

# Koen Hufkens



Senior Scientist

# Koen Hufkens

Roots in numerical & landscape ecology, now covering various topics from ecophysiology to remote sensing in function of **macro-ecology** research.

Worked at various institutions in the US (~6yr Boston University and Harvard), and in the EU (University of Ghent, INRAE, ETH, now University of Bern).

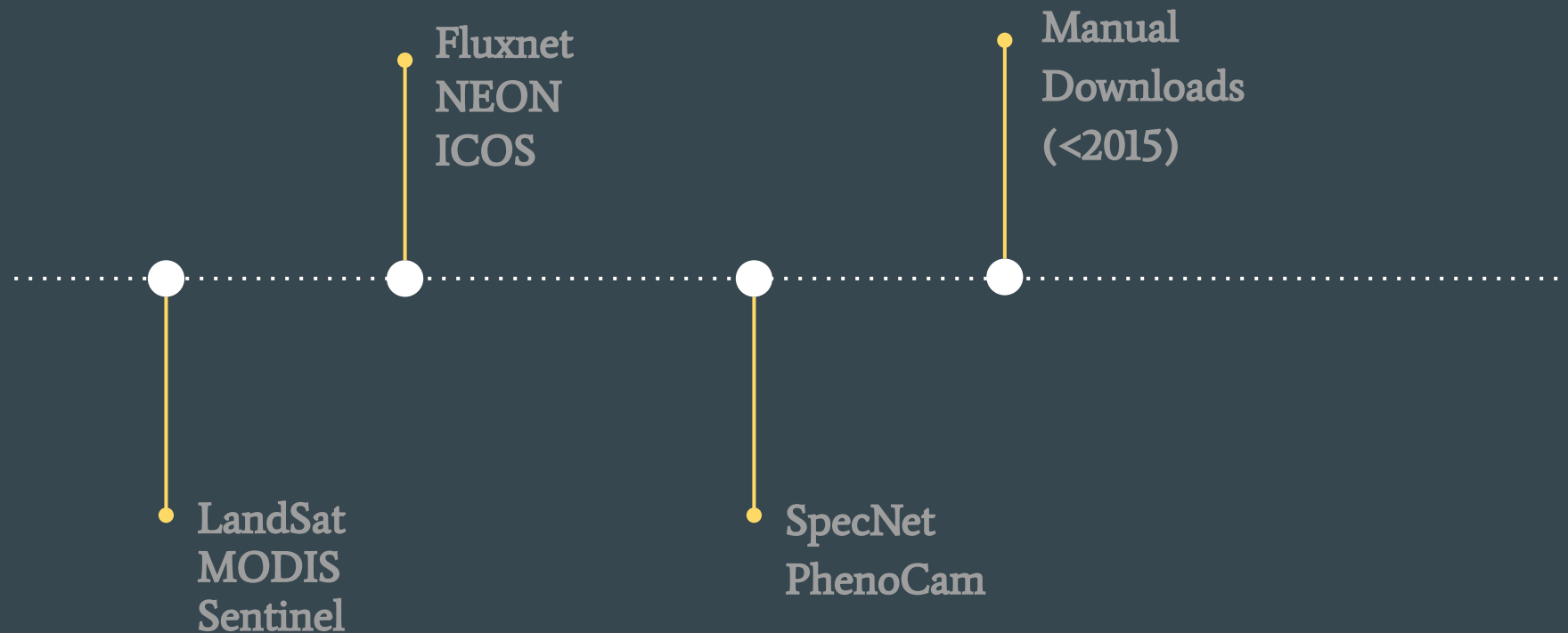
Recipient of an EU Marie Skłodowska-Curie Action and led a multi-institution research projects on eco-climatological data recovery (proxy measures) funded by the Belgian science policy office.

Currently, at the University of Bern, Geocompation and Earth Observation group

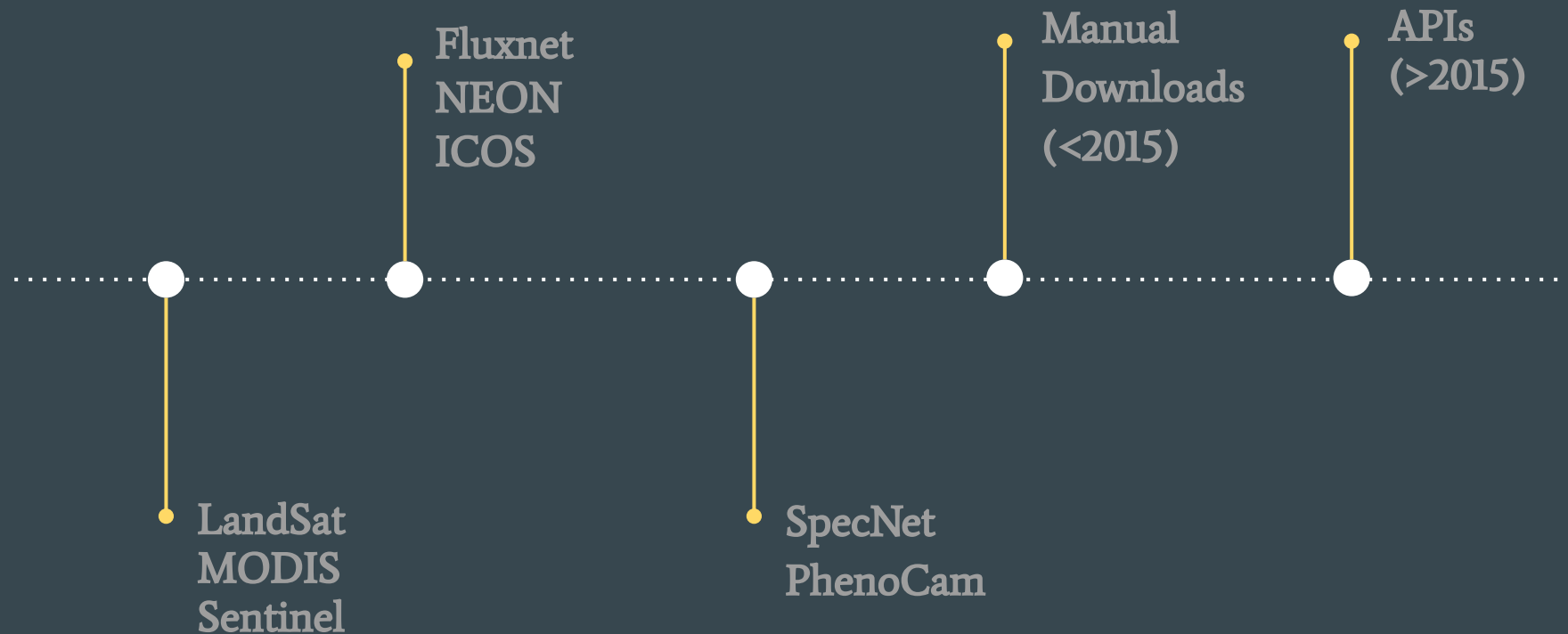
# Goals

- Have **fun** doing **science**
  - **Facilitate science**
    - Software development
    - System integration
  - Big science, **small budgets**
  - Teach these skills
    - **Community** building
-

# Ecology is not data poor, but (often) poorly integrated



# Ecology is not data poor, but (often) poorly integrated



# Lost time

- Data integration is hard
  - Stumbling block ECR
- **Isolated efforts**
  - **FluxnetLSM**
  - **FluxDataKit**
- No community support
- Teach these skills

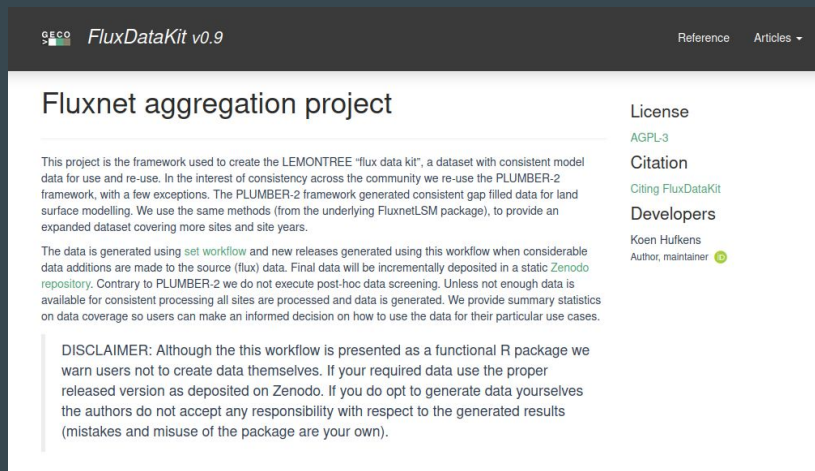
---

# Facilitate data integration

Requires API interfaces which are easy (e.g. R based MODISTools, ecmwfr, amerifluxr, packages etc).

Requires knowledge transfer, and open sharing of common or similar workflows.

Does not necessarily need “frameworks”, as hard to maintain.



The screenshot shows the GitHub repository for FluxDataKit v0.9. The page has a dark header with the repository name and version. Below the header, the title "Fluxnet aggregation project" is displayed. The main content area contains a description of the project, which is a framework for creating the LEMONTREE "flux data kit". It mentions the use of the PLUMBER-2 framework and the FluxnetLSM package. A disclaimer is also present, stating that the authors do not accept responsibility for the results generated by the package. On the right side, there are links for the license (AGPL-3), citation, and developers (Koen Hufkens).

FluxDataKit v0.9

Reference Articles

## Fluxnet aggregation project

This project is the framework used to create the LEMONTREE "flux data kit", a dataset with consistent model data for use and re-use. In the interest of consistency across the community we re-use the PLUMBER-2 framework, with a few exceptions. The PLUMBER-2 framework generated consistent gap filled data for land surface modelling. We use the same methods (from the underlying FluxnetLSM package), to provide an expanded dataset covering more sites and site years.

The data is generated using [set workflow](#) and new releases generated using this workflow when considerable data additions are made to the source (flux) data. Final data will be incrementally deposited in a static [Zenodo repository](#). Contrary to PLUMBER-2 we do not execute post-hoc data screening. Unless not enough data is available for consistent processing all sites are processed and data is generated. We provide summary statistics on data coverage so users can make an informed decision on how to use the data for their particular use cases.

**DISCLAIMER:** Although this workflow is presented as a functional R package we warn users not to create data themselves. If your required data use the proper released version as deposited on Zenodo. If you do opt to generate data yourselves the authors do not accept any responsibility with respect to the generated results (mistakes and misuse of the package are your own).

**License**  
[AGPL-3](#)

**Citation**  
[Citing FluxDataKit](#)

**Developers**  
Koen Hufkens  
Author, maintainer



# We are not software developers

*But can learn some of their practices to enhance research*



# Open Questions?

- Data integration needs?
- **Infrastructure** required?
- Community integration?
- How to teach / engage?
- Systems approach
- Awesome R/python scripts

---

# Contact

**Koen Hufkens**

[koen.hufkens@gmail.com](mailto:koen.hufkens@gmail.com)

<https://geco-group.com>

