onward

How to be a data science rock star in the geoscience world

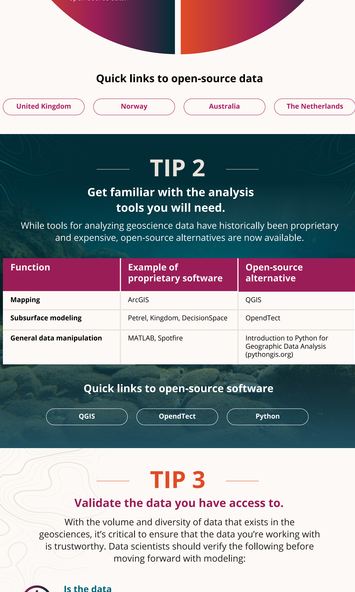
Tips and tricks to come to grips with geoscience data

Analyzing and interpreting geoscience data will be a critical requirement to accelerate the energy transition. This tip sheet is designed to guide data scientists who are new to understanding, validating, and working with geoscience data.

4 TIPS TO "GEOSCIENCE UP" YOUR DATA SCIENCE

TIP 1

Get familiar with the data formats you will need to work with.



Quick links to open-source data

United Kingdom | Norway | Australia | The Netherlands

TIP 2

Get familiar with the analysis tools you will need.

While tools for analyzing geoscience data have historically been proprietary and expensive, open-source alternatives are now available.

Function	Example of proprietary software	Open-source alternative
Mapping	ArcGIS	QGIS
Subsurface modeling	Petrel, Kingdom, DecisionSpace	OpendTect
General data manipulation	MATLAB, Spotfire	Introduction to Python for Geographic Data Analysis (pythongis.org)

Quick links to open-source software

QGIS | OpendTect | Python

TIP 3

Validate the data you have access to.

With the volume and diversity of data that exists in the geosciences, it's critical to ensure that the data you're working with is trustworthy. Data scientists should verify the following before moving forward with modeling:

Is the data current?

When was the data created? For certain foundational observations, such as reports from geoscientists, the date may not matter. In most other cases, data should be recent.

Is the data credible?

Who collected the data? Geological data often has a qualitative element to it. Different geologists may interpret a rock origin's differently, for example. The expertise of the scientist(s) who collected the data is important in evaluating how reliable the source is.

Is the data complete?

Where are there missing values, and why? Take note of any hints that the data description provides as to how to deal with missing data. Should you use imputation techniques or not?

TIP 4

Do your data due diligence but trust your gut (or the geoscientist you are working with).

Visualize the data through mapping, charting, or other specific geoscience visualizations. For high-dimensional data, consider using a dimensionality reduction technique like t-SNE or MDS.

Be cautious about imputing missing values. There may be an important reason why data is missing. In some cases, such as highly nuanced geochemistry data or data about how systems have behaved, the rule is to not impute. In other cases, such as geophysical data, it may be fine, but check the metadata to be sure.

Control data quality through exploratory analysis and visualization. For geophysical data, think about where noisy data could be coming from: complex geology, data collection, or processing. These give you hints about what to investigate next.

Compare against related data sets, if available. Validating data sets at different scales against each other — such as satellite imagery and rock samples — is a great way and discover new connections.

Embrace generative AI. It may be helpful in finding patterns in data, combining highly distinct data sets, or suggesting options for modeling and analysis.

Check the results against logical and geological truth after modeling. Model explainability methods SHAP or LIME can help with understanding the importance that your model places on each feature.

A word from our Director of Challenges, Nate Suurmeyer

Nate's five rules for being a data science rock star in the geoscience world:

1. Understand where the data comes from.
2. Understand how the data has changed over time.
3. Ensure the data is in a standardized, open format so it's not lost as software evolves.
4. Look for gaps and opportunities in the history of who has used the data and how.
5. Spend time with the data, make sure it feels right, and trust your gut.

"Oftentimes you don't have enough data or your data is so biased that actually figuring out what's next needs to happen in the imagination, and that's OK. Remember, oil was first discovered in the minds of people."

Nate Suurmeyer,
Director of Challenges, Onward

Learn more about our mission at:

ThinkOnward.com

Jump right in! Explore our platform and join a challenge:

A platform for change

onward