



Evaluation of a Distributed System for Cloud-Based Data Quality Control Services

Eugene Burger¹, Emily Osborne², Linus Kamb³, Madison Soden⁴, Claudia Schmid²

¹NOAA Pacific Marine Environmental Laboratory, Seattle WA, USA; ²NOAA Atlantic Ocean Marine Laboratory, Miami FL, USA; ³University Washington CICOES/PMEL, Seattle WA, USA; ⁴University of Miami/CIMAS/AOML, Miami, FL, USA

Quality Control



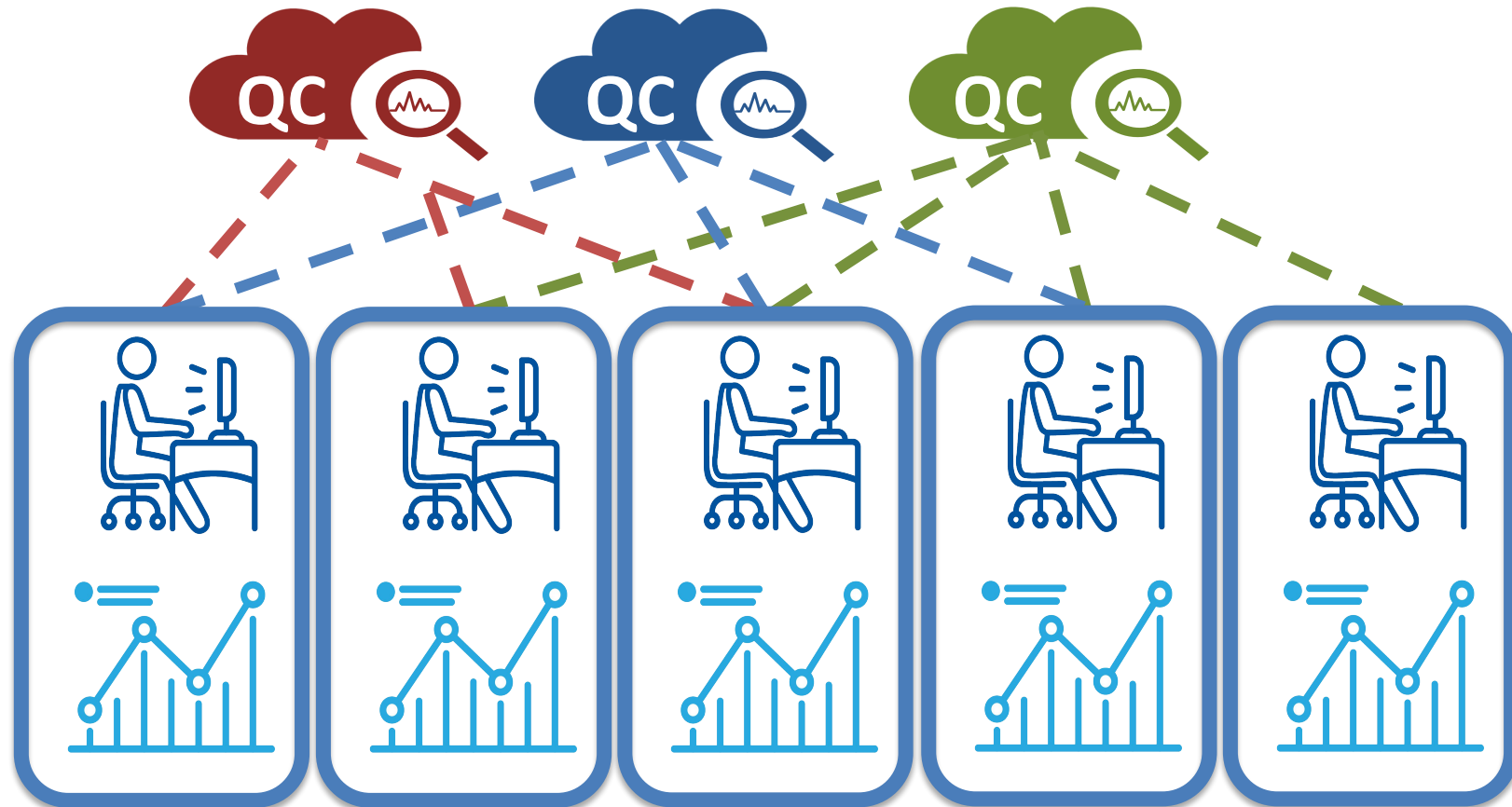
The assessment of the correctness of measurements and observations is an important element in the data lifecycle.

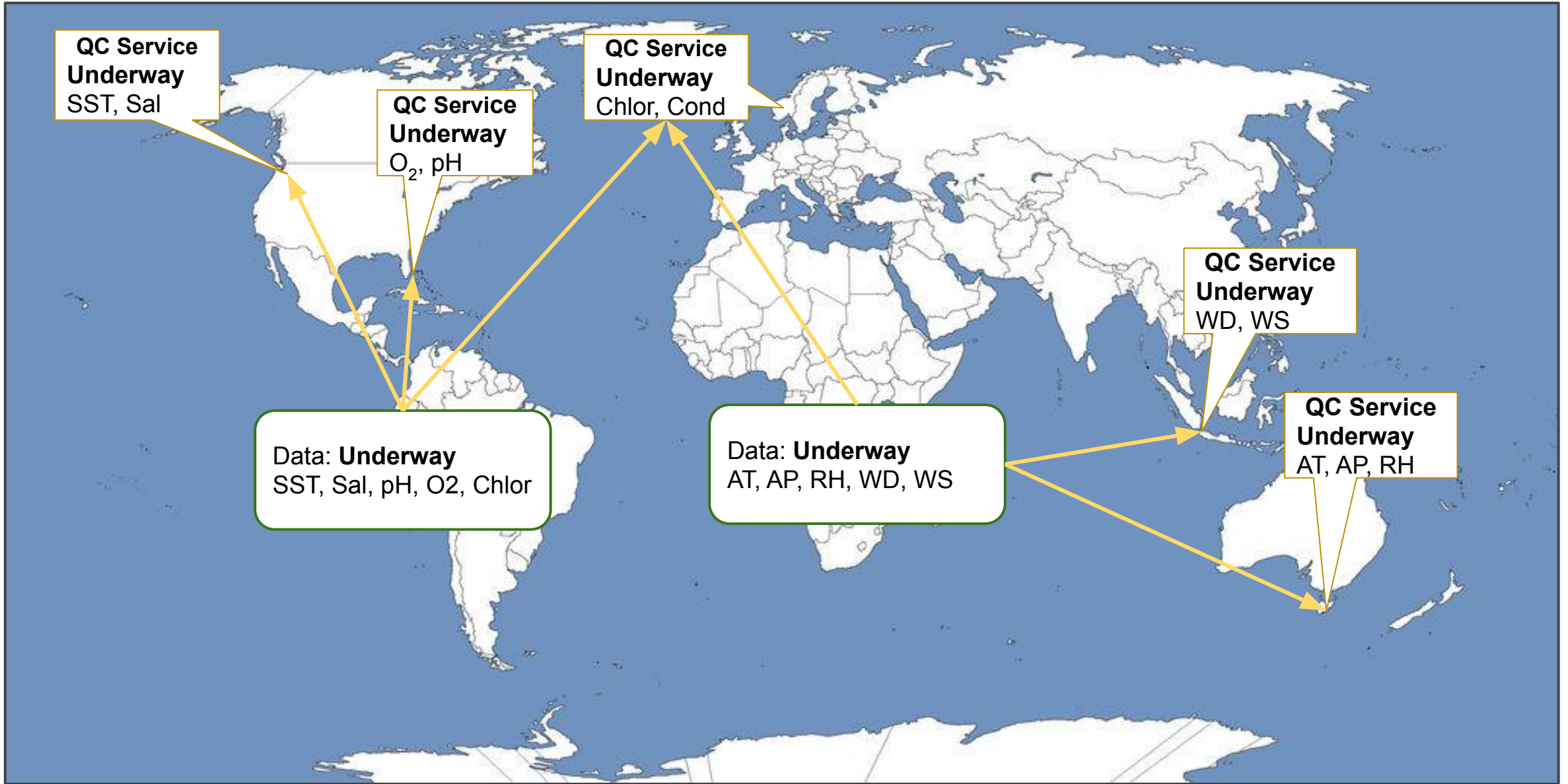
Data quality information supports **data integrity**

Data integrity is key to **scientific integrity**

Project Objective

A federation of *complementary* DQCaaS solutions that, in their totality, will provide level-0 QC solutions to a full range of well established QC routines





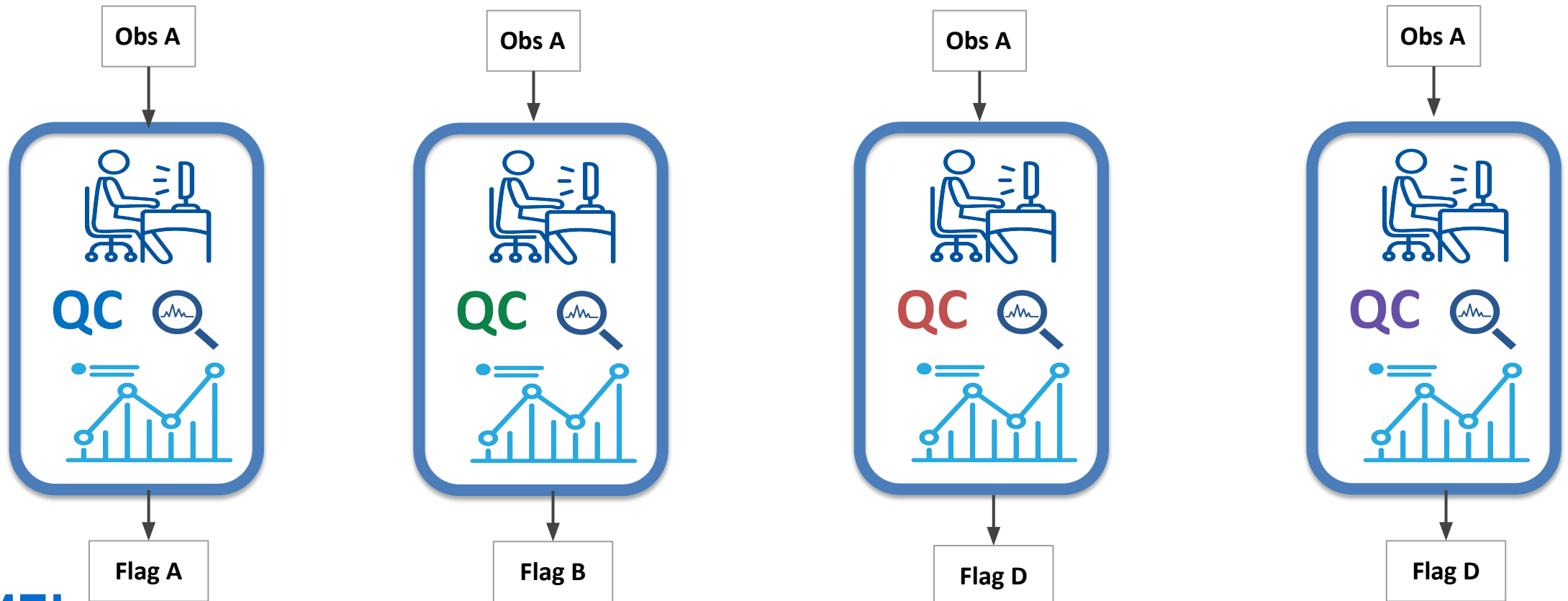
Quality Control Software Duplicated

- Duplicative software maintenance
- Software changes from original
- Required diligent version control over multiple instances

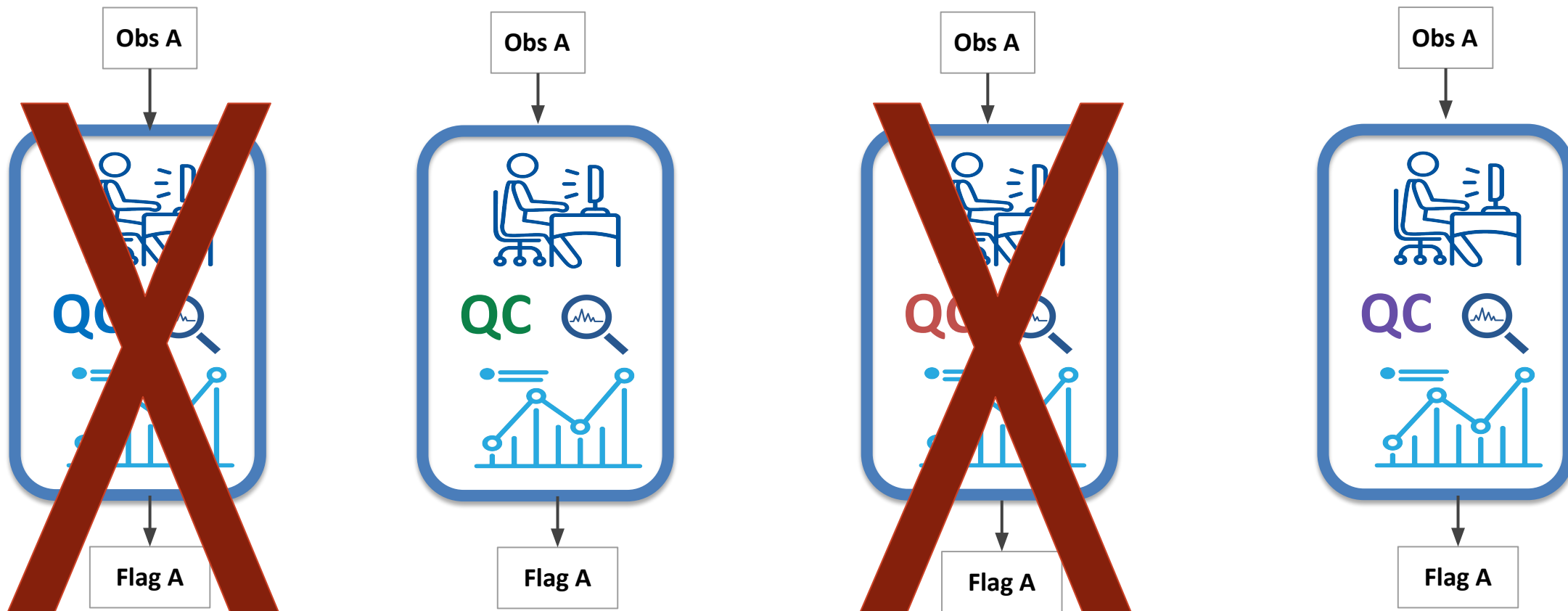


Dissimilar QC Routines

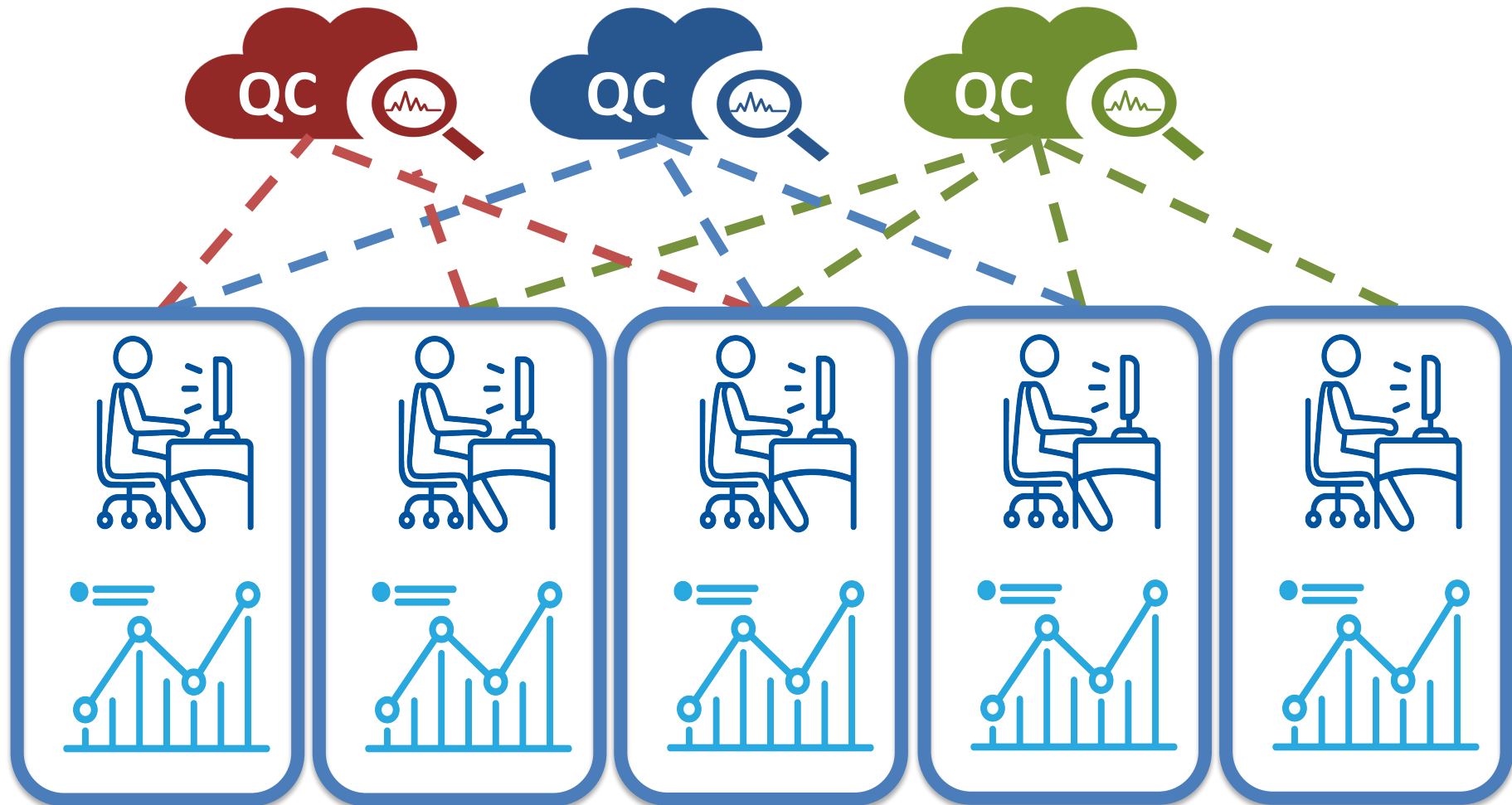
- Different approaches to the same QC objective
- Same objective, potentially different outcomes



Lacking Capacity Infrastructure



Objective



Project

Collaborators

- NOAA/OAR/AOML
- NOAA/OAR/PMEL
- Bjerkness Climate Data Center, Bergen, Norway

Tasks

1. Identify QC routines
2. Develop specification
3. Sample code
4. Test

NOAA's Atlantic Oceanographic
and Meteorological Laboratory

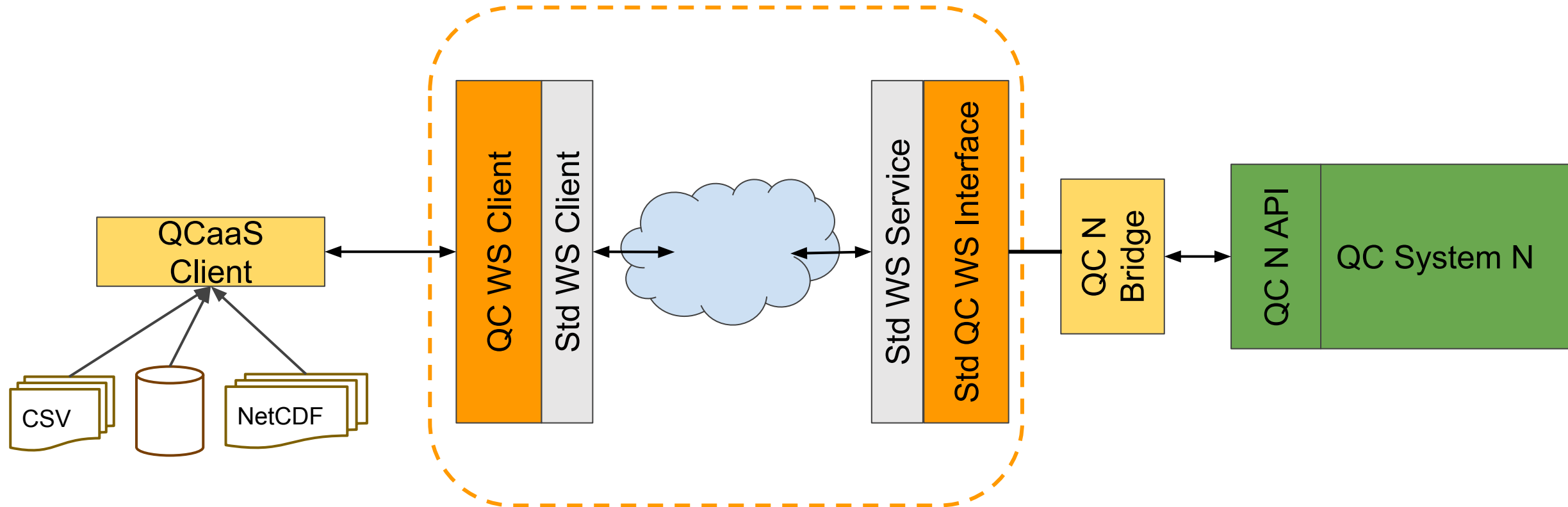
AOML

PMEL

Pacific Marine Environmental Laboratory



Detailed Web-based QC invocation

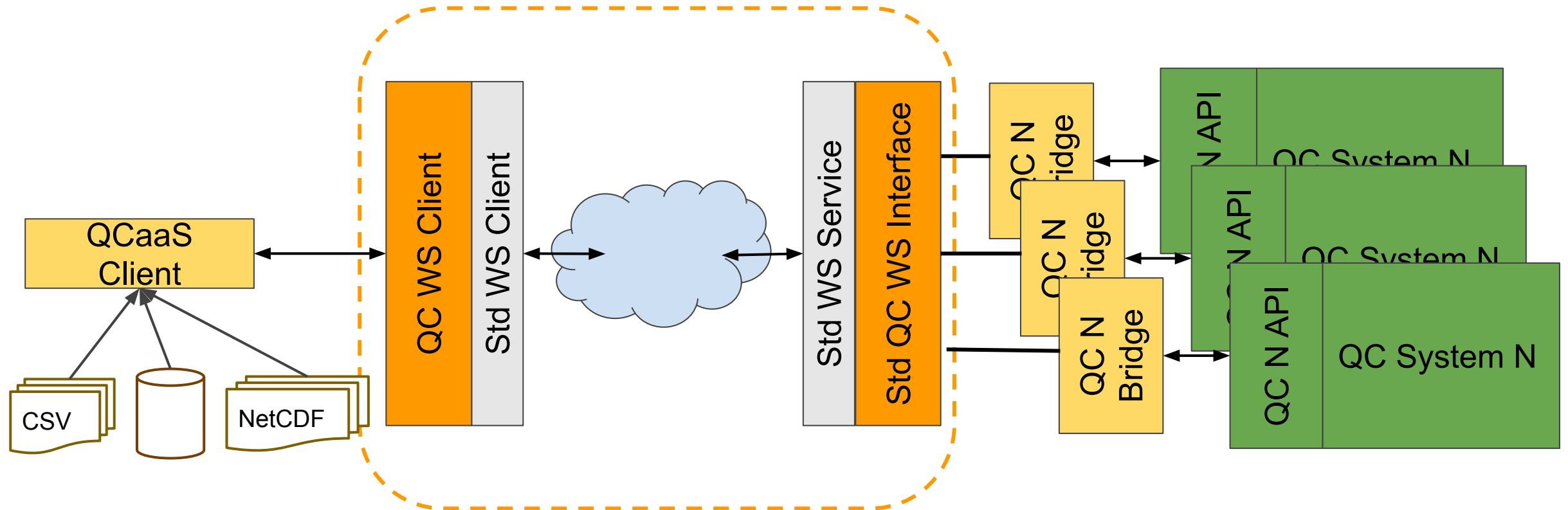


 QC Provider development

 Co-development

 PMEL development

Detailed Web-based QC invocation



QC Provider development

Co-development

PMEL development

QC Web Service Information

Service foundations

- QC Services should apply community accepted QC routines
- QC services will not harvest data
 - Data will be processed, and used in the response

Security

- No firm decision on authentication

Service categories

- Service Metadata
- QC Service

Request format

- RESTful API for invoking QC Web Services

QC Web Service Metadata

- Service Metadata
 - List of variables that can be QC'd
 - Required supplemental variables
 - Information about the QC routines / process
 - Configuration and Control parameters

Service information Sample Message

REQUEST: GET <https://data.pmel.noaa.gov/ws/qc/info.json>

RESPONSE:

```
{
  "service_info": {
    "name" : "PMEL/QARTOD",
    "version" : "0.1",
    "location" : "https://data.pmel.noaa.gov/ws/qc/"
  },
  "method" : {
    "name" : "QARTOD",
    "description" : "Standard tests for a variety of conditions.",
    "uri" : "https://ioos.noaa.gov/project/qartod/"
  },
  "standard_vocabulary" : {
    "name" : "Standard Variable Name Vocabulary",
    "uri" : "https://standards.com/oceanographic_variables"
  },
  "flagging_method" : {
    "name" : "QARTOD",
    "description" : "QARTOD defined QC flags",
    "uri" : "https://cdn.ioos.noaa.gov/media/2020/07/QARTOD-Data-Flags-Manual_version1.2final.pdf"
  },
  "qc_tests" : [
    {
      "name" : "range_check",
      "location" : "https://data.pmel.noaa.gov/ws/qc/range"
    },
    {
      "name" : "spike_test",
      "location" : "https://data.pmel.noaa.gov/ws/qc/spike"
    },
    {
      "name" : "drift_test",
      "location" : "https://data.pmel.noaa.gov/ws/qc/drift"
    }
  ],
  "legal" : {
    "privacy_statement" : "We will not expose your information.",
    "data_retention_policy" : "We will not retain any data."
  }
}
```

REQUEST: GET <https://data.pmel.noaa.gov/ws/qc/qartod/range>

RESPONSE:

```
{  "test_info" : {
    "name" : "range_test",
    "description" : "Check for questionable and
unacceptable data values.",
    "version" : "0.1a" },
  "method" : {
    "method_name" : "QARTOD range test",
    "method_description" : "Checks data values against
climatology.",
    "method_URI" :
"https://ioos.github.io/ioos\_gc/range/" },
  "accepted_variables" : [
    { "name" : "temperature",
      "supported_units" : "decimal_degrees_C" },
    { "name" : "salinity",
      "supported_units" : "PSU" },
    { "name" : "pH",
      "supported_units" : "" },
    { "name" : "dissolved_oxygen",
      "supported_units" : "umol/kg, mg/L, mL/L" }
  ],
  "supplemental_variables" : [
    { "name" : "latitude",
      "supported_units" : "decimal_degrees_north",
      "required" : "true" },
    { "name" : "longitude",
      "supported_units" : "decimal_degrees_180,
decimal_degrees_360",
      "required" : "true" },
    { "name" : "sample_time",
      "supported_units" :
"ISO8601,seconds_since_1970",
      "required" : "true" },
    { "name" : "sample_depth",
      "supported_units" : "meters_below_surface",
      "required" : "false" }
  ],
}
```

```
  "configuration_options" : [
    { "property_name" : "max_reasonable",
      "property_type" : "decimal",
      "property_description" :
"The max reasonable value above which will be
considered questionable.",
      "property_required" : "false" },
    { "property_name" : "min_reasonable",
      "property_type" : "decimal",
      "property_description" :
"The min reasonable value below which will be
considered questionable.",
      "property_required" : "false" },
    { "property_name" : "max_questionable",
      "property_type" : "decimal",
      "property_description" :
"The max questionable value above which will be
considered unacceptable.",
      "property_required" : "false" },
    { "property_name" : "min_questionable",
      "property_type" : "decimal",
      "property_description" :
"The min questionable value below which will be
considered unacceptable.",
      "property_required" : "false" }
  ]
}
```

QC Web Service Invocation

Invoke QC Service

- POST
 - Command and control parameters
 - Data description
 - Data
- RETURNS
 - Service metadata (abbreviated)
 - Invocation metadata
 - Flagged data
 - Results provide all necessary information to replicate request

Invocation Message Sample

POST: <https://data.pmel.noaa.gov/ws/qc/range>

```
{  "variable" : {
    "name" : "dissolved_oxygen",
    "units" : "umol/kg" },
  "configuration" : [
    "some_param" : "42",
    "other_param" : "0.5" ],
  "data" : {
    "data_variables" : [
      { "name" : "dissolved_oxygen", "units": "umol/kg" } ],
    "supplemental_variables" : [
      { "name" : "latitude", "units", "decimal_degrees_north" },
      { "name" : "longitude", "units" : "decimal_degrees_180" } ,
      { "name" : "sample_time", "units" : "ISO8601" } ],
    "rows" : [
      { "row_id": "1",
        "values" : 297.4, -127.34, 23.67, "20200708T145630Z" },
        ...
      ]
  }
}
```

Response Message Sample

```
{ "timestamp" : "20210808T145630Z",
  "service_info" : {
    "name" : "PMEL/QARTOD",
    "version" : "0.1",
    "location" : "https://data.pmel.noaa.gov/sdiq/ws/qc/qartod" },
  "test_info" : {
    "name" : "range_check",
    "version" : "0.1b",
    "location" : "https://data.pmel.noaa.gov/sdiq/ws/qc/qartod/range/0.1b" },
  "configuration" : [
    "some_param" : { "value" : "42", "source" : "user" },
    "other_param" : { "value" : "0.5", "source" : "user" },
    "min_value" : { "value" : "250", "source" : "default" },
    "max_value" : { "value" : "400", "source" : "default" } ],
  "flagged_values": {
    "variables" : [
      { "name" : "row", "units": "" },
      { "name" : "dissolved_oxygen", "units": "umol/kg" },
      { "name" : "latitude", "units", "decimal_degrees_north" },
      { "name" : "longitude", "units" : "decimal_degrees_180" } ,
      { "name" : "sample_time", "units" : "ISO8601" }
      { "name" : "dissolved_oxygen_qc" } ],
    "values" : [
      { 1, 297.4, -127.34, 23.67, "20200708T145630Z", "3" },
      { 2, 295.2, -127.34, 23.66, "20200708T150130Z", "3" },
      { 3, ...
      ...
    ]
  }
}
```

Challenges

Deploy on prem (AOML) and Commercial cloud (PMEL)

Commercial Cloud deployment

- Serverless hosting approach
 - AWS Lambda
- NOAA is required to use NOAA IP space for lambda endpoint
- Issues with AWS to accommodate this requirement
- Resulted in delay in commercial cloud deployment

Service Evaluation and Status

Evaluation

1. Invocation message format
2. Response message
3. Network/message latency
 - a. We will also evaluate cloud mirror services to reduce latency

Status

- Test client and server components completed
- On-prem implementation completed at PMEL
- Awaiting resolution of commercial cloud issues

Interested?

We welcome participants to this development

- Have a community accepted QC service
- Expertise on QC of a particular domain
- Test QC endpoints
- Contribute to message specification

Github repository - <https://github.com/NOAA-PMEL/QCaaS>

Eugene Burger - eugene.burger@noaa.gov

Linus Kamb - linus.kamb@noaa.gov



PMEL | Pacific Marine Environmental Laboratory

THANK YOU
QUESTIONS?