Defining a standardised approach on the collection and communication of geolocation coordinates for deforestation-free commodity value chains

GEOSPATIAL DATA SHARING PROTOCOL
STAKEHOLDERS' COMMENTS LOG

27th October 2023

Introduction

The first version of the GDS Protocol was developed between June and September 2023 with the support of the Food and Agriculture Organisation (FAO) under the Forest Data Partnership, a program aiming to halt and reverse forest loss from commodity production by improving global monitoring and supply chain tracking.

During this project phase, a first draft was consolidated by Preferred by Nature's team, assisted by an expert group composed of representatives from different organisations, including sustainability certification schemes and commercial organisations. Representatives of the group spanned both the forestry and agricultural sectors. The expert group was convened over a 2-day workshop at the end of August 2023, to work through technical options for the Protocol.

The resulting consolidated draft went through a wide and open consultation process in September 2023, in order to gather feedback from all interested parties. It was sent directly to over 50 organisations targeted as potentially interested parties, or users, of the Protocol, including large traders in key forest-impact commodities. The draft Protocol was also shared on Preferred by Nature's LinkedIn page and was made available via our website and press-releases.

This is the comments' log for the September 2023 consultation round.

Comment #	Stakeholder group	Comment/feedback	Response	If and how it was used to update the protocol
-5 Topic: General - Scope				
1	Consultants	Maintain the current narrow focus and avoid broadening scope beyond geospatial data transfer - it is not suggested to include various mega data elements.	At this stage, the Protocol kept a narrow focus and dictates the bare minimum of metadata that must be included. This does not prevent Protocol users to agree on additional optional metadata. Adding other required metadata may be discussed and added in later Protocol version according to the needs.	N/A
2	Government and intergovernmental bodies	Data-related issues are already being introduced during the data collection phase and not so much during data sharing. How do you address possible problems potentially coming from data collected with insufficient quality and not harmonized in terms of content, risking putting "apples and pears" together?	Although it is well understood that data quality is important and the "garbage in garbage out" scenario must be addressed, data quality and validating data were not in scope for this initial phase of the project. Data validation may be considered in the further stages and the Protocol adjusted accordingly.	N/A
3	Sustainability standards systems	We know that many countries at origin have Data Protection Laws in place, hence any effort to collect geo-location data must be in line with national laws, norms, and procedure to ensure that farmer's data in protected, particularly personal data.	The data collection part is not in scope of the Protocol. Moreover, the Protocol does not make it mandatory to maintain information about the supplier network trading the goods in scope (e.g. farmers' identities and personal data). Protocol users must agree between themselves which additional metadata they will communicate to each other and must consider Data Protection Laws.	No
-5 Topic: General – Data security				
4	N/A (aggregated)	- Ensure there is effective security of data along the supply chain. - The protocol should address how data is protected in transit and at rest (as stored on a repository like Sharepoint, OneDrive, NAS), if it is intended to be used broadly (outside of just EUDR) and in cases where data sharing may include ownership names, personal, or non-public information. - The primary concern within our sector revolves around the	Data security was not in scope for this initial phase of the project. Data security elements may be considered in the further stages and the Protocol adjusted accordingly.	No

		security of information and data exchanged between Sender and Receiver organizations, as well as the potential of misuse of it. This apprehension not only pertains to secure data storage by the Receiver but also extends to the prevention of any potential misuse of geolocation data, which is considered commercially sensitive information. Such data misuse could manifest in various ways, including the unauthorized prediction of market dynamics and commercial strategies of the Sender organization by the Receiver organization (or thirds). Furthermore, it is crucial to contemplate the vulnerability of data in editable formats, particularly in the context of potential data manipulation or tampering between the Sender and Receiver.		
		Topic: technical - identifiers		
5	Sustainability standards systems	The first statement assumes that these public registries maintain uniqueness but what if the data travels beyond the initial context? While the contract may specify the registry used, will the name of the registry travel with the geospatial data to ensure provenance is maintained? E.g., A national registry lists Home Depo with Unique ID: XPJF6. A global registry that lists IKEA with Unique ID: XPJF6. These are two different registries, but the Unique ID is the same for both Home Depo and IKEA.	The Protocol is not making it mandatory to maintain supply chain network information alongside the Geospatial Data, which is why it is not mandatory to declare which registry is being used for organisation identifiers. However, it is recommended to add this as optional metadata, so that this information may travel beyond the initial context, if necessary.	Yes - a clarification sentence has been added to the Protocol (2.1 guidance)
6	Government and intergovernmental bodies	The protocol discusses unique identifiers for transactions, sender and receiver. How will those identifiers be assigned and by whom? Will there be a digital data-sharing platform where users retrieve a unique ID during registration and new id's are assigned to every transaction that is done on the platform?	Identifiers must be assigned and agreed upon by Protocol users themselves. It is not intended to provide for a unique ID generation digital tool, nor to host individual transactions on a specific platform. It is expected that many due diligence and trading platforms will continue being used by relevant organisations.	No
		Topic: technical - JSON		

Comment #	Stakeholder group	Comment/feedback	Response	If and how it was used to update the protocol
7	N/A (aggregated)	Would GeoJSON be used as the format for encoding collections of simple geographical features and then JSON for the non-spatial attributes, or would JSON be used for the full "envelope"? Are GeoJSON and JSON are both acceptable?	The use of a Spatial Indexing System makes the use of JSON sufficient. Geospatial data is communicated through relevant cell IDs and not geometries. GeoJSON is not needed, but there is nothing preventing Protocol users from also using GeoJSON.	Yes - clarification added to the Protocol (2.4 guidance)
Topic: technical – spatial indexing system				
8	N/A (aggregated)	What is the background on using "spatial indexing" such as S2 or H3and what is its added value? Isn't it adding complexity to something that is supposed to be simple? Why is the Sharing Protocol referencing Spatial Indexing rather than defining raw Geospatial geometries such as defined in ShapeFiles and GeoJson?	Index technology for geospatial data offers advantages in terms of efficiency, scalability, uniformity, and integration with modern data systems, especially when dealing with large-scale and dynamic spatial datasets. It is a good way to fix many issues associated with projection, as it passively enforces data quality issues (e.g., having a geometry that is not a closed polygon). Indexing also minimizes the size of data files as it converts Geospatial Data to numeric values. Making operational search (inclusion, collision, overlap), using pure GIS information can be extremely slow: a geo indexing system will be way faster as shown in both Google and Uber application. It also works well for organisations with scarce GIS capabilities.	N/A
9	N/A (aggregated)	Comments on S2 and H3 systems: - Preferred use of H3 based on use cases that prioritized smoother visualization of the results. - .NET Core libraries can support better with H3 - https://github.com/pocketken/H3.net - S2 is more complex and resource-intensive - H3 spatial resolution is more consistent and simpler - H3 open-source community is more active - S2 is more accurate, but we are referring to data that have been mainly collected with devices with a low accuracy - Both systems work well for organizations with scarce GIS	There is currently no global consensus on an indexing system. Using either indexing system will function. S2 and H3 are two the two acceptable spatial indices, each having specific advantages.	Yes, the protocol is not prescriptive in which to use.

Comment #	Stakeholder group	Comment/feedback	Response	If and how it was used to update the protocol
		capabilities. - S2 and H3 seem to have limited language support - H3 us not hierarchical: hexagons cannot be fully embedded into other hexagons, which is the case for squares. S2 is better to build a full pavement and lossless system.		
Topic: technical – data structure				
10	Digital service providers	It seems weird to have all the envelopes for each node in the supply chain and then a single spatial index - it seems this should be a many to many relationships rather than a many to 1, especially when it's the one node at the point of harvest that is actually requiring the GeoSpatial data.	This element has been considered and is likely to be handled by additional API development (e.g. recursively reading envelopes to extract each node participation to the final geometry), outside of the Protocol content.	No
11	Digital service providers	Where did the 2m come from? This may be challenging in some geographies due to terrain challenges etc. Suggestion that some mathematical analysis to look at an average resolution over a number of points may be able to smooth out where there challenging data collection areas.	This element came from finding a balance between regulatory requirements (e.g. EUDR 6 digit for coordinates) and what is reasonable, feasible and meaningful in the context of addressing deforestation issues. This requirement should be further tested and discussed by stakeholders, and further Protocol versions may update this figure accordingly.	No
12	Digital service providers	ID, sender and receiver fields: a data type of uuid might be better than an int because it means that systems can generate values without the danger of the value "clashing" with another system that might create the same integer value. A uuid is always unique from the moment it gets created.	This proposition should be further tested and discussed at later stages. Randomness of uuid allocation from different computers may be inconvenient, unless a uuid generation system using the internal clock of the computer is being used. Further Protocol versions may update this accordingly.	No
13	Digital service providers	The overall payload could get potentially very large if the Sources field has many complex nested graphs of envelope structures represented as JSON. Would pointers via uuids to envelopes in the Sources field be more efficient, allowing a software application to "get" more envelopes via uuids only when necessary?	This proposition should be further tested and discussed at later stages. This would only work if a registry of envelope is provided. By using ID in the envelope instead of nested envelopes, it would affect the "stand alone" property. Further Protocol versions may update this accordingly.	No

Comment #	Stakeholder group	Comment/feedback	Response	If and how it was used to update the protocol
14	Digital service providers	Dict [str, Any (hashable)] – "Any" seems very broad because software systems will need to interpret this data and place it internally into some sort of data type. Perhaps expressing "Any" as JSON might suffice. But perhaps a stricter definition of supported data types might be better, such as supporting text, date/time, int, decimal, base64-encoded binary (e.g. for images), and so on.	The formulation of the metadata structure in Annex 1 may be further tested and refined. It could be something like Dict [str, Union [datetime. date, str, int,…]. In general, whatever the internal system needs will be envelope-compatible as long as it's hashable. Further Protocol versions may update this accordingly.	No
		Topic: general – next steps		
15	Consultants	Requirement to pilot test the Protocol across various commodities, geographies and scales (i.e. grower, trader, manufacturer, brand owner, including large companies and SME) to ensure it really is "fit for purpose" i.e. fully operational between supply chain actors and EUDR aligned.	Pilot testing is indeed extremely important but was not in the scope of the current project phase. Organisations interested in participating in testing are welcome to contact us.	N/A
16	Sustainability standards systems	Who is the owner or entity responsible for the maintenance and upkeep of the standards? Once "published" and in use, if actors have recommendations for further adjustments, where should those be sent, what is the expected frequency of updates, how can users be notified or find updates, etc.?	The Protocol ownership and governance still need to be collectively defined and implemented. We welcome all interested organisations to submit their suggestions and actively participate in these next steps.	N/A