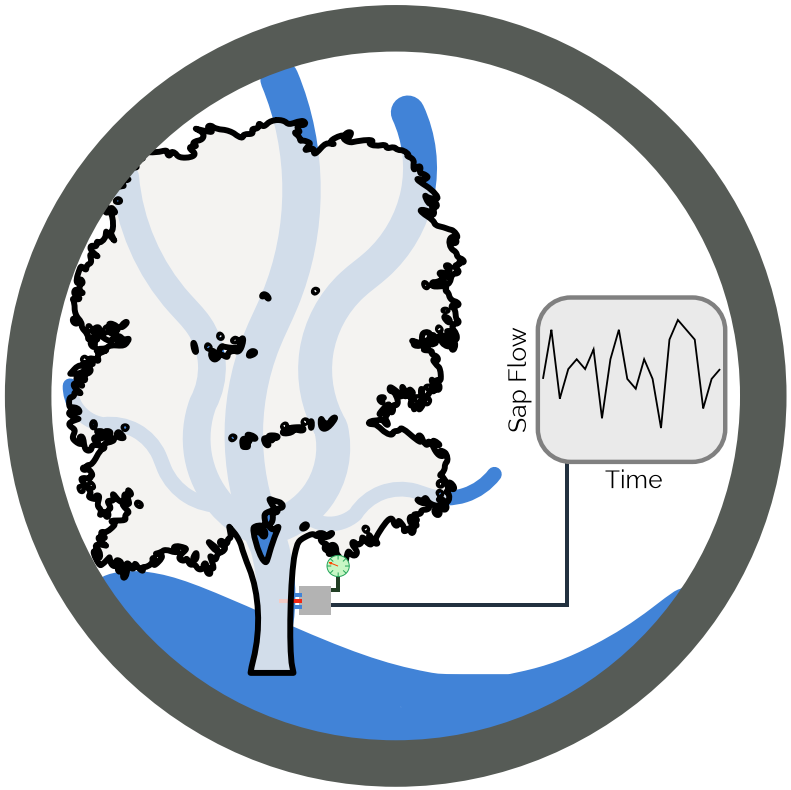
Compiling a global database of sap flow measurements: the SAPFLUXNET data workflow

XIV MEDECOS & XIII AEET meeting
Ecoinformatics: data science brings new avenues for ecology
Symposium

Víctor Granda, Rafael Poyatos, Roberto Molowny-Horas, Maurizio Mencuccini, Kathy Steppe & Jordi Martínez-Vilalta

CREAF
Centre of Ecological Research and Forestry Applications

Introduction



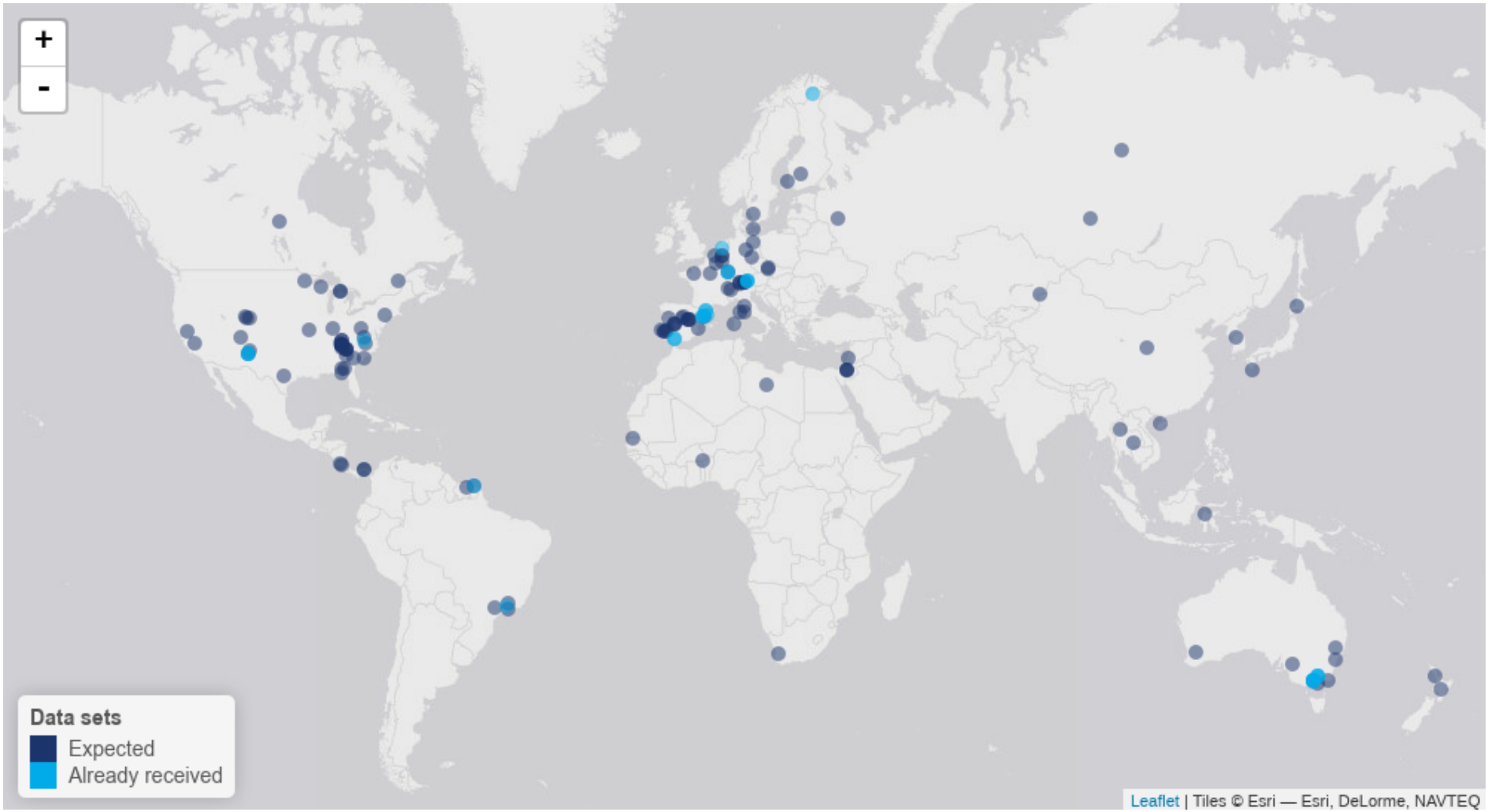
Different **thermodynamic methods** to determine sap flow using heat as a tracer sap movement

Proxy of the movement of water between the soil-plant-atmosphere continuum.

Allows **upscaling** from stem to plant and landscape level.

The time is ripe for a global database

The **SAPFLUXNET** initiative is building the first global database of plant-level sap flow measurements to analyse the environmental and physiologycal factors driving tree- and stand-level transpiration



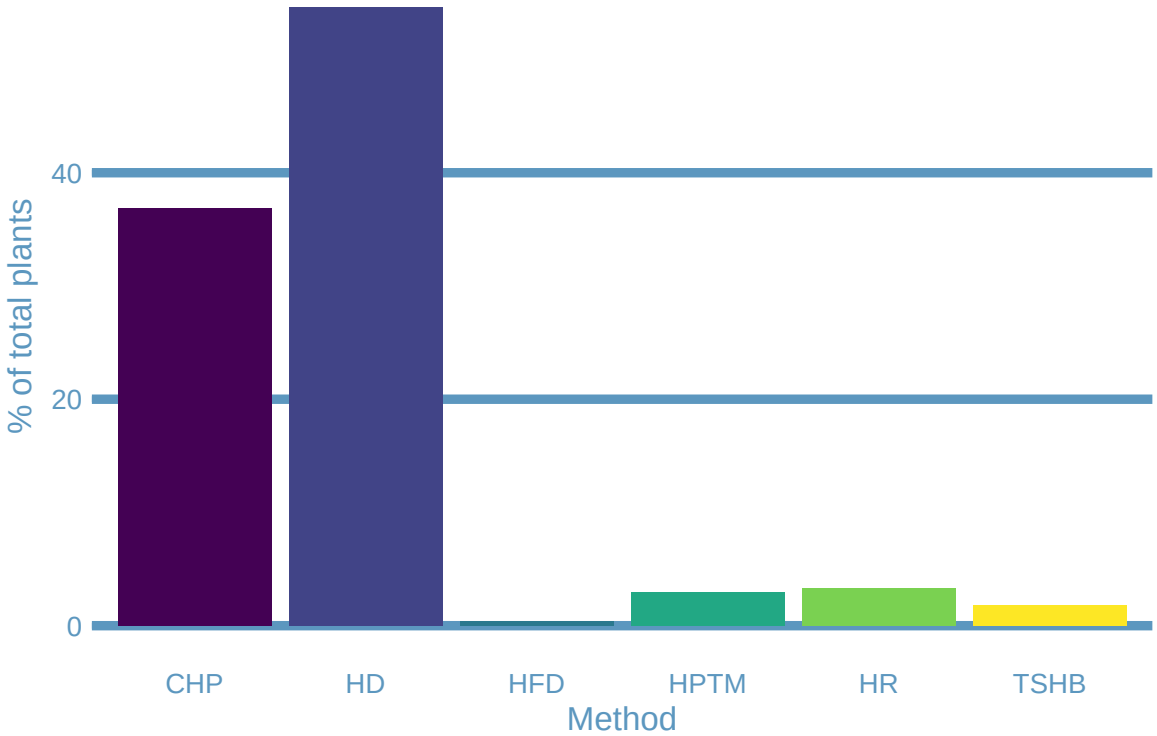
Target datasets

- Stem or whole-plant level
- Field conditions
- Sub-daily intervals
- Environmental data available (RH, Ta, PAR...)
- Abundant metadata (site, stand, plant, species and environmental)

Data Characteristics

High data complexity:

- substantial methodological variability



Data Characteristics

High data complexity:

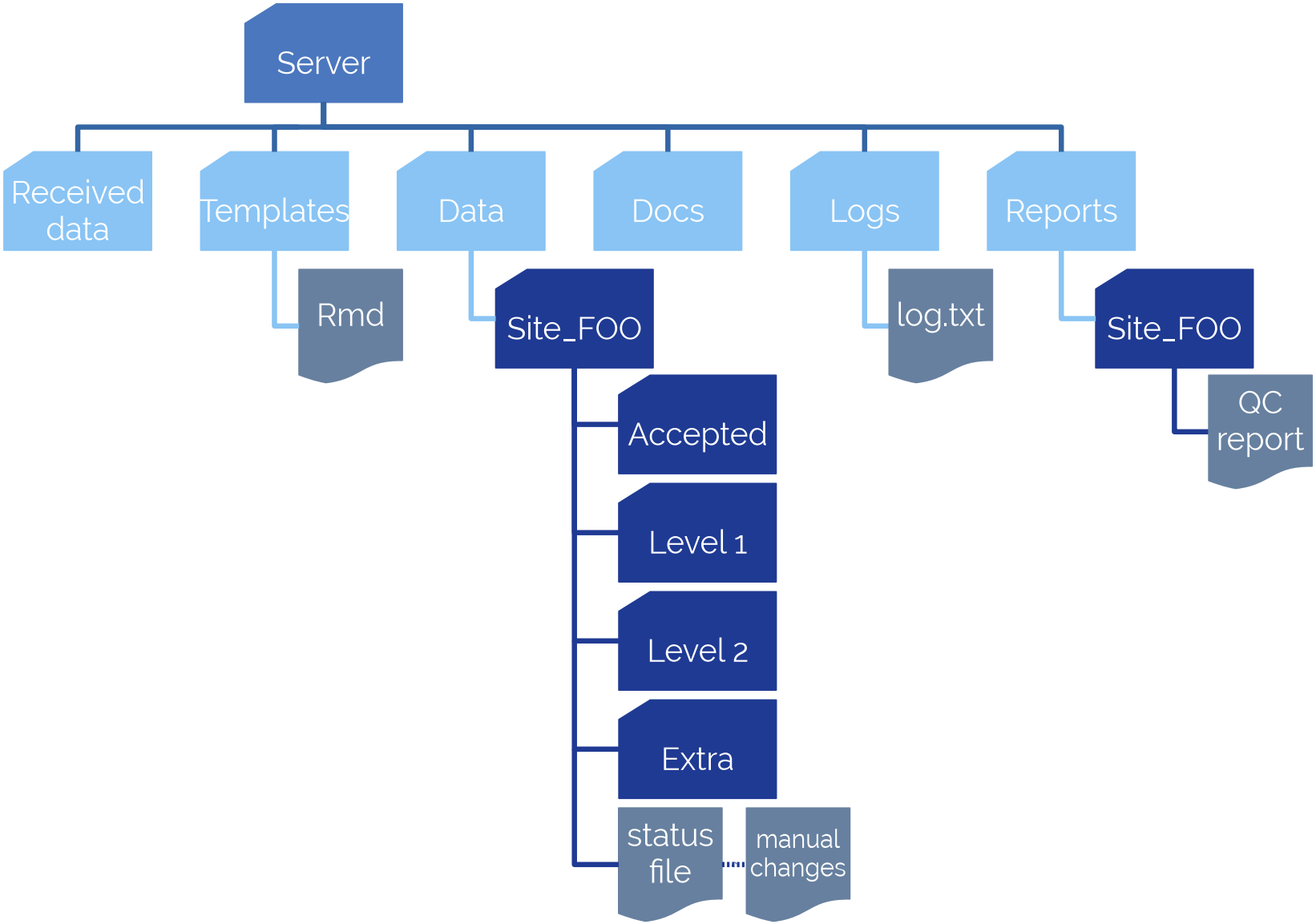
- Large datasets and ancillary metadata

Metadata	Items
Site	20
Stand	16
Species	4
Plant	24
Environmental	16
— — —	— —
Total	**80**

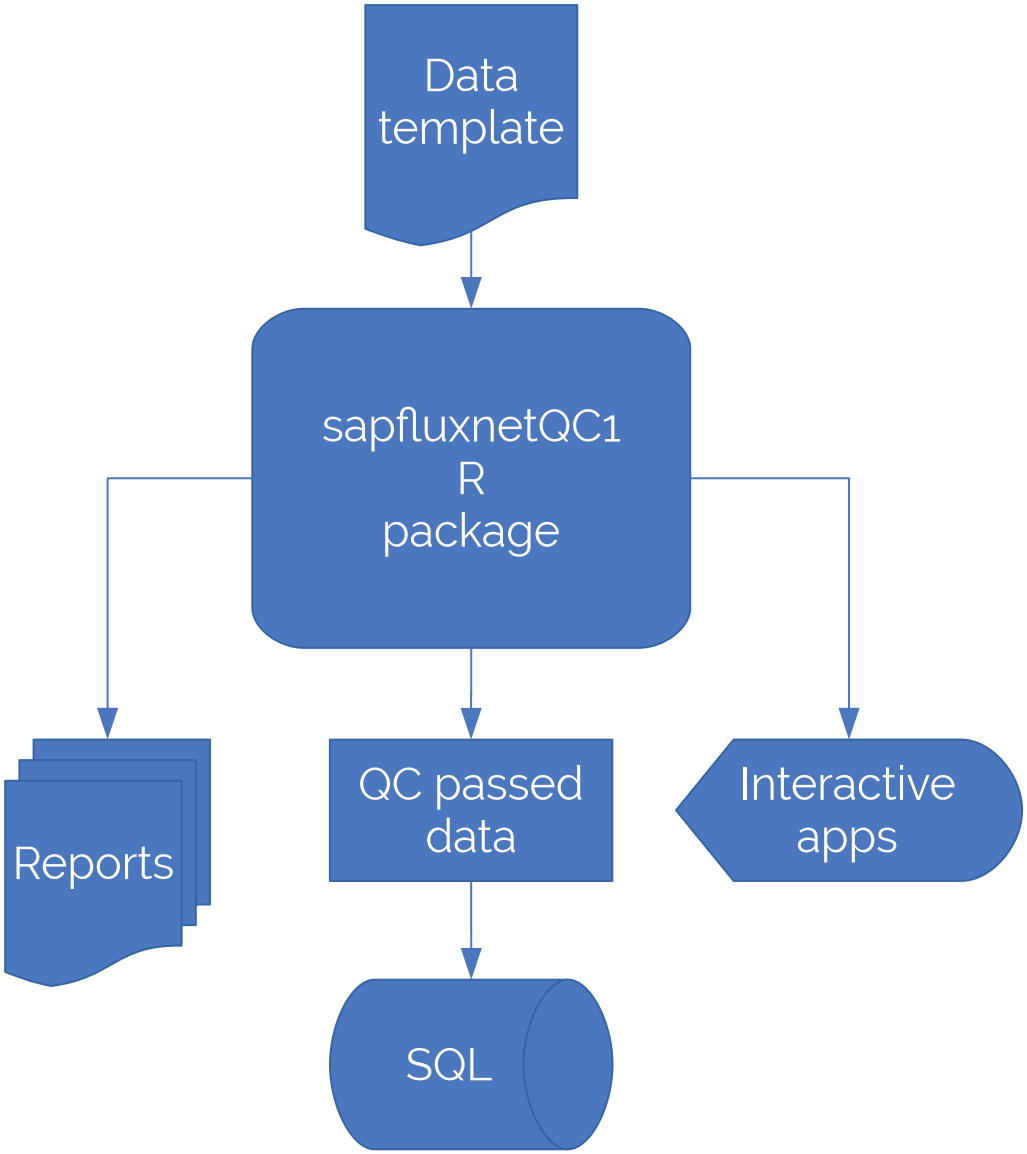
How to assay quality and store data?

We need **semi-automatic**, **reproducible** and **robust** checks to ensure the quality of the submmited datasets. Also, we need to store the data in a way that allows all essential information to be **available** in order to use the data in checks and analyses

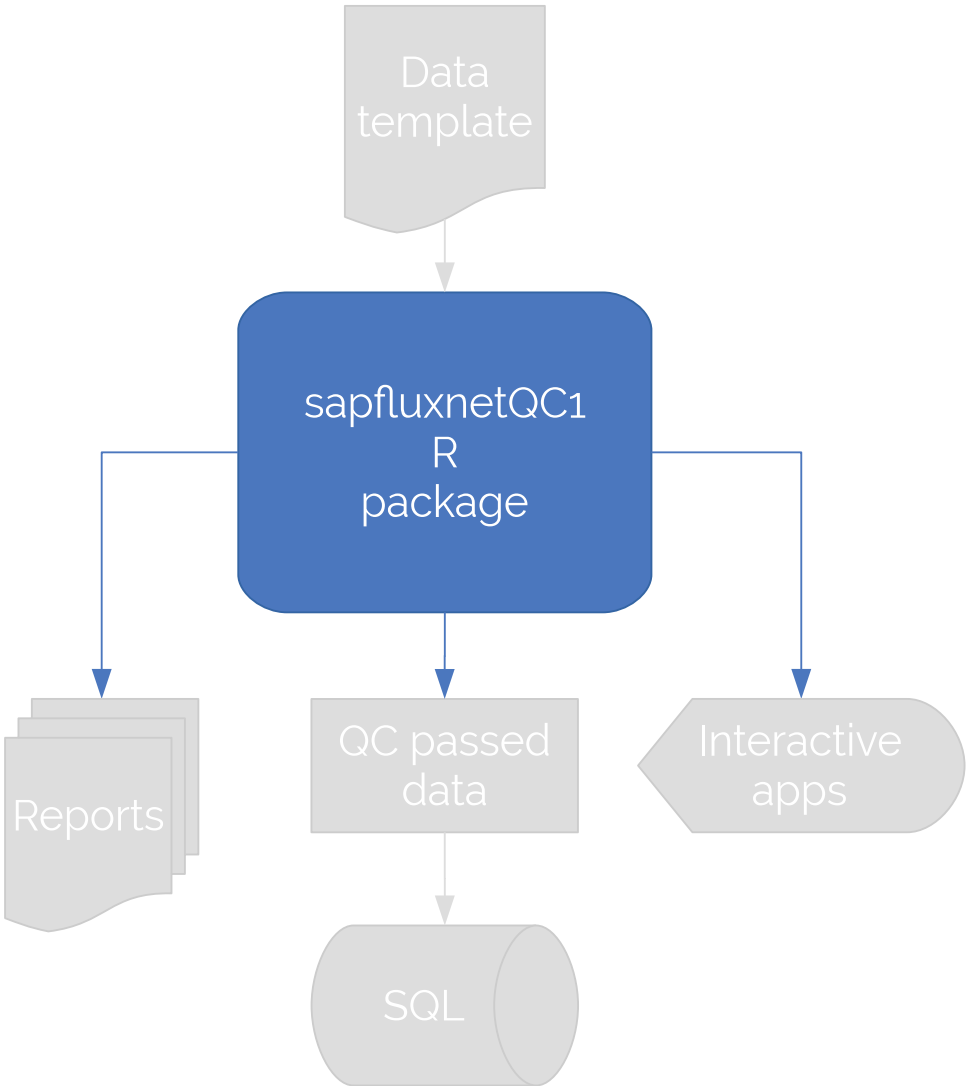
SAPFLUXNET Server Structure



SAPFLUXNET Data Flow

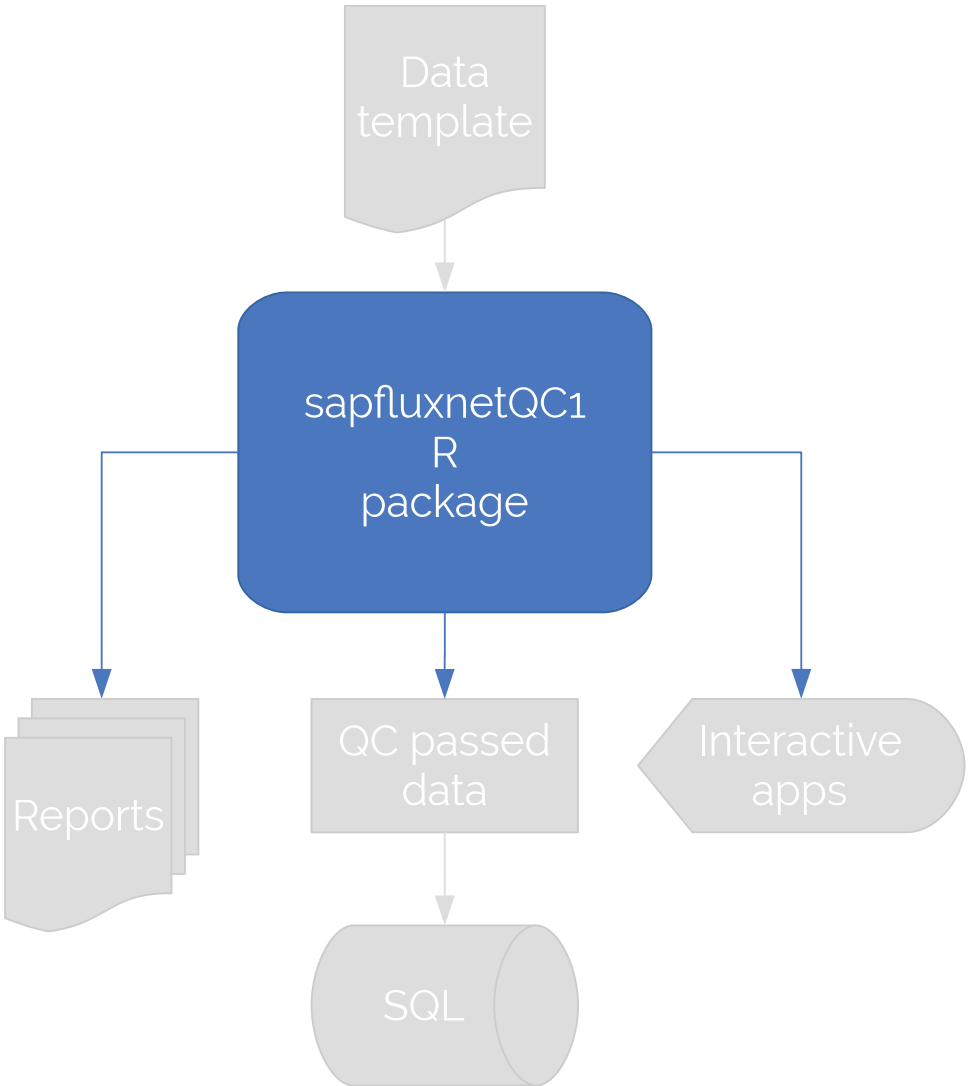


SAPFLUXNET Data Flow



1. Automatic Quality Control checks (QC)
2. Automatic report generation
3. Storing data in special object (SfnData S4 class)
4. Interactive functions allowing fine control of quality control

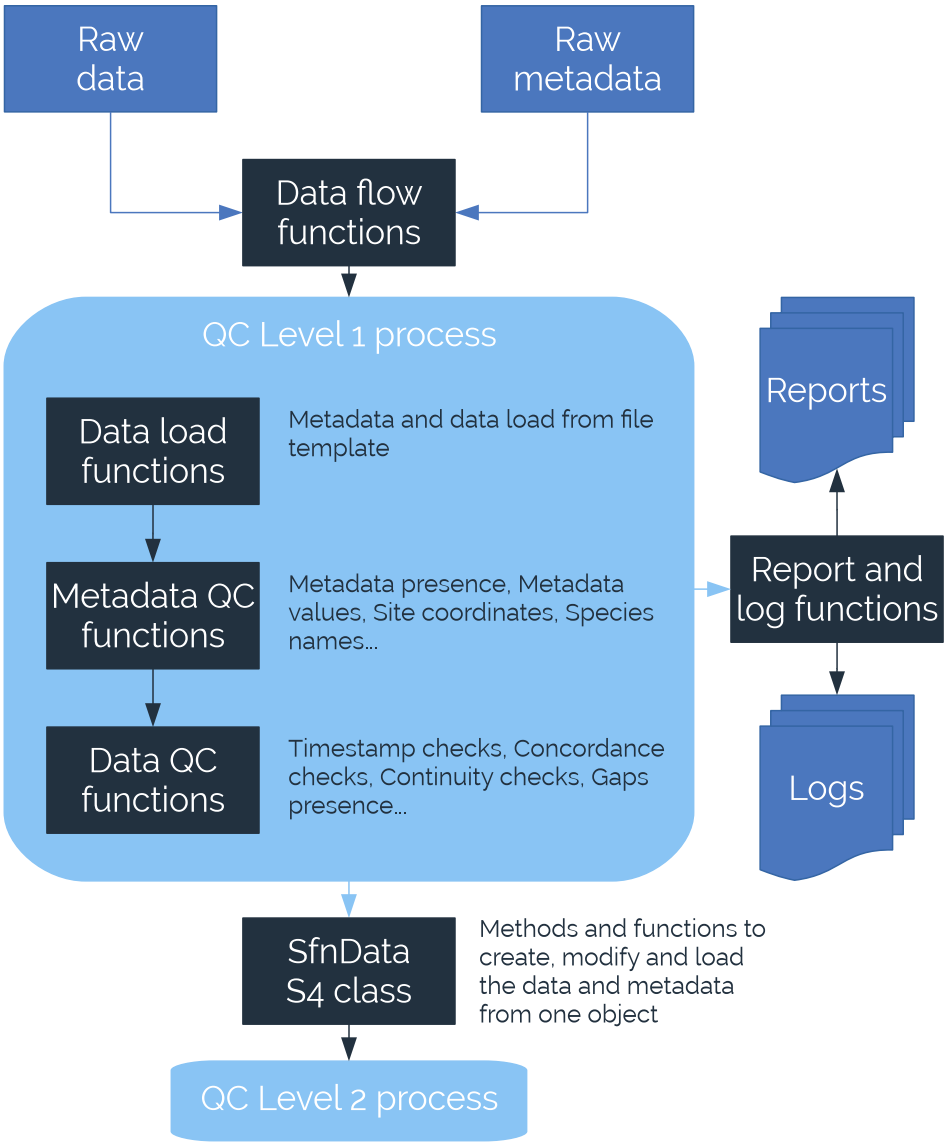
SAPFLUXNET Data Flow



Benefits of R as environment

1. Open
2. Reproducible
3. Easy maintenance and update
4. Easy integration with web and SQL technologies

SAPFLUXNET Data Flow



QC Level 1

In charge of general metadata and data quality checks:

- Presence/Absence of metadata and data variables
- Metadata values correctness (i.e coordinates, species names...)
- Sapflow and environmental data correctness (format, timezone, gaps, continuity...)
- Uniformization and unit transformations of data (solar time, sapflow at different levels...)