CODATA Executive Committee Recommendations Regarding Task Group Proposals 2016

The work of Task Groups is one of the important ways through which CODATA achieves its strategic objectives. Task Groups should contribute directly to one of the priority areas as articulated in the CODATA Strategy: http://dx.doi.org/10.5281/zenodo.50343

Task Groups proposals are invited every two years and are selected by the General Assembly. Proposals are assessed by external reviewers and by the Executive Committee which presents recommendations to the General Assembly.

For the CODATA General Assembly in 2016, a total of **fifteen** proposals were received.

Eight proposals for new Task Groups were submitted:

1. Agriculture Data, Knowledge for Learning and Innovation
2. Building Foundational Training in Research Data Science
3. Citizen Science and the Validation, Curation, and Management of Crowdsourced Data (with WDS)
4. Coordinating Data Standards amongst Scientific Unions
5. Globalising Open Science & Access to Research Software (with GYA)
6. Open Data Strategy in Africa
7. Protecting Historical Data
8. Materials Data in IoT and Big-Data Era

One proposal was submitted that represents a significant refocusing by an existing group.

9. Practice and Impact of Digital Data Citation http://www.codata.org/task-groups/data-citation-standards-and-practices

Six existing Task Groups applied for renewal:

10. Advancing Informatics for Microbiology (TG-AIM) http://www.codata.org/task-groups/advancing-informatics-for-microbiology
11. Data at Risk (DAR) http://www.codata.org/task-groups/data-at-risk
12. Earth and Space Science Data Interoperability (ESSDI) http://www.codata.org/task-groups/earth-and-space-science-data-interoperability
13. Linked Open Data for Global Disaster Risk Research (LODGD) http://www.codata.org/task-groups/linked-open-data-for-global-disaster-risk-research
14. Preservation of and Access to Scientific and Technical Data in Developing Countries (PASTD) http://www.codata.org/task-groups/preservation-of-and-access-to-scientific-and-technical-data-in-for-with-developing-countries-pastd
15. Science and Management of Physical Objects in the Digital Era http://www.codata.org/task-groups/management-of-physical-objects

On the basis of peer review and its own assessment, the Executive Committee makes the following recommendations to the General Assembly, which it hopes will be considered attentively.

CODATA has a limited budget for Task Groups, although Groups are expected to find other sources of funds and CODATA endorsement should assist with this. The Executive Committee asks the GA to consider that the more TGs are endorsed, the more thinly the CODATA financial and secretariat support is spread. The Executive Committee recommends the total amount available for TG funding for FY16-17 be 40,000 euros. Therefore, a maximum of 8-10 TGs is optimal in these circumstances.

It is the responsibility of the Executive Committee to make budget allocations to those TGs which are approved by the GA. Budgetary support will only be allocated if the Executive Committee is satisfied that the TG will deliver, that changes requested of the TG's proposed workplan have been made and will be executed.

CODATA Task Group Proposals: Executive Committee Recommendations

In what follows, we give a brief summary of the TG's proposed outputs and make some comments on strengths and weaknesses based on the Executive Committee's own considerations as well as the external reviews. **In some cases, the Executive Committee recommends that the proposal be endorsed on condition of certain changes being made. The Executive Committee will determine the support available from CODATA accordingly.**

As noted above, the Executive Committee judges that in consideration of CODATA's budget and the resource of the Secretariat, the optimal number of Task Groups that CODATA should have is **eight**, with a maximum of **ten**. A total of **nine** Task Groups are recommended for endorsement below.

1. Agriculture Data, Knowledge for Learning and Innovation

The team proposes a holistic approach to promoting open science and open, interoperable data in agricultural research, with a workplan including workshops and training activities with institutions in the East African Region. The team is strong and builds on a number of important CODATA contacts in Kenya in particular: however, it is important that membership is expanded more broadly in the East African Region. The involvement of KALRO, FAO and the World Bank is valuable. The proposers have already secured a grant of US$ 33,684 on which the proposed TG will build. This is a strong proposal, but it will be essential that this initiative cooperate closely with the African Open Data Platform initiative.

RECOMMENDATION: ENDORSE, on condition that the TG to collaborate with the African Open Data Platform Initiative

2. Building Foundational Training in Research Data Science

The proposal is to further develop and extend a programme of short courses in 'research data science': outputs include a refined curriculum, online training materials and programme of short courses and a business model for sustainability. This proposal builds on the work of a joint CODATA-RDA Working Group which completes its term in Sept 2016. The track record of the WG is impressive: see information page http://www.codata.org/working-groups/research-data-science-summer-schools; programme for the first school in Trieste http://indico.ictp.it/event/7658/other-view?view=ictptimetable; photos https://www.flickr.com/photos/145909074@N04/albums. ICTP and TWAS are very supportive and have agreed that the foundational school should run annually at ICTP for the next three years (next edition July 2017); agreement has been reached to run a school at the ICTP satellite in Sao Paolo in December 2017. The team is exploring opportunities to run schools in India and in various locations in Africa and Latin America. This is a proposal of considerable importance to CODATA's capacity building strategy.

RECOMMENDATION: STRONGLY ENDORSE, on condition that the TG to collaborate with the African Open Data Platform Initiative

3. Citizen Science and the Validation, Curation, and Management of Crowdsourced Data (with WDS)

This is proposed as a joint CODATA-WDS Task Group. The principal output will be a White Paper and journal article on data validation in citizen science, as well as a guide to best practices for data validation, curation, and management that will be promoted in the citizen science community. The proposition of a joint CODATA-WDS TG is very welcome. The proposal is well focused and there is a very high likelihood of their delivering what they say they will deliver. Indeed, one of the reviewers urged that they should be more ambitious in their outputs! However, the Executive Committee is pleased to see an achievable and well-targeted activity on this issue.

RECOMMENDATION: ENDORSE

4. Coordinating Data Standards amongst Scientific Unions

The TG will encourage and work with data groups or commissions in a number of International Scientific Unions to help develop and promote coordination of relevant data standards. Such work is of fundamental importance and core to CODATA's mission. This is a pragmatic proposal which aims to leverage the existing work on data standards being done by Scientific Unions, and bring them together to minimise duplication of effort. It will act as a linking and coordinating group, which will be useful. There is strong support from ICSU for CODATA's taking such a role.

RECOMMENDATION: STRONGLY ENDORSE

5. Globalising Open Science & Access to Research Software (with GYA)

This is proposed as a joint CODATA-GYA (Global Young Academy) Task Group. The principal outputs will be the launch and promotion of the GYA report on Access to Research Software; the group will then develop a database of discipline-specific FOSS for research and promotion of FOSS awareness and creation of FOSS educational events around the report and database. The opportunity to collaborate with the Global Young Academy and with the GYA's Open Science Working Group is welcome. Availability of research software and the accessibility and usability of FOSS is an important issue for research and for research data. The applicability of this work is wide, but far more so in Lower and Middle Income Countries. Unfortunately, the Executive Committee and reviewers felt that the workplan would need to be developed in substantially more detail for the proposal to be convincing. Reviewers questioned also whether the most effective focus would be the labour intensive task of creating a database or registry of free and open source research software or development of strategy and policy advice around access to research software. The Executive Committee suggests that there may be other ways than the current TG proposal to collaborate with the GYA Working Group. If endorsed, this TG must work closely with the African Open Data Platform initiative.

RECOMMENDATION: DO NOT ENDORSE

6. Open Data Strategy in Africa

The proposed group will a) develop a framework for Open Data in Africa; b) customize open data policies, principles and practices for Africa; c) develop and propagate an open data management curriculum. The proposal correlates very closely with the lines of activity of the African Open Data Platform initiative. The funding request is high and the activities may be too ambitious for the group as currently composed. The best way of securing outputs and impact may be to incorporate this group with the African Open Data Platform initiative and invite the proposers also to take leadership roles in other activities, most notably 'Building Foundational Training in Research Data Science' and 'PASTD'.

RECOMMENDATION: DO NOT ENDORSE, but incorporate the activities and invite the proposers to engage with the African Open Data Platform initiative and with other relevant TGs proposals if endorse (specifically Building Foundational Training in Research Data Science' and 'PASTD')

7. Protecting Historical Data

This proposal overlaps with the objectives of the Data at Risk TG. The proposers should be invited to incorporate with Data at Risk (if endorsed). An attempt should be made to conduct the proposed visit to the NOAA-NCEI warehouse in Boulder under CODATA DAR auspices.

RECOMMENDATION: DO NOT ENDORSE, but invite the group to incorporate within and collaborate with Data at Risk

8. Materials Data in IoT and Big-Data Era

The proposed deliverables are a) an international workshop or symposum collocated with an SIP International Symposium or an RDA Plenary; b) a workshop in Japan; c) an organized session or international workshop at SciDataCon 2018. The high level objectives are bold, but the outputs which should lead to these outcomes are very limited and preliminary. A CODATA TG should have more ambition than to run a domestic workshop and a workshop at the next conference. The workplan beyond this has scarcely been sketched. Unfortunately, this is an essentially unfinished proposal and as such cannot in good faith be endorsed. Reviewers and the Executive Committee can see the value in the high level objectives. It is recommended therefore that the group should work with CODATA Japan and with the international CODATA community to organize the proposed workshop and other activities which should *inter alia* lead to a more convincing proposal in two years time.

RECOMMENDATION: DO NOT ENDORSE, but urge CODATA Japan to organize the domestic workshop and other activities in order to prepare a more convincing proposal in two years' time

9. Practice and Impact of Digital Data Citation

This is a refocusing of a very successful CODATA Task Group. The Group's outputs have been among the key achievements of CODATA over the last 4 years: *Out of Cite, Out of Mind* Report http://doi.org/10.2481/dsj.OSOM13-043; contribution to Joint Declaration of Data Citation Principles https://www.force11.org/group/joint-declaration-data-citation-principles-final; International Series of Data Citation Workshops: http://bit.ly/data-citation-workshops. Independent reviewers regarded this as core to the CODATA mission, a flagship. The proposed outputs are a synthesis report on the *outcomes* of the various workshops held during the previous phase of the Task Group, followed by an international workshop on the relation of data to impact factors and assessment measures for research contribution. The findings of the workshop will be published in a White Paper and/or journal article. Taking into account impact factors and assessment measures is an important turn, but the current work to lobby for implementation data citation as standard practice should not be regarded as complete or finished.

RECOMMENDATION: STRONGLY ENDORSE, but request that continue also with international series of advocacy and stakeholder workshops

10. Advancing Informatics for Microbiology (TG-AIM)

The TG has been involved with a number of WDCM initiatives. In is not clear what the TGs outputs are over the last term and in particular how the activities of the TG are distinct from what WDCM would be doing anyway. In so disciplinary specific an area, a CODATA TG must bring something tangible to the table as well as producing something from which other research areas may draw lessons. The Executive Committee does not see a strong case for continuation. Future collaboration with WDCM would be valuable, but it should be reframed in terms of active work on standards, vocabularies and training where CODATA can genuinely enhance desired activities and outputs can be identified.

RECOMMENDATION: DO NOT ENDORSE

11. Data at Risk (DAR)

The TG has been active and outputs include a journal special edition http://www.sciencedirect.com/science/journal/22142428/6/supp/C and a workshop on the Rescue of Data At Risk to be held at UCAR in Boulder prior to International Data Week http://www.codata.org/task-groups/data-at-risk/dar-workshops. The Task Group proposes a special edition in the Data Science Journal, drawn from the papers at the UCAR workshop, Guidelines for

Data Rescue, a book 'Science Data from the Past – a world heritage in peril', a catalogue and 'classification index' of data-sets at risk and a workshop on aspects of data rescue to be held in an LMIC.

The Executive Committee thinks the TG has made good progress but that a major opportunity will be missed if the Group does not, in the next term, produce a major advocacy document with policy recommendations for funders, institutions and practitioners. To achieve this, the Executive Committee suggests that the group broaden its leadership and membership, and actively involves co-chairs to lead subtasks. The TG should also take advantage of the CODATA support structure, including online meetings and the wiki to develop the catalogue. Additionally, the proposers of the 'Protecting Historical Data' TG should be invited to participate.

The Executive Committee advises the TG to use the very successful model of the Data Citation TG: this would be to develop a set of policy recommendations and a procedural manual for institutions and then to disseminate these through an international series of workshops. The Executive Committee recommends endorsement on condition that the Task Group's work programme is revised along these lines.

RECOMMENDATION: ENDORSE, on condition that the workplan is refocussed as described and that co-chairs are encouraged to lead subtasks.

12. Earth and Space Science Data Interoperability (ESSDI)

The TG has worked actively in the geophysical and disaster analysis community. Recent activities include internationalization of a number of conferences and further development of the Historical Atlas of the Earth's Magnetic Field. Proposed in the renewal request is a second edition of the Historical Atlas and development and data filling of the multidisciplinary GIS for geoscience data with sophisticated tools for data analysis. The Task Group provides a strategically important link with an ongoing funded project, which will host a CODATA Eurasia conference in 2017 in Russia.

In a disciplinary specific area, a CODATA TG must produce something from which other research areas may draw lessons. The Executive Committee would, therefore, require greater emphasis upon more generic issues of interoperability and data analysis / data mining and how these issues relate to other and interdisciplinary research areas. Exploring the modalities for publishing data sets and fitting them to be reused effectively outside the immediate community would be one important way of addressing this need. Furthermore, this Task Group is well placed to contribute to the strategically important activity of coordination of data standards with international scientific unions and leverage its strong relationship with a number of the 'geo' unions. The Executive Committee recommends endorsement on condition that the Task Group's work programme is revised along the lines suggested above in order to focus more on issues of interoperability and the promotion of standards.

RECOMMENDATION: ENDORSE, on condition that the workplan is refocused as described

13. Linked Open Data for Global Disaster Risk Research (LODGD)

The Task Group proposes a series of three White Papers, the first of which 'Gap Analysis on Open Data Interconnectivity for Global Disaster Risk Research ' is currently out for public review. The second on 'Next Generation Disaster Data Infrastructure' is underway and will be published in 2017. The third, on 'National Policy Study on Disaster Data around the World' will be the main focus of work for the period of 2016-18. The TG has developed strong links with IRDR and with UNISDR and is a strategically important contribution to the post-Sendai framework for Global Disaster Risk Reduction.

RECOMMENDATION: STRONGLY ENDORSE

14. Preservation of and Access to Scientific and Technical Data in Developing Countries (PASTD)

The Task Group has had considerable achievements in the past: notably an extremely successful and influential training and policy workshop in Kenya in August 2014. That blueprint for twinning training and policy activities is valuable also. For renewal, the TG proposes further policy and training workshops, promoting and expanding the online services of the global change research data publishing and repository (hosted by the Chinese Academy of Sciences) and implementation guidelines and best practices/case studies for 'Open Data in a Big Data World' in developing countries, perhaps as a special issue of the Data Science Journal.

The Executive Committee considers that to be effective, the Task Group needs to ensure that existing personnel are still engaged and to mobilize new personnel. Furthermore, it is imperative that the Task Group, if approved, contribute coherently to CODATA's strategic objectives and in particular to the African Open Data Platform initiative. If the TG can 1) mobilise appropriate personnel to contribute 'implementation guidelines' and a 'good practice guide' that show genuine sensitivity to the circumstances of developing countries; 2) help deliver training based around these materials; and do these things in concert with the African Open Data Platform initiative, then it can continue to make a great contribution to achievement of CODATA's mission. Additionally, it is recommended that the proposers of the 'Open Data Strategy in Africa' TG proposal are invited to participate, and where appropriate given leadership roles in subtasks – this should not be difficult as there is already some overlap in participants. The Executive Committee recommends renewal on these conditions.

RECOMMENDATION: ENDORSE, with requirement to incorporate new personnel and to align activities closely in support of the African Open Data Platform Initiative

15. Science and Management of Physical Objects in the Digital Era

The Task Group is seeking renewal. It tackles an interesting issue. The proposed outputs are of interest and are ambitious: and article in the Data Science Journal on the goals, requirements, and challenges of an 'Internet of Scientific Samples'; a white paper on same topic of the 'Internet of Scientific Samples' to advocate greater investment for targeted developments in digital sample management; a web site that catalogues and links to software tools for digital sample management, ontologies, online tutorials and videos about sample management best practices, and other resources. This is promising and precisely the sort of outputs which the Executive Committee would expect of this group. Unfortunately, at the time of seeking renewal the TG has little to show for the last two years' existence. There *is* considerable enthusiasm in this group and this has manifested itself in sessions organized for SciDataCon 2016. But there is concern that the TG leadership may not have the focus or available bandwidth to ensure delivery of the promising outputs described above. Although the Executive Committee generally encourages collaborative groups (e.g. with RDA, WDS etc), there are concerns that the focus of this group on outputs should not be unnecessarily dissipated by the activity of also seeking RDA WG status.

RECOMMENDATION: DO NOT ENDORSE

Summary of Executive Committee Recommendations in an Ordered Table

1	Coordinating Data Standards amongst Scientific Unions	Closely aligned with CODATA strategy. **STRONGLY ENDORSE**
2	Building Foundational Training in Research Data Science	Closely aligned with CODATA strategy. **STRONGLY ENDORSE**
3	Practice and Impact of Digital Data Citation	Closely aligned with CODATA strategy. **STRONGLY ENDORSE**
4	Linked Open Data for Global Disaster Risk Research (LODGD)	Closely aligned with CODATA strategy. **STRONGLY ENDORSE**
5	Agriculture Data, Knowledge for Learning and Innovation	Valuable new proposal, aligned with CODATA strategy. **ENDORSE**
6	Citizen Science and the Validation, Curation, and Management of Crowdsourced Data (with WDS)	Valuable new proposal. **ENDORSE**
7	Preservation of and Access to Scientific and Technical Data in Developing Countries (PASTD)	Can be more closely aligned to CODATA strategy and achieve greater impact **ENDORSE SUBJECT TO CONDITIONS**
8	Earth and Space Science Data Interoperability (ESSDI)	Can be more closely aligned to CODATA strategy and achieve greater impact **ENDORSE SUBJECT TO CONDITIONS**
9	Data at Risk (DAR)	Can be more closely aligned to CODATA strategy and achieve greater impact **ENDORSE SUBJECT TO CONDITIONS**
10	Globalising Open Science & Access to Research Software (with GYA)	Valuable collaboration, but workplan not strong enough **DO NOT ENDORSE**
11	Open Data Strategy in Africa	Recommend incorporate with other activities. **DO NOT ENDORSE**
12	Science and Management of Physical Objects in the Digital Era	TG has not delivered enough in the last term **DO NOT ENDORSE**
13	Advancing Informatics for Microbiology (TG-AIM)	TG has not delivered enough in the last term **DO NOT ENDORSE**
14	Materials Data in IoT and Big-Data Era	Proposal incomplete and workplan not strong enough **DO NOT ENDORSE**
15	Protecting Historical Data	Recommend incorporate with other activities. **DO NOT ENDORSE**

Relation of TGs to the CODATA Strategy.

The Executive Committee recommends that the General Assembly endorse 9 of the 15 Task Groups proposed. The table below shows how these nine Task Groups would contribute to the priority areas identified in CODATA's strategy.

Promoting data principles, policies and practice	Advancing the frontiers of data science	Building capacity for Open Data and Open Science
Practice and Impact of Digital Data Citation	Coordinating Data Standards amongst Scientific Unions	Building Foundational Training in Research Data Science
Data at Risk (DAR)	Linked Open Data for Global Disaster Risk Research (LODGD)	Agriculture Data, Knowledge for Learning and Innovation
Citizen Science and the Validation, Curation, and Management of Crowdsourced Data (with WDS)	Earth and Space Science Data Interoperability (ESSDI)	Preservation of and Access to Scientific and Technical Data in Developing Countries (PASTD)