Carbon-ML – The Language of Carbon – White Paper DRAFT

Version: 1.90 updated April 25, 2022

Author(s): Lynn Connolly, lynn.connolly@carbon-ml.org

Jonathan Hollander, jh@carbonfinancelab.com

Nick Gogerty, ng@carbonfinancelab.com

Goal of document: An overview of Carbon-ML ecosystem,< Carml> language features and functions use cases / benefits and governance suggestions.

Using this document: This document has high level examples and business cases and is not intended to be overly proscriptive or exhaustive of implementations or states tools, actors or deployments.

Fully enumerated <CaRML> examples and detailed reference implementation are kept in an online Github Repository.

Table of Contents

About Us

The Carbon-ML project is developing an open-source ecosystem and language to support ongoing declarations of measurements for embodied carbon in any product or service using global and local contexts.

Carbon-ML is incubated by Carbon Finance Labs in partnership with Oxy Low Carbon Ventures with a goal to evolve into an independently governed project with multiple private, public sectors actors spanning industries, technology providers, data consumers and geographies.

Carbon Finance Labs	A finance and technology incubator creating climate change solutions. Our impact comes from a global network of resources and knowledge built over decades spent in the carbon, finance and technology sectors.
OXY LOW CARBON VENTURES	Oxy Low Carbon Ventures, LLC (OLCV), a subsidiary of Occidental, Petroleum

The Vision

As the world becomes increasingly concerned with carbon emissions accelerating environmental and climate impacts, there is an increasing emphasis on understanding the amount of carbon emissions produced during the entire lifecycle of a product or service.

Therein lies a problem. Data about the embodied carbon in the buying, producing or selling products or services is hidden and stuck across supply chains. Currently, the measuring, reporting, tracking and declaration of embodied carbon within a product or service is a mess and doesn't flow. There are numerous measurements, policies, mandates, systems, etc. designed to measure carbon emissions, but lack of consistency between them hinders carbon data flow. These systems and metrics are designed to report carbon at the company level, not at the product or service level.

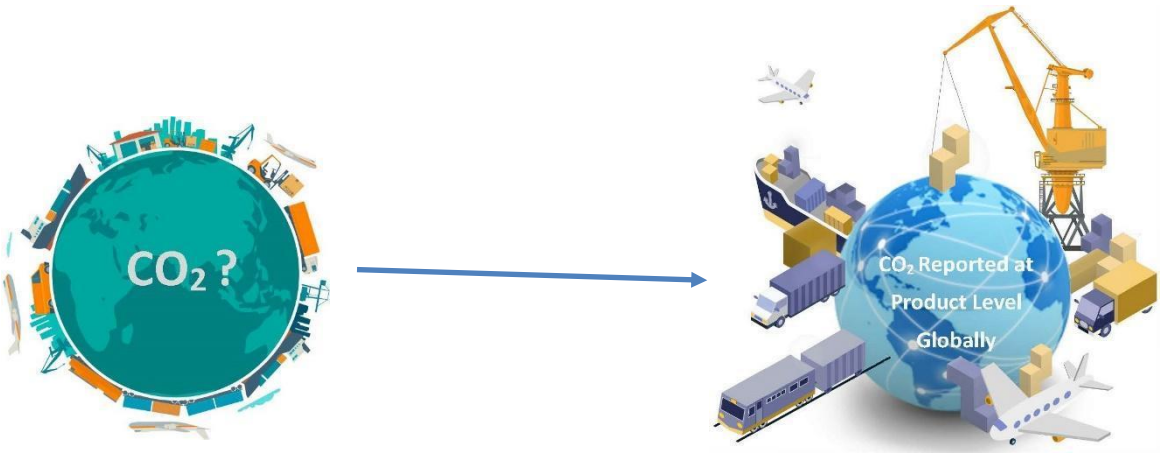
Our vision is for embodied carbon data to flow freely between all product or service related system interactions across supply chains, while maintaining useful contexts. Everywhere there is a price tag, there should be an embodied carbon (CO2e[1]) tag equivalent.

This can be created by:

- developing an open-source global ecosystem of actors:
- an extensible schema leveraging existing product and service taxonomies,
- enabling trusted and visible declarations of embodied carbon
- across all products or services at all points and actors across supply chains.

The target industries, products, services, and users are everyone and anyone – all industries, products and services; and related inventors, designers, developers, manufacturers, wholesale purchasers, sellers, buyers, consumers, monitors, regulators, transporters, researchers, etc. The world.

Our goal is ever standardized contexts for embodied carbon where the embodied product carbon is declared using standardized context that can be understood at any point by any person.



[1] Generally speaking in this document when "carbon is mentioned" it should be considered to be CO2e, the GWP of related gases all normalized to CO2e using the appropriate GHG GWP convention.

Operationalizing a Vision

Tackling the world of all products and services, and every person interacting with a product or service at every point along the supply chain is an enormous undertaking. A grandiose vision that requires a scaled, thoughtful, and phased approach to become a reality.

We know making this a reality means incorporating the work of others and building an open-source evolving solution. We work with others to create consortia of like-minded actors each with roles, use cases and responsibilities to provide input on development, usage and acceptance for fit for purpose solutions.

The first issue/open question is: 'is there a systems language to provide context for something that can be easily adopted globally.'

We found our answer by looking at the world of accounting and business reporting.

The Motivation is Faster High Quality Data Flow

The world of accounting and business reporting of economic flows has XBRL and iXBRL, an open international standard for digital business reporting, managed by a global not for profit consortium, XBRL International.

XBRL is an open-source, freely available global framework of accounting standards used for the exchange of business information. XBRL was developed in 1998 with the latest version introduced in 2003. It continues to evolve as additional XBRL modules, tools and language extensions are developed to be plugged delivering new functionality and use cases. XBRL is open-source, scalable, extensible, adaptable for countries and/or regions, continues to evolve and has widespread global adoption in private and public sectors.

Part of its functionality is in enabling the flow of automated reporting, data sharing, regulatory filings, etc. and reduction of bad data, and radically shrinking the need for conversions and integrations etc. In short, XBRL accelerates economic activity and data flows.

As an example, accounting concepts such as depreciation, goodwill and intangible assets among others are interpreted or presented in a standardized context relative to accounting frameworks at the declared presentation layer. The core data is clearly declared and transparent, but the resultant displayed interpretation may vary based on reporting or other requirements such as GAAP v. IFRS accounting standards for example.

Adaptation for Embodied Carbon

Embodied Carbon (the carbon added with economic value-added processes) currently has no formal language of **accounted** or declared expression for items / tags that would provide shared context such as:

- Why the embodied carbon declaration was made,
- Who made the embodied carbon declaration and who verified it,
- The quality of measurement (uncertainty) associated with data and measurement,
- What is being declared (product or service),
- How much embodied carbon is being declared,
- When was the embodied carbon declaration made,
- Where was the embodied carbon declaration made,
- Are there other embodied carbon factors that should be contextualized.

In addition, current fuzzy marketing concepts such as “green,” “carbon neutral,” etc. without being formally defined are used to denote multiple activities. This causes issues in determining the quantitative benefits or impacts when assessing purchasing decisions for industrial or individual consumption or compliance across a supply chain that more formalized attestations or declarations could provide.

Compounding this are attempts to establish carbon offset programs related to the embodied carbon of a product or service, where today services such as bundling credits/offsets or other market-based tools are only informally referred to as “tonnes” of X.

A formal expressed language and context of embodied carbon processes allows formal attestation to be made and managed in structured but open-source schema, similar to how XBRL evolved. It is hoped that all parties across supply chains can declare in a standardized context the carbon associated with their purchasing, value add and selling activities.

We believe that transparent CO2e signaling which is standardized, machine readable, and extensible can accelerate coordinated choices, actions, and consensus on normative activities such as reporting across supply chains. This data can initially create awareness of embodied carbon and then millions or billions of deeper actions and choices to mitigate and reduce CO2e as efforts to create new higher valued products or services become more certain financially, reputationally, and from regulatory reporting perspectives.

Our Goals and Outcomes

Our goal is to make embodied carbon data sharing & flow easy.

This likely requires:

1. An ever-evolving ecosystem of actors.
2. Systems comprised of extensible schema referencing existing and evolving product or service taxonomies.
3. An ecosystem that is trusted and visible, open-source, adaptable for easy implementation, globally adoptable, and technology agnostic.
4. Creating structured data declarations of measured embodied carbon (CO2e).

This enables the creation of:

1. Structured data declaring the measured CO2e associated.
2. Supporting every product or service at every point along the supply chain – upstream and downstream.
3. Information about the why, who, what, how, when, and where of the CO2e declaration, among other data reference points.

Empowering:

Private and public sector actors that produce, consume, govern, and track declarations of CO2e with maintainable, reasonable and standardized context for any product or service across supply chains.

Ecosystem partners to develop carbon related message types and related solutions such as displaying a product's or service's CO2e on the product's or service's label/description so that companies, wholesalers, consumers, suppliers, governments, etc. can make more informed choices.

Possible outcomes from structured data declarations of CO2e in products and services may include:

- **Influencing major actor purchasing changes** along the supply chains such as corporate purchasing, customer pressure, and government policy.
- **Having economic impacts** with new goods & services to drive growth in the voluntary carbon market.
- **Encouraging new behaviors** through better understanding of CO2e, conversations and trust, and resulting in new choices being made.
- **Empowering Buy Clean initiatives** beyond those drivers of Buy Clean initiators.

Carbon-ML Ecosystem and <CarML>

The Carbon-ML Ecosystem:

Carbon-ML is envisioned as an evolving ecosystem of:

- **Global participants** & observer consortia, (private companies, govt, regulators, NGO etc.)
- **Extensible schemas** of taxonomies maintained & versioned in an open but rigorous way ,
- **Taxonomies of data** & contexts, from existing sectors & verticals
- **Open mix of technology, languages and platforms** for expression and management
- **Governance principles** for organization & growth to maximize data flow & <CarML> use

The ecosystem participants form consensus-based decisions around the use of extensible schemas supported by existing product or service taxonomies, or extended/developed from scratch (only if needed). These decisions inform outcomes enabling standardized, trusted and visible declarations of CO2e in every product or service where <CarML> context is maintained across the supply chain.

A GitHub site for Carbon-ML is used for collaborative development of the Carbon-ML ecosystem, building consensus, with existing taxonomies, and eventually realizing a globally adoptable standard language for CO2e measurement declarations that is technology agnostic.

The primary open-source technical markup language to be used within the Carbon-ML ecosystem is <CarML>. However, the Carbon-ML ecosystem can accept multiple technical languages as long as the resultant standardized CO2e declaration front-end display/label and back-end taxonomy (context) reference is maintained.

What is <CarML>

<CarML> is an open-source extensible markup language for CO2e declarations being defined for the Carbon-ML Ecosystem. Currently, <CarML> can be expressed in XML or JSON, and is adaptable for other technical languages that ecosystem participants may want to use. The realization of the standardized context and <Key, Value> pairs. <CarML> maintains context through the use of <Key, Value> pairs for the why, who, what, how, when and where, attestation and embedded carbon credit information (if needed) of CO2e declarations. It enables the structured machine communications / declarations of CO2e that provides the structure and standard for CO2e declarations at each point along the supply chain. <CarML> is meant to evolve, and is not fully proscriptive of any technology, solution, and/or interpretation.

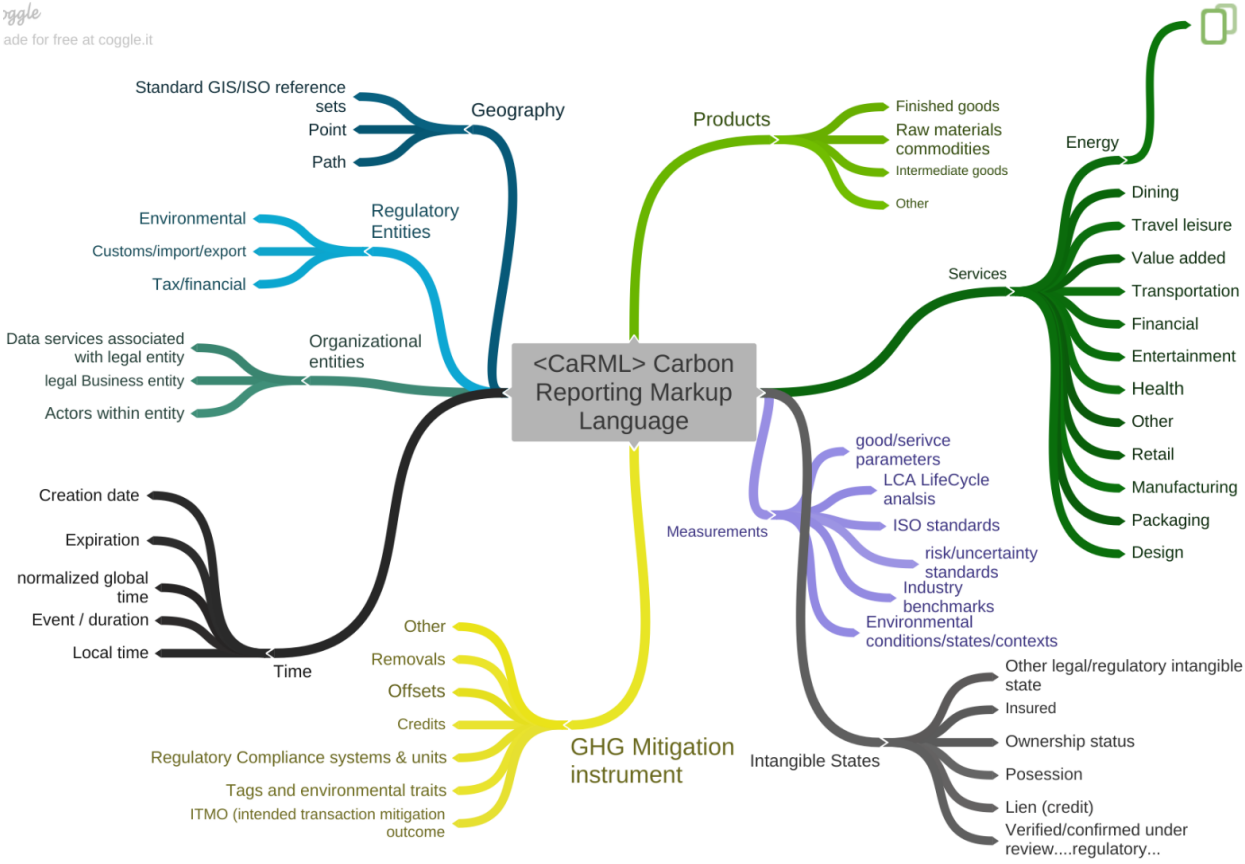
In short:

- <CarML> enables shared context for reporting embodied carbon data objects.
- <CarML> is initially envisioned as open-source XML or JSON, and a standard for how product/service information about CO2e is created, processed, distributed, declared and shared.
- <CarML> creates interactive / intelligent CO2e data
- <CarML> is essentially a Unique ID for CO2e information for products and services

- <CarML> is useful for many product/service handoffs, such as:

	Business	Government	Machine	Consumer	Other
Business	B to B	B to G	B to M	B to C	B to Other
Government	G to B	G to G	G to M	G to C	G to Other
Machine	M to B	M to G	M to M	M to C	M to Other
Consumer	C to B	C to G	C to M	C to C	C to Other
Other	Other to B	Other to G	Other to M	Other to C	Other to Other

Carbon-ML Ecosystem and Schema Representation(example)



Schema (Branch) and Taxonomy (Node) Principles:

Within the Carbon-ML Ecosystem are several **classifications of schemas**, each providing information leading to **a standardized, maintainable context** based CO2e declaration.

Contained within the schemas are **taxonomies representing individual entities / products / services / measurements / etc**. that lead to an eventual unique description at the **branch end node**.

Extensibility and Reference Schema principles:

- Where possible inherit or use existing global standards and regulations
- Clarity over completeness, and defensibility over strict accuracy
- If in doubt leave it to the interface and keep things light
- Utilize other schemas and taxonomies where possible
- Support multiple technical languages and platforms
- Open-source whenever possible, not proprietary
- **Solution (hacker) Oriented**. If a formal <CarML> message type or taxonomy is not supported, extend it then create an informal extension that works and submit a merge request. Languages are a do-ocracy! What works and gets used and lives. All else is pruned.

Schema related Taxonomy = Intelligent Context:

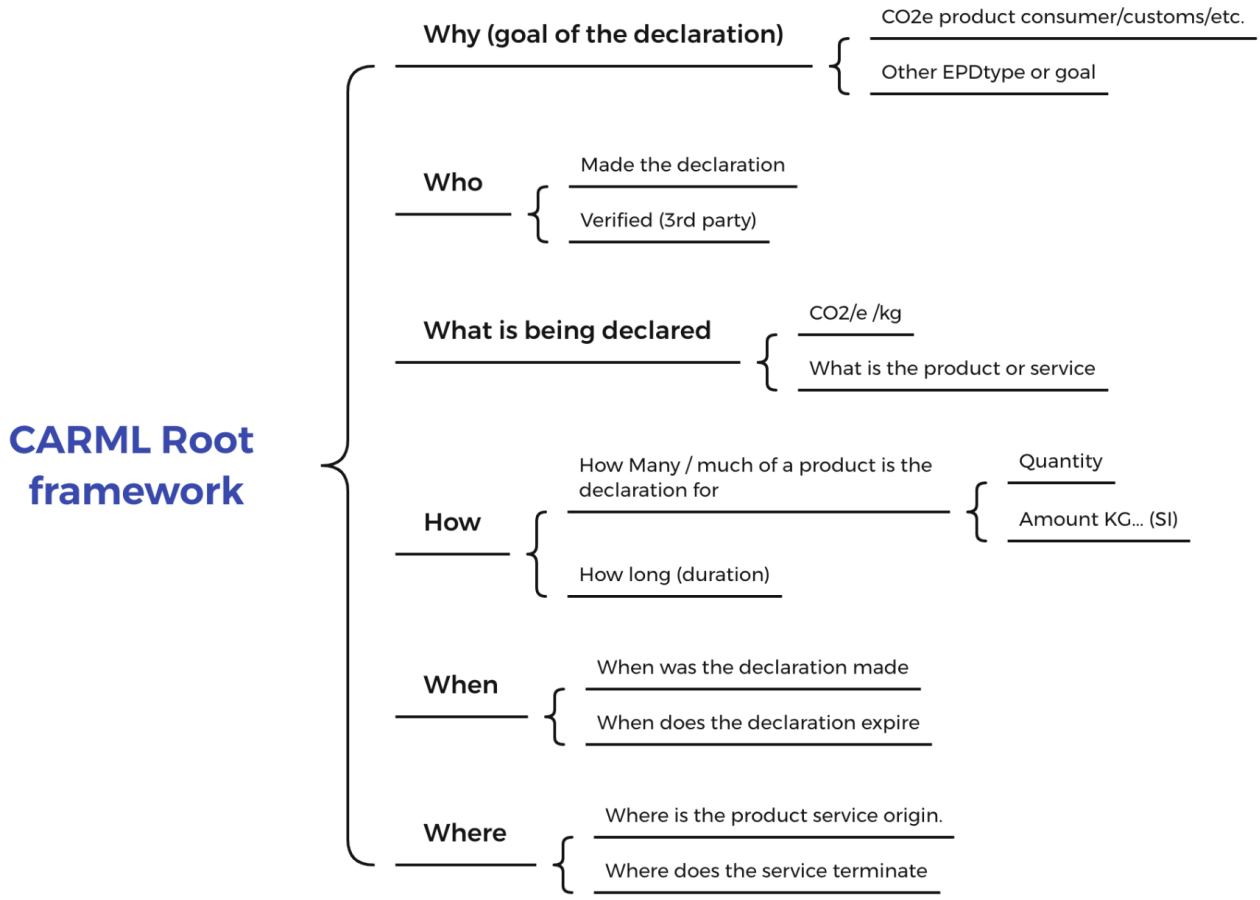
- Repurpose existing industry & govt. schemas, taxonomies, and data contexts
    - Re-use data from other systems leading to faster integrations & acceptance
    - Data should retain value / context and be exchangeable between external systems
    - <CarML> CO2e data objects with context are to be shared and utilized to automate forms submissions and reporting
        - Continual machine learning may enable automatic changes in CO2e flows with contextual inference

Customizable for Regional and Local Variants:

Examples of possible regional / local variants to be reflected in the end branch/schema using <CarML>:

- Context and maintenance can occur at the global or local level providing a static reference, or operating information such as a global boundary or local internal information that may be changed/updated over time
- Internal locations plant (private/public)
- Shift manager/lot number (private/public)
- Purchase Order#’s (public/private)

Carbon-ML Example Schema Branch / Taxonomy Node Framework



- <CaRML> is designed to be an open extensible markup language supporting a collection of extensible schema for structured machine communication and declarations about the carbon CO2e associated with all economic activities at the individual product or service and/or service level.
- The <CaRML> extensible schema is meant to evolve and extend as a framework, not being fully proscriptive of any one solution or interpretation.
- <CaRML> will be transparent with machine readable CO2e signaling in order to accelerate reporting across supply chains, creating awareness of carbon and then enable efforts to mitigate and reduce CO2e to create new higher valued product or services and services, and with <CaRML> become more certain financially, reputationally and from regulatory reporting perspectives.
- There is no <CaRML> schema end solution. <CaRML> is extensible and is designed such that a small part of the tool or tags can be implemented and still be useful.
- Early <CaRML> schemas will be shaped with input from key stakeholders in the ecosystem and build on existing schemas and taxonomies.

What is Meant by a <CarML> Message Type?

The <CarML> message type itself is a structure, a standardization of how embodied carbon within any product or service at any point along the supply chain is reported. The reporting of CO2e is accomplished through <CarML> message types that can be customized based on who is using, receiving, and/or acting-on the CO2e information contained within the message type. The PoC (proof of concept) generic message type discussed following is one such example.

The <CarML> message type is simply a unique descriptor of the CO2e in a format determined by the usage of the message type. We are not inventing a new technology language, but rather incorporating existing technology, languages, structures, taxonomies, schema, etc. to create something newly useful. A new standardization, for structured reporting, a new way of taking all the variations of how CO2e is currently reported and referred to, and developing a universal CO2e standard - a common, globally accepted way of reporting and referring to embodied CO2e – the <CarML> message type.

The <CarML> message types represent the encapsulation of this standard. And, they are being designed and developed to be highly flexible, customizable, scalable, portable, among others.

Examples of <CarML> message types may include: wholesale, manufacturing or retail required reporting formats based on specific industries; customs or regulatory required information; carbon credits RFQs; or just LCI data - the list is endless.

<CarML> Message Type Inputs

A <CarML> Message Type is developed by constructing some or all of the <CarML> elements incorporating the why, who, what, how, when, and where; subject to the requirements of the consumer of the data object pertaining to the declaration.

Complete and partial declarations are allowed using only a single fact to be declared or a whole statement of embodied carbon within the product or service. The compliance or verification of input type, string check for integrity would be a dependency built into the object generating and object consuming endpoints/systems of the <CarML> Message Type.

The overall use would be to ascribe <CarML> compliant attributes to an object/process which can be accessed and verified – basically an input string check of the <CarML> Message Type in order to separate well formed signal from unstructured noise.

This would be structured data, global in context, and drives the point of “how to declare carbon” with the syntax constructed so the <CarML> Message Type may even be mapped for further checks against compliance and completeness for the user.

A <CarML> Message Type:

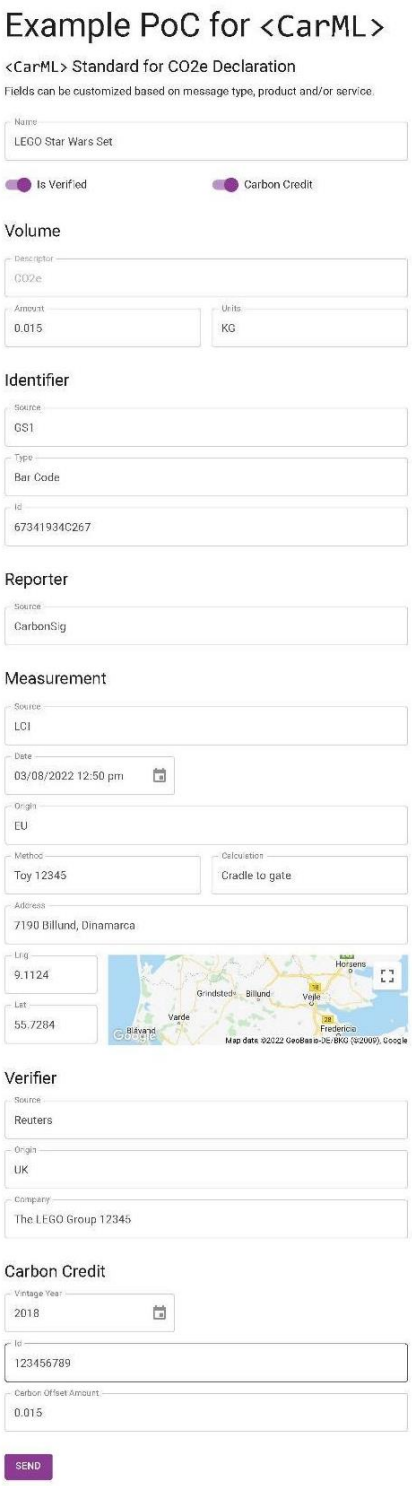
- Can be customized by the user(s) and adapted informally for regional norms/system conventions
- Can use any data declarative convention (current Proof of Concept demonstrates XML and JSON)

The underlying <CarML> Message Type Technology Design from back-end to front-end display:

- Can call on proprietary databases while maintaining their confidentiality and privacy
- Includes a “translation layer” for the use of multiple technology languages
- Employs best practices around privacy and cybersecurity

The Proof of Concept of the <CarML> Message Type can be found at: https://github.com/Carbon-ML-org/form-poc

An example of the <CarML> Message Type <declaration> display is:



Description of the <CarML> Message Type <Declaration> Inputs

Name (general)

The name of the product and/or service the <CarML> Message Type is referencing.

Why

Why the CO2e declaration or statement is made and/or updated.

The Why is a process description regarding why the event, declaration of measurement and/or change in measurement of CO2e happened or is being made. Why was the CO2e for a product and/or service added, altered, updated, or changed among others.

The Why may include non-carbon add events that directly correspond to the CO2e declaration and may have caused a restatement and/or format revision. Examples would include legal state changes/assignments, logical changes/assignments, specifically a duty paid, package certified, audit completed, auditor verified, etc.

These can include changes in the legal or informational state/context of a good or service but involve no physical change or ownership change that would include a transfer of environmental rights and claims in a system of record. These also include re-statement for regimes which may use different CO2e representational conventions.

What (specific)

What the CO2e declaration references. What product and/or service is being referenced and identified.

Example reference and identification schema could include:

- GS1 Barcodes
- ISO codes
- SIC or standard descriptions of service definitions which may have regulatory, legal or industry standard descriptors analogous to SIC codes down to 9 digits
- Life Cycle Assessment (LCA) references
- 3rd party publicly available schema or taxonomy

The <CarML> Message Type includes fields to input the Source identification schema, taxonomy, or other; the type of identification; and the unique ID reference for the product and/or service.

Amount: The amount of CO2e expressed in kg associated with the product and/or service. The data fields may also include whether this amount is an absolute measurement, or was reference data or a combination of primary and reference data used.

Who (functional role)

Basically, Who made the CO2e declaration. Elements can include who owned, measured, declared, referenced, verified etc. the CO2e in a product and/or service at this point along the supply change. The

Who would comprise any entity within the Carbon-ML ecosystem with a defined role. An entity can be a company, individual, governance organization, reference data set, etc. Examples of Carbon-ML ecosystem entities and related roles for context include among others:

Who would declare, measure, reference CO2e:

- Corporations
- Commodity producers
- Service providers
- Governments
- Regulators
- NGO's and IGO's
- Researchers
- Verifiers such as auditors
- Consumers

Ecosystem Roles:

- Owners of a product or service
- Consumer of a product or service
- Procurement agents
- Taxonomy or Schema maintainer/provider
- Carbon registrar or issuer
- Verifier or reviewer
- Product innovators
- Individual viewer or observer
- Technical developer
- Regulatory system
- Procurement system

How

How was this fact about CO2e assessed, derived, etc. What was the measurement system and specific methodology within that system that was used. What was the resultant CO2e amount. How were the boundary conditions of the LCI determined. How was the quality of data determined (primary, reference etc.)

When

When did the CO2e measurement occur, when was it declared. The When involves a point in time event such as an ISO standard time convention likely linked to UTC. Time can be expressed as relative elapsed time such as relative to the location (where) something occurred or as an absolute reference to an event.

Examples of When (Time):

- Process, start, end, completion
- Product or service event
- Time of system and data entry update
- Product or service or process expiration
- Service initiation

- Service completion
- Time start of process / time end of process (for travel planes etc. mapped to UTC)

Where

Where did the activity, the CO2e declaration event, occur. The Where can be a point or service route, a point along the supply chain, etc. The Where can be identified through the use of GIS/ISO standards for maps and geo locations or polygons.

Carbon Credits/instruments associated with declarations

Carbon Credits/instruments utilized as instruments supporting claims of embodied carbon reductions.

Carbon Credits directly assigned to the product and/or service and used as direct offsets can also be included on the form. In addition to the carbon offset amount, descriptors and identifiers related to the carbon credit such as type, compliance/voluntary, quality, vintage, IREC registry, state such as retired or pending, future, and owner can also be included.

Other instruments or claims supporting embodied carbon data declarations may be regulatory or private sector attestations such as GO's, REC's, SRECs, IRECs, C-capsule etc. Quality declarations etc.

<CarML> Message Type Technical Design and Implementation

(Notes: this section reference the PoC on the GitHub site with examples of message types, RFQ, etc. and explains how multiple and even non-<CarML> languages can be supported)

Building a <CarML> message type

There are three main technical elements that work together to build/create and express each <CarML> Message Type. These are the Back End Databases, the Domain or "translation layer" that houses the business objects/data definition dictionary and supports API calls and endpoints, and then there are Front End UI/UX interfaces which express <CarML> message types for users.

The "translation layer" will also include automated data field checks to ensure the input is expressed in the correct format; however, detailed compliance checks for reasonableness will be conducted when the message type has been completed for review before sending.

The technical implementation is designed to be adaptable for multiple languages and configurations, scalable, flexible, and customizable among others.

Various scenarios were considered when designing the logical model such as:

- Connectivity to proprietary databases
- Regional and Industry differences in Label/Field names (context) but having the same definition (content)
- Preferences for different technology languages such as JSON and XML
- Differences in Message Type content based on who is using it and/or who is it for
- A UI/UX that can be fully automated or used for manual input of content
- The roles and responsibilities of each industry Carbon-ML consortium in the development of guiding principles and Message Types for their respective industry and usage

Technology Agnostic

Supports multiple languages through a "translation layer"

Maintains privacy of proprietary information

Supports Multiple Schema types

Adaptable to regional and country norms

Message Type Category Examples

The data fields to be included under the main categories (Name, Why, What, Amount, Who, How, When, Where, Verified, Environmental Mitigation Instrument) of each defined category of Message Types is to be determined by the representative Carbon-ML Ecosystem Industry/Service/Product Consortium. The individual Consortiums can define an appropriate number of message type categories and message types within those categories, as relevant to their industry/product/service and the supply chain markets/users they serve.

Once the data fields for each unique message type are determined, defined and approved, they will be merged and included to be formally supported in the <CarML> Domain Layer/Data Dictionary Layer which sits between the frontend UI/UX and backend reference databases. This layer also houses the relevant APIs and endpoints to access the correct data that corresponds to the defined data field.

This Domain Layer acts as a de facto “translation layer” so that multiple variations of the same data element/business object as defined on the frontend will be “translated” to a standardized API call / endpoints on the backend in order that the correct data from the relevant reference database is selected. This serves to reduce redundancy and increases performance and standardization.

Each Message Type Category can be customized based on the Consortium’s requirements and/or individual business or local variations, for example. In addition, other documents can be linked to each message type, as needed, for more detailed informational purposes.

For a particular product or service, multiple message types can be generated simultaneously, as needed, for different purposes. For example, a manufacturing item may require a general message type for supply chain partners, a regulatory message type for customs requirements, and a carbon credit RFQ (request for quote) to make the item net zero. These message types can all be generated at the same time from the <CarML> platform.

Some examples of Message Type Categories and how they may be customized include:

Consumer Goods (example)

Name: The name of the product.

Why: The why refers to why the CO2e declaration is being made, and for consumer goods it may be that the declaration is for informational purposes only.

What: The unique ID associated with the product. This may be several fields, basically a hierarchy structure of Category, Name, Unique ID such as a GS1 Barcode.

Amount: The amount of CO2e expressed in kg associated with the product. The data fields may also include whether this amount is an absolute measurement, or was reference data or a combination of primary and reference data used.

Who: Who made the CO2e declaration. This may be a CarbonSig report, an outside service provider, a defined internal department, etc.

How: How is the measurement/formula used for the determination of the CO2e. This may be an LCA analysis, internal measurement, from an outside service provider, etc. The fields should be representative of the methodology and there should be an appropriate number of data fields to impart an understanding of the calculation. Fields may include the Methodology name, types of inputs, type of calculation, etc.

When: When was the CO2e measurement conducted. This is a general date/time field for when the CO2e measurement was made.

Where: Where was the CO2e measurement conducted. The where may also be the location of the product and a general geolocator input such as longitude and latitude can be used.

Verified: Some CO2e declarations may require verification by third parties. If so, then these fields would refer to who verified the CO2e declaration and some type of identification fields for the verifier, such as a business name, license/unique ID, location, etc.

Environmental Mitigation Instruments: For a product to be labeled as net zero, an environmental mitigation instrument such as carbon credits may be purchased. If so, these fields would pertain to the carbon credits purchased, unique IDs, and some characteristics about the carbon credits such as vintage year, project, location, etc.

Manufacturing (example)

Name: The name of the item.

Why: The why refers to why the CO2e declaration is being made, and for manufacturing there may be a requirement from purchasers, industry requirements, regulators, etc. The why may consist of several standardized fields to select the reason why the CO2e declaration is being made.

What: The unique ID associated with the item. This may be several fields, basically a hierarchy structure of Category, Name, Unique ID such an industry standard classification and/or internal ID system.

Amount: The amount of CO2e expressed in kg associated with the item. The data fields may also include whether this amount is an absolute measurement, or was reference data or a combination of primary and reference data used.

Who: Who made the CO2e declaration. This may be a CarbonSig report, an outside service provider, a defined internal department, etc.

How: How is the measurement/formula used for the determination of the CO2e. This may be an LCA analysis, internal measurement, from an outside service provider, etc. The fields should be representative of the methodology and there should be an appropriate number of data fields to impart an understanding of the calculation. Fields may include the Methodology name, types of inputs, type of calculation, etc. There also may be regulatory requirements in terms of the information required regarding the CO2e measurement and how it was conducted.

When: When was the CO2e measurement conducted. This is a general date/time field for when the CO2e measurement was made.

Where: Where was the CO2e measurement conducted. The where may also be the location of the item and a general geolocator input such as longitude and latitude can be used.

Verified: Some CO2e declarations may require verification by third parties. If so, then these fields would refer to who verified the CO2e declaration and some type of identification fields for the verifier, such as a business name, license/unique ID, location, etc.

Environmental Mitigation Instruments: For an item to be labeled as net zero, an environmental mitigation instrument such as carbon credits may be purchased. If so, these fields would pertain to the carbon credits purchased, unique IDs, and some characteristics about the carbon credits such as vintage year, project, location, etc.

Regulatory Requirements (example)

Name: The name of the product/item/service.

Why: The why refers to why the CO2e declaration is being made, and for the Regulatory message type, the why would be for regulatory purposes and there should be a field for the agency who issued the requirement and the regulation ID/code.

What: The unique ID associated with the product/item/service. This may be several fields, basically a hierarchy structure of Category, Name, Unique ID such as industry standard classification, GS1 barcode, internal ID system, etc.

Amount: The amount of CO2e expressed in kg associated with the product/item/service. The data fields may also include whether this amount is an absolute measurement, or was reference data or a combination of primary and reference data used.

Who: Who made the CO2e declaration. This may be a CarbonSig report, an outside service provider, a defined internal department, etc. However, these fields may also need to correspond to recognized providers by the regulatory authorities.

How: How is the measurement/formula used for the determination of the CO2e. This may be an LCA analysis, internal measurement, from an outside service provider, etc. The fields should be representative of the methodology and there should be an appropriate number of data fields to impart an understanding of the calculation, and the required fields may be pre-defined by the regulatory requirements and associated agencies.

When: When was the CO2e measurement conducted. This is a general date/time field for when the CO2e measurement was made.

Where: Where was the CO2e measurement conducted. The where may also be the location of the product and a general geolocator input such as longitude and latitude can be used.

Verified: Some CO2e declarations may require verification by third parties and this may be a regulatory requirement. If so, then these fields would refer to who verified the CO2e declaration and some type of identification fields for the verifier, such as a business name, license/unique ID, location, etc.

Environmental Mitigation Instruments: For a product/item/service to be labeled as net zero, an environmental mitigation instrument such as carbon credits may be purchased. If so, these fields would pertain to the carbon credits purchased, unique IDs, and some characteristics about the carbon credits such as vintage year, project, location, etc.

Environmental Mitigation Instruments RFQ (such as Carbon Credit Request for Quote) (Example)

A RFQ may be a more streamlined message type with several fields not necessarily needed, but may require the inclusion of others.

Name: The name of the product/item/service.

Why: The why refers to why the CO2e declaration is being made, may not be needed.

What: The unique ID associated with the product/item/service. This may be several fields, basically a hierarchy structure of Category, Name, Unique ID such as industry standard classification, GS1 barcode, internal ID system, etc.

Amount: The amount of CO2e expressed in kg associated with the product/item/service. The data fields may also include whether this amount is an absolute measurement, or was reference data or a combination of primary and reference data used.

Who: Who made the CO2e declaration. This field may not be needed.

How: How is the measurement/formula used for the determination of the CO2e. This field may not be needed.

When: When was the CO2e measurement conducted. This field may not be needed

Where: Where was the CO2e measurement conducted. The where may also be the location of the product and a general geolocator input such as longitude and latitude can be used. This may be useful.

Verified: Some CO2e declarations may require verification by third parties. This field may not be needed.

Environmental Mitigation Instruments: For a product to be labeled as net zero, an environmental mitigation instrument such as carbon credits may be purchased. There may be several fields needed regarding the type of environmental mitigation instrument the product/item/service owner may want to purchase. Fields required may be type of instrument such as carbon credits, vintage year preference, verification preference, amount of purchase, etc. These fields may also be determined by the Carbon Credit broker/dealers as to what is required to supply a quote.

The RFQ can be electronically sent to several broker/dealers to obtain quotes and other details. However, Carbon-ML can only provide an RFQ message type, for the transaction to be concluded, this would occur directly between the two parties and outside of the Carbon-ML Ecosystem.

Financial Markets Environmental Credit Risk (example)

This message type would correspond to the information required as part of an assessment of the environmental risk a company/service organization may be vulnerable to and may also be part of the overall credit risk assessment of a company/service organization.

Name: The name of the product/service and company.

Why: The why refers to why the CO2e declaration is being made, and for this message type it may be needed as part of an overall credit risk assessment for the company/business line where more detailed inputs for individual products and services would be useful. There may also be other required fields regarding any regulatory and/or governmental requirements regarding CO2e or other environmental factors for the product/service/industry etc. that would need to be included.

What: The unique ID associated with the product/item/service. This may be several fields, basically a hierarchy structure of Category, Name, Unique ID such as industry standard classification, GS1 barcode, internal ID system, etc.

Amount: The amount of CO2e expressed in kg associated with the product/service. The data fields may also include whether this amount is an absolute measurement, or was reference data or a combination of primary and reference data used.

Who: Who made the CO2e declaration. This may be a CarbonSig report, an outside service provider, a defined internal department, etc.

How: How is the measurement/formula used for the determination of the CO2e. This may be an LCA analysis, internal measurement, from an outside service provider, etc. The fields should be representative of the methodology and there should be an appropriate number of data fields to impart an understanding of the calculation. For a credit risk assessment, there may be required fields and/or a more detailed understanding of the CO2e calculation that would need to be included.

When: When was the CO2e measurement conducted. This is a general date/time field for when the CO2e measurement was made.

Where: Where was the CO2e measurement conducted. The where may also be the location of the product and a general geolocator input such as longitude and latitude can be used.

Supply Chain and Usage Considerations: For a credit risk assessment, there may be some requirements regarding the use of the product/service and/or supply chain tracking. Data fields such as supplier and consumer, locations, etc. may be helpful. Any additional data fields can be added to message types as determined by the related Consortium and users of the message types.

Verified: Some CO2e declarations may require verification by third parties. If so, then these fields would refer to who verified the CO2e declaration and some type of identification fields for the verifier, such as a business name, license/unique ID, location, etc.

Environmental Mitigation Instruments: For a product/service to be labeled as net zero, an environmental mitigation instrument such as carbon credits may be purchased. If so, these fields would

pertain to the carbon credits purchased, unique IDs, and some characteristics about the carbon credits such as vintage year, project, location, etc.

Transportation Services (example)

This message type may have several purposes such as an input to a product or another services CO2e declaration, a measurement for passenger transportation services, etc.

Name: The name of the transportation service.

Why: The why refers to why the CO2e declaration is being made, and for transportation services it may be a requirement of the companies using those services, regulatory requirements, informational purposes, etc.

What: The unique ID associated with the service. This may be several fields, basically a hierarchy structure of Category, Name, Unique ID such as industry standard classification and/or some type of internal unique code. This can also be standardized by the Consortium.

Amount: The amount of CO2e expressed in kg associated with the service. The data fields may also include whether this amount is an absolute measurement, or was reference data or a combination of primary and reference data used.

Who: Who made the CO2e declaration. This may be a CarbonSig report, an outside service provider, a defined internal department, etc.

How: How is the measurement/formula used for the determination of the CO2e. This measurement and how it is calculated may be unique for each transport service type. The data fields, inputs, calculations, etc. can all be customized based on transport type as needed. For example, the data fields and measurement calculation for aviation, ocean shipping, railroads, trucks, etc. may be very different. These fields for “how” the measurement was calculated can be standardized as several broad categories and then a drop-down can be accessed or the message type for each transport type can be customized. Regardless, the data fields should be representative of the methodology and there should be an appropriate number of data fields to impart an understanding of the calculation. Fields may include the Methodology name, types of inputs, type of calculation, etc.

When: When was the CO2e measurement conducted. This is a general date/time field for when the CO2e measurement was made.

Where: Where was the CO2e measurement conducted. The where may also be the initial location of the service and a general geolocator input such as longitude and latitude can be used.

Verified: Some CO2e declarations may require verification by third parties. If so, then these fields would refer to who verified the CO2e declaration and some type of identification fields for the verifier, such as a business name, license/unique ID, location, etc.

Environmental Mitigation Instruments: For a service to be labeled as net zero, an environmental mitigation instrument such as carbon credits may be purchased. If so, these fields would pertain to the

carbon credits purchased, unique IDs, and some characteristics about the carbon credits such as vintage year, project, location, etc.

Use Case Examples

GS1 Labeling

- Carbon-ML supporting GS1 would allow for the “labeling” of approximately 100 million consumer products or services out of the box and be integrated with many inventory/supply chain systems.
    - Processing managers can compare based on Carbon quality
    - Consumers can make more informed decisions

International Trade / Customs and Border Agents

- Carbon-ML supporting international trade/customs and border agents, policies such as the EU Carbon Border Adjustment Mechanism (CBAM)
    - More accurate and standardized assessment of reporting of embodied carbon in product or services in line with customs carbon border policies

Government and State Regulators

- Carbon-ML supporting Government and State regulators' understanding of carbon related data, as standardized data allow for better comparability and tracking of embodied carbon within products or services and services.
    - Better assessment of procurement processes and service provider selections
    - Better assessment of legislation, regulations, and enforcement

Financial Markets / Investment Decisions

- Carbon-ML supporting financial markets investment decision making by providing more accurate tracing and tracking, and comparable representations of embodied carbon within product or services and services by companies.

Materials / Supplies Purchasing Decisions

RFQ (credits)

RFP

Proposed Ecosystem Participants

The proposed Carbon-ML Ecosystem includes participants from industry verticals and from organizations that provide leading practices and standards across multiple industries and services such as regulators, technology service providers, NGOs etc.

Each individual industry and/or sub-industry has its own Carbon-ML Consortium, as it is recognized that there are differences in the CO2e reporting requirements by industry.

For organizations that provide leading practices and standards across multiple industries and services, they will be asked to be active participants in multiple Carbon-ML Consortia and provide their leadership and expertise in these areas.

Global Organizations Providing Expertise and Standards for Multiple Industries

- Create use cases and best practices
- Provide leadership and guiding principles
- Contribute taxonomies, schema, datasets
- Design of <CarML> message types

Industry Associations and Actors

- Create <CarML> message types/share/consume
- Create use cases/best practice examples

Industry Oversight Organizations

- Contribute taxonomies, extension suggestions
- Message types for industry needs
- Technical tools for message checking

Existing Embodied Carbon Working Groups, Academics and Others

- Best practices/principles for environmental integrity and declarations
- LCI declaration standards
- Message types associated with LCI/LCA work

Regulatory and Governance Oversight

- Provide best practices
- Taxonomic structures related to governance
- Message types for compliance needs
- Technical tools for message compliance checking

Technology Experts/Organizations

- Technical guidance advisory for best practices
- Develop 3rd party services & tools for creating, compliance checking, passing, parsing and expressing <CarML> data and message types

Technical Design Architects & Ecosystem Thinkers

- Meta level thinking about deep design principles for scale associated with:
    - Governance (open source / fit for purpose)
    - Schema & taxonomy
    - Tools & implementations

Initial Schema (Taxonomy) Categories and Industries

Categories (from Sphera Product or service Sustainability (GaBi) databases):

- Oil & Gas
- Retail
- Automotive
- Building & Construction
- Chemicals & Petroleum
- Consumer Goods & Packaging
- Education & Academic Research
- Electronics
- Energy & Utilities
- Finance
- Food & Agriculture
- Healthcare & Life Sciences
- Hydrogen
- Industrial Product or services
- Information Technology
- Metals & Mining
- Transportation

Consumers use cases for <CaRML> message types

- Purchasing/procurement managers looking to ingest and compare product or services on Carbon quality for buy clean biases or compliance issues to reduce scope 1 & 2 emissions
- CSR (Corporate Social Responsibility) managers/reporting and systems in companies looking to simplify data gathering
- Product or service innovators looking to create declared or lower carbon product or services
- Cities, industrial processes looking to track goods and services consumed easily
- State, National regulators wishing to ingest Carbon related data quickly and easily around product or services and services
- Customs and border agents looking to assess carbon border policies
- Infrastructure managers seeking to inventory initial CO2e and ongoing management/maintenance CO2e
- Researchers/NGOs performing analysis
- Procurement agents (RFPs (Request for Proposal))
- Consumers and consumer-facing web front ends tracking and consuming narratives about carbon for specific audiences/languages
- Govt's: National, Regional & local regulatory, customs and border agencies looking to track provenance, NDC's etc. of embodied carbon in goods and services

Producers use cases for publishing <CaRML> message types

- Manufacturers & service producers
- Govt. Regulators / agencies
- NGOs
- Verifiers, accounting firms etc.
- Commodity producers
- Customs agents
- Researchers

Governance Framework

Governor role/ committee	Technical Advisory	Industry representatives	Government regulatory	Software industry	Environmental groups
Function Responsibility	Technical function ease of use/implementation	Representation and correct taxonomy schema	representation for usability and extensions	Ease of deployment & utility	Integrity of statements, use cases
Governance responsibility /lever	Branching/tooling/expressions	Taxonomy selection, versioning etc	Schema and representation choices fit for purpose / advocacy	usability / advocacy	compliance checks for declarative integrity for object templates/representations
Observers	monitor developments, post comments during open consultations	monitor developments, post comments during open consultations	monitor developments, post comments during open consultations	monitor developments, post comments during open consultations	monitor developments, post comments during open consultations
Committee members	vote on changes to schema, taxonomic updates, versions, deprecations, message types	vote on changes to schema, taxonomic updates, versions, deprecations, message types	vote on changes to schema, taxonomic updates, versions, deprecations, message types	vote on changes to schema, taxonomic updates, versions, deprecations, message types	vote on changes to schema, taxonomic updates, versions, deprecations, message types
Board level					

Ongoing development of Principles and Guidelines

(This section is about development principles. Governance to be built out and include a brief paragraph explaining at the current moment Carbon-ML centralized during incubation and actively looking for founder participants to join consortia providing compliance, governance, and oversight.)

Usability (ease of deployment/consumptions) has priority over everything. If a design choice must be made between robustness, completeness etc. versus usability, usability should win. There is no <CaRML> schema endpoint solution. <CaRML> is extensible and so is design such that a small part of the tool or tags may be implemented.

Framework for supporting multiple schemas:

Early schemas to include shaping and input from the key stakeholders from Carbon instruments, industry verticals, government compliance and reporting, NGO stakeholders.

Appendix A: Carbon-ML Ecosystem - Definition of Terms

Carbon-ML Ecosystem

Extensible Schema

Taxonomies

Context

Roles

Types of Schemas / Taxonomies

Appendix B: Guiding Principles

Carbon-ML combines measuring, tracking and tracing embodied carbon emissions for any product or service with the development of an open-source global ecosystem comprised of extensible schemas from existing taxonomies. We incorporate principles from climate, product or service, sustainability, and technology taxonomies.

1. **Actionable:** In general, understanding carbon through a common language is an integral step towards making informed decisions. Through clearer visibility of embodied carbon in individual products or services along supply chains, a more sustainable environment and economy may be supported. The Carbon-ML Ecosystem should provide this focus. (re-work)
2. **Environmental Integrity:** The Carbon-ML ecosystem of <CaRML> language, actors, systems, and declaration should support the highest level of environmental integrity, transparency and acknowledge choices/tradeoffs when these are made such as carbon v. biodiversity under DNH (do no harm) principles.
3. **Goal focused:** The Carbon-ML Ecosystem should have **clearly defined goals, scope, and objectives; be transparent, scalable, robust, and adaptable; be developed with open-source code that is technology neutral, freely available, that maintains a clarity that allows for ease of third-party verification. (Multiple principles)**
4. **Beneficial:** The Carbon-ML Ecosystem should provide a benefit and value to stakeholders, participants, industry, users, end-product or service consumers, and others in furthering the understanding of embodied carbon in a standardized, practical, easy to understand open-source format incorporating the characteristic of technical simplicity.
5. **Open integration:** The Carbon-ML Ecosystem should integrate work already completed or under development, thereby incorporating existing global standards, metrics and methodologies, and product or service and industry taxonomies; and provide flexibility of structure for future integrations and evolution.
6. **Collaborative:** The Carbon-ML Ecosystem evolve through active collaboration,
7. **Iterative & Ongoing:** progressing iteratively with the understanding that adoptability and usability is prioritized over strict perfection, through both a holistic and detailed lens, and effective optimization and automation.
8. **Global access:** The Carbon-ML Ecosystem should incorporate global partners encompassing each of the Ecosystem categories and general oversight including technical experts, setting defined objectives and responsibilities, tracking measurable actions and contributions, and adding to further development of the Ecosystem, all within the stated scope and governance protocols.
9. **Objective metrics:** The Carbon-ML Ecosystem should be objective with standardized measures and metrics, a measurement and performance indicator reporting system aligned with defined goals and objectives, and a monitoring and management process which includes input from appropriate stakeholders.
10. **Open Governance:** The Carbon-ML Ecosystem should have open and robust governance using consistent policies, processes and definitions; that incorporates global policies and regulations, and provides for metrics, monitoring, and revisions.
11. **Locally relevant & adaptable:** Where possible and not undermining environmental integrity, The Carbon-ML Ecosystem should be adaptable to local, regional, and country based norms and specificities. The declaration of CO2e at each branching point, for each schema and related taxonomy tree, should be adaptable and adoptable for all regions globally, allowing for regional changes to each resulting branch/stem/leaf.