



DATA INTEGRATION PLATFORM

Data Harmonization for Action

80% of data scientists' (and scientists') time is spent on data preparation. This generally accepted rule of thumb is a major reason why we are not yet seeing the results we need from AI and data science projects. Data isn't accessible. Or it is riddled with errors. Or the files/fields are in different formats. Consolidating your data into a central location is only one piece of the puzzle. Harmonizing data from disparate sources is critical to making it actionable via exploration and analysis.

Automate experimental data harmonization

Flexible

tailor data schema based on your needs and preferences

Productized

compatible with Allotrope + purpose-built for life sciences



Scalable

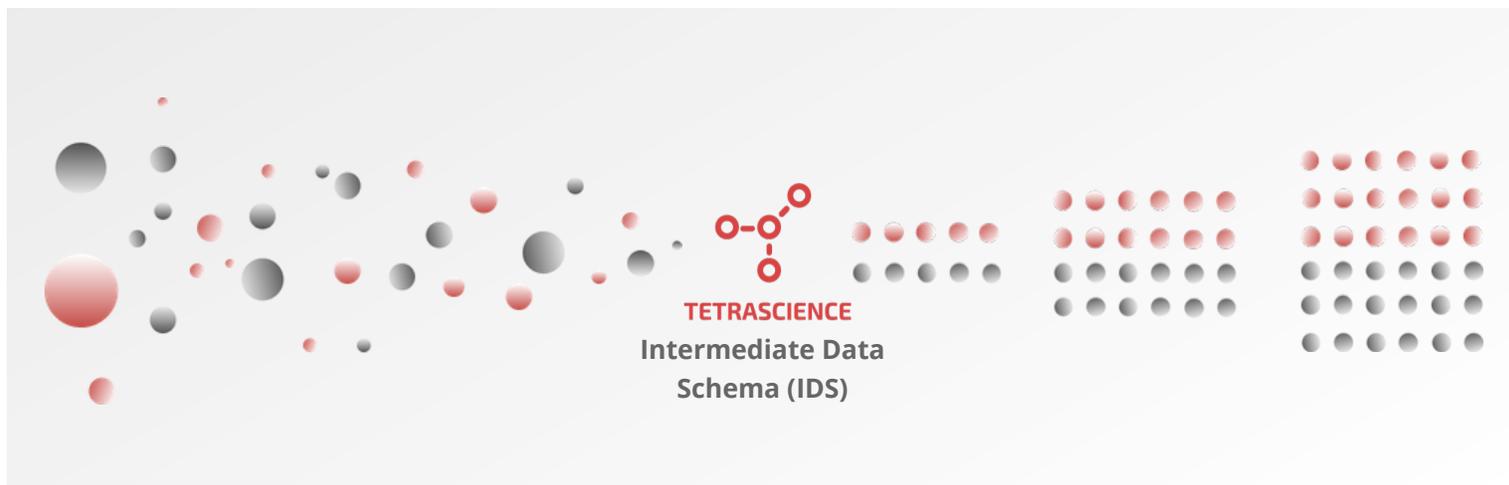
compatible with web applications and data science

Secure

retains raw data and audit trail for reproducibility

TETRASCIENCE INTERMEDIATE DATA SCHEMA (IDS)

Instrument data. CRO/CDMO data. ELN data. Moving all your raw data to a central location only solves part of the problem. Date and time formats are still different. Instrument readouts are reported with a different number of digits after the decimal point. Don't spend any more time manually aligning data. The TetraScience Intermediate Data Schema (IDS) serves as the universal adapter, automatically harmonizing all of your experimental data.



Built for compatibility

Designed to be vendor agnostic and future-proof:

- readable by both humans and machines
- compatible with the world of web application and data science
- leverages JSON schema, a well-defined and popular framework to enforce structure and data integrity
- defines the derivative SQL tables and indexing rules
- provides flexible references to large data sets

Data includes

as much data as possible, such as:

- user
- system
- method
- results
- datacubes
- logs
- vendor RAW file

Enables data consumption options

Use the TetraScience IDS as-is, or further transform to your schema/ontology of choice. You can even define your own IDS. Once data is in the IDS, the hardest part is done. Now you can focus on what to do with your data.

- Compatible with Allotrope Foundation ontology
- Supports knowledge graphs and RDF graphs

ADVANCED DATA MODELING AND STANDARDIZATION

Here are further examples of how the TetraScience IDS handles complex data

Data cubes for multi-dimensional matrices

```

{
  "datacubes": [{
    "name": "3D chromatogram",
    "description": "More information (Optional)",
    "channel_id": "123412",
    "another property": "you can decide",
    "measures": [{
      "name": "intensity",
      "unit": "ArbitraryUnit",
      "value": [
        [111, 222, 333, 444, 555],
        [111, 222, 333, 444, 555],
        [111, 222, 333, 444, 555]
      ]
    }],
    "dimensions": [{
      "name": "wavelength",
      "unit": "Nanometer",
      "scale": [180, 190, 200]
    }, {
      "name": "time",
      "unit": "MinuteTime",
      "scale": [1, 2, 3, 4, 5]
    }
  ]
}]
}

```

Example	Measure	Dimension
HPLC	1. Detector intensity	1. Wavelength 2. Retention time
Raman Spectroscopy	1. Intensity	1. Wavenumber shift
Plate Reader	1. Absorbance 2. Concentration	1. Row position 2. Column position
Mass Spectrometer	1. Intensity	1. Mass charge ratio 2. Time

Large multi-dimensional arrays or images

```

{
  "result": {
    "cell": {
      "viability": {
        "value": 0.2,
        "unit": "Percent"
      }
    }
  },
  "image": {
    "type": "s3file",
    "key": "s3://bucket/image1.png",
    "version": "dWS9o0BR5l1SYp4n...",
    "checksum": "2e453f52..."
  }
}

```

	Key Use Case	Comments
Parquet	Big data, data science, cloud computing	Data science + big data applications
HDF5	Group disparate files in one file	E.g. images + excel files together
Instrument raw format	Vendor specific software or analysis tools. Importing files back into vendor software	Keep images and instrument binary files in original format for domain specific tooling and ecosystem

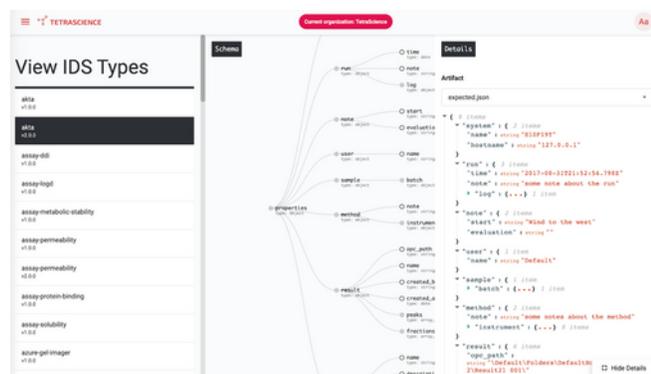
ADDITIONAL INFORMATION

Tell me more about how to build my own Intermediate Data Schema

Build your own data model and we will handle the "plumbing" for you.

The TetraScience Data Integration Platform:

- performs automated validation to ensure the schema is designed properly, without error. For example, not using multiple data types for the same field, incorrect IRI format, etc.
- manages the IDS, keeping track of different versions
- provides a view of the IDS structure; share and collaborate with your colleagues
- automatically indexes your files such that you can search every data field inside IDS
- enables customized search queries based on IDS



Now that your
experimental data is
accessible and harmonized,
the only question left is...
what will you do with it?



TetraScience is the leader in transforming the Digital Lab. We provide advanced data engineering capabilities that make life sciences R&D data truly accessible and actionable. More than 80 leading pharmaceutical and biotech companies rely on our cloud-native Data Integration Platform and Lab Monitoring System application to automatically centralize and harmonize their experimental data, preparing it for data science + AI, detecting anomalies, and pushing alerts. Activate the flow of your data.

www.tetrascience.com
hello@tetrascience.com
@TetraScience  

294 Washington St.
11th Floor
Boston, MA 02108