).
- **Publishers need to be able to document data restrictions and quality control applied** – Some responses acknowledged that data publishers may choose or be obliged to publish only a reduced subset of available data. Similarly data publishers may be able to document quality control processes applied prior to publication. Publishers should not be pressured to suggest greater precision than is actually the case. In all of these cases, metadata should be provided in standard form to clarify these aspects.

- **Publishers need to be able to respond to comments and evaluation processes** – Various comments addressed possible enhancements to ensure that data publishers are able to respond to feedback and fitness-for-use indicators, potentially with a delay period before these indicators appear publicly.
- **Avoid establishing barriers to possible future data types** – If requirements for fitness for use are defined too narrowly, they might serve as a barrier to GBIF receiving new data types, if these do not fit these requirements.
- **Enable experts to provide validation and annotation for data sets** – Several responses recommended that expert researchers, etc. should have opportunities to document their evaluation of datasets relevant to their work.

Question 3: Suggested enhancements to these processes

The following additional suggestions were received. See Appendix 3 for the detailed comments that underlie these summary points.

- Establish a separate service or facility for (advanced) fitness-for-use analysis
- Do not mix past (provider/publisher) endorsements with future (dataset) endorsements
- Introduce agreement from data publishers to allow sustainable archival of their data
- Track use of data outside scientific literature
- Allow Participants to tag datasets as “disputed”
- Detect and link suspected duplicate records from multiple sources
- Enable country to which record relates to check data for sensitivity
- Monitor correlation between endorsements and other fitness-for-use indicators

Appendix 1 – Comments received in response to question 1

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

- **The current endorsement process is critical for the strength and health of national Nodes –** Many responses expressed concern that changes to the endorsement process would undermine the role of existing GBIF Nodes and would reduce the focus in some countries on funding coordinated activity.
    - "I strongly feel on the opportunity that the existing mechanism gave to GBIF Participants to engage dialogue with researchers groups in their respective countries. All efforts should be made to keep this window open." (ARCOS)
    - "Only for publishers which countries who don't have a consolidated network or a node that can handle the support on data publishing, we agree to expanding the GBIF model. We are not agreeing in the case of publishers which countries have a solid network that support among the dataset publishing the orientation for IPT use, DwC comprehension, licencing and data quality." (Colombia)
    - "The success of national nodes is critical to the success of GBIF. To assign this responsibility to national nodes is a way of strengthening the national community and demonstrating the relevance of national nodes at the country level." (Costa Rica)
    - "How the countries will respond? They could think that having a node is less important than before." (Anonymous)
    - "Among German GBIF Nodes the response was mixed, with a majority perceiving this as a weakening of the Node's position and thus not in favour of this change." (Germany)
    - "In our experience, the endorsement procedure is the main mechanism for the nodes (i.e., for Mexico, the national biodiversity information system developed in CONABIO) to interact with data publishers and engage them with national policies (for example, considering data quality evaluation process)." (Mexico)
    - "I think that in case of well functioning Nodes their role as principal body for endorsement should be maintained." (Netherlands)
    - "However, we recommend that the national node should remain a central contact point in interaction with the national representation of such thematic expert communities." (Norway)
    - "However, the potential weakening of a country's national GBIF community is a concern." (Asencio)
    - "This will compromise the relevance of the Node at the national level. The existence of a National Biodiversity Information Facility comes into existence through financial commitments of our ministries both to GBIF (membership dues) and country level investment. National mandates exist to report on biodiversity data generated, coordinated and disseminated, with the National Node playing an important role, in these reporting obligations. Increased relevance of the Node to stakeholders and partner organisations means that more data is mobilised to support policy, decision making and the green economy." (South Africa)
    - "Within a country there may also be political issues at play and the removal of the endorsement requirement may result in these issues being accentuated. The ability to publish completely independently of the Node may undermine the case for the Node as part of the national infrastructure which may result in loss of support (at a time when there is a need to build up national support and investment). It is vital that we do what we can to support the current national infrastructures." (UK)
    - "The countries might find that the need of a national coordinating service (the node) is obsolete, and could stop supporting it, with consequent decrease of engagement to GBIF)." (Anonymous)
    - "However, the potential weakening of a country's national GBIF community is a concern." (Asencio)
- **The current endorsement process adds significantly to the quality of data published** – A number of participants emphasised the role of well-functioning Nodes in assisting data publishers and reviewing data to ensure high-quality data and clarify issues around ownership. One possible consequence of the proposed change would be to add significant quantities of data that have never been evaluated.
    - "Although the proposed changes have strengths, the traditional endorsement seems to have more strength to guarantee reliability and quality assurance of data." (Korea)
    - "De plus, le nouveau système ne participe pas à l'établissement de « communautés nationales GBIF » qui sont, à notre avis, très importantes dans ce type de projet, étant donné qu'elles jouent un rôle capital dans l'homogénéité et la qualité des données transmises." (Anonymous)
    - "In Spain, the GBIF national node has the best information about the quality of the data from national authors and should have the maximum capacity when valuing." (Anonymous)
    - "The initial endorsement by participant nodes is an important step and should not be overtaken because the close relationship between the nodes and publishers ensures that: publishers have fully understood the implications of being a GBIF publisher (by ensuring the understanding of Data Sharing Agreement and Data User Agreement); the data quality assessment of the first dataset allows the identification of errors that are reported to the publisher, together with strategies for correction." (Anonymous)
    - "There have been some instances recently where there were issues with the data ownership which were picked up as part of the endorsement process. If the model is being changed there does need to be a step in the process

where the publisher confirms that they are the owner (or have confirmation that they are acting on behalf of the owner) before the data are published." (UK)

  - "The idea of open endorsement is excellent; I am rather a bit unquiet about doing it after publishing as there may be information clogging. Indeed, if the data are not "safe", it may hinder the overall quality of the data portal and we may lose the user confidence. Otherwise, as no one is liable to the data quality, the endorsement could be skipped. I could contact four of our data providers, both of them would like to pass through the node validation before submitting for publication, three of them said anyway they have no/or poor connexion so the passage through MadBIF is needed." (Madagascar)
  - "Ce nouveau système, s'il est adopté, risque d'avoir comme conséquence une accumulation de données "non évaluées", et donc une diminution de la qualité des données. Etant donné la surcharge dont souffrent tous les scientifiques, il n'est pas sûr que ces derniers prendront vraiment le soin d'évaluer tous les nouveaux sets de données apparaissant sur GBIF.org, même si cela les concerne directement dans leurs projets. Ils seront plutôt fortement tentés de prendre l'option d'utiliser ces données "non évaluées" sous une forme ou une autre." (Anonymous)

- **National language needs** – An associated benefit of the current endorsement process is that data publishers work with Nodes that share the same language and can therefore provide more appropriate support.
  - "Interactions with data publishers are better when they are in their own language." (Anonymous)
  - "The communication between the publisher and the Secretariat (and other participants) will likely be made in English, losing the advantages of using native languages, which in some cases is important." (Anonymous)
- **Confusion and lack of clarity** – Some respondents raised concerns that confusion may arise from disagreement about particular data sets and that it will be difficult to present this information clearly to users
  - "Confusion can result from the fact that two participants could give endorsements in different directions, i.e., one participant outside of the the geographic context gives endorsement, but the direct national participant does not. This can lead to bad perception by the user of the value of the resource; different countries/regions require different mobilization strategies – Secretariat will not be able to prioritize all needs in its contacts with the publishers;" (Anonymous)
  - "Es importante que la entrada de conjuntos de datos estén referenciados con un código que indique que son datos "no evaluados". Es necesario que estos datos vayan siempre acompañados de sus respectivos metadatos para que el usuario pueda discernir la conveniencia de usarlos." (Andorra)
  - "If endorsement will move from the National Nodes to other levels (other Nodes, organisations, affiliates) the criteria for endorsement should be standardised in a way. Any dataset may receive endorsements from any number of participants"" (point 7). I think we add a confusing level of subjective information here." (Netherlands)
  - "Need to define a better mechanism for the framework for endorsement including how endorsements from different sources (on the dame data) can be presented meaningfully to the user." (UK)
- **Increases in work load** – The proposed changes effectively include a shift from endorsing only data publishers to endorsing individual data sets. Some respondents indicated a concern that this may introduce an unmanageable amount of additional work, particularly for the GBIF Secretariat.
  - "We wonder if GBIF has enough budget to create a team that can handle publishing requests and accompaniments (that might take months) for all the potential publishers in the world and their respective languages." (Colombia)
  - "How the secretariat will manage? It is time consuming" (Anonymous)
  - "This proposal definitely has some advantages but we should careful that we do not compromise quality as more resources will be required with the endorsement of individual data sets." (Ghana)
  - "The proposed changes will overload the Secretariat with more tasks of monitoring, reporting, selecting and contacting relevant participants." (Anonymous)
  - "I fully support these proposals. One possible concern is resource constraints at the GBIF Secretariat level that may hinder the ability to monitor new datasets for signs of abuse." (Asencio)
- **Adopt the proposed changes, but ONLY for publishers in non-Participant countries** – Recognising the importance of strengthening GBIF Nodes, and recognising the value for GBIF of engaging with data holders outside current GBIF Participant countries, some respondents suggested using different processes for the two situations.
  - "The process of adding data publishers from countries that are not members of GBIF should be done as proposed by the background document to simplify the way of integrating new data publishers to the GBIF network and encourage other countries to be part of GBIF." (Costa Rica)
  - "We suggest not abandoning the obligation to endorse Data Publishers before their first contribution for countries which signed the MoU." (Switzerland)

- "Only for publishers which countries who don't have aconsolidated network or a node that can handle the support on data publishing we are agree in for expanding the GBIF model." (Colombia)

- **Retain the current model, but expand the role of other Participants in endorsing publishers without a relevant GBIF Node** – Several GBIF organisational Participants already endorse data publishers, and in some cases, GBIF country Nodes have hosted data for countries that are not yet members. These approaches could be expanded to enable endorsement of unaligned data publishers.
  - "However, there should be a mechanism for making connection to and communicating with the local (nearest) GBIF community or thematic node." (Finland)
  - "We think that it would be better if these non gbif members data publishers connect their data via an existing node: thematic one or gbif member country which is close by language, community or regional localisation." (Anonymous)
  - "We think that the possibility to endorse datasets through associate participant organisations is insufficiently promoted as a way for providers from non-member countries to contribute. In Europe and the Mediterrenian area, for example, the BioCASE secretariat as the GBIF Node for the Consortium of European Taxonomic Facilities, has endorsed datasets from non-GBIF countries." (Germany)
  - "South Africa understands the importance of ensuring that non GBIF participants may want to share data and that it would be of value to fill data gaps– but proposes that other mechanisms can be sort to ensure that this data becomes accessible. This might include: 1. That datasets could connect with regional /thematic initiatives, and in this way data becomes accessible; 2. That a host is sought from amongst the GBIF participants that can host the data from non-GBIF members." (South Africa)
- **Maintain separate endorsement processes for data publishers and data sets** – There are different strengths in the two processes. Maintaining both may be the best solution.
  - "However, the a posteriori approval of single datasets cannot and should not replace the endorsement of a new Data Publisher by the relevant national node (or an international partner organization), as this represents a simple but efficient step to promote exchanges at national (institutions, authorities) and international level (secretariat), and hence the homogeneity and quality of the shared content. From our perspective, the two levels of endorsement (Data Publishers and Datasets) are complementary and should both be offered." (Switzerland)
- **Relevant Nodes should be notified automatically when relevant data publishers or data sets are added** – Nodes should always at least be given the opportunity to make early contact with new data publishers or when an existing publisher adds new data.
  - "A node should be informed when a new data set from an already endorse data publisher arrives." (Anonymous)
  - "One German node was in favour of opening up the contribution process but suggested that (i) GBIF national nodes would be informed about "free" providers from their countries" (Germany)
  - "The proposal includes a mechanism to automatically notify the respective country and thematic organisation node when relevant new datasets are added. We believe this notification service to the nodes to be an important part of the new endorsement model." (Norway)
  - "Implement an automatic alert system that informs nodes when an update or a new dataset is published by an endorsed publisher from that node." (Anonymous)
- **Introduce minimal requirements that must be satisfied for any new data set** – The addition of large numbers of new data sets may create excess work load, which might be relieved if data publishers were required to organise data only into larger data sets.
  - "Since the proposal already recognizes the complexity of the process to endorse each dataset individually when a given publisher publishes many "small" datasets; is it possible to impose restrictions on the size of datasets? Here, I envisage a situation where a publisher would be encouraged to wait until s/he has a big dataset to contribute and/or aggregate the data in some logical ways before publication." (ARCOS)
  - "Still some minimal requirements should be clearly defined to avoid overloading GBIF with problematic datasets (for example if somebody offers thousands of small datasets....)." (Netherlands)

Appendix 2 – Comments received in response to question 2

- **Indicating fitness-for-use for data sets may create divisions –** Some responses expressed concern that the proposed approach would lead to a perceived division between "good" and "bad" data publishers and might lead to conflict and withdrawal of data. This was considered a particular concern if subjective assessments are permitted.
    - 'Ces indicateurs soustendent le niveau de confiance applicable aux données et participent de la fiabilité des données. Le problème c'est de savoir sur quels critères va-t-on se baser pour définir les indicateurs. Il ya aussi à craindre que finalement il y ait des bons et des mauvais noeuds GBIF , ce qui instaurerait une discrimination dans les informations sur la biodiversité, alors que chaque noeud est sensé fournir des données d'occurrences et d'observations avérées de son territoire national .' (Central African Republic)
    - 'This would be an incestuous activity which will inflict enduring wounds upon the collections community. Data quality standards are useful and important. GBIF software and services to enable publishers to assess their compliance and to attain them would be highly-valuable. Having a system of peer review of data sets based on 'authoritative' professional opinions, is a power play and self-destructive infrastructure for an organization chartered to engage and aggregate all repositories of biodiversity documentation. Dump the plan to use solicited or unsolicited reviews in the calculation of merit/status/suitability indicators for data sets. Don't make this a competitive arena to attain collection supremacy, standing and accreditation. Use indicators to motivate publishers to improve their data, not to categorize and label them with permanent online reviews. GBIF should not create infrastructure to evaluate and rank collection databases against one another as if they were restaurants or hotels on TripAdvisor, but rather to point out the strengths and highlights of respective collections, provide attainable goals, with status indicators, which they can reach for algorithmically-assessable higher levels of data quality, and encourage them to invest in data improvement activities with software and services support to assist.' (Beach)
    - "If the indicators are based on professional opinion, such a system will lead to stratified classes of publishers and de-motivation for publishers of 'underdeveloped' data sets. Creating tiers of data set access and visibility based reviewer comments will create a competitive arena leading to GBIF-sponsored collections elitism. Algorithmically assessing the completeness and quality of data sets, even perhaps adding fitness-for-use interpretations on top of those measures, which were impartially and consistently applied, would be much more constructive and productive." (Beach)
    - "I think that ranking data sets according their fitness for use would lower the willingness of some publishers to participate; anyway, users could normally appreciate by themselves. But it could be interesting to provide to the users a way to give a feedback on individual records liability as in some case even with trustful data providers I found that there could be errors in geolocalisation, e.g.). Both of them did not like the idea of data ranking but agree that it could be interesting to get a feedback for each record, one suggested that this feedback should be addressed directly to the data provider." (Madagascar)
    - "By principle, all data in GBIF should be considered at the best quality available and possible. Defining categories of datasets (unevaluated, endorsed?) violates this principle, and ranks data in categories where publishers will not like to be seen. Quality should be assessed in a separate process (see below)" (Anonymous)
- **Indicators need to be clear and objective** – Many responses supported the need for useful indicators of the fitness-for-use of different data sets, but emphasised the need that such indicators should be meaningful and helpful to users, and not problematic or contentious for publishers. In particular there was repeated emphasis on the need for these indicators to avoid subjective and controversial judgments (see also some of the comments above regarding the risk that indicators might cause divisions).
    - "I have concerns about fitness of use indicators based on endorsements(see below) This is a good idea if they are limited to factual statements surrounding the scope, standard compliance and data consistency of the data provided. It will be very difficult to use endorsements of fitness for use by others as a meaningful and consistent measure and apply it comprehensively over data sources. At best endorsements would be patchy (what does the absence of an endorsement really mean?); or at worst misleading (data is useful for one particular usecase, but not for another and the details of use leading to the endorsement are not known)." (Paton)
    - "Oui : il faut que ces aptitudes soient les plus factuels possibles (obligation de format et de documentation)" (David)
    - "The GBIF infrastructure is a publishing mechanism and engaging with communities that generate and publish data so that it can be used and reused, is a delicate process. Fitness for Use indicators should be based on metrics and we advise that datasets are not ranked in a qualitative way which may compromise further data publishing by data providers/publishers." (South Africa)
    - "If endorsements of data by others is to be adopted then there needs to be clear criteria as to what constitutes 'good'. Providers could offer assessments on data quality as they see it- again using clear criteria." (Paton)"
    - "Also different types and compositions of a dataset should have different measures. How to consider the size or type of a data-set? For example: dynamic vs static data, monitoring, tracking, collection- data and finally datasets that are a mixture of these?" (Finland)

- "We regard usability indicators as positive in principle, but this needs a follow-up consultation with a more concrete proposal of a set of indicators." (Germany)
- "However, It is important to distinguish the levels of attached information to a dataset. From neutral statistical information on one site to the subjective judging by GBIF participants / data users on the other side, and in between there can be a whole suit of "fitness-for-use" indicators. It would be good to make an inventory of what current GBIF data-users use as indicators of Fitness-for-use." (Netherlands)
- "Probably a useful tool but in need of considerable development and a difficult one to achieve. As always in similar situations independent judgements must be assured." (Sweden)
- "The issue is more how these fitness for use indicators are derived and made consistent across different sources of them" (UK)
- "On the one hand there is an urgent need for evaluation of data quality / fitness for use of GBIF datasets, if GBF datasets are to be used as direct input for international initiatives like IPBES. On the other hand, it is an essential part of the work of scientists to evaluate the fitness for use of any dataset they use to answer a specific research question. This is not and cannot be the job of a research infrastructure like GBIF. However, here we do need to distinguish technical quality (adherence to format, completeness etc.) from the actual content. Technical quality can be controlled and evaluated by GBIF itself." (Germany)
- "It is a good idea to allow user comments/annotations on records and datasets. However, this would require moderation to ensure the additions are constructive and relevant. The ability to use filters to differentiate between unevaluated and evaluated data is important." (Asencio)
- "I'm not sure it will be useful. When looking for data I prefer to get all there is available and then filter it for my own purpose. If it doesn't create a lot of work it might be worth trialing." (Groom)"
- "If we should rank data set in some way, we could perhaps do it according the accuracy of the data. I mean to flag that this record/or data set fits for an habitat type or a province, but not for use requiring more precision." (Madagascar)

- **Publishers need to be able to document data restrictions and quality control applied** – Some responses acknowledged that data publishers may choose or be obliged to publish only a reduced subset of available data. Similarly data publishers may be able to document quality control processes applied prior to publication. Publishers should not be pressured to suggest greater precision than is actually the case. In all of these cases, metadata should standardly be provided to clarify these aspects.
    - "Before users are allowed to report problematic datasets or records, there should be adequate measures in the (DwC) standards to express the original quality of withheld data. It should contain also the reason why the detailed data can't be exposed or why the locality information is published with lower resolution. The stated reasons might be for example: a) sensitivity, b) commercial/legal reasons or c) the original data in detail is not in a format that can be expressed with current DwC-standard (gridded or tracking data)." (Finland)
    - "It is also important to note that data published may be different from data held, that may potentially be made available for suitable clients (detailed coordinates etc.). The German GBIF provider community mirrors the basic controversy of seeing GBIF as a data discovery tool (and thus, provide all data almost regardless of quality) and the position that only quality controlled data can be submitted. We agree on the position that the latter should be ensured at least for observational datasets, which cannot be controlled by going back to the physical specimens. So here we do need what are essentially metadata provided by the data provider on what quality controls were made prior to publishing." (Germany)
    - "It should be clearly stated that fitness-for-use measures serves all data usage and is not about "quality competition". False accuracy is actually a bigger problem than publishing low resolution data and stating with it how (and under which terms) to get hold of the withheld data." (Finland)
- **Publishers need to be able to respond to comments and evaluation processes** – Various comments addressed possible enhancements to ensure that data publishers are able to respond to feedback and fitness-for-use indicators, potentially with a delay period before these indicators appear publicly.
    - "We also think that scientists should be encouraged to publicly comment on datasets, with regard to their suitability for their research task – again with the provision that providers can react and/or respond to this." (Germany)
    - "Je pense qu'il faut d'ailleurs laisser aussi l'accès à celles « mal évaluées » et proposer une option de « demande d'évaluation supplémentaire » au déposant." (David)
    - "We posit that GBIF should publish the results of automated error controls, but should give data providers a reasonable time span to try and correct the errors." (Germany)
    - "Publishers should be informed instantly when a resource published by them is reviewed. A clarification regarding the expectactions of such a review is needed, in order to have publishers informed on how their dataset will be reviewed." (Anonymous)

- **Avoid establishing barriers to possible future data types** – If requirements for fitness-for-use are defined too narrowly, they might serve as a barrier to GBIF receiving new data types, if these do not fit these requirements.
    - "Attention toutefois à ce que ces aptitudes ne brident pas trop la « diversité » des types de données envisageables dans le futur" (David)
- **Enable experts to provide validation and annotation for data sets** – Several responses recommended that expert researchers, etc. should have opportunities to document their evaluation of data sets relevant to their work.
    - "I would especially welcome options for expert validation and annotation of data. E.g. in BioFresh we envisaged to harvest all European fish data from GBIF and build a curated dataset from this. A mechanism to feed such expert curation back to GBIF would make this kind of effort more valuable." (BioFresh)
    - "Some of the issues related to fitness-for-use are first discovered when the data is used in a scientific study. Allowing scientists to share such issues that they discover could become a valuable complement to endorsement and fitness-for-use indicators developed by thematic expert communities and/or national nodes. Mechanisms such as dataset- and record-level annotation services should be explored as a complement to fitness-for-use indicators." (Norway)
    - "The scientific relevance and fitness-for-use for biodiversity datasets are often more appropriately evaluated in a thematic expert community. Endorsement and fitness-for-use annotations from a thematic expert community could perhaps contribute to provide even more trust in GBIF-mediated data than the endorsement by the national GBIF node." (Norway)
    - "I wonder if the endorsement system couldn't work like the peer review of publications, rather than the more passive approach suggested. This might generate a lot of work for the community in the short term, but in longer term the endorsements are likely to be more reliable." (Groom)

Appendix 3 – Additional suggestions received

- **Establish a separate service or facility for (advanced) fitness-for-use analysis**
  - "I think the (advanced) fitness-for-use analysis is something for a separate GBIF service and / or facility? A fitness for use analysis should be interactive and the user should be able to fine-tune their analysis, this means computing power and advanced algorithms, that is a project on its own." (Netherlands)
- **Do not mix past (provider) endorsements with future (data set) endoresements**
  - "The proposal will mean 'active' endorsement of datasets and (one assumes) that this will involve some checks of the data quality and of the metadata. Past endorsements should not be mixed with these. I recommend only endorsements after the proposed changes are implemented are published through the portal." (SMEBD)
- **Introduce agreement from data publishers to allow sustainable archival of their data**
  - "One German node was in favour of opening up the contribution process but suggested (ii) that by submitting their data, the free providers agreed to have their data archived and sustainably provided by GBIF." (Germany)
- **Track use of data outside scientific literature**
  - "Implement a data use reporting system for reporting uses, especially other than published scientific papers, which are being monitored by GBIF." (Anonymous)
- **Allow Participants to tag data sets as "disputed"**
  - "In addition to endorsing or not endorsing a dataset, participants should be given the option to tag a dataset as "disputed", as a flag to users and the secretariat. The aim should of course be to remove this flag when the issue has been resolved or deemed invalid." (Norway)
- **Detect and link suspected duplicate records from multiple sources**
  - "Identification of duplicates on both dataset- and record-level remains a particular important challenge when allowing easier entry of biodiversity datasets into the GBIF-infrastructure. It is easy to imagine a situation where different data owners maintain and provide complementary information about the very same primary biodiversity data entities. While we wait for persistent identifiers, alternative mechanisms to identify and link suspected duplicates (and records with potentially complementary information about the same entity/occurrence) might be useful to explore." (Norway)
- **Enable country to which record relates to check data for sensitivity**
  - "Also an "opposite direction" feedback mechanism to allow (not mandatory) check-up (or approval) from the country and/or region where the data is originally collected to prevent problematic/sensitive data from being published in full detail as part of data repatriation. IUCN local red-lists should be run against the data sets, and if there are matches, detailed locality information should be withheld." (Finland)
- **Monitor correlation between endorsements and other fitness-for-use indicators**
  - "It will be interesting to test the correlation between "endorsement" indicator and other fitness for use indicators of datasets." (Israel)