Comments on "Guidance for regulating digital platforms: a multistakeholder approach Draft 1.1"

By Guy Berger (personal capacity), 9 January 2023

General:

1. The title of the document includes reference to "a multistakeholder approach". *However, the subsequent Guidance could benefit a lot from more sustained follow-through of this theme.*

2. The introductory section states that the guidance is for actors seeking to regulate, co-regulate and self-regulate digital platforms. *This could also produce benefit if it includes reference to "multi-stakeholder engagements" as a further modality in regulation, and further if there is more sustained follow-through of all four considerations through-out the document. Also very important here would be upfront recognition of budgetary and capacity constraints on regulation of such a complex communications ecosystem, and then to highlight, in this light, the value of light-touch approaches that allow for a multi-modal regulatory ecosystem to thrive.*

3. There are quite a number of repetitions in the document. It is suggested to *retain only the reference that at its most relevant instance*. This would shorten, simplify and clarify the structure and content of the Guidance. In particular, the pre-ambular 5 pages could be condensed into an abstract, and content re-allocated to the actual proposed Guidance that starts midway through page 5. At present, it is not clear why some points are in the first five pages and not later, or vice versa, or why there is some duplication.

Why UNESCO section:

4. This section can benefit from reference to the endorsement by the UNESCO General Conference of the concept of "Internet University" and the four R.O.A.M principles. First, this demonstrates important earlier Member State work on digital governance issues, and their achievement of consensus via the CONNECTing-the-dots conference in 2015, as well as a volume of publications and programmatic work in the area. Second, the frame retains its relevance to the issue of guidance for Platform regulation. For example, the principle of Openness reinforces the stress on transparency, and can be explicitly noted when the latter is discussed. The principle of Accessibility reinforces the points on data access, MIL and disability. *In other words, an already approved approach by UNESCO adds legitimacy and value, and can be profitably and explicitly referenced at relevant points (even in footnotes at times).*

5. It is suggested to unpack "freedom of expression" in a sentence or two, since this has a bearing on regulation. In particular, it could be valuable to reference the rights to *press freedom* (as the freedom to publish to an audience, which is the benefit of social media – see See UNESCO World Trends Reports and World Press Freedom declarations), and *access to information* (which may include access to data that is of public interest). Furthermore, it is important to signal that components of free expression necessarily include *pluralism* (to encourage regulation to enable, not restrict, a diversity of views and languages), as well as that, in order to be a meaningful right for all, there needs to be access to online communications as well as a range of literacies. Transparency around paid expression (*advertising* and *sponsorship*) is also a fundamental feature for effective free expression and for information as a public good.

Leaving "freedom of expression" at a very general level misses an opportunity for the Guidance to have more practical resonance. In addition: a specific reference could also be made that warns explicitly of the risk that state-regulation, especially in the absence of other modalities of regulation,

can lead to arbitrary and over-reached restrictions on freedom of expression. This will avoid misunderstandings about, and abuses of, UNESCO's Guidance.

Independent Regulation:

5. This important section risks duplication with the subsequent section. *The two could be integrated in one place.*

6. What merits elaboration in handling this part of the Guidance is how regulation (i.e. state-authorised regulation) relates to co-regulation and self-regulation (following the promise of the Introduction). *It would be valuable to show how these may be nested within an overall governance system, and also how multi-stakeholder participation constitutes a vital fourth component.*

7. The wording of clause 11[1] can be misleading, in that it could imply that a regulator/system sets the overall goal in isolation. This could be *reworded* to highlight (a) that the overall function for the regulatory systems should be defined by accountable instances such as legislatures, and (b) that such a mandate is best developed through multi-stakeholder engagement. Further, there could be a point (c), to the effect regulators themselves can benefit from expertise and legitimacy in terms of ongoing multistakeholder engagement in terms of how they implement, interpret and review their operations.

The benefits of this guidance section

8. A sentence could be profitably added that shows how the Guidance on content moderation is not being presented in isolation of other relevant dimensions of Platform governance. For example, *"This Guidance presents elements that can articulate with other aspects of Platform governance, such as those that deal with issues like ownership, taxation, inter-operability, privacy and data protection".* Else, it could look perhaps that UNESCO is silo-blind ☺

Further, a sentence can be suggested as well to the effect that "*Effective implementation of this Guidance would require co-ordination amongst the applicable different regulatory authorities at national level at least, and such co-ordination would do well to engage the relevant stakeholders such as the platforms themselves, as well as elements from civil society, technical community, etc."*

9. Section 15 could be reworded to avoid the repetition of "guidance" across the chapeau sentence and 15.1 and 15.2

Section One: the goal of regulation

10. This section could helpfully highlight that regulation can (and should) go beyond a primarily *restrictive* perspective for platforms and communications, and recognise its complementary potential to play an *enabling* role. For example, regulation can have a goal to restrict the transmission of hate speech, *and* a goal to promote media and information literacy amongst those receiving such content. It can have a goal to place graduated restrictions for larger scale platforms *and* it can have a goal to support a pluralism of smaller platforms. Unless this point is made, a one-sided view can prevail over a more dialectical one.

11. Further, the section could do well to mention that a regulatory system should not aspire to micro-regulate all online content, but rather aim to set realistic parameters within which co-

[1] This guidance is principle-based – *with the regulator (or the regulatory system when there is more than one regulatory entity or body), setting the overall goal for regulation w*hich the digital platform services must fulfil (my italics).

regulatory, self-regulatory and multi-stakeholder regulatory practice all play their key roles in protecting freedom of expression while combating content that potentially harms human rights. More valuable substance could then be give such as signalling the *issue of codes of practice and codes of conduct* as evolving tools within the *co-regulatory* part of the wider ecosystem.

12. Clause 18 gives an example of governments being transparent about requests to "remove and block". This point is repeated elsewhere in the document (eg. 23.4), and is also expressed there more widely than only removal and blocking. Suggest to delete here, rather than allow the possible impression that less dramatic requests from government do not rise to the threshold of needing transparency.

13. The last sentence in clause 18 is somewhat cryptic: "The regulator should also be able to scrutinize the scope of requests to ensure adequate balance between illegality and freedom of expression". Is there indeed a "balance" to be struck here, what would it be? It is unlikely that a regulator (even independent) will play the role of a judicial body to condemn a government request for platforms to sanction content prima-facie illegal content – even if the law concerned fails to meet international standards for freedom of expression.

14. Clause 21. Sub-numbering incorrectly changes to 22.2, etc.

15. Clause 22.5. This focuses on oversight of the impact of regulation, which is a good point. A distinct point, that could be provided earlier eg. after 22.2, is to have independent oversight of Platform transparency. *In other words, companies should not "mark their own homework", but have independent audits for submission to the regulatory system.*

Section Two – Fulfilling the goal

16. Clause 23.1. It is not evident why these rights alone are identified. It is suggested to also add *"autonomy, dignity, reputation, privacy, association and policy participation"*

17. Clause 23.5 is tricky terrain. Perhaps UNESCO may wish to avoid this. For governments to guarantee that all content removals have to be subject to due process of law can cripple Platforms (especially smaller ones) with SLAPPs. Each platform, like each media house, has a right to not publish content (or comments) that do not accord with its T&Cs (or choice of editorial line). The public has no right to have their content on every Platform. The only legal right they have is to consumer protection: did the Platform misapply its T&Cs in regard to removal of a given content? Here, consumer protection law already exists in most countries, and does not need to be specifically added as part of specifically platform regulation. Perhaps this point can be explicitly made – *that a regulatory system can profitably articulate with consumer protection regulation to deal with some of the problems with Platforms*.

18. Clause 25. The phrase - community standards - could benefit from being put in quotation marks. It is an ideologically-loaded rather than neutral term. The reality is that the standards do not usually emanate from users (but from owners), and that there is not actually a singular "community" implicated in a given platform.

19. Clause 25.4. This implies that Platforms should be giving attention to ensuring a "plurality aof views and opinions". This is tricky terrain, and UNESCO may wish to avoid it. It suggests that each platform should have a regulated responsibility for a plurality of views and opinions. However, pluralism of views and content diversity are usually approached as a *characteristic of the sector as a whole, and not a legal requirement for each platform* (except public broadcasting). It would be preferable to remove this reference and reword – or elsewhere indicate - that the aim of the

regulatory system should be to ensure pluralism across the online comms sector as a whole, including eg. by providing support for decentralised and minority-language platforms. This is an important enabling role for a regulatory system.

Content management policies

20. Clause 27. It is not clear what is meant by "types of content". If this is a reference to formats – eg. visual content, vs verbal content, for example, it may be more complicated than the Guidance possibly implies. To police written text for deliberately misspelt words/names or dog whistles is complex, but it is even more challenging for memes, live streaming or metaverse interactions.

21. Clause 27.1. The reference to "protected characteristics" is understood, but at the same time if there could be a reference to a UN interpretation *of what merits protection*, this would align to international standards.

22. Clause 28.2. It is not fully evident that identifying behaviours that promote mis-disinfo involves a trade-off with privacy and anonymity. Platforms assess such behaviours using IP addresses and patterns of content promotion amongst these. I have not seen cases where they make the associated IP addresses public nor link them to specific actors (in ways that may violate the privacy of the latter without public interest justification).

23. Clause 28.3. For all Platforms to report on how they identify synthetic content that misinforms, is quite demanding technologically. Is the issue not rather that there is misinformation (measured against a standard of truth and context), rather than how this content may have been created?

24. Clause 29. This wording implies that all previous items (28.1 through to 28.5) merit having restrictions on virality. However, 28.5 refers to state-linked accounts. But, eg. UNESCO would support that science-based anti-Covid messaging from government, should be allowed to go viral unhindered. State content may be questionable or not; the key thing is disclosure of provenance. Hence the wording of 29 needs tweaking to exclude 28.5.

25. Clause 30. Platforms can inform users when content is removed, but the meaning of doing soe also when content "subject to content moderation" needs unpacking. Depending on how "moderation" is defined, ALL content on most platforms can be seen to go through a form of moderation. Algorithmic ordering and recommendation systems de facto give green lights, yellow lights and red lights to content in a way that advantages and disadvantages (to the extent of exclusion). Even on eg. Mastodon, with a purely chronological feed, users may report items that violate T&Cs to the moderator of the server they use, and the particular content may be removed, the poster expelled, or whole servers may be de-federated. In this light, it could be better to spell out beyond removal, cases like: when content is expressly labelled; restricted in terms of commentary or re-sharing or advertising association; given special limits in terms of amplification or recommendation (as distinct from "organic/algorithmic" amplification and recommendation).

26. Clause 32: Grammar and spelling errors here. Additionally, the sentence that "where possible", users should have access to a platform rep in their own country, does not give much guidance to regulators. It could possibly be more useful to state that "regulators could require that at a scale of users or impact, to be defined, an available platform must have a local presence". It is for regulators to assess what is possible, and for the platform to decide whether to operate in the jurisdiction or not accordingly, if they do not wish to have a local presence.

27. Clause 33.2. This para is arguably the place where the last sentence[2] under Clause 34 would be better located. Additionally, a fairly brief reference to "a systematic risk assessment" is perhaps insufficient in terms of guidance. Instead, it could be elaborated possibly that *"Platforms should conduct annually a general risk assessment of potentially damaging content, and report on what mitigation measures they adopted with what impact on human rights*". To this, it could be added: *"As elaborated later, further specific risk assessments can be called for when there either predictable and fast-breaking events that merit consideration".* While much in the Guidance will not necessarily be "new" for regulators, the issue of requiring human rights risk assessments of Platforms is something to which *UNESCO could give much more flagging*. This important component, while not yet universally recognised as a key component for regulation, enables regulators to remain at arms-length from the complexities and granular detail of content moderation, and correctly puts more legal onus on the platforms themselves to follow the Ruggie Principles. It does beg questions about what constitutes an effective risk assessment and how such gets rolled out, and assessed by regulators. *Guidance could do well to give at least another clause on these issues*. This is even though recognising that this high-level topic will indeed benefit from separate and more concrete follow-up at implementation level which UNESCO is already already working on. At present, however, *there is insufficient information being given here*.

28. Clause 33.4: Amongst the options cited for responses "providing alternative information….", it would be important to mention "de-monetization" in terms of advertising.

29. Section 35: This is one section that addresses advertising in some depth. However, it could be helpful to regulators, both advertising regulators and electoral regulators*, to tackle the issue of truth in advertising*. As is known, Meta consciously does not moderate political adverts for disinformation. Twitter, which recently decided to accept political adverts, will probably do the same. *A regulatory system may well need to address such a laissez-faire approach in view of ensuring electoral integrity. Would UNESCO have Guidance for this?*

30. Clause 37.1 – suggest to delete. The point is already implicit earlier.

31. Clause 38 on "data access". It is not clear what is meant by "*stable*" access. There is no inherent reason why access should be only for data that is "*aggregated*" (which is quite a general term), as distinct from data that is "anonymised".

32. Clause 47. It is not clear what is meant by "*minimum safety requirements*". The term "safe" is also used in clause 6. Public safety is a criterion that can be part of a three part-test for legitimately limiting expression according to international standards. One can distinguish this, however, from other realms: being safe in terms of having the rights of children being protected, or safe in terms of having other rights protected (eg. rights to dignity, equality, political participation). A generic use of the term safety, which conjures up only a vague threat landscape, is not very helpful for specific regulatory guidance, and also opens the way for exploitation*. It is recommended that UNESCO think about whether it wishes to use the word "safety" in this Guidance, given that the term can indeed be used to overly-restrict freedom of expression, as evident in debates about the UK online safety bill.*

33. Clause 49: The proposed *power to summons Platforms* deemed to be not be compliant as currently formulated opens the risk to cases of subjective, arbitrary and selective regulation. It is suggested instead that the Guidance propose a public inquiry into possible cases (so as to avoid Regulators "deeming") that the company could be summonsed without clear prior process. The

[2] The platform should be reflecting on how any product or service impacts upon user behaviour and not just on the aim of user acquisition or engagement

second part of this Clause also seems to call out for attention. It gives Regulators the power only to "*recommend* a set of measures". If an entity is not compliant with regulations, should it not receive *instructions* and then possible *sanctions* for continued violation? Would mere recommendations constitute regulation, if there are no teeth? On the other hand, to avoid potential abuse, could it be proposed in the Guidance that the Regulatory system should have *the power to take a Platform to court* in the event of continued non-compliance? This would be preferable to implying that a regulator itself should (in the words of the Clause) make "a judgement" - even if "evidence-based"? It is better, probably, to envisage the Regulator making assessments, and leaving it for Courts to make "judgements". For the sake of protecting freedom of expression, it may be better to not have regulators giving assessments and imposing fines in the case of online communications, but rather keep the courts as custodians of rule of law.

34. Footnote 16 could deleted without loss.