

## Open Science Story: Developing open classification tools for transposable elements

*ServiceDesk Facilities*



News

article\_published\_on\_label  
June 24, 2025

WUR researchers Like Fokkens and Anna Fensel, along with WUR alumnus Sibbe Bakker, are developing open tools to better classify and study elusive 'jumping genes' across all species. The collected data must be FAIR: findable, accessible, interoperable, and reusable.

Jumping genes is the popular name for transposable elements (TE). Their ability to copy and relocate within a genome has historically made them difficult to detect and describe accurately and systematically, posing a significant challenge for research.



Transposable elements are essentially the genome's dark matter

Like Fokkens (Assistant Professor at the Laboratory of Phytopathology, Plant Sciences)

Fokkens forms a team with two other researchers: Sibbe Bakker (research assistant, Bioinformatics, Faculty of Science, Utrecht University) and Anna Fensel (Professor in Artificial Intelligence, Plant Sciences). The three are united by a common interest in shedding light on some of that dark matter. The team brings together multiple disciplines: Fokkens is a fungal bioinformatician, Bakker is a plant bioinformatician, and Fensel is an AI and data scientist. By combining Fokkens and Bakker's biological insights with Fensel's expertise in data structuring and interoperability, the team is developing open tools that can classify and compare transposable

elements across species and databases.

## Why is your work necessary?

TEs have seen renewed interest from the scientific community in recent years as important drivers of evolution and valuable biotechnology tools for inserting genetic material. However, as they have traditionally been studied within the context of specific model systems, TE data for species outside of these systems is limited. Fokkens explains: "If you want to study TE's within a model system, that's fine. But if you want to study TE's themselves, with comparisons, it's frustrating. Databases all describe them differently, and searching across them requires a lot of extra work".

And more work seems to be on the horizon. Recent improvements in genome sequencing and assembly have made transposable elements easier to detect. These advances mean more detailed genomic data, including previously hard-to-detect repetitive sequences, are now within reach. "The technology is advancing quickly, and the scientific community is generating more data on transposable elements than ever", Bakker points out. This wealth of data also enables more thorough characterisation of TE 'behaviour' in host genomes, such as when and where they jump. This is currently lacking and would greatly improve their ability to be used as tools.

## How are you using open science?

"The challenge now is making sure we have a way of working with this influx of new data through classifying and comparing it with what we already have." That is why the researchers want to make a FAIR classification system, providing clear ontologies and open, standardised methods to categorise TEs and their dynamics.

If you want to have an apples-to-apples comparison, you first need a clear definition of what an apple is,"That's what we want to do

Like Fokkens

A major focus for the team is interoperability, a core component of FAIR data. Instead of replacing existing databases, the goal is to link existing databases with broader genomic archives, ensuring that fragmented datasets can interact. "To make transposable element datasets truly interoperable," Fensel explains, "we need structured metadata that allows researchers and automated systems to search and analyse this data efficiently".

## What do you need to succeed?

While the technical aspects of transposable element classification are key to this project, its success relies on engagement from the wider research community. TE Hub, an existing community-led platform for transposable element researchers, is important in bringing specialists together and facilitating discussions on standardisation. A community-wide approach ensures that the classification system reflects the needs of different research fields, rather than being shaped by just a handful of perspectives.

The next challenge will be securing funding and dedicated time to transform community ideas into usable infrastructure.

This kind of infrastructure benefits everyone working on TEs, but developing it requires significant effort

Sibbe Bakker

The three have applied for an Open Science NL Infrastructure Grant, which would provide the resources needed to implement their classification system at scale.

As they await the results of their application, the team continues working with the research community to refine the project's foundations, ensuring that the classification system meets the needs of scientists across different fields. Regardless of the outcome, their commitment to open science remains at the heart of their work, building a more connected and accessible future for transposable element research.

*This Open Science Story is based on an interview with Like Fokkens, Anna Fensel and Sibbe Bakker by Ben Excell, Community Manager Open Science & Education.*