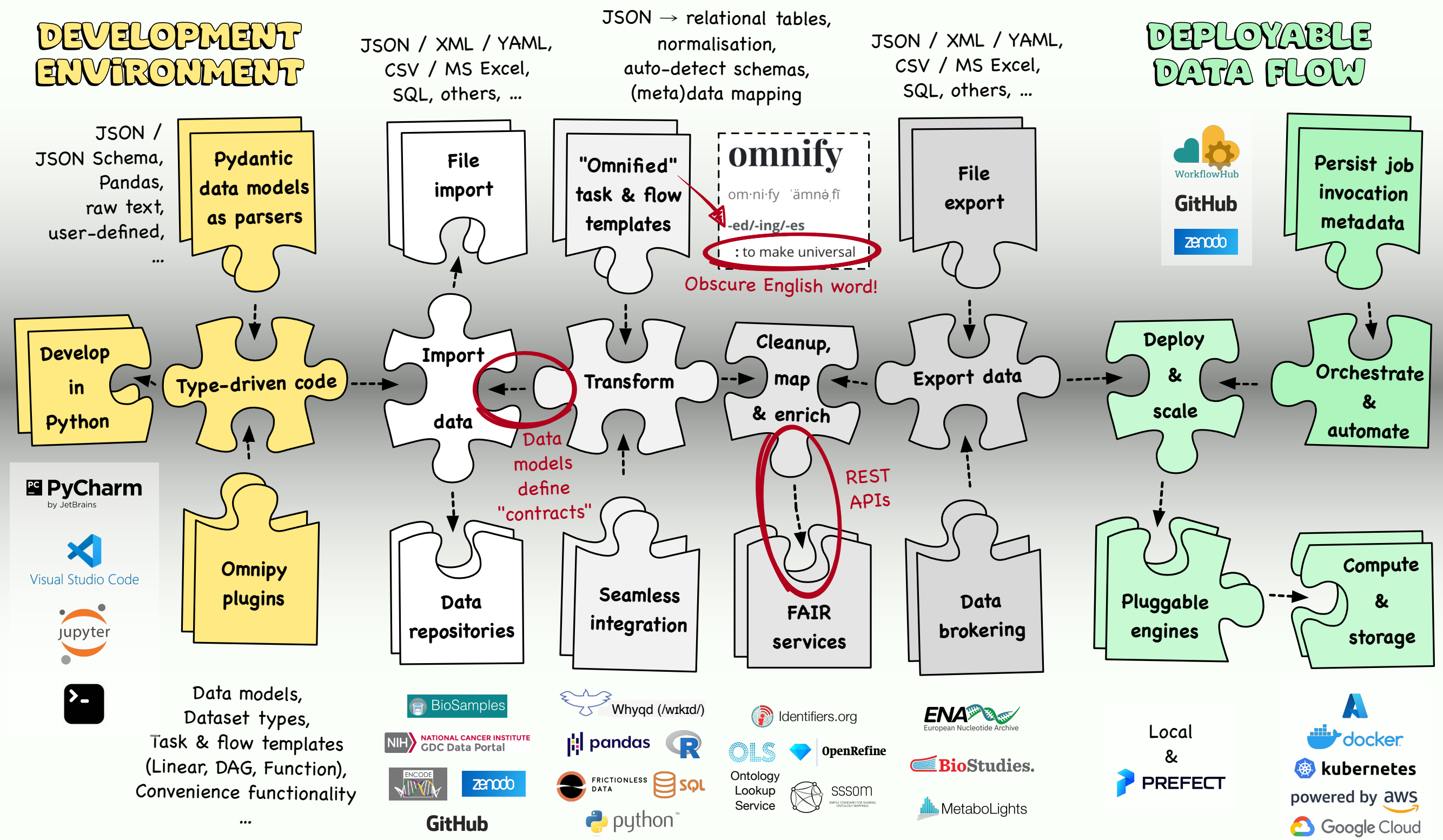
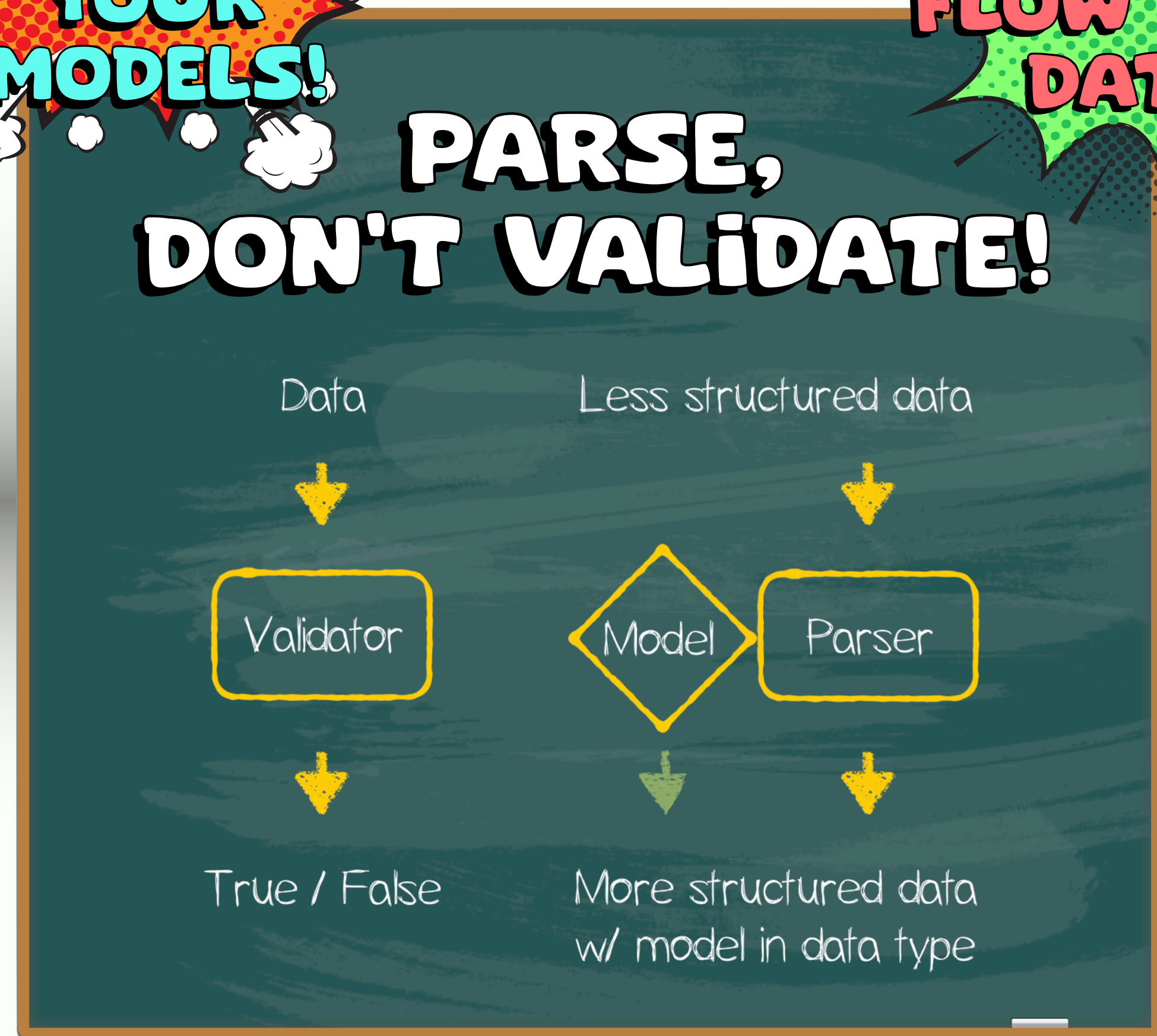




omnipy

PARSE TO MODELS AT THE LINKS IN THE FLOW



(Note: Most illustrated integrations are yet to be implemented)

# - TURNING THE TABLES WITH OMNIPY

[HTTPS://OMNIPY.READTHEDOCS.IO/](https://omnipy.readthedocs.io/)

SVEINUNG GUNDERSEN,  
UNIVERSITY OF OSLO / ELIXIR NORWAY

The time is ripe to challenge some fundamental assumptions on how to wrangle your data! Most RDA members will agree that data models, or metadata schemas if you want, are key to interoperability. Unfortunately, they have some inherent issues. In most cases, data models are implemented either too loosely – as free-for-all spreadsheets that just invite chaos – or too rigidly – as strict validation routines that do not allow a single mismatch! In both cases, the meticulous and mundane work of mapping real-world data to the models requires a significant amount of effort.

The slogan "Parse, Don't Validate" is making rounds in programming circles – but the main concepts are not new. It is a well known programming guideline to allow large variation in the inputs, but produce stringent outputs with a guaranteed structure. This is the main concept behind "parsers". Dynamic

programming languages like Python have excelled for years in parsing capabilities, but lack of strict typing has limited the possibility of stringency in the outputs. However, cutting-edge developments in the Python language now allow the implementation of the best of both worlds – now provided to you in a single package named Omnipy.

Omnipy is a new Python library that offers a systematic and scalable approach to research data and metadata wrangling. It allows researchers to import data in various formats and continuously reshape it through typed transformations. For large datasets, Omnipy seamlessly scales up local test jobs for remote deployment on a variety of compute infrastructures.

We aim to shape Omnipy into a community effort by encouraging integrations with other solutions.