



CVPIA Data Management Strategy

Megan Cook, CVPIA Science Coordinator

Sadie Gill, FlowWest



— BUREAU OF —
RECLAMATION

Central Valley Project Improvement Act (CVPIA)

- Co-implemented by the U.S. Fish and Wildlife Service and Bureau of Reclamation
- Establishes abundance goals for Chinook salmon, *O. mykiss*, and green and white sturgeon
- Established a science-based prioritization and adaptive management process, founded on Structured Decision Making



CVPIA Science Integration Team (SIT)



COLLABORATIVE



TRANSPARENT



ADAPTIVE

Near-term Restoration Strategy



Restoration Actions for Chinook salmon



Information Needs for Chinook, *O. mykiss*, and sturgeon



Five-year Timeline



Annual Updates

CVPIA Data Management Strategy

Transparency

- Make all model code, data, and analyses open source and available to the public

Accessibility

- Publish data in accessible formats in a single location

Transparency via GitHub

- Models, source code, data available
- Version controlled collaboration



A screenshot of a web browser displaying the GitHub profile page for the organization 'Central Valley Project Improvement Act - Open Science Collaborative'. The browser's address bar shows 'github.com/CVPIA-OSC'. The page header includes navigation links for 'Pull requests', 'Issues', 'Marketplace', and 'Explore'. The organization's profile information is visible, including its logo and a description: 'Facilitating open science practices for fisheries management in California's Central Valley'. Below this, there are tabs for 'Repositories' (7), 'Packages', 'People' (7), 'Teams', 'Projects', and 'Settings'. The 'Pinned repositories' section features three cards for 'fallRunDSM', 'springRunDSM', and 'winterRunDSM', each with a description and a language indicator (R). A search bar for repositories is present, along with filters for 'Type: All' and 'Language: All', and a 'New' button. The 'EDlutils' repository is highlighted with a green line graph, showing metadata workflow helpers for CVPIA funded research. The 'Top languages' section shows R and HTML. The 'People' section shows 7 members with profile pictures.

Data Accessibility

- Metadata standard for consistent metadata (EML)
- Tools to assist data stewards
- Centralized data repository (EDI)

The screenshot shows a web browser displaying the article 'Creating EML Using cvpiaEDIutils' on the EDIutils website. The page has a dark blue header with navigation links for 'EDIutils 0.0.0.9000', 'Reference', and 'Articles'. The article title is 'Creating EML Using cvpiaEDIutils'. Below the title, there is a 'Contents' sidebar with links for 'Our Example', 'Creating EML', and 'Putting It All Together'. The main content area starts with the text 'The following libraries are needed to create a working EML document.' followed by a code block containing the R library loading commands: `library(cvpiaEDIutils)`, `library(tidyverse)`, `library(readxl)`, and `library(EML)`. Below this, it says 'In this document, we create an EML file which has the opportunity to append the following elements:' followed by a list of EML elements: `- eml`, `- dataset`, `- creator`, `- contact`, `- associated party`, `- title`, `- short name`, `- abstract`, `- keyword set`, `- intellectual rights`, `- license`, `- methods`, `- maintenance`, `- coverage` (with sub-elements: `- geographic coverage`, `- temporal coverage`, `- taxonomic coverage`), `- data table` (with sub-elements: `- physical`, `- attribute list`), and `- methods`. At the bottom, there is a section titled 'Our Example' with introductory text about using data from John Hannon.

EDIutils 0.0.0.9000 Reference Articles

Creating EML Using cvpiaEDIutils

Contents

- [Our Example](#)
- [Creating EML](#)
- [Putting It All Together](#)

The following libraries are needed to create a working EML document.

```
library(cvpiaEDIutils)
library(tidyverse)
library(readxl)
library(EML)
```

In this document, we create an EML file which has the opportunity to append the following elements:

```
- eml
- dataset
  - creator
  - contact
  - associated party
  - title
  - short name
  - abstract
  - keyword set
  - intellectual rights
  - license
  - methods
  - maintenance
  - coverage
    - geographic coverage
    - temporal coverage
    - taxonomic coverage
  - data table
    - physical
    - attribute list
  - methods
```

Our Example

We will use the data provided by John Hannon to show an example of how to put the metadata together in a way which can be uploaded to the EDI data portal website. The excel sheet of which Hannon's data resides is found in "~/data-raw/template/template.xlsx". This file can be edited with your data and used with this document to create your own working EML

Centralized Data Repository

➤ Environmental Data Initiative (EDI)



The screenshot shows the homepage of the EDI Data Portal. The browser address bar displays 'portal.edirepository.org/nis/home.jsp'. The page features a navigation menu with 'HOME', 'DATA', 'TOOLS', 'HELP', and 'LOGIN'. A search bar contains the text 'salmon' and a magnifying glass icon, with a link to 'ADVANCED SEARCH' below it. The main content area is titled 'Welcome to the EDI Data Portal' and contains two paragraphs of text. To the right, there is a line graph titled 'Contributed Data Package Growth' showing cumulative data packages from 2013 to 2020. Below the graph, there are two summary statistics: 'Contributed Data Packages' and 'Total Data Packages (including EcoTrends and Landsat)'. The browser window title is 'Data Portal - Home | Environmental Data Initiative (EDI) - Google Chrome'.

Welcome to the EDI Data Portal

Data are one of the most valuable products curated by the Environmental Data Initiative (EDI). Data and metadata derived from publicly funded research are made available through this website with as few restrictions as possible, and on a non-discriminatory basis. In return, EDI expects users of data to act ethically by contacting the data provider prior to using it in any published research. In accordance with professional etiquette, data accessed from this website should be cited appropriately when used in a publication. A digital object identifier (DOI) is provided for each dataset to facilitate citation.

The EDI Data Portal contains environmental and ecological data packages contributed by a number of participating organizations. Data providers make every effort to release data in a timely fashion and with attention to accurate, well-designed and well-documented data. To understand data fully, please read the associated metadata and contact data providers if you have any questions. Data may be used in a manner conforming with the license information found in the "Intellectual Rights" section of the data package metadata or defaults to the EDI Data Policy. The Environmental Data Initiative shall not be liable for any damages resulting from misinterpretation or misuse of the data or metadata.

Contributed Data Package Growth

— Unique — All Revisions

Date	Unique	All Revisions
Jan 2013	0	0
Nov 2013	~2,000	~4,000
Sep 2014	~3,000	~6,000
Jul 2015	~4,000	~8,000
May 2016	~5,000	~10,000
Mar 2017	~6,000	~12,000
Jan 2018	~7,000	~14,000
Nov 2018	~8,000	~16,000
Sep 2019	~9,000	~18,000
Jul 2020	~10,000	~22,000

Contributed Data Packages
Unique: 7924; All Revisions: 23272

Total Data Packages (including EcoTrends and Landsat)
Unique: 44290; All Revisions: 75041

Contact Information

Megan Cook

megan_cook@fws.gov

Sadie Gill

sgill@flowwest.com