Coordinated Assessment Data Exchange (CAX)

Flow Configuration Document

Version 1.0a

October 4, 2016

**Revision Log**

| **Date** | **Version No.** | **Description** | **Author** | **Reviewer** | **Review Date** |
| --- | --- | --- | --- | --- | --- |
| 12/31/2014 | 1.0 | Draft Final | Greg Wilke, PSMFC |  |  |
| 10/04/2016 | 1.0a | Date changed to correspond with posting to Exchange Network and Revision Log added. | Greg McNelly, ECOS | Greg Wilke, PSMFC | 10/4/16 |

Table of Contents

[1.0 Introduction 4](#_Toc463355718)

[2.0 CAX Overview and Flow Summary 4](#_Toc463355719)

[2.1 Flow Identification 5](#_Toc463355720)

[2.2 DES Data Tables 6](#_Toc463355721)

[3.0 Virtual Node Flow Description 6](#_Toc463355722)

[3.1 Schema Information 6](#_Toc463355723)

[Schema Structure 6](#_Toc463355724)

[*Diagram of the CAX v1.0 Schema Components* 7](#_Toc463355725)

[3.2 Data Service Information 7](#_Toc463355726)

[4.0 Data Submission Information 10](#_Toc463355727)

[4.1 Data Exchange Standard 11](#_Toc463355728)

[4.2 Maintaining and Updating the Data Exchange Standard (DES) 11](#_Toc463355729)

[4.3 Partner Data Sharing Agreements and Controls 12](#_Toc463355730)

[4.4 Automated Data Processing 15](#_Toc463355731)

[4.5 Manual Data Sharing 16](#_Toc463355732)

Coordinated Assessment Data Exchange (CAX)

Flow Configuration Document

# 1.0 Introduction

The Coordinated Assessments (CA) Project is an effort to develop efficient, consistent, and transparent data-sharing among the co-managers (fish and wildlife agencies and Tribes) and regulatory/funding agencies (BPA & NOAA) of the Columbia River Basin (CRB) for anadromous fish related data. The project has been coordinated by the Pacific Northwest Aquatic Monitoring Partnership (PNAMP) with support from the Pacific States Marine Fisheries Commission (PSMFC) StreamNet project. Oversight is provided through the Coordinated Assessments Planning Group (CAPG) which consists of representatives from the states, tribes and federal agencies. Facilitation and support are provided through the Coordinated Assessments Core Team. The CA Project involves nearly a hundred biologists and data managers across the Columbia River Basin (CRB), representing agencies and tribes with an interest in anadromous fish management and recovery.

The initial effort of the CA Project has focused on the development of a Data Exchange Standard (DES) for four salmon and steelhead population (VSP) indicators including natural origin spawner abundance, natural origin smolt to adult return ratio (SAR), and natural origin recruits (adult and juvenile) per spawner. The initial DES will be expanded as additional indicators are agreed upon by the participants in the CA Project. Over time, the DES will include most of the indicators of highest importance to fisheries management and restoration decision makers.

The Data Exchange Standard (DES) is used to define the contents of a centralized Coordinated Assessments Exchange (CAX) data repository at StreamNet, where data can be submitted by data providers through automated or manual processes, depending on the capabilities of the individual providers. Public access to the CAX is provided through a virtual node on the EPA Data Exchange Network. Data providers will develop individual Flow Configuration Documents (FCD) as appropriate to describe their intentions for data sharing as part of the CA Project. All CA Project documents are available on the PNAMP website: <http://www.pnamp.org/project/3129>.

# 2.0 CAX Overview and Flow Summary

The Coordinated Assessments Exchange (CAX) is being developed to support the flow of data from the federal, state and tribal data collection and management agencies to users of salmon and steelhead data and indicators in the Columbia River Basin. The CAX is supported through a data repository at StreamNet that is automatically updated from individual data providers. Each data provider will define their data sharing processes through individual Flow Configuration Documents, Appendices to this document, with the expectation that data will eventually be automatically transferred in a format consistent with the DES for all data that is available and released for use by the data provider. Development of the CAX has relied on considerable support from Bonneville Power Administration (BPA) for data base development within the agencies and tribes and from the Environmental Protection Agency (EPA) for data flow development between the agencies and tribes and the StreamNet project (Figure 1).

****

**Figure 1.** Schematic of Coordinated Assessments Data Exchange (CAX) basic architecture.

## 2.1 Flow Identification

Flow Name: Coordinated Assessments Data Exchange (CAX)

Flow Description: Exchange of Pacific Northwest salmon and steelhead indicators and metrics

Flow Steward: Pacific States Marine Fisheries Commission

 StreamNet Project

 205 SE Spokane Street, Suite 100

 Portland, Oregon 97202

 (503) 595-3100

 [www.streamnet.org](http://www.streamnet.org)

Flow Contact: Greg Wilke

GWilke@psmfc.org

## 2.2 DES Data Tables

* **A1. NOSA Table -** This table describes information concerning natural origin spawner abundance (NOSA) as defined by the Coordinated Assessments project. "Spawner abundance" refers to the number of fish that actually spawn, not necessarily the total number of fish returning to a spawning area -- all pre-spawning mortality has already been accounted for in the numbers represented in this table.
* **A2. SAR Table -** This table describes information concerning smolt to adult ratios (SAR) as defined by the Coordinated Assessments project. Smolt to adult ratios are specific to the smolt and adult locations described in each row of data.
* **A3.** **RperS Table -** This table describes information concerning recruits per spawner (R/S) as defined by the Coordinated Assessments project. Recruit per spawner ratios are specific to the locations and seasons described in each record of data. This table can include the number of juvenile or adult recruits as measures, or full life cycle productivity. That is, "recruit" can be defined at any life stage.

3.0 Virtual Node Flow Description

(*CAX to EPA*)

Virtual Node Implementation: the project proposes to create data publishing services and new data flows using the EPA Virtual Node services. The CAX flow is a new flow for the Exchange Network. Use of Virtual Node will avoid need for installation of a local Node at StreamNet. Instead, data providing partner clients can interact with the Virtual Node to aggregate data in a cloud based repository for data set sharing and re-publishing as described above. In addition to these Exchange Network Priorities, the CAX project will produce several re-usable components of relevance to other entities managing salmonid data, this includes all of the environmental and wildlife Agencies and Tribes on the west coast including Alaska.

# 3.1 Schema Information

## Schema Structure

The CAX schema describes the Natural Origin Spawner Abundance (NOSA), Smolt to Adult Ratio (SAR), and Recruits per Spawner (RperS) metrics and indicators that compose the CAX node at present. The following diagram illustrates the major components of the CAX schema. Double blocks represent components that may be repeated. Each data type stands alone and represents an optional component.

**

## Diagram of the CAX v1.0 Schema Components

## 3.2 Data Service Information

Three data services are defined for the CAX flow on the Exchange Network. The first service, CAX.GetNOSA\_v1, returns a detailed set of Natural Origin Spawner metrics and indicators that match a set of optional criteria. These data represent total number of spawners for the year in the population. The second service, CAX.GetRPerS\_v1, returns a detailed set of Recruits Per Spawner metrics and indicators that match a set of given criteria. These data represent the number of recruits per spawner for each brood year, again by entire population. The third service, CAX.GetSAR\_v1, returns a detailed set of Smolt to Adult Ratio metrics and indicators that match a set of given criteria. These data represent the ratio of the number of adults that returned to the number of smolts that produced them for each migration year, by entire population.

These are the out-facing EPA Exchange Network Services for ad-hoc Query network data publishing on the CAX Virtual Node.

**CAX.GetNOSA\_v1**

**Data Service Type:** Query

**Data Service Parameters, Order, and Format:**

|  |  |
| --- | --- |
| **Parameter Name** | **CommonName** |
|  Position | 1 |
|  Data Type | xsd: CommonNameType |
|  Cardinality | Zero or multiple, comma delimited list |
|  Max Length | 1000 |
|  Wildcard | Treat search string(s) as “equal”. No wildcard character supported  |
| **Parameter Name** | **ESU\_DPS** |
|  Position | 2 |
|  Data Type | xsd: ESUType |
|  Cardinality | Zero or multiple, comma delimited list |
|  Max Length | 1000 |
|  Wildcard | Treat search string(s) as “equal”. No wildcard character supported  |
| **Parameter Name** | **Run** |
|  Position | 3 |
|  Data Type | xsd: FishRunType |
|  Cardinality | Zero or multiple, comma delimited list |
|  Max Length | 500 |
|  Wildcard | Treat search string(s) as “equal”. No wildcard character supported  |
| **Parameter Name** | **RecoveryDomain** |
|  Position | 4 |
|  Data Type | xsd: RecoveryDomainType |
|  Cardinality | Zero or multiple, comma delimited list |
|  Max Length | 1000 |
|  Wildcard | Treat search string(s) as “equal”. No wildcard character supported  |
| **Parameter Name** | **WaterBody** |
|  Position | 5 |
|  Data Type | xsd: String  |
|  Cardinality | Zero or one |
|  Max Length | 100 |
|  Wildcard | Treat search string as equal with beginning or trailing wildcard \* allowed |

**Payload Format:** The service returns an XML file conforming to the Index.xsd schema

**Data Service-level Business Rules:** Zero or more parameters may be provided.

**XML Header Usage:** The EN Header is not implemented for CAX Query service.

**CAX.GetRPerS\_v1**

**Data Service Type:** Query

**Data Service Parameters, Order, and Format:**

|  |  |
| --- | --- |
| **Parameter Name** | **CommonName** |
|  Position | 1 |
|  Data Type | xsd: CommonNameType |
|  Cardinality | Zero or multiple, comma delimited list |
|  Max Length | 1000 |
|  Wildcard | Treat search string(s) as “equal”. No wildcard character supported  |
| **Parameter Name** | **ESU\_DPS** |
|  Position | 2 |
|  Data Type | xsd: ESUType |
|  Cardinality | Zero or multiple, comma delimited list |
|  Max Length | 1000 |
|  Wildcard | Treat search string(s) as “equal”. No wildcard character supported  |
| **Parameter Name** | **Run** |
|  Position | 3 |
|  Data Type | xsd: FishRunType |
|  Cardinality | Zero or multiple, comma delimited list |
|  Max Length | 500 |
|  Wildcard | Treat search string(s) as “equal”. No wildcard character supported  |
| **Parameter Name** | **RecoveryDomain** |
|  Position | 4 |
|  Data Type | xsd: RecoveryDomainType |
|  Cardinality | Zero or multiple, comma delimited list |
|  Max Length | 1000 |
|  Wildcard | Treat search string(s) as “equal”. No wildcard character supported  |
| **Parameter Name** | **Brood Year** |
|  Position | 5 |
|  Data Type | xsd: String  |
|  Cardinality | Zero or multiple, comma delimited list |
|  Max Length | 100 |
|  Wildcard | Treat search string(s) as “equal”. No wildcard character supported |

**Payload Format:** The service returns an XML file conforming to the Index.xsd schema

**Data Service-level Business Rules:** Zero or more parameters may be provided.

**XML Header Usage:** The EN Header is not implemented for CAX Query service.

**CAX.GetSAR\_v1**

**Data Service Type:** Query

**Data Service Parameters, Order, and Format:**

|  |  |
| --- | --- |
| **Parameter Name** | **CommonName** |
|  Position | 1 |
|  Data Type | xsd: CommonNameType |
|  Cardinality | Zero or multiple, comma delimited list |
|  Max Length | 1000 |
|  Wildcard | Treat search string(s) as “equal”. No wildcard character supported  |
| **Parameter Name** | **ESU\_DPS** |
|  Position | 2 |
|  Data Type | xsd: ESUType |
|  Cardinality | Zero or multiple, comma delimited list |
|  Max Length | 1000 |
|  Wildcard | Treat search string(s) as “equal”. No wildcard character supported  |
| **Parameter Name** | **Run** |
|  Position | 3 |
|  Data Type | xsd: FishRunType |
|  Cardinality | Zero or multiple, comma delimited list |
|  Max Length | 500 |
|  Wildcard | Treat search string(s) as “equal”. No wildcard character supported  |
| **Parameter Name** | **RecoveryDomain** |
|  Position | 4 |
|  Data Type | xsd: RecoveryDomainType |
|  Cardinality | Zero or multiple, comma delimited list |
|  Max Length | 1000 |
|  Wildcard | Treat search string(s) as “equal”. No wildcard character supported  |

**Payload Format:** The service returns an XML file conforming to the Index.xsd schema

**Data Service-level Business Rules:** Zero or more parameters may be provided.

**XML Header Usage:** The EN Header is not implemented for CAX Query service.

# 4.0 Data Submission Information

(*Data Source to CAX*)

The project will allow virtual sharing of data sets across State/State, State/Tribal, and Tribal/Tribal jurisdictions. Virtual sharing configuration will allow authorized partner access to datasets without the need to access partner local systems. Because this capability will be provided across jurisdictions it will support improved regional environmental management across the entire Columbia River Basin which is of priority concern. The project will aggregate and publish CAX data from a cloud based data repository allowing consumption of this data by client applications and private/public websites. Data will be published in both standard EN format and as EN REST based services.

## 4.1 Data Exchange Standard

The key to success of the CAX has been the development of an agreed upon Data Exchange Standard (DES) describing the data elements and their structure needed to support the exchange of four VSP indicators and supporting metrics: natural origin spawner abundance, smolt to adult return ratio, and recruits per spawner (adult and juvenile). The DES was developed with wide participation of the larger working group, first through an extensive pilot program to document data flows and availability of the indicator and supporting metrics conducted with Oregon, Washington, and Idaho state agencies, six Columbia River Basin tribes, and one tribal coalition (Columbia River Inter-Tribal Fisheries Commission). This was followed by intensive focus on refining the draft DES by a development team consisting of data management and resource management experts. The draft DES was then vetted and approved by the CA Working Group for implementation. Three tables were developed in the original DES:

* **A1. NOSA Table -** Natural origin spawner abundance as defined by the CA Project.
* **A2. SAR Table -** Smolt to adult return ratios (SAR) as defined by the CA Project.
* **A3. RperS Table -** Recruits per spawner (R/S) as defined by the CA Project for both juveniles and adults

## 4.2 Maintaining and Updating the Data Exchange Standard (DES)

1. The CA planning team develops a list of additional possible DES to develop, and these are prioritized for the next five years through collaborative regional discussion. New regional DES development then is undertaken in priority order. Priority is given to high level indicators that will best serve the region if standardized and shared across sovereign and institutional boundaries.
2. The CA planning team identifies persons to invite to help establish DES, and guidance for how and when DES are developed in a collaborative process.
3. Once direction on data types is established, StreamNet staff will call together collaborative work groups of professionals to design the DES for each type of data. This process will include a definition of terms to avoid any ambiguity in the DES. This process will continue until a general consensus is reached on the DES. If consensus cannot be reached, the issue will be reported to the CA Planning Team. The Planning Team will resolve the issue and report back to the work group with direction.
4. If/when consensus is reached on a DES the group will publish the standard to StreamNet and give notice to a broad audience of the intent to adopt the DES (see current process for PSMFC staff work product detail; ftp://ftp.streamnet.org/pub/streamnet/Projman\_files/ExchangeFormat/CurrentDraft/DES-Change-Process.pdf)
5. The date of the posting will be noted in the publication. A shakedown period of one month will allow any previously unidentified issues with the DES to be brought forward.
6. If issues are brought forward in this shakedown period, the working group will meet to discuss and resolve. If resolution occurs, the group will return to step 3 and repeat the process. If no resolution is possible, they will return to step 2 and repeat the process.
7. If no additional issues are brought forward during the shakedown period, the DES will be published on StreamNet, with the date of adoption, and be titled the “StreamNet DES Standard” followed by the date: OO-00-20XX. The definition of terms will be published on monitoring methods.org. This standard will remain in effect until the CA Planning Group calls for reconsideration of DES standards.
8. In general, this period should be approximately one year, or until the CA planning group determines that revised DES standards are needed due to input and requests from concerned professionals. The StreamNet staff will solicit the CA Planning group and other professionals at approximately annual intervals to determine the need to reopen the DES for review.
9. Standards will be revised by repeating this process, starting with 2) above.
10. The CA Planning Group and the StreamNet staff will try and synchronize revisions of standards to minimize disruptions and simplify data management to the greatest extent possible. StreamNet staff will serve as central collectors of DES comments and revision requests, and will maintain a record of issues with the DES standards during the period between adoption and the next review. This record will be presented to the collaborative work groups (2 above) that periodically review standards for modification.
11. StreamNet staff will publish any revised DES standards on StreamNet and any change in term definitions on monitoring methods.org, with an updated date to reflect the new revision.

## 4.3 Partner Data Sharing Agreements and Controls

The CAX database on the EPA node will serve as a secure, shared repository for indicator and metric level data. The CAX database (defined by the DES) will be available on the EPA Exchange Network as a virtual node. It will not be available on StreamNet. The current understanding on access to this data is that EPA requires registration and approval before access to the node is granted. This will be confirmed and implemented as part of the CA project. The CA planning group will be tasked with generating a list of approved organizations and individuals that have access to start with, as well as a process for granting access upon request.

Metadata, including citations to individual agencies or tribes, biologists, databases, and reports, will be linked to the indicator and metric level data on the node. At the current time, it is not known if this can be accomplished technically through adding metadata links that would auto generate a list of citations for each metric in the DES. This is the preferred option for assuring that data contributors are properly acknowledged and referenced by data users. If this is not technically feasible, than a more “generic” listing of agencies or tribes, biologists, databases, and reports will be linked to the metric level data on the EPA node and prominently displayed when data is accessed and downloaded. This issue will be investigated by the XCT for feasibility and technical issues. The option for assuring that data contributors are properly acknowledged and referenced by data users will depend on their findings.

Indicator Level Data

This Data Sharing Agreement, EULA, and policy are intended for use with Coordinated Assessment (CA) indicator level data sets that have passed QA, and are ready for dissemination online through StreamNet and the EPA node.

A Specific web page or pages will be designated for the CA indicator only data with reference to data providers on StreamNet. This will be the same information presented on the EPA node. Users attempting to access these pages will be presented with a EULA requiring agreement before access to the site is granted. The proposed EULA language is;

*IMPORTANT! BE SURE TO CAREFULLY READ AND UNDERSTAND THE TERMS AND CONDITIONS SET FORTH IN THIS END-USER LICENSE AGREEMENT. YOU WILL BE ASKED TO REVIEW AND EITHER ACCEPT OR NOT ACCEPT THE TERMS OF THE EULA. YOU ARE NOT AUTHORIZED TO ACCESS OR USE THE DATA IN THIS DATABASE UNLESS AND UNTIL YOU ACCEPT THE TERMS OF THIS EULA.*

*In order to access, share, or use these datasets you agree to;*

1. *Acknowledge that data in the StreamNet database may be updated at any time if revised information becomes available.*
2. *Corrections to data may at times be necessary, data contributors will strive to make in a timely manner, but are under no obligation to do so. Notification of corrections and/or data updates will be posted on the StreamNet website.*
3. *Data contributors, StreamNet, PSMFC, and BPA accept no responsibility or liability for the accuracy of this data or for the uses and/or interpretations to which this data may be put.*
4. *Data contributors have uploaded links to the Metadata records and associated publications that may discuss the limitations and proper use of the data. It is your responsibility to understand and comply with these restrictions if you use this data.*
5. *Database managers may take measures to restrict access to this data if violations occur, to ensure violations are not repeated.*
6. *Data users should be careful to properly credit individuals and agencies when using data, and should follow reputable standards for the use and interpretation of scientific information.*
7. *Users must assume responsibility to determine the usability of the data for their purposes. Data users acknowledge the StreamNet database is dynamic. Records can be added, deleted and changed on a frequent basis.*
8. *Before publication; The Data User is expected to contact the appropriate Data Contributor and to maintain a record of contact via email and if necessary phone prior to significant use of data in any publication in order to:*
9. *Obtain verification, data use limitations and context for data through metadata records.*
10. *Secure appropriate permissions prior to submission for publication by sending requests to data owners and cc’ing the data publishers or peer review team.*
11. *Arrange appropriate acknowledgements, citations and/or authorships.*
12. *Failure to comply with the requirements may result in denial of access to restricted data files in future requests. (Data owners have 90 days to respond to data publication reviews). Data users will be notified of any accusations of failure to comply and have opportunities to defend your action to ensure access.*
13. *Please note that a disclaimer for any liability incurred by imposition of the policy, and for enforcement responsibility of the policy by the data repository manager and funding agency, will need to be applied. StreamNet would endeavor to design and maintain access to the data for only those that agreed to comply with the agreements for each data set, but would not accept any liability for the legality of any DSA, for any compliance failure, or any ownership of an enforcement function. Any requirements imposed for data sharing by an agency or funding entity would be a contractual matter and not the purview of StreamNet.*

Upon accepting this EULA, data users will be allowed free access to the Indicator data in the CA database. The expectation is that this data has been derived, quality checked, shared between multiple entities, and generally collected through the expenditure of public funds. Misinterpretation of the actual data (as opposed to cause and effect relationships, responsibilities, etc.) should be limited at this level.

Metadata, including protocols used to calculate the indicators for each population, will be available on this StreamNet website. This metadata will list the source agency, scientist, publication, and other information as references for the user. For each population, a “generic” list of each agency or tribe, including contact information, will be generated as user’s download data.

Summary Table

|  |  |  |
| --- | --- | --- |
|  | **Input** to DSA form (to be filled out by data contributor/creator/provider) | **Output** – what data requester sees on the website, or when clicks ‘Download’ |
| **Indicator Data:** **StreamNet Database****Open Access with EULA user acknowledgement** | Indicator data downloaded from CAX database onto StreamNet dedicated web page.  Metadata including protocols and citations will be included. No stipulations for required interaction with consumers are necessary. See EULA. | Data users acknowledge policy and basic data sharing agreement by “clicking” before access. A list of all data contributors by population is generated at download. See EULA. |
| **Metric Data:** **EPA Node****Restricted access with Password and second EULA user acknowledgement** | Data uploaded to CAX consistent with DES. Metadata including protocols and citations included for each metric. All data has been approved for sharing via EPA EN and QA/QC completed. | Access available through EPA EN and limited to approved participants through EPA process. Metadata, including citations, is linked to each metric (if possible) or generated by population at download.  |

## 4.4 Automated Data Processing

1. Identification

In order to ensure that only an authorized user from an organization may insert and update CAX records for that organization, a unique identifier is assigned to each individual data submitter. The individual unique identifier, called the API Key, is included along with every data record submitted to the CAX to be inserted or updated. This identifier is assigned to an organization/individual by StreamNet upon request by an organization to begin data submissions. All record requests received at the CAX, either read or write, are authenticated to ensure the individual/organization is authorized to perform the requested transaction.

1. Submission Format

Data is submitted to the API for insert and update operation in DES XML format. Each API call corresponds to a single record in one of the DES Data Tables (A1. NOSA Table, A2. SAR Table, A3. RperS Table**)**

1. Data Submission

StreamNet provides a RESTful Application Programming Interface (API) for CA data to be submitted to the CAX in Index.xsd schema format. The interface allows for table/record level read, write, update, delete operations for CAX data and each operation type can be allowed or denied based on the assigned API Key.

[StreamNet CAX API documentation](http://www.streamnet.org/api.html)

1. Validation

All CA data records submitted to the API for insertion or update are validated against the Index.xsd schema format definition. Records that pass validation are inserted or updated in the CAX and immediately return a “success” code to the submitter. Records that fail validation are not updated or inserted, but will return a “fail” code to the submitter with a detailed error message describing the issue(s) with each record that failed to be submitted.

1. Security

All communication with the API is conducted over secure HTTP (HTTPS). This ensures that all data transmissions to or from the API are encrypted over the internet.

## 4.5 Manual Data Sharing

The Data Exchange Standard (DES) describes the data to be exchanged regardless of the technology being used. If data systems are still in development or a provider wishes to submit data manually, the DES will allow submission of that data in Excel, Access, or other third party format. StreamNet staff, as managers of the CAX, will assist data providers with whatever level of support is necessary to submit and upload their data to the data repository.