

# Session 3: Data use cases

## Frictionless DarwinCore

# *André Heughebaert*

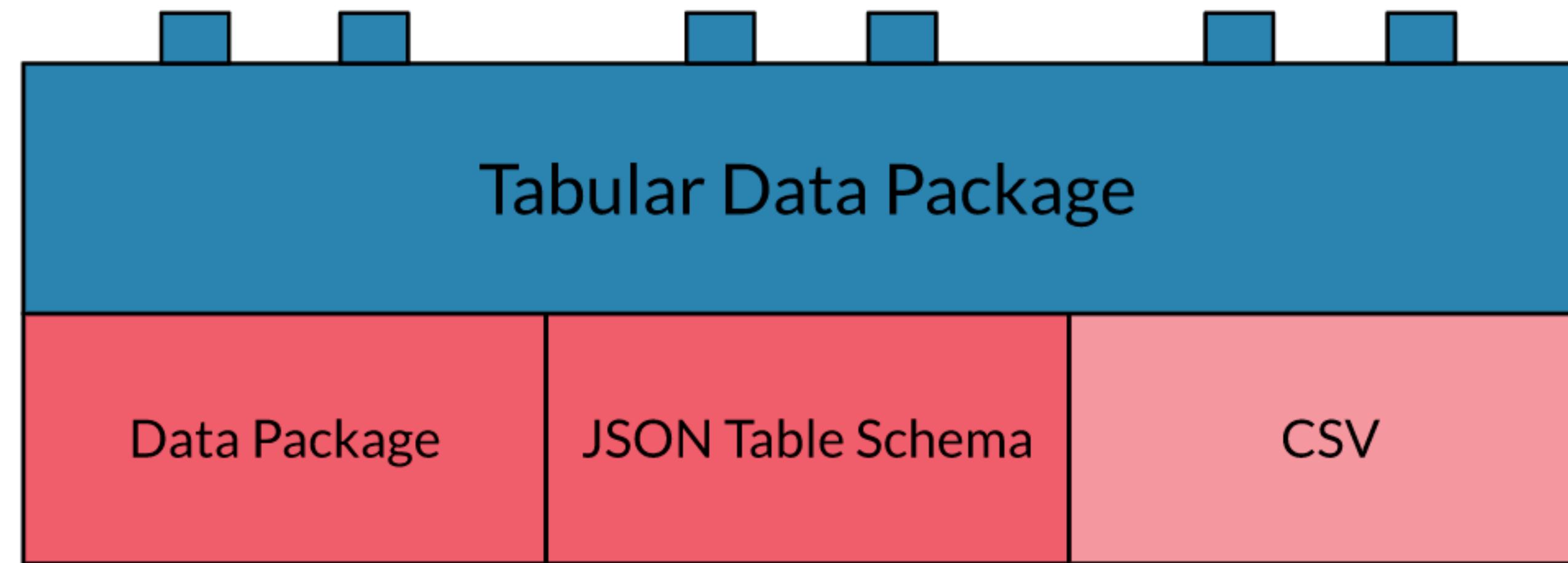
# FRictionless DATA



**Frictionless Data** is one of the core projects at Open Knowledge Foundation, whose aim is to reduce friction in working with data, with a goal to make it effortless to transport data among different tools and platforms for analysis.



## Tabular Data Package



<http://frictionlessdata.io/>

Web apps

[OpenRefine](#)

[DatHub.io](#)

[goodtables.io](#)

[DataPackage Creator](#)

CLI

Goodtables

Library

R

Python

Go

Javascript

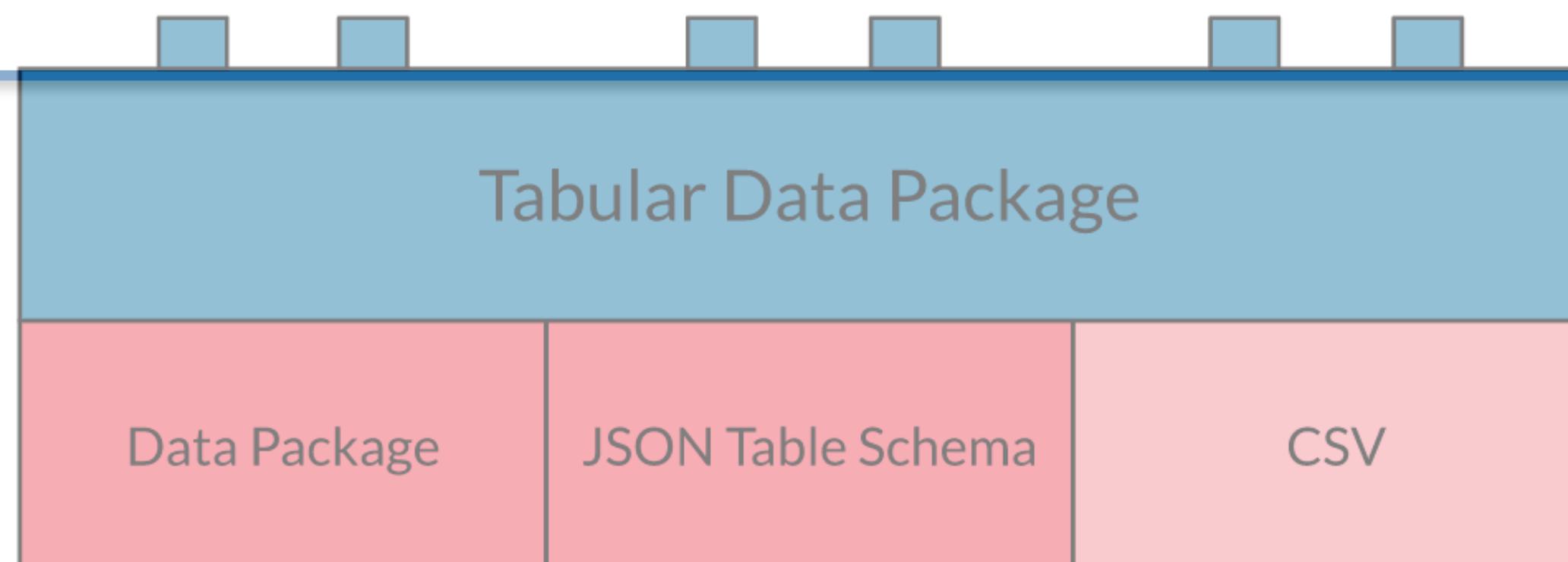
Ruby

Pandas

SQL

Julia

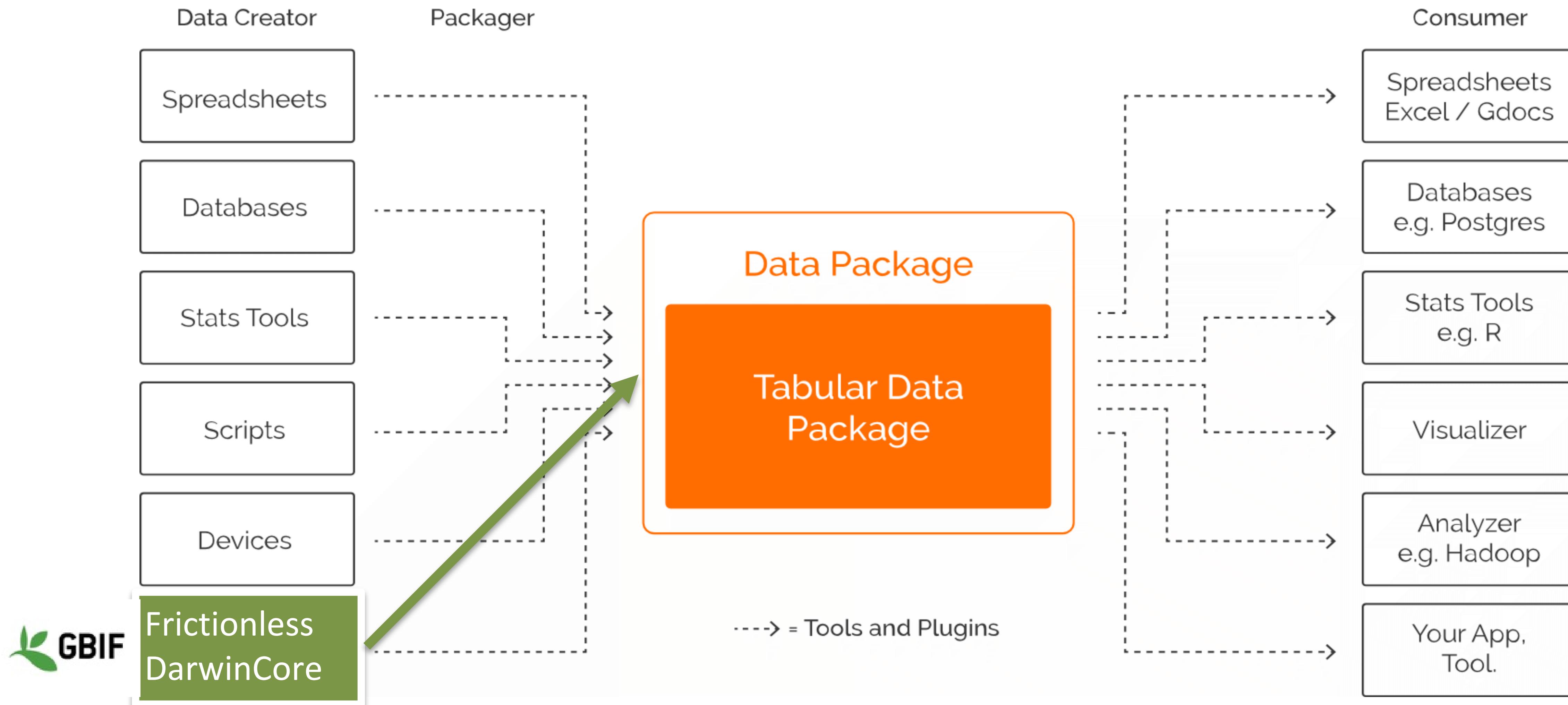
## Tabular Data Package



# HOW DOES FRICTIONLESS DIFFER FROM DARWIN CORE?

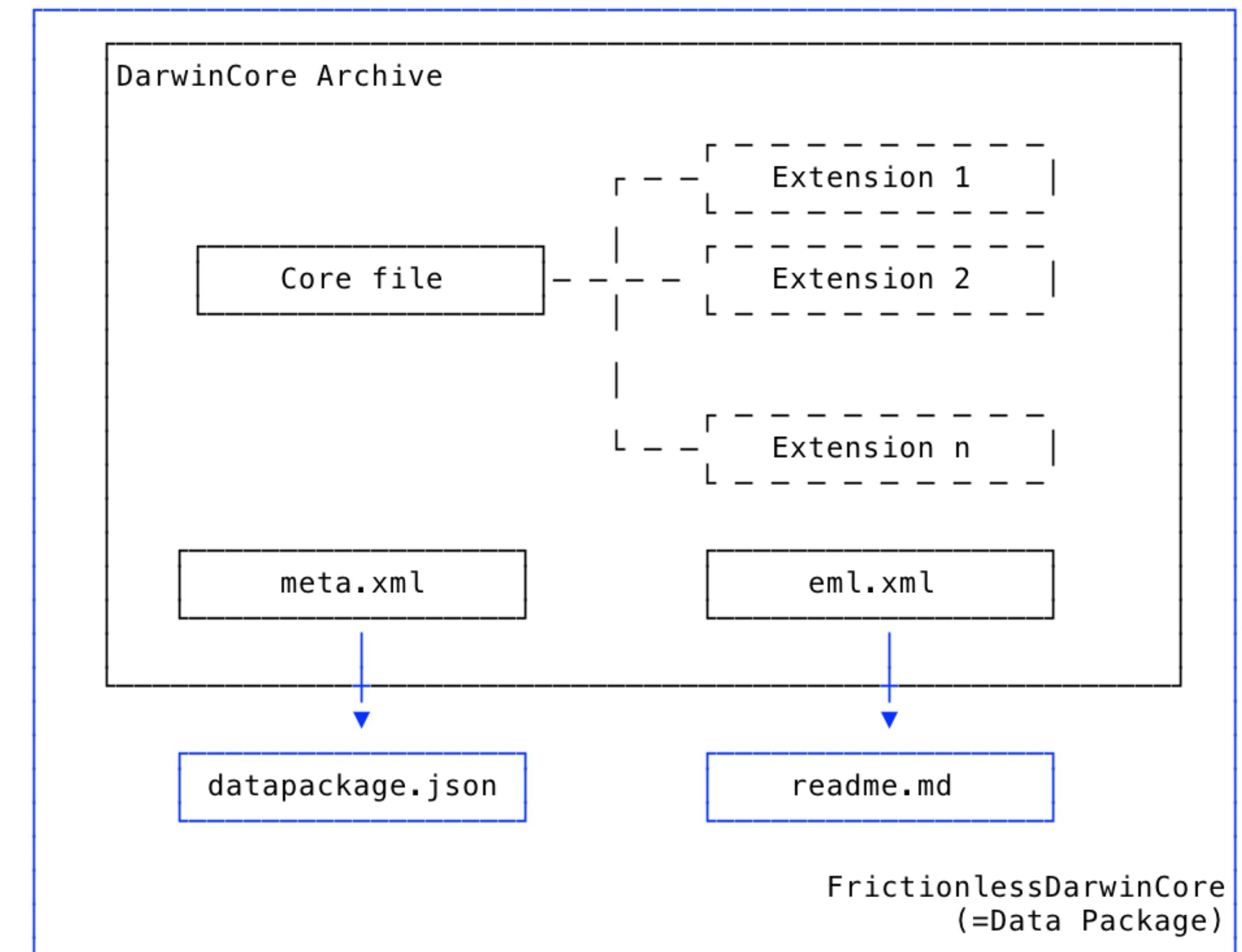
- CSV(data) + JSON(schema)
- Domain agnostic (eg Fiscal data package)
- Truely relational
- Allow explicit constraints on columns such as:
  - **startDayOfYear** integer { "minimum": 1, "maximum": 366}
  - **decimalLatitude** number { "minimum": -90.0, "maximum": 90.0}
  - **countryCode** string { "minLength": 2, "maxLength": 2}

# FRictionless DARWIN CORE



# FRICTIONLESS DARWIN CORE

An Open Source Python Library (and CLI) that converts your DwC Archives into Frictionless Data Packages.



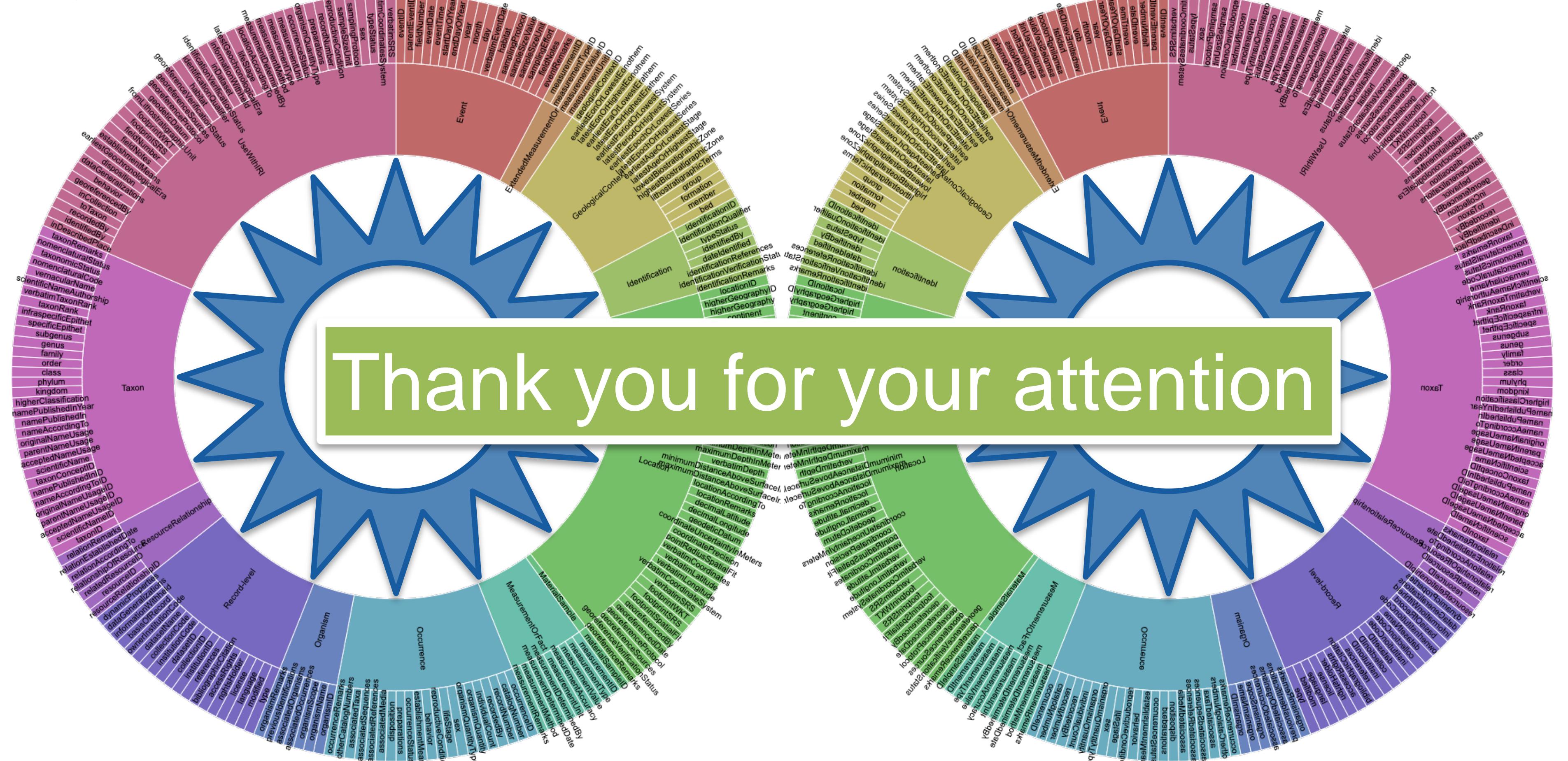
How to contribute?

<https://github.com/frictionlessdata/FrictionlessDarwinCore>

# GOODTABLES VALIDATION REPORT

- **countryCode** [105,123] [maximum-length-constraint] The value "BEL" in row 105 and column 123 does not conform to the maximum length constraint of "2"
- **language** [470,36] [maximum-length-constraint] The value "Various" in row 470 and column 36 does not conform to the maximum length constraint of "2"
- **coordinateUncertaintyInMeters** [470,142] [type-or-format-error] The value "707.1" in row 470 and column 142 is not type "integer" and format "default"
- **basisOfRecord**: [1440,64] [enumerable-constraint] The value "ObservedSpecimen" in row 1440 and column 64 does not conform to the given enumeration: "[PreservedSpecimen', 'FossilSpecimen', 'LivingSpecimen', 'MaterialSample', 'Event', 'HumanObservation', 'MachineObservation', 'Taxon', 'Occurrence']"

<https://github.com/frictionlessdata/goodtables-py>



# Session 3: Data use cases Frictionless DarwinCore

André Heughebaert