



Environmental Analysis Program

Pomona College
Biogeochemistry Laboratory

SOP-L807

Instrument Data Export and File Management

Standard Operating Procedure

SOP Number	SOP-L807
Revision	0.1
Author(s)	Marc Los Huertos
Date	2026-06-26
Organization	EA Program, Pomona College

FOR INTERNAL USE ONLY

This document is the property of the Environmental Analysis Program at Pomona College.
Unauthorized distribution is not permitted.

Table of contents

1 Purpose	2
2 Scope	2
3 Responsibilities	2
4 Definitions	2
5 Safety and Hazards	3
6 Equipment and Materials	3
6.1 Instruments and Export Formats	3
6.2 Transfer Media	3
7 Procedure	4
7.1 General Export Workflow	4
7.2 Instrument-Specific Export Instructions	4
7.2.1 Shimadzu UV-Vis (UV-1900i / UVProbe)	4
7.2.2 Thermo iCAP ICP-OES (Qtegra)	4
7.2.3 Shimadzu TOC-L (TOC-Control)	5
7.2.4 YSI ProDSS Field Sensor	5
7.2.5 Bio-Rad QX200 ddPCR (QuantaSoft)	5
7.3 File Transfer to RStudio Server	5
7.4 Automated Import Scripts in R	5
7.5 Data Integrity Checks	6
7.6 Archival and Backup	7
7.7 Instrument Log Cross-Reference	7
8 Quality Control and Acceptance Criteria	7
9 Waste Disposal	8
10 References	8
11 Training Record	8
12 Appendix A: File Hash Verification	8

13 Appendix B: Instrument Data Format Quick Reference

8

1. Purpose

This SOP standardizes the procedures for exporting raw data from laboratory instruments, transferring files to the data management system, and organizing instrument outputs for downstream analysis. It ensures that raw instrument data are preserved without modification, properly documented, and accessible for reproducible analysis (Gandrud, 2015).

2. Scope

- **Applies to:** All analytical instruments in the EA Program laboratory that produce digital data files, including spectrophotometers, ICP-OES, TOC analyzers, ion chromatographs, balances with data output, pH/conductivity meters with logging, and field sensors that store data internally
- **Does not apply to:** Manual readings recorded directly in lab notebooks (see [SOP-L010](#)); drone imagery (see [SOP-L805](#)); bioinformatics sequencing data (see [SOP-L705](#))
- **Location:** EA Program laboratory (SGM 133)

3. Responsibilities

Role	Responsibility
Lab PI	Approve file management protocols; ensure backup compliance
Lab Manager	Maintain instrument computers; manage shared storage; train users on export procedures
Analyst / Student	Export data following this SOP; verify file integrity; commit to version control
Data Reviewer	Confirm that raw files match lab notebook entries and are properly archived

4. Definitions

- **Comma-separated values (CSV):** A plain-text file format in which values are separated by commas and rows by line breaks; universally readable; the preferred format for sharing tabular data between instruments and analysis software.
- **Data provenance:** The documented history of how data were collected, processed, transformed, and stored; enables reproduction of results and detection of errors; captured through lab notebooks, metadata files, and Git commit history.
- **File naming convention:** A standardized pattern for naming files that encodes date, instrument, project, and analyst information without spaces or special characters (e.g., YYYY-MM-DD_InstrumentID_ProjectCode_UserInitials_raw.csv); required for all EA Program instrument export files.

- **Instrument export file:** The raw data file generated directly by instrument software (e.g., Shimadzu LabSolutions .lcd or .txt export, Isodat .dxf export, HOBOWare .csv export); must be saved unmodified as the archival primary record.
- **Metadata:** Data describing other data; includes instrument settings, calibration records, sample identifiers, units, processing steps, and analyst notes; stored alongside raw data in a companion metadata file.
- **Raw data:** The unprocessed data as produced by the instrument or analyst at the time of measurement; must never be altered; all subsequent processing produces derived data files that preserve the original.

5. Safety and Hazards

Data export is a computational activity with no physical hazards. When interacting with instruments (connecting USB drives, accessing instrument PCs), follow the safety protocols in the relevant instrument SOP.

6. Equipment and Materials

6.1. Instruments and Export Formats

Instrument	Native Format	Preferred Export	Notes
Shimadzu UV-Vis (UV-1900i)	.spc / .csv	CSV	Export from UVProbe software
Thermo iCAP ICP-OES	.rst / .csv	CSV	Export from Qtegra software
Shimadzu TOC-L	.toa / .csv	CSV	Export from TOC-Control
Dionex ICS Ion Chromatograph	.cdf / .csv	CSV	Export from Chromeleon
Mettler Toledo balance	Direct serial output	CSV via terminal log	Use logging script
YSI ProDSS multiparameter	.csv / .dat	CSV	Download via KorDSS software
Hach DR3900 spectrophotometer	Internal memory	CSV via USB	Export from instrument menu
Bio-Rad QX200 ddPCR	.qlp / .csv	CSV	Export from QuantaSoft
PANalytical XRF	.csv / .xls	CSV	Export from SuperQ

6.2. Transfer Media

Item	Purpose	Notes
USB flash drive (lab)	Sneakernet transfer	Dedicated drive; virus-scanned

Item	Purpose	Notes
Network shared folder	Direct instrument PC transfer	If instrument PC has network access
RStudio Server upload	Upload to analysis environment	Via Files pane or scp

7. Procedure

7.1. General Export Workflow

The same sequence applies to all instruments:

1. **Complete the analytical run** and verify results on the instrument software
2. **Export raw data** from the instrument software to CSV format (or the most open/portable format available)
3. **Name the file** following the convention: `[instrument]_[method]_[YYYY-MM-DD]_[batch].[ext]` (e.g., `icpoes_metals_2026-05-21_batch01.csv`)
4. **Transfer the file** to the project's `data/raw/` directory using USB or network transfer
5. **Verify the transfer** by opening the file and checking row counts, column headers, and a few spot-checked values against the instrument display
6. **Record the export** in the lab notebook: file name, instrument, date, number of samples, and any export settings
7. **Commit to Git** with a descriptive message

Critical Rule

Never modify raw instrument files after export. If corrections are needed, create a new file in `data/processed/` and document the changes in your processing script. The raw file is your permanent record.

7.2. Instrument-Specific Export Instructions

7.2.1. Shimadzu UV-Vis (UV-1900i / UVProbe)

1. In UVProbe, open the data file (`.spc`)
2. File > Save As > CSV
3. Select “Wavelength and Absorbance” columns
4. Save to the USB drive as `uvvis_[method]_[YYYY-MM-DD].csv`

7.2.2. Thermo iCAP ICP-OES (Qtegra)

1. In Qtegra, navigate to the completed sequence
2. Right-click the sequence > Export Results
3. Select CSV format; include sample IDs, element concentrations, and QC flags
4. Save to the USB drive as `icpoes_[method]_[YYYY-MM-DD]_[batch].csv`

7.2.3. Shimadzu TOC-L (TOC-Control)

1. In TOC-Control, open the sample table for the completed run
2. File > Export > CSV
3. Include sample name, injection number, area, concentration, and CV
4. Save as `toc_[method]_[YYYY-MM-DD].csv`

7.2.4. YSI ProDSS Field Sensor

1. Connect the ProDSS to a computer via USB
2. Open KorDSS software
3. Select the data log for the sampling event
4. Export as CSV, including date/time, site ID, and all parameters
5. Save as `ysi_[project]_[YYYY-MM-DD].csv`

7.2.5. Bio-Rad QX200 ddPCR (QuantaSoft)

1. In QuantaSoft, open the completed experiment
2. Options > Export CSV (Well Data or Cluster Data)
3. Include well ID, sample name, target, concentration, positive/negative droplet counts
4. Save as `ddpcr_[target]_[YYYY-MM-DD].csv`
5. Also export the amplitude data for gating verification

7.3. File Transfer to RStudio Server

After exporting to USB or local machine:

```
# Option 1: scp from local machine to RStudio Server
scp icpoes_metals_2026-05-21_batch01.csv \
  username@rstudio.pomona.edu:~/project/data/raw/

# Option 2: Use the RStudio Files pane
# Click Upload > Browse > select file > OK
```

7.4. Automated Import Scripts in R

Create a standardized import script for each instrument type. This ensures consistent parsing across analysts and batches:

```
#' Import ICP-OES data from Qtegra CSV export
# '
#' @param file_path Path to the exported CSV file
#' @return A tibble with standardized column names
import_icpoes <- function(file_path) {
  raw <- read_csv(file_path, skip = 2) # Skip Qtegra header rows
```

```
raw %>%
  rename_with(tolower) %>%
  rename_with(~ gsub(" ", "_", .)) %>%
  mutate(
    date_analyzed = as.Date(date_analyzed, format = "%m/%d/%Y"),
    across(where(is.character) & !c(sample_id, sample_type),
           as.numeric)
  ) %>%
  # Flag values below detection
  mutate(across(
    where(is.numeric),
    ~ if_else(. < 0, NA_real_, .)
  ))
}

# Usage
icp_data <- import_icpoes("data/raw/icpoes_metals_2026-05-21_batch01.csv")
glimpse(icp_data)
```

```
#' Import TOC data from TOC-Control CSV export
#'
#' @param file_path Path to the exported CSV file
#' @return A tibble with standardized column names
import_toc <- function(file_path) {
  raw <- read_csv(file_path)

  raw %>%
    rename_with(tolower) %>%
    rename_with(~ gsub("[^a-z0-9]", "_", .)) %>%
    select(sample_name, injection, area, concentration_mg_l = conc,
           cv_percent = cv) %>%
    filter(!is.na(sample_name))
}
```

7.5. Data Integrity Checks

After importing, run basic integrity checks:

```
# --- Row count matches instrument ---
cat(sprintf("Rows in file: %d\n", nrow(icp_data)))
# Compare with number of samples in lab notebook

# --- No fully blank rows ---
stopifnot(!any(rowSums(is.na(icp_data)) == ncol(icp_data)))

# --- Sample IDs match the run sequence ---
```

```
expected_ids <- c("Blank", "Std1", "Std2", "Std3",
                 paste0("S", sprintf("%03d", 1:20)),
                 "CRM", "Dup")
setdiff(expected_ids, icp_data$sample_id)
# Should be empty if all expected samples are present

# --- Calibration standards in expected range ---
icp_data %>%
  filter(grepl("^Std", sample_id)) %>%
  select(sample_id, starts_with("fe_")) %>%
  print()
```

7.6. Archival and Backup

All raw instrument files must be archived in three locations:

Location	When	Responsible
GitHub repository (data/raw/)	Immediately after export	Analyst
Lab instrument PC	Retained on instrument	Lab Manager
External backup drive	End of each semester	Lab Manager

Tip

Large binary files (e.g., raw .spc or .qlp files that exceed GitHub's file size recommendations) should be stored on the external backup drive and the network share, not in Git. Only the exported CSV goes into the repository.

7.7. Instrument Log Cross-Reference

Every data export event must be recorded in the instrument log (see [SOP-L007](#)). The log entry should include:

1. Date and time of the analytical run
2. Analyst name
3. File name(s) exported
4. Number of samples (including QC samples)
5. Any instrument warnings or anomalies noted during the run
6. Cross-reference to the lab notebook page number

8. Quality Control and Acceptance Criteria

- Raw files must be unmodified after export (verify with file hash if needed: `md5sum filename.csv`)

- File names must follow the naming convention in Section 6.1
- Row counts and sample IDs must match the run sequence recorded in the lab notebook
- Calibration data in the export must match the calibration curve approved during the run
- Refer to [SOP-L802](#) for analytical QC criteria (blanks, duplicates, recoveries)

9. Waste Disposal

Not applicable (computational procedure).

10. References

Gandrud, C. (2015). *Reproducible research with R and RStudio* (2nd ed.). CRC Press.

11. Training Record

Training is documented in the EA Program training log. Before exporting instrument data independently, personnel must:

1. Read this SOP in full
2. Complete training on the specific instrument (see the instrument-specific SOP)
3. Demonstrate a complete export, transfer, verification, and commit workflow under supervision
4. Be signed off by the PI or lab manager

12. Appendix A: File Hash Verification

To confirm that a file has not been modified after export, compute and record an MD5 hash:

```
# On Linux / Mac / RStudio Server terminal
md5sum data/raw/icpoes_metals_2026-05-21_batch01.csv
# Record this hash in the lab notebook alongside the file name
```

```
# In R
tools::md5sum("data/raw/icpoes_metals_2026-05-21_batch01.csv")
```

13. Appendix B: Instrument Data Format Quick Reference

Instrument	Header Rows to Skip	Delimiter	Encoding	Date Format
Shimadzu UV-Vis	1	Comma	UTF-8	MM/DD/YYYY

Instrument	Header Rows to Skip	Delimiter	Encoding	Date Format
Thermo iCAP	2	Comma	UTF-8	MM/DD/YYYY HH:MM
Shimadzu TOC-L	0	Comma	Shift-JIS (convert to UTF-8)	YYYY/MM/DD
Dionex ICS	3	Comma	UTF-8	MM/DD/YYYY HH:MM
YSI ProDSS	0	Comma	UTF-8	MM/DD/YYYY HH:MM
Bio-Rad QX200	0	Comma	UTF-8	YYYY-MM-DD
PANalytical XRF	Variable	Tab or comma	UTF-8	DD/MM/YYYY
