Digital Accountability can re-legitimate Multilateralism

Florian Cafiero, Jean-Philippe Cointet, Grégoire Mallard

Digital Accountability can re-legitimate Multilateralism

Florian Cafiero[1,2] Jean-Philippe Cointet[3,4] Grégoire Mallard[2]

[1] Université Paris Sciences et Lettres
[2] Center for Digital Humanities and Multilateralism, Graduate Institute Geneva
[3] médialab, Sciences Po Paris
[4] INCITE, Columbia University

Correspondence: florian.cafiero@chartes.psl.eu

Multilateralism is undergoing a profound crisis of effectiveness and legitimacy. [1, 2] Skepticism stems not only from governments, as illustrated by rising patterns of member state criticism and withdrawal,[3, 4] but also from citizens, who express lower confidence in international organizations and increasingly perceive global governance as distant and democratically deficient.[5, 6] In such a context, restoring confidence is essential: multilateralism must show, in public, how decisions are made and on what grounds. We argue that strengthening public accountability—anchored in timely, intelligible records of negotiation—is a precondition for rebuilding trust and making multilateral cooperation deliver on its promises.

There are, of course, good reasons to preserve a culture of diplomatic discretion: negotiators need protected spaces to explore options, test compromises, and signal privately without immediately locking themselves into public positions. Yet practices of secrecy that were designed to preserve room for manoeuvre during negotiations have gradually been extended far beyond what is necessary. Collective decisions are thus made behind doors that often remain closed long after the risks of disclosure fade. Not surprisingly, when governments that distrust democratic accountability negotiate with each other on the multilateral scene, they reproduce and even amplify practices of secrecy found at the national level. After the twentieth century opened with a call for open negotiation, secrecy is now getting overused [7].

This lack of openness not only reinforces perceptions of weak public accountability, it also prevents systematic historical reconstruction of multilateral bargains and creates an opacity in which "alternative" narratives—including conspiratorial ones—can more easily circulate and undermine the legitimacy of international organizations and the negotiations they host. The result is predictable: in the absence of reliable and faithful records, speculation, sometimes amplified by algorithmic means, fills the gaps, and the system's legitimacy suffers[8].

To help foster this change, multilateralism does not necessarily need to adopt a radically transparent and open framework. At a time when massive digitization projects of multilateral archives are conducted [9], and when IOs produce millions of digital born records and documents, we claim that creating public accountability through proper diplomatic digital data curation is direly needed. *Digital accountability* refers to this disciplined, time-bound approach to releasing the records of diplomatic negotiations, which should enable citizens to track how diplomatic outcomes were reached. By publishing versioned drafts, agendas, minutes where they exist, along with basic metadata once talks conclude and risks subside, international bodies would make their decisions intelligible [10], auditable, and teachable to a wide variety of publics. The publics, whether conceived as an assemblage of national publics or a global citizenry of cosmopolitan individuals, desire and deserve information delivered in a transparent way in order to trust governments. Reviving the alliance between multilateral organizations and citizens is the main payoff. A secondary dividend is that such records would finally allow today's analytical

methods developed by scholars in diplomatic studies, international relations and computational social sciences to serve the public interest rather than speculate around missing traces.

A compact for ex post transparency. A workable equilibrium is within reach. Protect live negotiations; then release what explains the outcome.

- *Time-bound release.* Publish agendas, non-sensitive minutes, and *versioned* negotiating drafts with redlines within 6–12 months after talks close, with longer delays only for a concrete, reviewable security rationale.
- *Minimal metadata.* Give every document a stable identifier, date, provenance, language tags, and a link to previous and subsequent versions. If material is withheld, log its existence and the reason.[11]
- *Multilingual care.* Align translations and flag contested terms. Translation shapes meaning; documenting choices is part of accountability.
- *Cross-venue linkage.* Light-touch issue tags and cross-references should connect related texts across institutions and over time. Bargaining does not respect institutional borders.

These are well-understood record-keeping practices in research and procurement,[11, 12] and they fit diplomacy's constraints: no real-time disclosure, targeted redaction when harm is plausible, and priority for the right to *understand* decisions taken in the public's name.

What is missing now. Today's diplomatic record is often scattered, unversioned, and hard to link. Even where access exists, the absence of stable identifiers and public change logs makes it difficult to reconstruct how specific phrases traveled, who made the first move, or when a redline softened. However, we understand that reality is far from practice. IOs are at multiple stages of archival development, ranging from those without archival services due to years of neglect or limited resources, to those developing state-of-the-art records management services and innovating with AI. Turning best archiving practices into a reality may seem even more unrealistic for multilateral organizations facing an unprecedented budget crisis. However, we highlight three examples below that can illustrate how modest documentary reforms could restore/enhance the picture in three typical multilateral sites.

UN Security Council consultations. The Council's guidance recognizes informal consultations for which no verbatim public record exists.[13, 14] After a resolution is adopted, the public sees the final text and sponsor names but not the sequence of redrafts that explains why particular terms survived. Post-embargo release of principal drafts including dates, sponsors, and machine-readable redlines would enable clause-level histories of how sensitive language was negotiated.

UNFCCC negotiations. Climate diplomacy unfolds across open and closed formats.[15] Over multiple sessions, formulations migrate and cross-references proliferate.[16] Without structured version histories, it is nearly impossible to see how precedent is invoked, contested, and transplanted. Publishing session-by-session versions with stable identifiers would make the life cycle of key clauses visible.

EU trilogues. Informal negotiations among Parliament, Council, and Commission have long raised concerns about uneven access and linkage.[17–20] Standardized identifiers, consistent metadata (participants, dates), and public change logs for non-sensitive items, released once acts are adopted, would enable the reconstruction of compromises and the evaluation of reforms.

Legitimate concerns, workable answers. Three risks recur and can be addressed without sacrificing accountability.

Chilling effects and security. We suggest to publish data ex-post according to a tiered release protocol. Live talks are protected; sensitive annexes can be partially released or redacted with documented reasons. Consultation with affected parties reduces the chance of harm. The goal is reproducible understanding, not play-by-play disclosure.

Manipulation by digital tools. When digital systems are trained on diplomatic records or used to assist negotiators, officials, or publics in drafting and analysis, they must be governed carefully. Any system that helps generate text or simulate pressure on the basis of these corpora or other data should retain prompts and outputs, keep audit logs, avoid optimizing targeted persuasion toward identifiable individuals, and undergo independent red-team testing.[21, 22] Clear uncertainty reporting is part of the duty of care.

Inequality. Dominant languages and well-resourced institutions already crowd out others.[16, 23] Funding for translation, inclusion of small states and civil society in curation, and explicit support for underrepresented regions are needed so that openness does not entrench imbalance.

Five commitments international bodies can adopt now.

1. **Release on a clock.** Agendas, non-sensitive minutes, and versioned drafts with redlines published within shortly after negotiations close.

2. **Make changes traceable.** Stable identifiers, dates, provenance, language tags, and public change logs for every released item.[11]

3. **Align languages.** Provide aligned multilingual texts and short notes on contested terms and translation choices.

4. **Link across arenas.** Tag issues and cross-reference related documents across institutions and time.

5. **Normalize prospective evaluation.** Encourage research groups to publish their processing pipelines alongside results.[24, 25]

What openness buys us—after the embargo lifts. Only once the record exists can society learn from it. Releasing versioned drafts and minimal metadata would allow researchers, journalists, and citizens to follow the making of compromise rather than guess at it. In trade, similar calibration between transparency and confidentiality has already been explored to support legitimacy while protecting sensitive exchanges.[26, 27]

Importantly, openness is not an end in itself. It is the prerequisite for responsible analysis. With real, governed corpora, the tools we already have can finally be useful to servepublic understanding. The semantic longitudinal analyses of resolutions, reports, and drafts can revealhow issues rise and fall and how concepts migrate across venues.[23, 28–30] Clause-level tracking can reveal legal diffusion and identify hubs of normative influence.[31] Network views can link discourse to roles and coalitions.[32] Monitoring can flag rhetorical shifts around humanitarian emergencies that deserve scrutiny, not spin.[33, 34] Prospective forecasts of agenda outcomes or coalition alignments can be registered in advance and scored after the fact, creating a norm of honest error and cumulative learning rather than hindsight wisdom.[22, 24] And generative systems can assist in drafting communiqués sensitive to precedent and cultural nuance—*if and only if* their inputs come from curated, versioned records rather than unverified collections of texts scraped online.[25, 35–37]

Conclusion. Multilateralism will not capitalize on the public trust it has gained for a century and fend off current governmental encroachments unless it succeeds in publicly sharing the reasons of its successes and failures. Releasing, on a schedule, the versioned traces that explain how collective decisions came to be is the first step. Without reliable documentary infrastructures, both human commentators and algorithmic systems are pushed toward conjecture, over-interpretation, and sometimes outright fabrication. Do this, and diplomacy becomes legible; public debate becomes informed; and the analytical tools of our time can illuminate rather than invent. The choice is not between secrecy and spectacle. It is between a well-governed documentary infrastructure that enables accountability and a patchwork of heterogeneous fragments of disconnected texts that invites conjecture.

References

[1] Jens Steffek. Triangulating the legitimacy of international organizations: Beliefs, discourses, and actions. *International Studies Review*, 25(4):viad054, 2023. doi: 10.1093/isr/viad054.

[2] Henning Schmidtke and Tobias Lenz. Expanding or defending legitimacy? why international organizations intensify self-legitimation. *The Review of International Organizations*, 19(4):753–784, 2024. doi: 10.1007/s11558-023-09498-0.

[3] Hylke Dijkstra and Farsan Ghassim. Are authoritative international organizations challenged more? a recurrent event analysis of member state criticisms and withdrawals. *The Review of International Organizations*, 2024. doi: 10.1007/s11558-024-09557-0.

[4] Inken von Borzyskowski and Felicity Vabulas. Public support for withdrawal from international organizations: Experimental evidence from the US. *The Review of International Organizations*, 19(4):809–845, 2024. doi: 10.1007/s11558-024-09539-2.

[5] Lisa Dellmuth, Jan Aart Scholte, Jonas Tallberg, and Soetkin Verhaegen. The elite–citizen gap in international organization legitimacy. *American Political Science Review*, 116(1): 283–300, 2022. doi: 10.1017/S0003055421000824.

[6] Farsan Ghassim. Perceptions of a global democratic deficit: An international survey experiment. *Perspectives on Politics*, 22(4):1224–1247, 2024. doi: 10.1017/S1537592724000987.

[7] Grégoire Mallard. *Fallout: nuclear diplomacy in an age of global fracture*. University of Chicago Press, 2019.

[8] Thomas Hale, David Held, and Kevin Young. *Gridlock: Why Global Cooperation Is Failing When We Need It Most*. Polity, Cambridge, 2013.

[9] Colin M Wells. Total digital access to the league of nations archives: Digitization, digitalization, and analog concerns. In *Archiving Conference*, volume 16, pages 12–16. Society for Imaging Science and Technology, 2019.

[10] Rebecca Adler-Nissen and Alena Drieschova. Track-change diplomacy: Technology, affordances, and the practice of international negotiations. *International Studies Quarterly*, 63 (3):531–545, 2019.

[11] Open Contracting Partnership. Open contracting data standard (ocds): Release schema, identifiers, and guidance on redaction and versioned publication, 2020. URL `https://standard.open-contracting.org/latest/en/schema/reference/`. Guidance on rolling releases, stable identifiers (OCID), versioning, and documenting withheld material.

[12] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Nikolaus Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(160018), 2016. doi: 10.1038/sdata.2016.18. URL `https://www.nature.com/articles/sdata201618`.

[13] UN Dag Hammarskjöld Library. Meeting records — un security council (research guide), 2025. URL `https://www.dagdok.org/w/dd/en/un-system/security-council/meeting-records`. States that preparatory work occurs in informal consultations for which no public record exists.

[14] United Nations Security Council. Note by the president of the security council s/2024/507 (working methods), 2024. URL https://main.un.org/securitycouncil/sites/default/files/2025/Note%20507_2024_Rev.pdf.

[15] UNFCCC. How-to-cop: Overview of negotiating formats, 2024. URL https://unfccc.int/sites/default/files/resource/How-to-COP.pdf. Explains open vs. closed meetings (informal negotiating formats).

[16] Tommaso Venturini, Nicolas Baya Laffite, Jean-Philippe Cointet, Ian Gray, Vinciane Zabban, and Kari De Pryck. Three maps and three misunderstandings: A digital mapping of climate diplomacy. *Big Data & Society*, 1(2), 2014. doi: 10.1177/2053951714543804.

[17] European Ombudsman. Ombudsman calls for more transparency in eu law-making "trilogues", 2016. URL https://www.ombudsman.europa.eu/en/press-release/en/69214.

[18] Gijs Jan Brandsma. Transparency of eu informal trilogues through public council meetings: The dynamics of voting and institutional transparency. *Journal of European Public Policy*, 26(7):912–928, 2019. doi: 10.1080/13501763.2018.1488887.

[19] Marius Spoerer. The european ombudsman's role in access to documents. *ERA Forum*, 2022. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC9465659/. Open access version.

[20] ClientEarth. Transparency of trilogues: briefing and access-to-documents practice, 2025. URL https://www.clientearth.org/media/54tnqlkr/clientearth-briefing-on-transparency-in-trilogues.pdf.

[21] A. Pan, J. S. Chan, A. Zou, et al. Do the rewards justify the means? measuring trade-offs between rewards and ethical behavior in the machiavelli benchmark. In *International Conference on Machine Learning*, 2023.

[22] J. P. Rivera, G. Mukobi, A. Reuel, et al. Escalation risks from language models in military and diplomatic decision-making. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 2024.

[23] Martin Grandjean. The paris/geneva divide: A network analysis of the archives of the international committee on intellectual cooperation of the league of nations. In Elisabet Carbó-Catalan and Diana Roig Sanz, editors, *Culture as Soft Power: Bridging Cultural Relations, Intellectual Cooperation, and Cultural Diplomacy*, pages 65–98. De Gruyter, Berlin; Boston, 2022. doi: 10.1515/9783110744552-004.

[24] Ian Gray and Jean-Philippe Cointet. Multilateralism of the marginal: Least developed countries and international climate negotiations, 1995–2016. *American Journal of Sociology*, 129(3):796–855, 2023. doi: 10.1086/726812.

[25] Florian Cafiero. Datafying diplomacy: How to enable the computational analysis and support of international negotiations. *Journal of Computational Science*, 71:102056, 2023. doi: 10.1016/j.jocs.2023.102056.

[26] Robert Wolfe. Letting the sun shine in at the wto: How transparency brings the trading system to life. Staff working paper ersd-2013-03, World Trade Organization, Economic Research and Statistics Division, 03 2013. URL https://www.wto.org/english/res_e/reser_e/ersd201303_e.pdf. Discusses limits in practice for sensitive negotiating documents.

[27] Petros C. Mavroidis. The evolving understanding of transparency in the wto. Columbia Law & Economics Working Paper, 2015. URL `https://scholarship.law.columbia.edu/faculty_scholarship/788`.

[28] Sejung Park, Dahoon Chung, and Han Woo Park. Analytical framework for evaluating digital diplomacy using network analysis and topic modeling: Comparing south korea and japan. *Information Processing & Management*, 56(4):1468–1483, 2019. doi: 10.1016/j.ipm.2018.10.021.

[29] Rafael Mesquita and António Pires. The references of the nations: Introducing a corpus of united nations general assembly resolutions since 1946 and their citation network. *Journal of Peace Research*, 2024. doi: 10.1177/00223433241254997.

[30] Slava Jankin Mikhaylov, Alexander Baturo, and Niheer Dasandi. Words to unite nations: The complete united nations general debate corpus, 1946–present. *Journal of Peace Research*, 2024. doi: 10.1177/00223433241275335.

[31] S. Uddin, H. Lu, W. Alschner, et al. An nlp-based novel approach for assessing national influence in clause dissemination across bilateral investment treaties. *PLOS ONE*, 19(3): e0298380, 2024. doi: 10.1371/journal.pone.0298380.

[32] Steffen Eckhard, Ronny Patz, Moritz Schönfeld, and Hilde van Meegdenburg. International bureaucrats in the un security council debates: A speaker-topic network analysis. *Journal of European Public Policy*, 30(2):214–233, 2023. doi: 10.1080/13501763.2021.1998194.

[33] Alex Knapp. Speaking of civilians: Automated text analysis of the united nations' framing of complex humanitarian emergencies. *International Interactions*, 50(2), 2024. doi: 10.1080/03050629.2024.2318559.

[34] Diana Ingenhoff, Gaia Calamai, and Efe Sevin. Key influencers in public diplomacy 2.0: A country-based social network analysis. *Social Media + Society*, 7(1), 2021. doi: 10.1177/2056305120981053.

[35] Joanna Wojciechowska, M. Odrowaz-Sypniewski, M. W. Smigielska, et al. Deep dive into the language of international relations: Nlp-based analysis of unesco's summary records. In *3rd Workshop on Computational Linguistics for the Political and Social Sciences*, 2023.

[36] Meta Fundamental AI Research Diplomacy Team (FAIR). Human-level play in the game of diplomacy by combining language models with strategic reasoning. *Science*, 378(6624): 1067–1074, 2022. doi: 10.1126/science.ade9097.

[37] Volker Stanzel and Daniel Voelsen. Diplomacy and artificial intelligence: Reflections on practical assistance for diplomatic negotiations. Technical report, Stiftung Wissenschaft und Politik (SWP), 2022.