Proposal to add Environmental Consideration Information to CycloneDX v1.6 Model Cards data

Matt Rutkowski
IBM, STSM Open Source Supply Chain Security

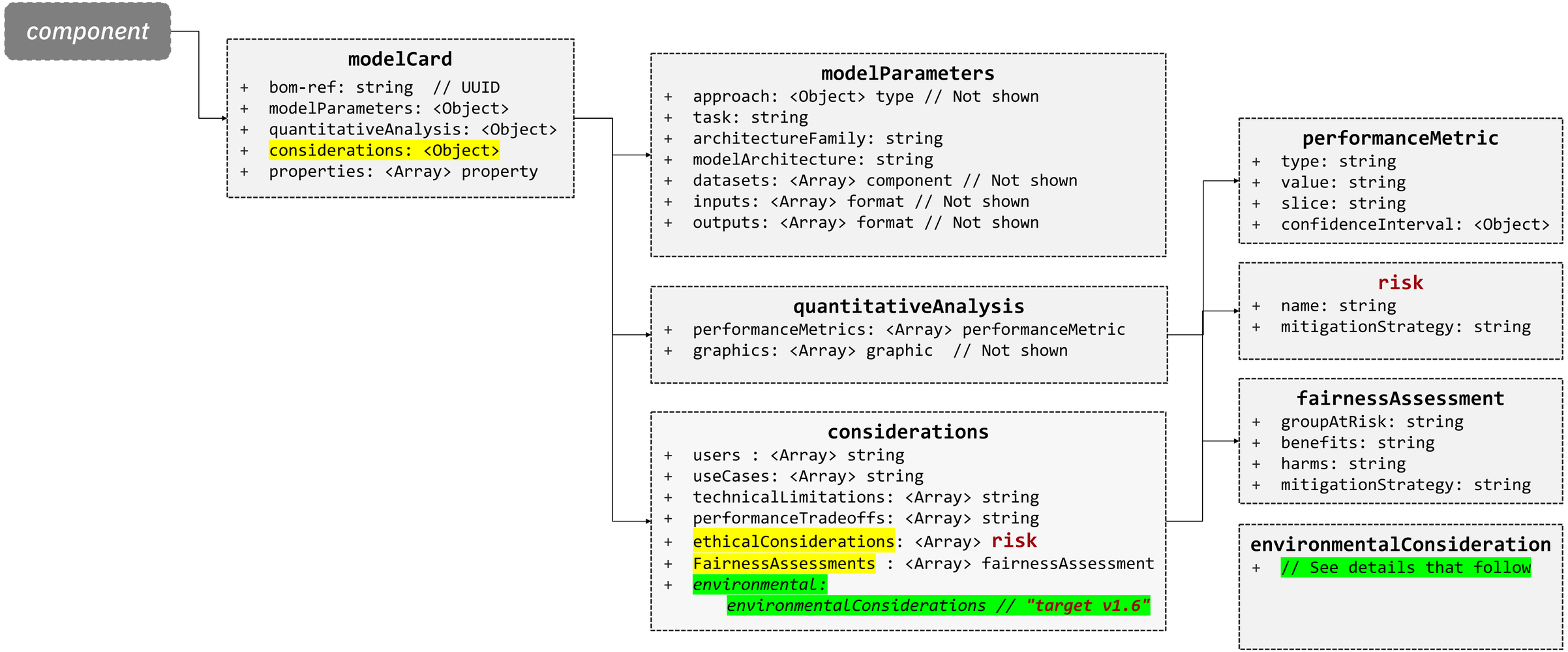
March 7, 2024

IBM

OWASP CycloneDX Machine Learning BOM (ML-BOM)

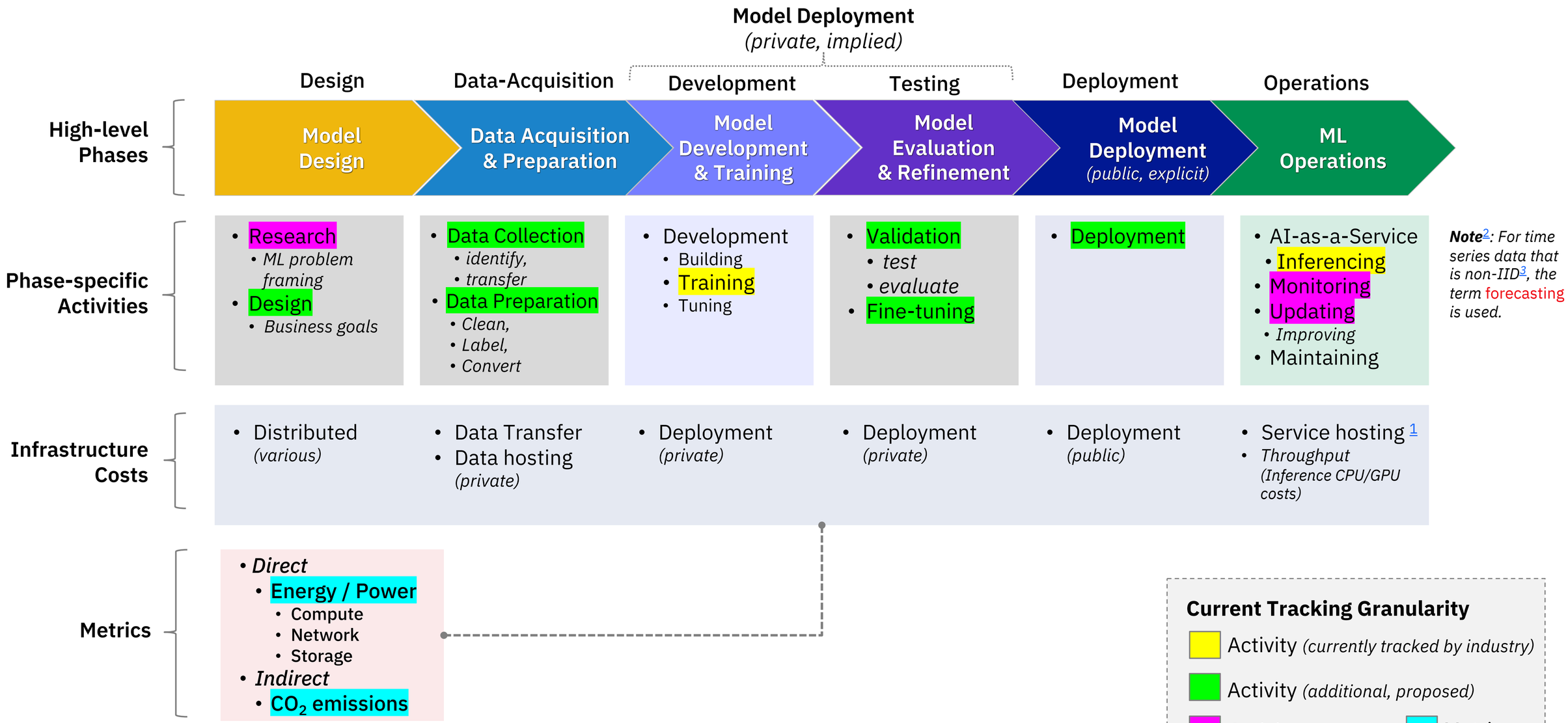
CycloneDX
https://cyclonedx.org/

Model Card Schema overview (not exhaustive)



https://cyclonedx.org/docs/1.5/json/#metadata_component_modelCard

AI CI/CD Lifecycle – Environmental Considerations / Impacts *(costs)*



Datatype: "environmentalConsiderations"

```
"environmentalConsiderations": {
  "type": "object",
  "title": "Environmental Consideration",
  "description": "Describes various environmental impact metrics the corresponding model has exhibited across its lifecycle.",
  "properties": {
    "energyConsumptions": {
      "title": "Energy Consumptions",
      "description": "Describes energy consumption information incurred for specified model lifecycle activities.",
      "type": "array",
      "items": {
        "$ref": "#/definitions/energyConsumption" // See next slide for type properties
      }
    } // FUTURE: Potential to add more consideration "types" each with different metric structures...
  }
},
```

Note: Allows room to add more kinds (types) of environmental considerations apart from energy (and/or its corresponding CO2 footprint) information

Datatype: "environmentalConsideration"

```
"energyConsumption": {
  "title": "Energy Consumption",
  "description": "Describes energy consumption information incurred for the specified model lifecycle activity.",
  "type": "array",
  "items": {
    "type": "object",
    "properties": {
      "activity": {
        "type": "string",
        "title": "Activity",
        "description": "An activity that is part of a machine learning model development or operational lifecycle.",
        "enum": [ "design", "data-collection", "data-preparation", "training", "fine-tuning", "deployment", "inference" ]
      },
      "energySources": {  // Note: multiple energy providers per-activity, For example, multiple cloud data centers.
        "title": "Energy Sources",
        "description": "The provider(s) of the energy consumed by the associated model development lifecycle activity.",
        "type": "array",
        "items": { "$ref": "#/definitions/energyProvider" }
       },
       "energyCost": {
         "title": "Energy Cost",
         "description": "The total energy cost associated with the model lifecycle activity.",
         "$ref": "#/definitions/energyMeasure"
       },
       "CO2CostEquivalent": {
         "title": "CO2 Equivalent Cost",
         "description": "The CO2 cost (debit) equivalent to the total energy cost.",
         "$ref": "#/definitions/co2Measure"
       },
       "CO2CostOffset": {
         "title": "CO2 Cost Offset",
         "description": "The CO2 offset (credit) for the CO2 equivalent cost.",
         "$ref": "#/definitions/co2Measure"
       }
    }
  }
},
```

Datatype for Model Lifecycle: "activity"

```
"activity": {
  "type": "string",
  "title": "Activity",
  "description": "An activity that is part of a machine learning model development or operational lifecycle.",
  "enum": [ "design", "data-collection", "data-preparation", "training", "fine-tuning", "validation", "deployment",
"inference" ],
  "meta:enum": {
    "design": "model design including problem framing, goal definition and algorithm selection.",
    "data-collection": "model data acquisition including search, selection and transfer.",
    "data-preparation": "model data preparation including data cleaning, labeling and conversion.",
    "training": "model building, training and generalized tuning.",
    "fine-tuning": "refining a trained model to produce desired outputs for a given problem space.",
    "validation": "model validation including model output evaluation and testing.",
    "deployment": "explicit model deployment to a target hosting infrastructure.",
    "inference": "generating an output response from a hosted model from a set of inputs."
  }
},
```

Datatypes for Datacenter: "energyProvider"

```
"energyProvider": {
  "type": "object",
  "title": "Energy Provider",
  "description": "Describes the physical provider of energy used for model development or operations.",
  "properties": {
    "name": {
      "type": "string",
      "title": "Name",
      "description": "The name of the energy provider."
    },
    "description": {
      "type": "string",
      "title": "Description",
      "description": "A description of the energy provider."
    },
    "address": {
      "$ref": "#/definitions/postalAddress", // See in later slide
      "title": "Address",
      "description": "The physical address (location) of the energy provider."
    },
    "energySource": {
      "type": "string",
      "enum": [ "coal", "oil", "natural-gas", "propane", "nuclear", "wind", "solar", "geothermal", "hydropower", "other" ],
      "title": "Energy Source",
      "description": "The energy source for the energy provider."
    },
    "energyProvided": {
      "$ref": "#/definitions/CO2Measure"
      "title": "Energy Provided",
      "description": "The energy provided by the energy source for the activity."
    }
  }
},
```

https://www.energy.gov/energy-sources

Datatype: "co2Measure"

```
"co2Measure": {
  "type": "object",
  "title": "CO2 Measure",
  "description": "A measure of carbon dioxide (CO2).",
  "properties": {
    "value": {
      "type": "number",
      "title": "Value",
      "description": "Quantity of carbon dioxide (CO2)."
    },
    "unit": {
      "type": "string",
      "enum": [ "tCO2eq" ],
      "title": "Unit",
      "description": "Unit of carbon dioxide (CO2).",
      "meta:enum": {
        "tCO2eq": "Tonnes (t) of carbon dioxide (CO2) equivalent (eq)."
      }
    }
  }
},
```

Datatype: "energyMeasure"

```
"energyMeasure": {
  "type": "object",
  "title": "Energy Measure",
  "description": "A measure of energy.",
  "properties": {
    "value": {
      "type": "string",
      "title": "Value",
      "description": "Quantity of energy."
    },
    "unit": {
      "type": "string",
      "enum": [ "kWh" ],
      "title": "Unit",
      "description": "Unit of energy.",
      "meta:enum": {
        "kWh": "kilowatt-hour (kWh) is the energy delivered by one kilowatt (kW) of power for one hour (h)."
      }
    }
  }
},
```

Datatypes for Datacenter: "postalAddress"

```
"postalAddress": {
  "type": "object",
  "title": "Postal address",
  "description": "An address used to identify a contactable location.",
  "properties": {
    "country": {
      "type": "string",
      "title": "Country",
      "description": "The country name or the two-letter ISO 3166-1 alpha-2 country code."
    },
    "region": {
      "type": "string",
      "title": "Region",
      "description": "The region or state in the country. For example, Texas."
    },
    "locality": {
      "type": "string",
      "title": "Locality",
      "description": "The locality or city within the country. For example, Austin."
    },
    "postOfficeBoxNumber ": {
      "type": "string",
      "title": "Post Office Box Number",
      "description": "The post office box number. For example, 901."
    },
    "postalCode": {
      "type": "string",
      "title": "Postal Code",
      "description": "The postal code. For example, 78758."
    },
    "streetAddress": {
      "type": "string",
      "title": "Street Address",
      "description": "The street address. For example, 100 Main Street."
    }
  }
}
```

Note: Can add a schema.org "contact point" in the future.
See: https://schema.org/ContactPoint