# DIVER APPLICATION
# ENVIRONMENTAL DATA SPECIFICATION

OVERVIEW

This document serves as an environmental data specification with details on the Data Categories, detail fields, and valid values for NOAA's DIVER (Data Integration Visualization Exploration and Reporting) application.[1] Though this data specification does not serve as an explicit template for data exchange, it highlights the DIVER common data models and data exchange methods. The DIVER application is an on-line data warehouse and set of query tools which contains environmental and project data collected in support of NOAA's role as a trustee for natural resources on cases and projects across the country. In addition to standardizing and integrating environmental data from many sources and providers, the DIVER application is the central environmental data management tool for NOAA's involvement in Natural Resource Damage Assessments (NRDA), including data collected and analyzed in response to the Deepwater Horizon Oil Spill. As part of the collection and distribution of data related to the Deepwater Horizon oil spill, NOAA is preparing the DIVER application to accommodate Gulf of Mexico restoration monitoring data.

v. 1.3

August 9, 2019

DIVER Development Team

---

[1] DIVER: https://www.diver.orr.noaa.gov. Developed by NOAA's Office of Response and Restoration in partnership with NOAA's Restoration Center.

# 1. DATA CATEGORIES

The DIVER data warehouse currently contains seven "Data Categories" for environmental data. Data Categories are overarching data models or database schemas for specialized environmental data types. Current data categories are described in Table 1, with details and tables in **Appendix 1a** and **Appendix 1b**.

**Table 1 Data Categories Description**

| DATA CATEGORY | DESCRIPTION |
|---|---|
| **Samples** | Environmental data typically from field collected samples and results from laboratory experiments, including contaminant chemistry, biological, and compositional analyses.[2] |
| **Field Observations** | Observational and ecological/biological monitoring and field measurements. |
| **Photos** | Images and their keywords and metadata, which may include geographic location. |
| **Telemetry** | Spatial/Geographic tracking data on the movement of tagged animal species (e.g., fish, marine turtles, and marine mammals). |
| **Instruments** | Information about devices (e.g., CTD Rosette; remotely operated vehicle; etc.) used to collect environmental data during a field study. |
| **Bioassay** | Toxicity data from studies conducted in a laboratory to determine the effects of contaminants or other stressors on specific species. DIVER contains toxicity data on field collected material and laboratory-only toxicity tests. |
| **Shoreline Cleanup and Assessment Technique (SCAT)[3]** | Survey data collected during an oil spill response effort to document shoreline oiling conditions. Note: this data category contains no publicly available data. |

# 2. CORE FIELDS

Each Data Category contains Core Fields, which provide high-level descriptions of key information and are consistent across different types of data. Core fields are used to match common values (or controlled vocabulary) across Data Categories. The core fields typically contain information about environmental sampling, measurements, observations, or instruments, which allow for users to search or query for all information related to a study, workplan, or data source. The core fields listed in Table 2

---

[2] Analytical chemistry data was previously organized into NOAA's Query Manager (QM) Database and accessible through the QM desktop application. All of the data has been ingested into DIVER and is accessible online through the DIVER Explorer tool. Fields and/or definitions in the Samples cube may reference Query Manager or QM.

[3] More information on SCAT: http://response.restoration.noaa.gov/oil-and-chemical-spills/oil-spills/resources/shoreline-cleanup-and-assessment-technique-scat.html

are generally required for data brought into DIVER, and are divided into User-Provided fields and DIVER-Created fields.[4] Some core fields (such as Region) may be identified from metadata.

### Table 2. Core Fields for Environmental Data Categories

| FIELD NAME | FIELD DEFINITION | FIELD SET WITHIN DIVER EXPLORER[5] |
|---|---|---|
| *User-Provided Fields* | | |
| Case/Activity | The name of the case incident or the activity used to collect data. | Case/Activity Overview |
| Collection Workplan | The workplan under which the field data were collected. | Case/Activity Overview |
| Region | Region | Case/Activity Overview |
| Workgroup | The Technical Working Group under which the field data were collected. | Case/Activity Overview |
| Workplan Topic Area | The main resources of focus of a Collection Workplan. | Case/Activity Overview |
| Workspace Name | Name of the Portal Workspace where data were entered. | Case/Activity Overview |
| Collection Form | The type of the data submission form used by the field team to submit raw field data. | Collection Summary |
| Collection Study Name | The name of the study under which the field data were collected. | Collection Summary |
| Data Category | General category of data collection (e.g., Samples, Bioassay, Field Observations, Instruments, Photographs, Telemetry, or SCAT). | Collection Summary |
| Data Classification | The purpose for which data was collected within the case incident or activity. | Collection Summary |
| Data Source | The originating owner of the dataset. | Collection Summary |
| Lead Sampler | Primary contact (name) for individual field questions. | Collection Summary |
| Source Type | General owner/source of the data (e.g., NRDA, Response, Responsible Party). | Collection Summary |
| Collection Matrix | The type of sample or record collected (e.g., Sediment, Water, Photograph, Wipe). | Field Data |
| Sample ID | Unique ID assigned to each sample by the field sampler. | Field Data |
| Station Description | Description of a station or grouping (by reach, section of river, embayment). | Field Data |

---

[4] For datasets that are referential (linked), or where core fields are unknown, incomplete records may be brought into DIVER.

[5] Field Set is assigned to each field to facilitate querying in DIVER Explorer by grouping similar fields together (e.g., all fields related to location, date, and time).

| FIELD NAME | FIELD DEFINITION | FIELD SET WITHIN DIVER EXPLORER[5] |
|---|---|---|
| Station/Site | Station or site identifier. This is often defined by the workplan and/or recorded by the field team, but may be standardized to database requirements. | Field Data |
| Area Description | General description or name of the area sampled (e.g., a particular reef or marsh). | Location/Date/Time |
| County/Parish | The County or Parish where the field collection was completed. | Location/Date/Time |
| Date | Data collection date, as year, month, and day. | Location/Date/Time |
| End Latitude | End Latitude in decimal degrees for trawls, transects or samples collected in a line. | Location/Date/Time |
| End Longitude | End Longitude in decimal degrees for trawls, transects or samples collected in a line. | Location/Date/Time |
| Start Latitude | Sample Latitude in decimal degrees, or start Latitude for samples collected by trawl, transect or in a line. | Location/Date/Time |
| Start Longitude | Sample Longitude in decimal degrees, or start longitude for samples collected by trawl, transect or in a line. | Location/Date/Time |
| State | The state where the field event took place. | Location/Date/Time |
| Analysis Category | General category of analysis performed (e.g., Plankton - Nekton, Visual Observation, Contaminant Chemistry). For additional detail, see Analysis Type and/or Analysis. | Results: All Data Types |
| Analysis Status | Status of samples in the analysis process as reported by laboratories or through results (e.g., Archived, Results Available, In Analysis Queue etc.). | Results: All Data Types |
| Analysis Type | Subcategory (i.e., type) of analysis performed, such as Biomass, Hematology, Genetics, etc. For additional detail, see Analysis. | Results: All Data Types |
| Review Status | Extent of data quality review performed. | Results: All Data Types |
| Sharing Status | Identifies extent of data distribution (e.g., Publicly Available). | Results: All Data Types |
| *DIVER-Created Fields* | | |
| Region ID | Region ID | Case/Activity Overview |
| Station Group | Predefined sets of grouped stations/locations | Case/Activity Overview |
| Workspace ID | Workspace ID | Case/Activity Overview |
| DIVER Dataset | DIVER's internal database table name | Collection Summary |
| File Collection ID | Record identifier for the corresponding DIVER file collection. | Collection Summary |

| FIELD NAME | FIELD DEFINITION | FIELD SET WITHIN DIVER EXPLORER[5] |
|---|---|---|
| Record ID | Identifier for each observation data sheet entered into the DIVER database. | Collection Summary |
| Trip ID | Identifier for tracking field collection events and the way data files were provided to the Data Management Team (one Trip ID per file collection or zip file). | Collection Summary |
| Component ID | Auto-generated unique component identification number. | Overview |
| Project ID | Auto-generated unique project identification number. | Overview |
| Image Id | Record identifier for a particular photograph. | Results: All Data Types |
| Link to Related Files | Link to source files for related data | Results: All Data Types |
| Photo URL - Midsize | Mid-sized image | Results: Photographs |
| Photo URL - Original | Original image | Results: Photographs |
| Photo URL - Thumbnail | Thumbnail sized image | Results: Photographs |
| QM Site ID | Identifier for a site in the Query Manager database | Results: Samples |

3. **DETAIL FIELDS**

Each Data Category contains dataset-specific fields in addition to the Core Fields. All fields (i.e., both Core Fields and Detail Fields) are grouped into Field Sets to help with data access and querying. Field Sets are groupings of fields into topics such as Overview, Location/Date/Time, and Results; fields that are output-only (not queryable) will not have a field set. Table 3 summarizes the number of fields in each Data Category, grouped by Category. Appendix 1a provides details about the Data Categories, core fields and detail fields, including specifications of primary keys in DIVER for different data categories. Appendix 1b contains spreadsheet tables that describe all fields for a specific Data Category. The tables include field definitions, a designation of Core Field or Detail Field, the field name as reported in comma-delimited and shape file outputs, and whether the field has a list of valid values. The concept of valid values is discussed in Section 5 of this document and in Appendix 2, which also contains definitions of valid values that are presented as codes within certain fields.

**Table 3. Count of fields in each Data Category, by Field Set**

| Field Set | Samples | Field Observations | Instruments | Telemetry | Photographs | Bioassay | SCAT |
|---|---|---|---|---|---|---|---|
| Case/Activity Overview | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| Collection Summary | 13 | 11 | 15 | 11 | 11 | 11 | 11 |
| Field Data | 7 | 10 | 4 | 6 | 4 | 5 | 9 |
| Location/Date/Time | 16 | 14 | 8 | 14 | 13 | 12 | 14 |
| Overview | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| Results: All Data Types | 10 | 9 | 7 | 8 | 7 | 7 | 8 |
| Results: Bioassay | 0 | 0 | 0 | 0 | 0 | 4 | 0 |
| Results: Instruments | 0 | 0 | 7 | 0 | 0 | 0 | 0 |
| Results: Observations | 0 | 3 | 0 | 0 | 0 | 0 | 0 |
| Results: Photographs | 3 | 3 | 3 | 3 | 7 | 3 | 3 |
| Results: Samples | 41 | 7 | 1 | 1 | 1 | 8 | 4 |
| Results: Telemetry | 0 | 0 | 0 | 2 | 0 | 0 | 0 |
| Taxonomic Categories | 18 | 18 | 0 | 17 | 0 | 18 | 0 |
| No Field Set (Not queryable, available in outputs only) | 21 | 5 | 6 | 5 | 5 | 31 | 15 |
| **Total Fields** | **141** | **92** | **63** | **79** | **60** | **111** | **76** |

*Notes.* See **Appendix 1a** and **Appendix 1b** for more information about the fields related to each Data Category. See **Appendix 3** for an expanded version of this table which presents a comparison of the fields related to particular data categories. The numbers here represent the total number of fields within each category.

### 4. SPATIAL ATTRIBUTES

The DIVER data warehouse has start and end coordinates (start latitude/end latitude and start/end longitude) integrated into records for each Data Category. Spatial attributes are not required for data integration but are required for display on the map in the DIVER Explorer query tool. Data with no spatial attribution will still be available in the query result and in the Table tab, but using a spatial filter will remove these results from the query. In addition to point coordinates, DIVER has the ability to natively store more complex geometries such as multipoint, multiline and polygons, however the data warehouse does not currently store polygons. Coordinates are natively held in Decimal Degrees, Horizontal Datum NAD83 and presumed Vertical Datum NAVD88.

### 5. VALID VALUES

Each of the DIVER Data Categories use standards, conventions, and values that are specific to the type of data.[6] For example, the Samples and Bioassay data categories are standardized based on the legacy Query Manager database structure. In addition, DIVER standardizes the use of -999.99 and -9 to indicate a particular meaning across data categories; -999.99 indicates that a value should not be expected (i.e., an analysis was not conducted), while -9 indicates that a value was expected but not received (i.e., a missing value). In order to assist the data provider in understanding which fields have associated valid values in DIVER, Appendix 2 (.xlsx) contains lists of valid values for DIVER fields with standardized values, such as Analysis Category and Collection Matrix. A dictionary of chemical names and codes is also included.

Currently, DIVER addresses the standardization of species names by referencing the Integrated Taxonomic Integration System (ITIS),[7] which contains authoritative taxonomic information on plants, animals, fungi, and microbes of North America. As needed, data managers may discuss species name standardization with data providers and data users.

The DIVER team continues to work with partners to build and update our vocabulary of valid values. The DIVER team works with data providers to ensure that valid values (or suggested values when appropriate) are relevant to their project and used to support effective data querying within DIVER. This may include working with data providers to modify valid values lists to include values not currently used in DIVER to better accommodate project-specific needs and terminology.

### 6. DATA SUBMISSION

The DIVER Data Warehouse can accept both structured and unstructured data as well as related files. Below we describe the processes by which DIVER accepts data, preferred data file formats, information to include with each data file, and best practices for generating a data file that matches DIVER data structures.

---

[6] The explicit type or format of a field when it is ingested by DIVER is also important. All fields, with the exception of date, are ingested as text. If a field exclusively contains numeric values, such as for the "Analysis Result" field, the field and values are converted to a numeric format on ingest.

[7] https://www.itis.gov/

PROVIDING DATA TO DIVER

NOAA typically uses File Collections within the DIVER Portal application (login required) to gather both data (e.g., spreadsheet or database files that match DIVER structures or contain sufficient information to allow DIVER data managers to transform the data) and related files (e.g., scanned field forms, GPS, field notes, photos, and other related files). All structured data files should be in one of the following formats: Microsoft Excel (XLS/XLSX), comma separated values (CSV), or database format (e.g., Microsoft Access). Unstructured data for inclusion to DIVER should be converted into an electronic data file in the specified format prior to upload; the following sections of this document contain guidance. Information contained in related files should be clearly linked to the data in the electronic data file, preferably by date and by *Location* identifier (site, station, transect, etc.) or by *Sample* identifier.

NOAA OR&R has created Electronic Data Deliverables (EDDs), also known as Data Templates, for contaminant chemistry and bioassay data which are used by laboratories and data partners to provide integrated field and laboratory data in a standardized manner. The DIVER team has also developed templates for delivery of standardized biological data, field observations, and non-analytical chemistry sample results. The most recent versions of these templates are available for download from the public DIVER **Forms and Templates** webpage.

The DIVER data team also works with data providers to create ETL procedures (Extract, Transform, and Load) from their data management system or data store as needed. DIVER data managers typically work with data providers and other data specialists to review and determine the best course of action to integrate the data into the DIVER data warehouse. This process can vary in complexity depending on the format of the incoming data.

PREFERRED DATA FORMATS

All environmental data uploaded to DIVER should be formatted according to one of the following formats. Data providers with structured non-tabular data (e.g., photos, video) or continuous monitoring data should contact NOAA to determine the best format. Metadata should be included with all environmental data, following the standards described below.

- **Format 1 (preferred):** Structured data that matches an available DIVER Data Template.
- **Format 2**: Structured data that does not match a DIVER template. The data file should be formatted into a single header row that displays variables as columns, with rows containing a single record of raw data (i.e., each row only contains one measured parameter,[8] outside of general descriptive fields like location and depth). For example, if at a given site on a given day, the count of recreational anglers, the count of recreational boaters, and the weather conditions were all noted, there would be three rows of data for that site/day.
- **Format 3:** Relational database with linked tables (e.g., a stations table, a sampling information table, and a results table). These must be relational databases and an entity-relationship

---

[8] In this document, the terms parameter and analysis are used interchangeably. The "Analysis" field is used to describe measured parameters.

diagram must be included.
- **Format 4:** Pivot table style. Each row represents a given location and date, and all parameters are included on the same row in individual columns. Using the previous example, there would be a single row of data for that site/day, with columns for count of recreational anglers, count of recreational boaters, and weather conditions.

BEST PRACTICES FOR SUBMITTING DATA FILES

The level, or grain, of data that are uploaded should be determined on a project-specific basis by data providers and NOAA data managers. Prior to upload to DIVER, all data should undergo the quality control and validation steps identified in work plans to ensure data quality objectives are met. Each dataset should provide contextual information and fully describe the results, including:

- **Where** the data were collected (station name, transect ID, coordinates in decimal degrees, coordinate system, upper and lower depth, depth units);
- **When** the data were collected (date and time if relevant, may also include tidal cycles or other relevant timing information – a separate analysis date may also be relevant);
- **How** the data were collected (collection method and equipment or gear type used – laboratory data may also include preparatory and analytical methods);
- **What** data were collected (parameter name, parameter units, indication of data quality and/or if the result was excluded from further analysis);
- Supporting information (e.g. species by scientific and common name, other visual observations, temperature); and,
- Information on **who** collected the data and **why**.

*Metadata*

DIVER creates spatial metadata for export with all data queries, which describes the specific query run and metadata at the DIVER (system) level. However, many archiving and data sharing protocols (e.g., NOAA's NCEI or data.gov) require metadata for individual datasets compiled to a database or dataset, rather than for the system as a whole. The exact requirements for these systems may vary, but it is anticipated that the project and spatial metadata described below can be used to systematically create the required formats.

Ideally, data submitted to DIVER should have metadata compliant with the ISO 19915-2 or FGDC CSDGM metadata standards. [9, 10] However, we often work with data submissions that do not meet this standard for various reasons. Metadata accompanying data submissions should include the following minimum information, and may be included in a separate ReadMe tab within the data file:

1. **File Name:** The actual name of the Shapefile, Excel spreadsheet, Access database, or other file type (ex. FWconcV3_tPAH.shp). This information will help Data Managers identify metadata associated with data when there are multiple datasets submitted.

---

[9] https://www.iso.org/standard/39229.html
[10] https://www.fgdc.gov/metadata/csdgm-standard

2. **File Title:** The name to be displayed in DIVER for the dataset or study (e.g., Freshwater tPAH Concentrations).
3. **Attribute Fields:** List the field names that represent the information being documented. Provide field definitions and the units where applicable.
4. **Collection Matrix:** The type of sample record collected (e.g., Sediment, Water, Photograph).
5. **Keywords:** Any Keywords that the data should be tagged with to make it easier to be searched. Keywords should be selected from established keyword lists such as the [Global Change Master Directory](#) or other established keyword lists. Additional keywords can also be used if the keyword is not part of one of the established lists.
6. **Purpose:** Describe the purpose of the data and why it was collected/created.
7. **Description:** Describe the data and what information it contains. Provide enough information so that a user has an understanding on the dataset and information.
8. **Data Source:** Source of the data. This should include the organization that collected the data. Also include any data caveats and information on data lineage and secondary uses. Information on where the original data resides or which study the data were collected under is useful for tracking data back to the source.
9. **Time Period:** Time period for the dataset.
10. **Contact Information (for internal use only):** PI contact information so that Data Managers can connect regarding questions with the data.
11. **Additional Comments/Notes:** Add any additional information about the data not captured in previous sections.

*Spreadsheet contents*

Where possible, data providers should follow the best practices outlined below to ensure each structured data file is in an easily understandable format with sufficient level of detail to be interpreted and transformed into the DIVER data structure. This is particularly important for data files that do not match an existing DIVER data template (e.g., Formats 2, 3, and 4).

● When defining values, refer to the appendices of this Data Specification to determine if a value already exists. If you describe a new value, choose a simple name. To indicate spacing, underscores "_" are preferred to hyphens "-" (hyphens correspond to several different characters depending on dash length). CamelCase may also be used to clarify long variable names.
● Report one piece of information in each cell. For example, do not include both results and units in the same cell (e.g., 3 g/mL). Instead, create separate columns for Results and Units.
● Date variables should contain a single date. To represent a date range, include separate columns for Start Date and End Date. Dates should be formatted either as MMDDYYYY or in three columns: Year, Month, and Day. This avoids confusion with auto-formatting and changing cell type.
● Be consistent in the words you use and their expression. For example, "Percent Cover" and "percent cover" and "% cover" and "%Cover" and "percent cover –veg" should be standardized to one value within the data file. Data managers prefer "percent cover" because it avoids capitalization *and* special characters and is not overly specific.

- Avoid special characters in variable names, units, and values.
- Include one data table per spreadsheet tab. All rows and columns of data should be contiguous. Graphics and charts should not be included in tabs that contain raw data.
- Avoid merged or blank cells. Write "Not Available" or a defined blank value in each blank cell.
- Do not use font type, font color, or highlighting to explain the data. Instead, options include adding another column to explain the concept you are attempting to differentiate or adding a notes column.
- Remove all extra spaces within a cell. For example, "Restoration " has an extra space after the word that should be deleted. Cells with a single space but otherwise no content should be replaced with "Not Available" or a defined blank value as indicated above.

Each dataset should include a data dictionary that explains each column of data and any controlled vocabularies that are used within the dataset. Definitions of data fields as well as information about what the codes within those fields represent and how values were calculated are needed in order to provide sufficient information for the data managers to integrate the data into DIVER, and for the end users to access the data and compare across datasets through queries in DIVER Explorer. Data dictionaries should contain an explanation of what each variable means; the measurement units; the collection method, if relevant; and if a controlled vocabulary or rule exists for this variable. For example, if a variable called "Environmental_Conditions" only has two permitted values, "wind temperature" and "air temperature," the data dictionary should explain that these are the only accepted values as well as define the values. Data dictionaries should indicate if a particular variable is calculated from other variables or from information not presented in the data file. When possible, data providers should review the field definitions and standardized valid values described within this document and its appendices for agreement. **Appendix 1b** presents field definitions and is organized by Data Category. **Appendix 2** presents valid values and guidance on standardized fields including field definitions for each of the DIVER Data Categories.

When applicable and possible, study notes should be included with the data submission. Study notes contain additional information regarding the dataset that may not be captured in the metadata and are common practice in DIVER for certain data types such as analytical chemistry data. Information captured in study notes for analytical chemistry data typically includes data source, contact information, station information, sample information, replicates information, summed parameters information, data caveats, known issues with the data, data restrictions, and other information that would be helpful to both the data manager and end user. NOAA provides a template on public DIVER's **Forms and Templates** page to help create study notes as part of the Samples and Bioassay data templates. Data managers review study notes and may add information to capture the edits and standardizations made to submitted data.

## 7. DATA ACCESS

The DIVER website provides access to data organized by region through the DIVER Explorer query tool. Users can search, filter, and download data using the DIVER Explorer query tool or view maps and spatial data in ERMA, NOAA's mapping and visualization application. Some activities and cases that NOAA works on are also highlighted in the Recent Datasets of Interest section on the DIVER website. The help materials provide guidance and examples of how to use the DIVER Explorer query tool.

DIVER Explorer provides three primary export options for a query under the Data and Export tab: (1) the user-created data table (which displays the columns and filters the user has chosen to display in the data table), (2) the complete data table with all fields used in the query, and (3) the full export package, which contains the complete data table including additional related fields, a readme file, and related supplemental files. Each of these data export options provide the user with a spreadsheet in CSV format and an ISO 19115-2 compliant metadata record in XML and HTML. Depending on the type of data queried, supplemental files that define qualifier codes, study notes, and study references may be included. For example, chemistry data outputs typically include a study notes file with more details about analytical methods, data treatment, and data caveats; a file that defines relevant qualifier codes by study; and a file that includes citations to study references. DIVER Explorer can also provide spatial data outputs in shapefile (SHP) and KML format including ISO 19115-2 metadata.

## 8. REFERENCES

OMB Executive Order M-13-13- The Open Data Policy- Managing Data as an Asset (and related)
http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf

NOAA Administrative Order 212-15- NOAA's Environmental Data Management Policy (and related)

http://www.corporateservices.noaa.gov/ames/administrative_orders/chapter_212/212-15.html

NOAA Environmental Data Management Procedural Directives

https://www.nosc.noaa.gov/EDMC/PD.all.php

ISO 19115 Geographic Information - Metadata

https://geo-ide.noaa.gov/wiki/index.php?title=ISO_19115_Geographic_Information_%E2%80%93_Metadata

ISO 19115-2 (Geographic Information - Metadata Part 2)

https://geo-ide.noaa.gov/wiki/index.php?title=ISO_19115-2_(Geographic_Information_%E2%80%93_Metadata_Part_2)

## 9. APPENDICES

Appendix 1a.  Information and description of DIVER Data Categories and Fields

Appendix 1b.  Spreadsheet with all DIVER Fields, definitions and overview of values by Data Category (.xlsx)

Appendix 2.  Valid Values for standardized fields and Chemical Dictionary (.xlsx)

Appendix 3.  Comparison of DIVER Fields across Data Categories