

Notebook-as-a-VRE: collaborative virtual labs to build digital twins of ecosystems



UNIVERSITY
OF AMSTERDAM

Gabriel Pelouze^{1,2}, Spiros Koulouzis^{1,2}, Koen Greuell^{1,2}, Nafiseh Soveizi², and Zhiming Zhao^{1,2}

vlic@lifewatch.eu, z.zhao@uva.nl

¹ LifeWatch ERIC, Virtual Lab & Innovation Center (VLIC), Amsterdam, The Netherlands

² Multiscale Networked Systems, University of Amsterdam, Amsterdam, The Netherlands

Complexity of ecosystems



Digital twin for predictive studies



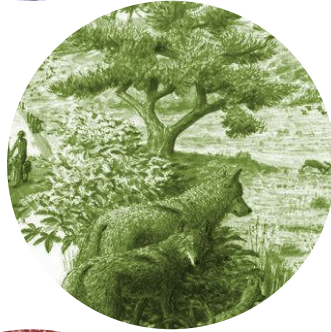
Physical system



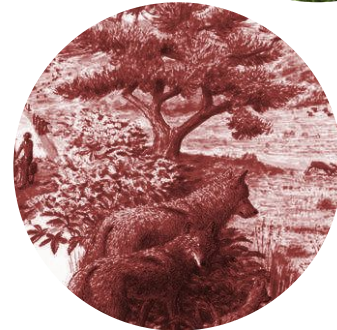
Digital twin



Scenario 1



Scenario 2

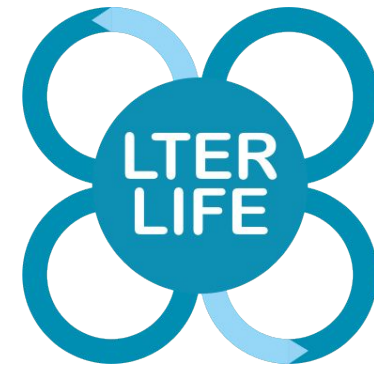
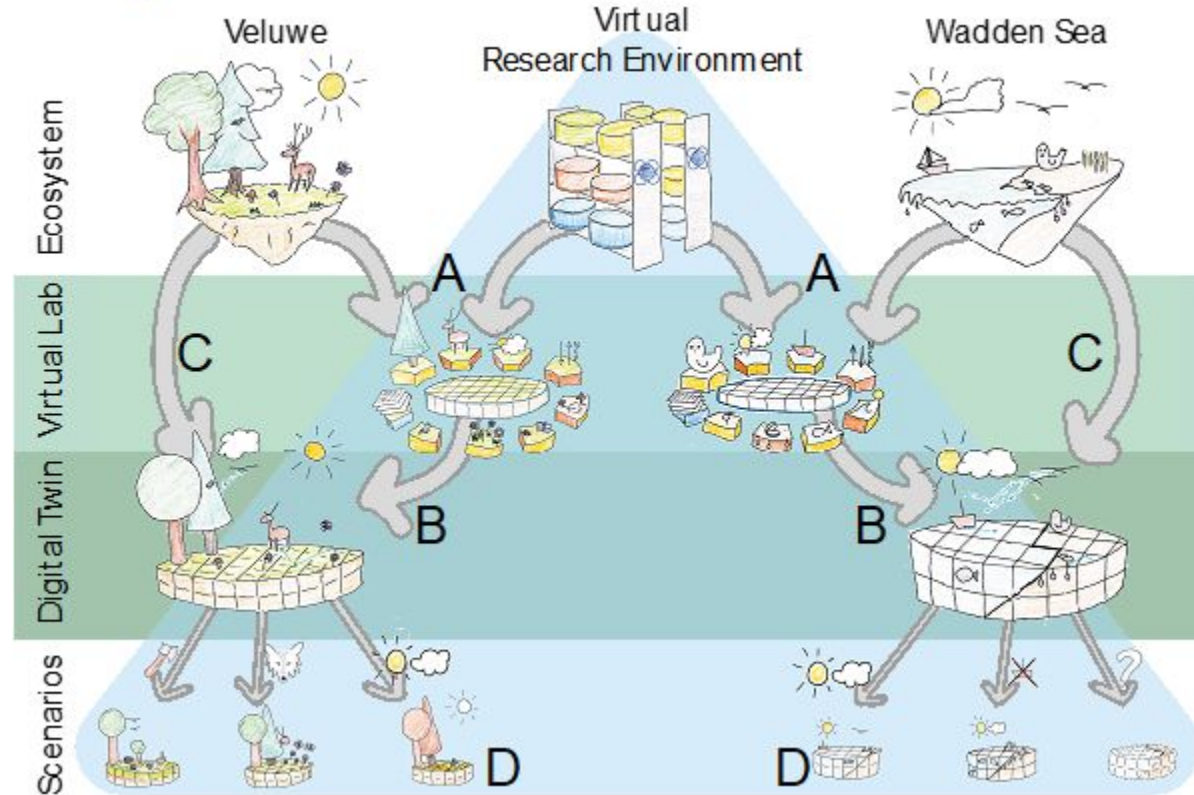


Scenario 3

But... scattered data and models

LTER-LIFE: a Dutch DT infrastructure project

Working with the LTER-LIFE infrastructure



<https://lter-life.nl/en>



National Institute for Public Health and the Environment
Ministry of Health, Welfare and Sport



UNIVERSITEIT
VAN AMSTERDAM



Utrecht
University

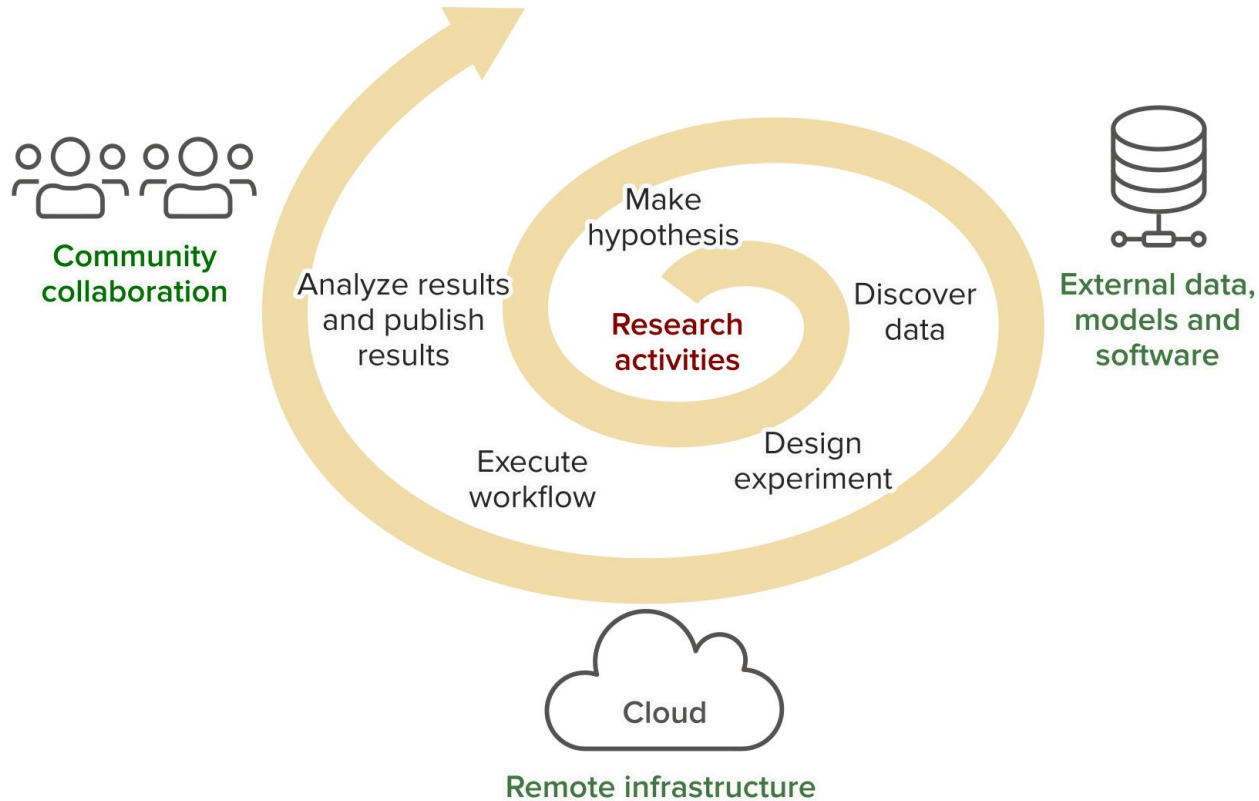


WAGENINGEN
UNIVERSITY & RESEARCH



university of
 groningen

Data-centric research lifecycle



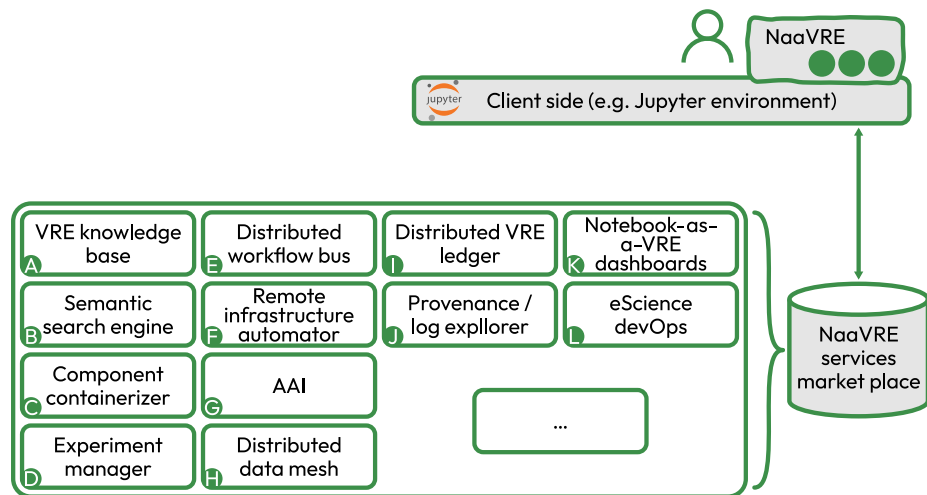
"Virtual Research Environments (VREs) provide user-centric support for discovering and selecting data and software services from different sources and composing and executing application workflows, ..." (Jeffery et al., 2020)

How to support data-centric experiments

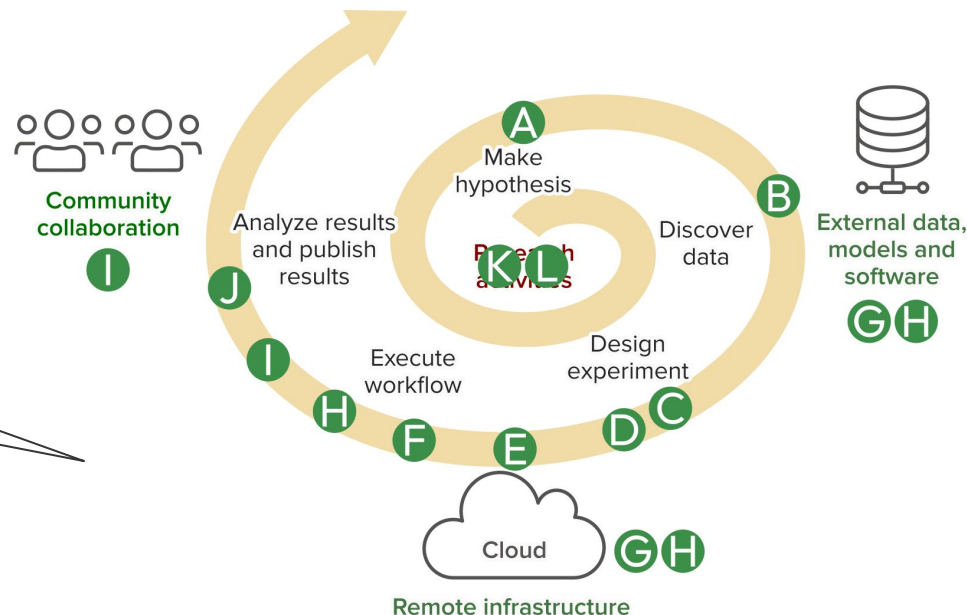
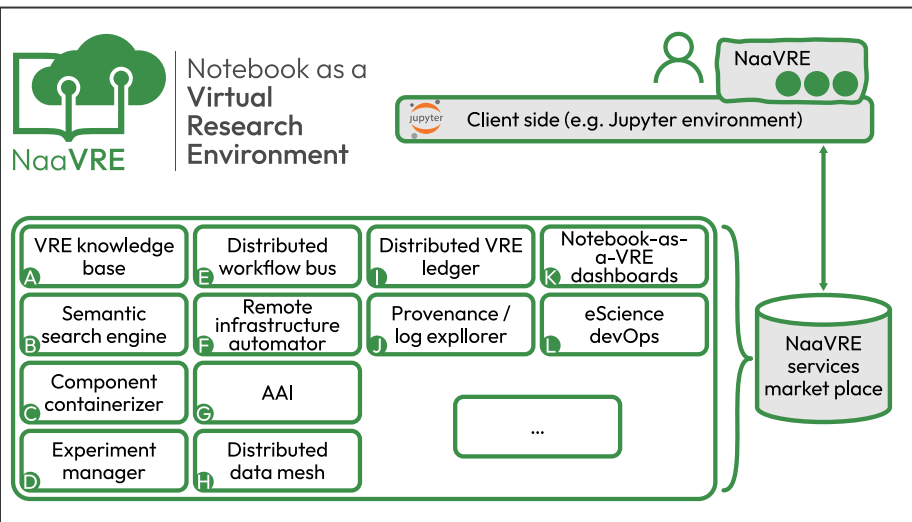
- **Computational notebooks**
 - Widely used by researchers
 - Documentation and visualization alongside code
 - Interactive code execution
 - → **developing and tweaking models**
- **Virtual research environment**
 - Collaboration
 - FAIR data, models and services
 - Remote infrastructure
 - Scaling up models (e.g. porting prototype to large-scale cloud environment)
 - → **sharing mature models & more**
- **Combined, they are NaaVRE**



Notebook as a
**Virtual
Research
Environment**

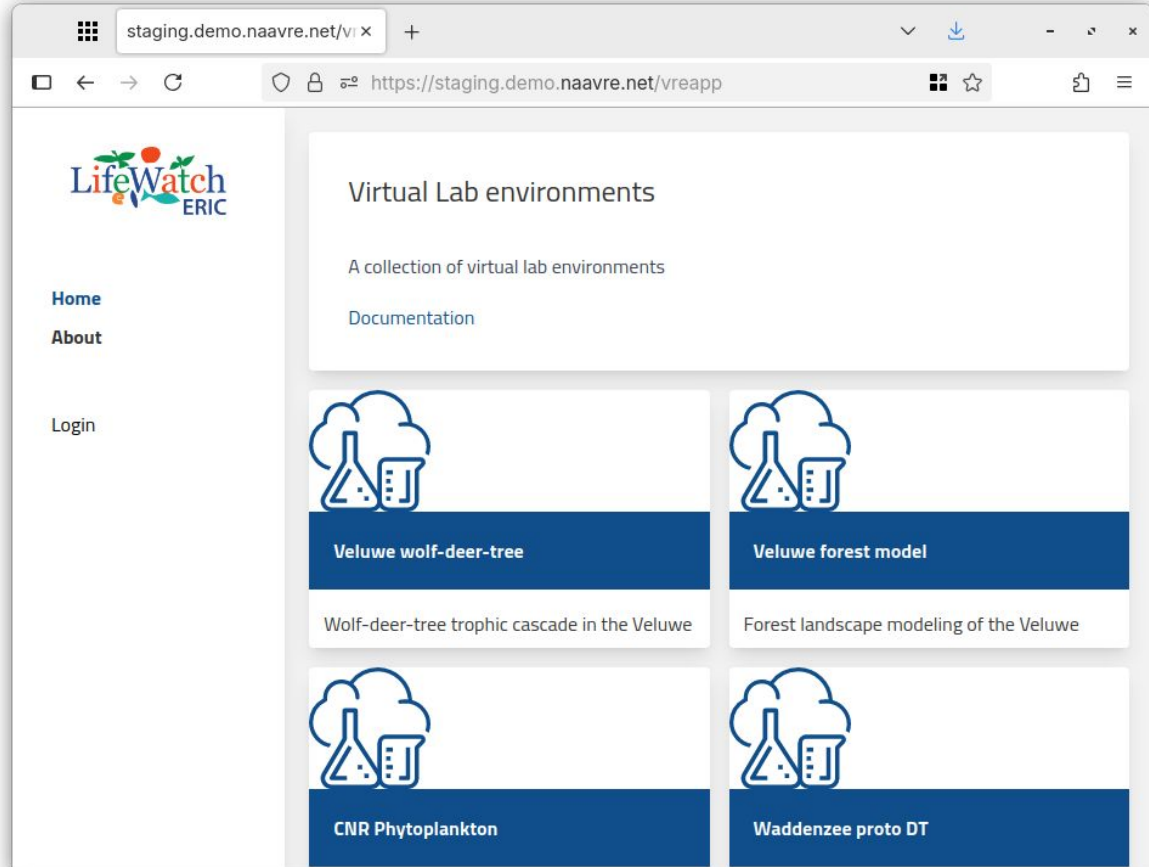
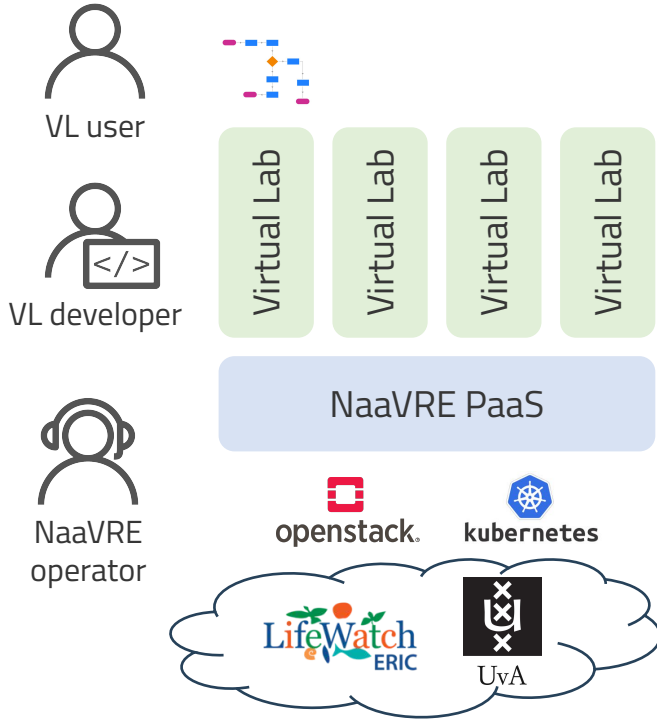


Research activities and Jupyter extensions



Zhao, Z., Koulouzis, S., Bianchi, R., Farshidi, S., Shi, Z., Xin, R., Wang, Y., Li, N., Shi, Y., Timmermans, J., Kissling, W.D.: Notebook-as-a-VRE (NaaVRE): From private notebooks to a collaborative cloud virtual research environment. Softw Pract Exp. spe.3098 (2022). <https://doi.org/10.1002/spe.3098>.

Notebook-as-a-VRE and Virtual Labs

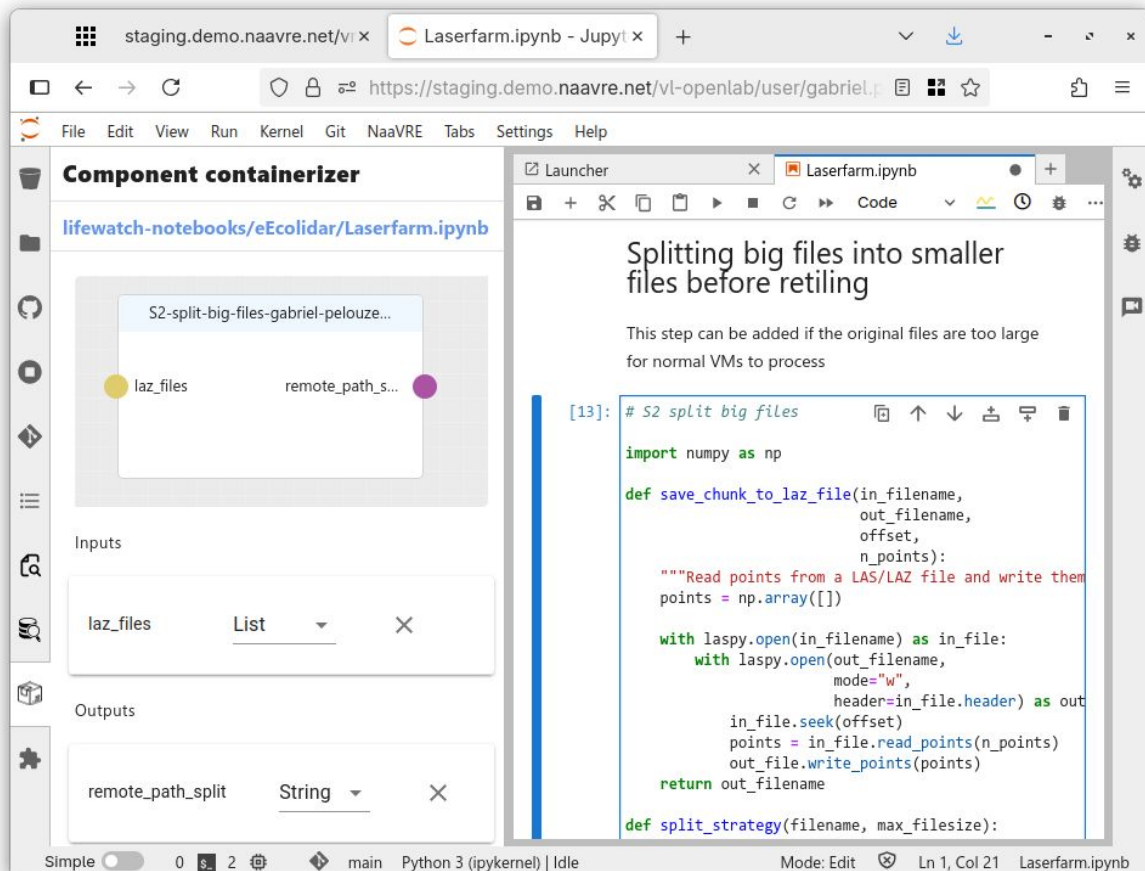


From notebooks to cloud workflows

1. PROTOTYPE MODEL IN NOTEBOOK 2. CONTAINERIZE CELLS

Encapsulate **user-selected cells** as standardized services and **containerize** them as **reusable components**.

Code cell →  docker



The screenshot displays a JupyterLab environment. On the left, the 'Component containerizer' sidebar shows a workflow component named 'S2-split-big-files-gabriel-pelouze...'. This component has two inputs: 'laz_files' (a List type) and 'remote_path_split' (a String type). The main area shows a code cell titled 'Splitting big files into smaller files before retiling'. The code cell contains the following Python code:

```
[13]: # S2 split big files
import numpy as np

def save_chunk_to_laz_file(in_filename,
                           out_filename,
                           offset,
                           n_points):
    """Read points from a LAS/LAZ file and write them
    points = np.array([])

    with laspy.open(in_filename) as in_file:
        with laspy.open(out_filename,
                        mode="w",
                        header=in_file.header) as out_file:
            in_file.seek(offset)
            points = in_file.read_points(n_points)
            out_file.write_points(points)

    return out_filename

def split_strategy(filename, max_filesize):
```

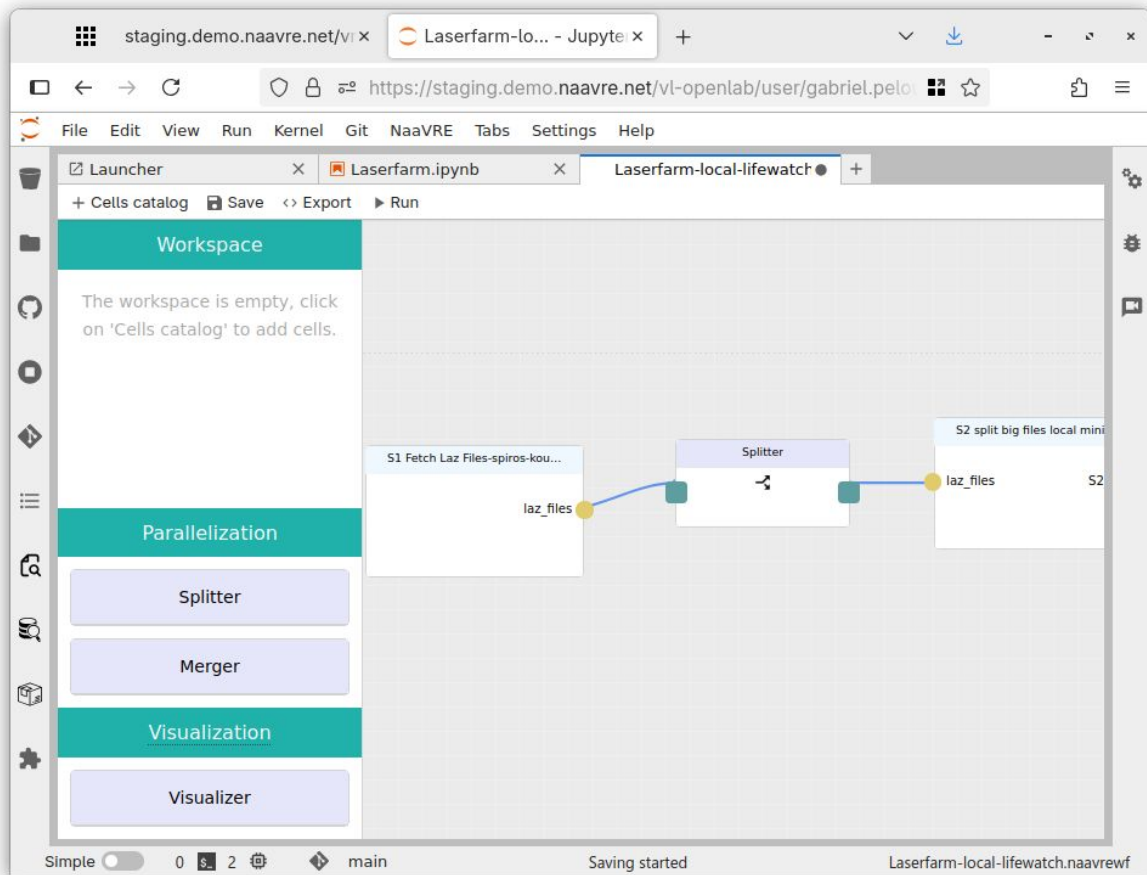
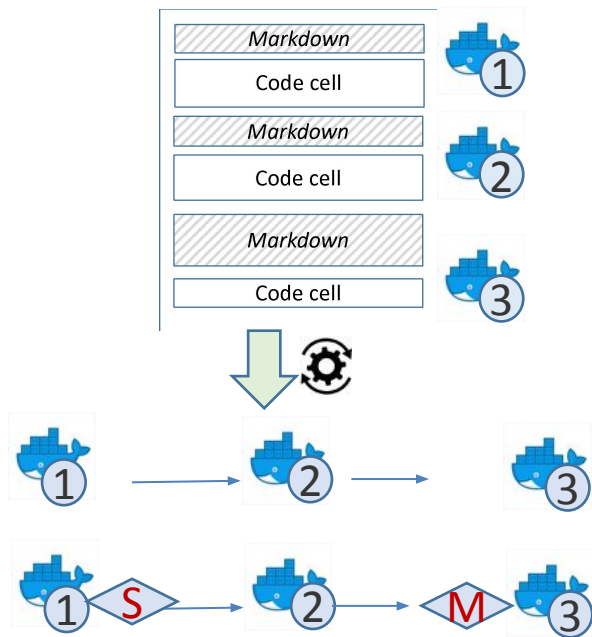
The interface also shows a menu bar with options like File, Edit, View, Run, Kernel, Git, NaaVRE, Tabs, Settings, and Help. The bottom status bar indicates the current mode is 'Edit' and the file is 'Laserfarm.ipynb'.

From notebooks to cloud workflows

3. COMPOSE A WORKFLOW:

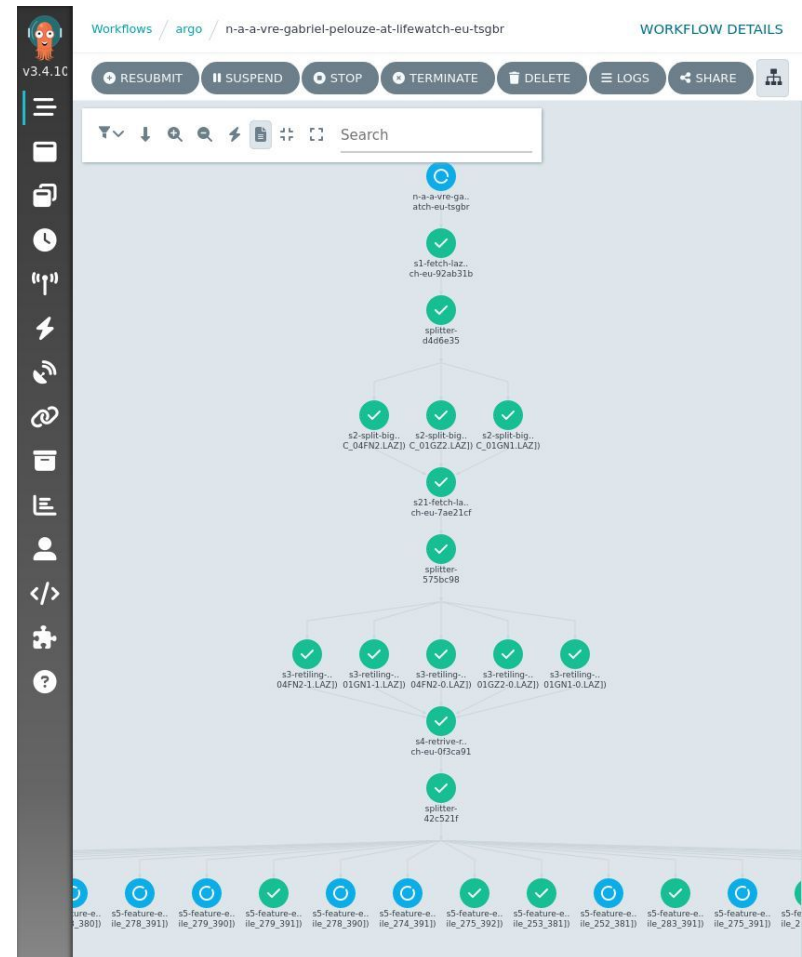
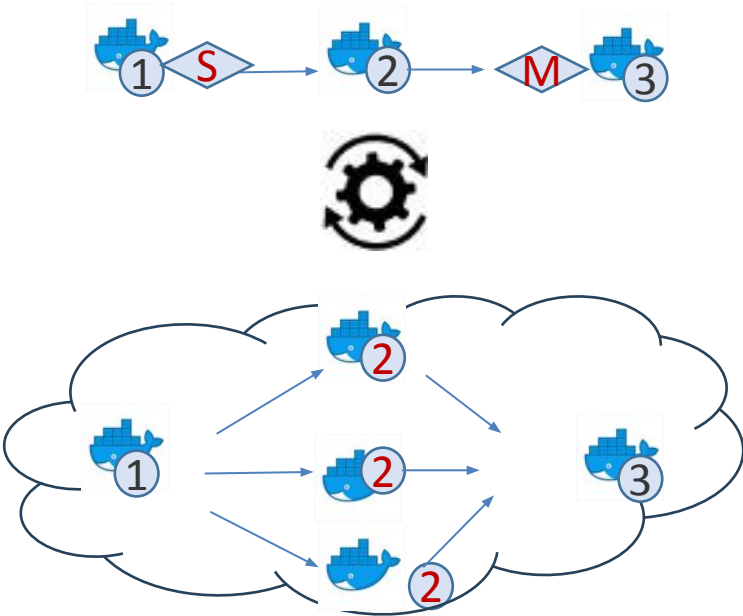
Customize data flow and input

An experiment is a workflow of services, with customized input data and infrastructure



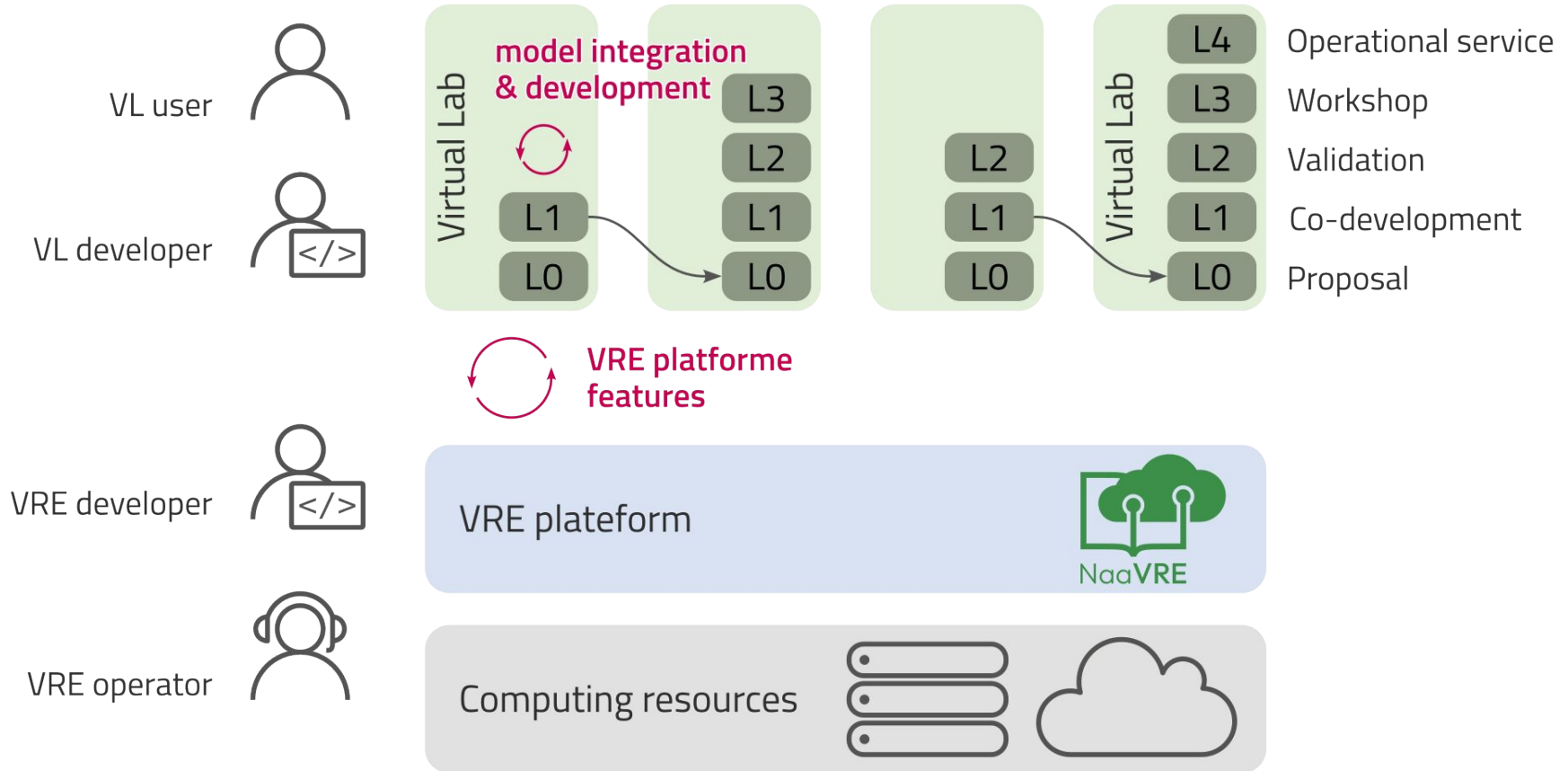
From notebooks to cloud workflows

4. RUN THE WORKFLOW: Automate the workflow components deployment and execution



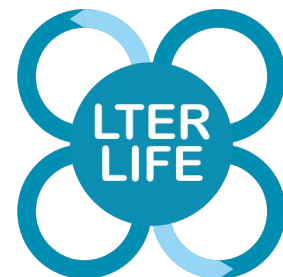
Zhou, H., Hu, Y., Ouyang, X., Su, J., Koulouzis, S., Laat, C., Zhao, Z.: CloudStorm: A framework for seamlessly programming and controlling virtual infrastructure functions during the DevOps lifecycle of cloud applications. Softw: Pract Exper. 49, 1421–1447 (2019). <https://doi.org/10.1002/spe.2741>

VRE and Virtual Lab co-development framework



Digital Twin Virtual Labs

- Veluwe (forest)
 - Bud burst prediction
 - Forest model
 - Wolf-deer-tree trophic interaction
- Waddenzee (intertidal zone)
 - Remote sensing of primary production



Digital-twin oriented developments and open challenges

Digital-twin

- Model and data FAIRification
- Model composition and runtime integration ("DT engine")
- AI in digital twins (e.g. knowledge graphs, machine learning, foundation models)

Supporting

- FAIR Digital Objects (RO-crate)
- Community engagement
- Software quality

See for yourself



Notebook as a
Virtual
Research
Environment



<https://naavre.net>

Learn, try, get in touch