An update on Yoda: using iRODS to manage data throughout your research

Lazlo Westerhof
l.r.westerhof@uu.nl

Facts & Figures

PROFESSORS

>700

FACULTIES

7+2
teaching institutes

STUDENTS

>35,000
>20,000 Bachelor
>11,000 Master
>2,510 International
117 Nationalities

Utrecht University

Strategic Themes

- Strong links between teaching & research
- Integrated approach to academic / scientific and societal issues
- Strong disciplinary research programmes
- Eleven interdisciplinary focus areas

**PATHWAYS TO SUSTAINABILITY**

DYNAMICS OF YOUTH

INSTITUTIONS FOR OPEN SOCIETIES

LIFE SCIENCES

Utrecht University

What is Yoda?

- System to preserve, share, archive and publish research data during several stages of the research process

- Integrated Research Data Management solution

- Important strategic development for reaching Research Data Management goals
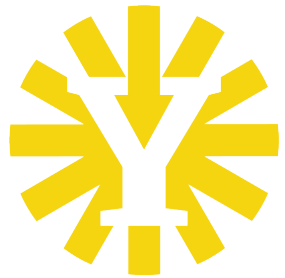
Organize your research data and work according to FAIR principles

Yoda

research data management

Utrecht University

History

- Institutional service, developed and maintained by Utrecht University
- Cross domain archive and repository
- Developed as open-source software
- Sustained funding by the board, delivered through university corporate services

- Production service as of 2015
- First data package published in 2017
- Presented at the 2018 iRODS User Group Meeting
  - https://irods.org/uploads/2018/irods_ugm2018_proceedings.pdf

Your research data and work according to FAIR principles, organize

Utrecht University

Utrecht University iRODS managed research data
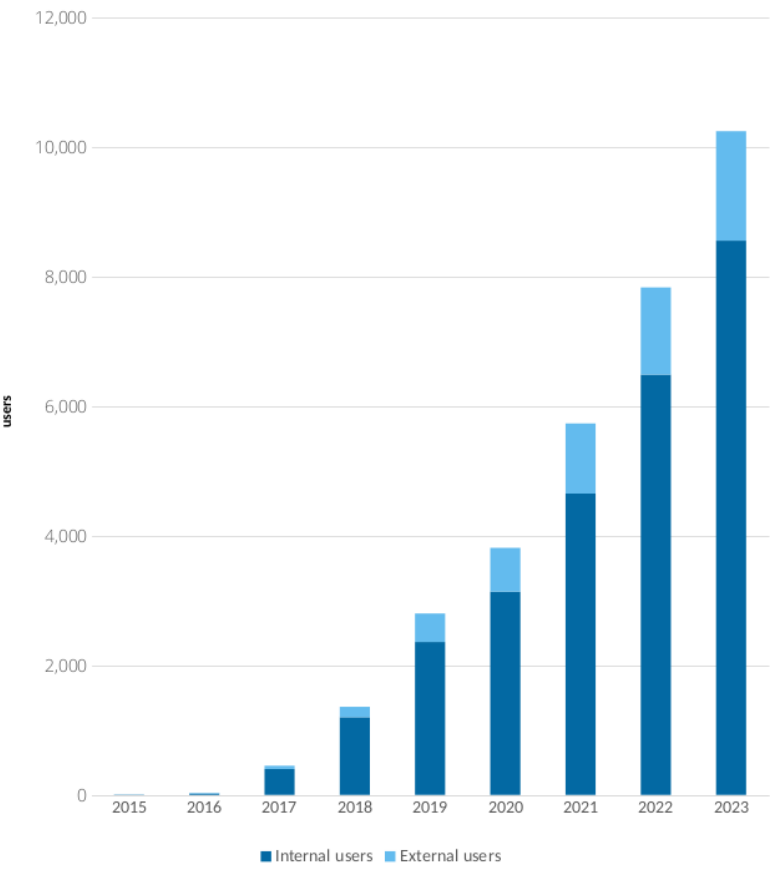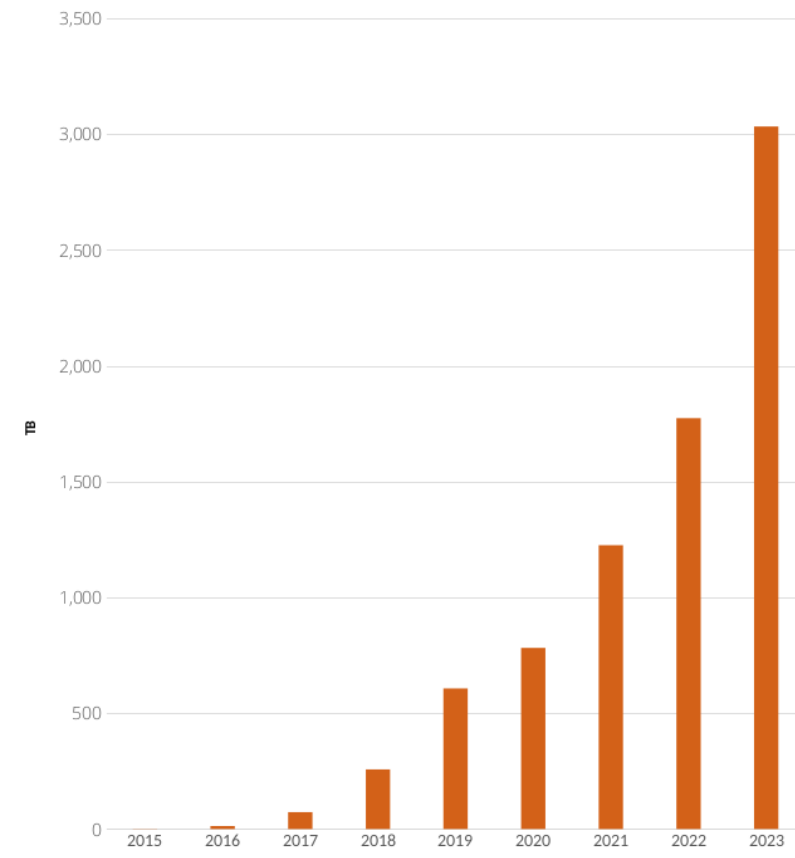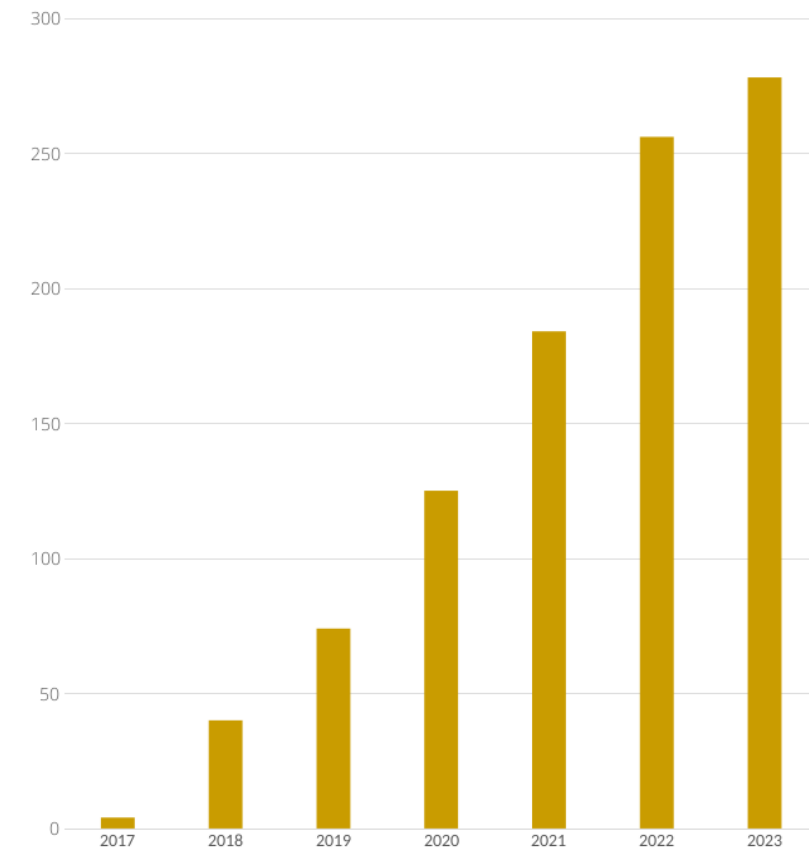
Yoda is build on iRODS
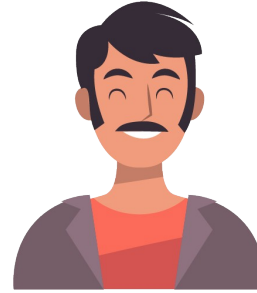
user interaction	**web portal** *https*	**network disk** *davrods*	**power-user** *iCommands*
configuration	Data policies and -services		
data integration	iRODS®		

Utrecht University

Yoda data compartments

A community comprises of multiple research groups (data compartments)

Per research group:
- members and group manager(s)
- metadata schema

Per community:
- appointed datamanager(s)
- cost calculation / invoicing

Community A

Research Group A

Research space

Vault space

Research Group B

Research space

Vault space

Community B

Research Group C

Research space

Vault space
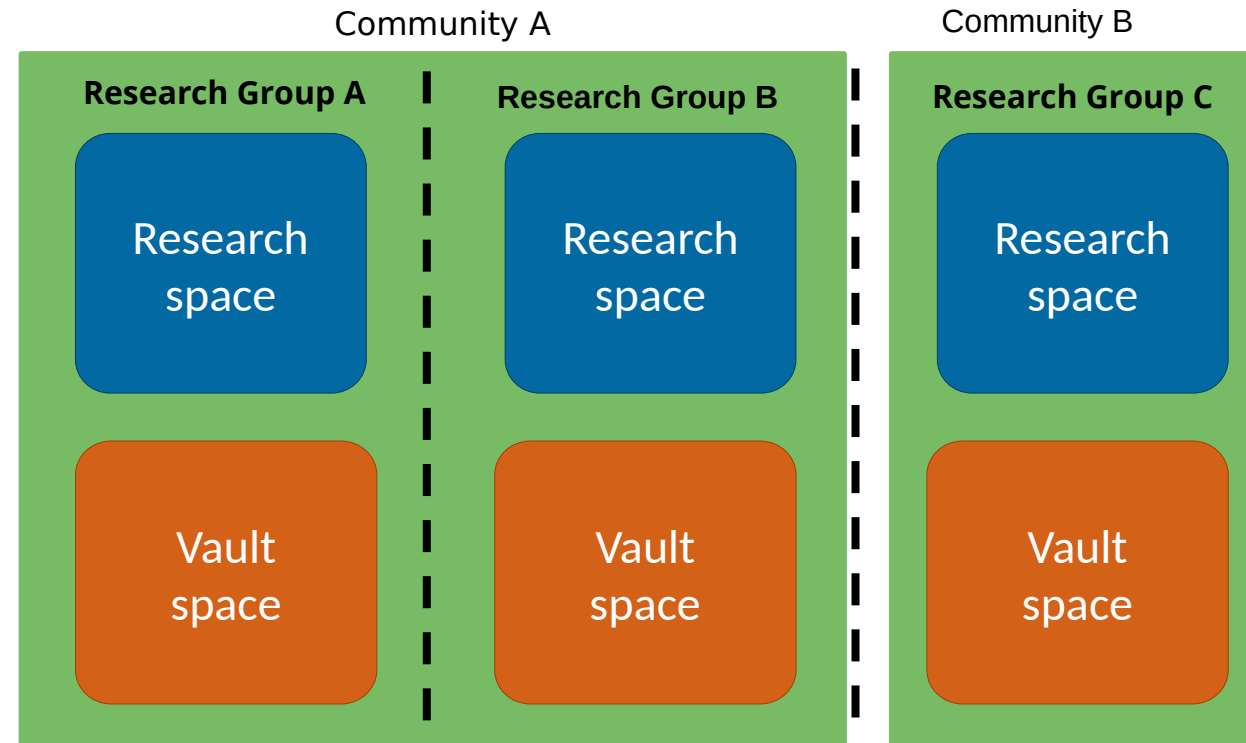
Each data compartment relates to an iRODS group
Community concept implemented as metadata on iRODS groups.

Utrecht University

Data deposit workflow

Research space

Vault space

data folder + metadata → **Submit** → **Approve** → **Secured** → data package

Researcher
requests to deposit data package

Data manager
checks if (meta)data complies with policies

System
deposits a copy in the vault

Utrecht University

FAIR publication workflow

Vault space

data package → **Submit** → **Approve** → **Published** →

DOI
landingpage
DataCite
OAI-PMH

Researcher
requests to publish
data package

Data manager
checks if (meta)data
complies with
publication policies

System
publishes the metadata and
provides internet access to
data if classified as "Open"

Utrecht
University

Yoda timeline

Year	Description	Yoda	iRODS
2015	Yoda portal and intake module for Youth project Group manager module	v0.4 - v0.9	3.3
2016	Yoda network disk (davrods module) Yoda portal supports dynamic modules	v0.9.7	4.0
2017	Research workspace, revisions, metadata form Vault archive, deposit workflow, statistics, data publication workflow (DOI)	v1.0	4.1.1
2018	Vault metadata operations, EPIC PID, External user provisioning (de/re)publication workflows, OAI-PMH harvestable	v1.1 - v1.4	4.1.1 - 4.2.5
2019	Metadata-schema management, dynamic metadata forms rendering Metadata form based on JSON schema, file up/download in portal	v1.5	4.2.5

Metadata

- XML → JSON
  - easier validation
  - more flexible
  - somewhat readable

- bidirectional conversion between JSON and iRODS AVUs
  - https://irods.org/uploads/2019/vanSchayck-Maastricht-JSON2AVU-slides.pdf

- generic way within iRODS to provide:
  - metadata templates
  - validation
  - user interaction

- metadata form defined in JSON schema
  - rendered with React
  - JSON output

Utrecht University

Python rule engine

- rewritten ruleset and developed generic wrappers
  - maintainability
  - performance
  - development speed
  - libraries and frameworks
  - available tooling

https://irods.org/uploads/2020/Westerhof-Smeele-UtrechtUni-Yoda_and_iRODS_Python_rule_engine_plugin-slides.pdf

Utrecht University

```
# iRODS rule language.
concat(*x, *y, *foo) {
    *foo = *x ++ *y;
}


# Equivalent Python rule.
def concat(rule_args, callback, rei):
    x, y = rule_args[0:2]
    rule_args[2] = x + y



# Yoda API Python rule.
@api.make()
def api_concat(ctx, foo, bar):
    return foo + bar
```

Yoda timeline

Yoda **iRODS.**

Year	Description	Yoda version	iRODS version
2015	Yoda portal and intake module for Youth project Group manager module	v0.4 - v0.9	3.3
2016	Yoda network disk (davrods module) Yoda portal supports dynamic modules	v0.9.7	4.0
2017	Research workspace, revisions, metadata form Vault archive, deposit workflow, statistics, data publication workflow (DOI)	v1.0	4.1.1
2018	Vault metadata operations, EPIC PID, External user provisioning (de/re)publication workflows, OAI-PMH harvestable	v1.1 - v1.4	4.1.1 - 4.2.5
2019	Metadata-schema management, dynamic metadata forms rendering Metadata form based on JSON schema, file up/download in Portal	v1.5	4.2.5
2020	Metadata format changed from XML to JSON, with JSON-AVU Python rule engine	v1.6	4.2.7

Python iRODS client

- migration from PHP client to Python client
- easier to build new workflows
- modules for communication with iRODS
    - connection manager module
    - API module
- modern web file uploads

https://irods.org/uploads/2021/Westerhof-UtrechtUniv-Retrospective_Migrating_Yoda_from_the_PHP_iRODS_Client_to_the_Python_iRODS_Client-slides.pdf

Utrecht University

```python
# Yoda API Python rule.
@api.make()
def api_concat(ctx, foo, bar):
    return foo + bar


# Callable from Flask frontend .
response = api.call('concat',
                    {'foo': 'test',
                     'bar': '123'})


# Callable from JavaScript frontend .
let str = await
    Yoda.call('concat',
              {'foo': 'test',
               'bar': '123'});
```
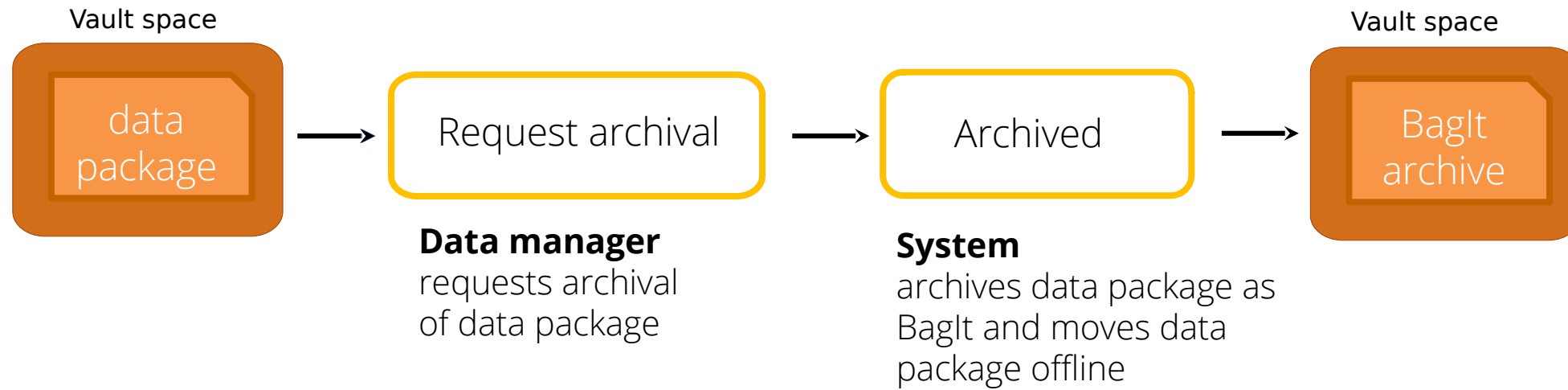
Yoda timeline

Yoda **iRODS**

Year	Features	Yoda	iRODS
2015	Yoda portal and intake module for Youth project Group manager module	v0.4 - v0.9	3.3
2016	Yoda network disk (davrods module) Yoda portal supports dynamic modules	v0.9.7	4.0
2017	Research workspace, revisions, metadata form Vault archive, deposit workflow, statistics, data publication workflow (DOI)	v1.0	4.1.1
2018	Vault metadata operations, EPIC PID, External user provisioning (de/re)publication workflows, OAI-PMH harvestable	v1.1 - v1.4	4.1.1 - 4.2.5
2019	Metadata-schema management, dynamic metadata forms rendering Metadata form based on JSON schema, file up/download in Portal	v1.5	4.2.5
2020	Metadata format changed from XML to JSON, with JSON-AVU Python rule engine	v1.6	4.2.7
2021	Yoda API, OIDC authentication Change from iRODS PHP client to Python iRODS client	v1.7	4.2.7
2022	Full theming support User settings & notifications	v1.8	4.2.11

Database

- PostgreSQL 9 → 15

- Performance improvements
    - Asynchronous processes
    - Many small files

- Prepare for OS upgrades

Data package archiving workflow

Vault space

data package

→

Request archival

Data manager
requests archival
of data package

→

Archived

System
archives data package as
BagIt and moves data
package offline

→

Vault space

BagIt archive

Utrecht University

Yoda timeline

| Year	Description	Yoda	iRODS
2015	Yoda portal and intake module for Youth project Group manager module	v0.4 - v0.9	3.3
2016	Yoda network disk (davrods module) Yoda portal supports dynamic modules	v0.9.7	4.0
2017	Research workspace, revisions, metadata form Vault archive, deposit workflow, statistics, data publication workflow (DOI)	v1.0	4.1.1
2018	Vault metadata operations, EPIC PID, External user provisioning (de/re)publication workflows, OAI-PMH harvestable	v1.1 - v1.4	4.1.1 - 4.2.5
2019	Metadata-schema management, dynamic metadata forms rendering Metadata form based on JSON schema, file up/download in Portal	v1.5	4.2.5
2020	Metadata format changed from XML to JSON, with JSON-AVU Python rule engine	v1.6	4.2.7
2021	Yoda API, OIDC authentication Change from iRODS PHP client to Python iRODS client	v1.7	4.2.7
2022	Full theming support User settings & notifications	v1.8	4.2.11
2023	Metadata per research group, DOI versioning Data package archiving workflow, database improvements	v1.9	4.2.12

Yoda consortium

Launched March 2023

Safeguard the development of Yoda
as a national RDM platform

Effective collaboration on:
- knowledge base
- support for researchers
- product development
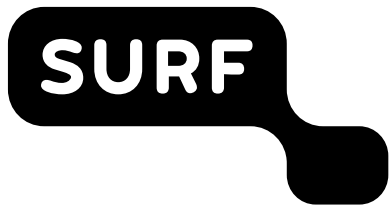
Future work

Upgrade to iRODS 4.3.x:
- Migrate ruleset to Python 3
- Migrate asynchronous processes to delayed rule engine

OS upgrades:
   - AlmaLinux 8
   - Ubuntu 20.04

Proof of concepts:
- iRODS S3 client

Docker test environment

- Containerized version of Yoda
  - Developed for local development
  - Continuous integration
  - Demonstrations

More information:

https://utrechtuniversity.github.io/yoda/development/docker-setup.html

```
$ git clone https://github.com/UtrechtUniversity/yoda.git
$ cd yoda/docker/compose
$ docker compose pull
$ ../up.sh
```

Utrecht University
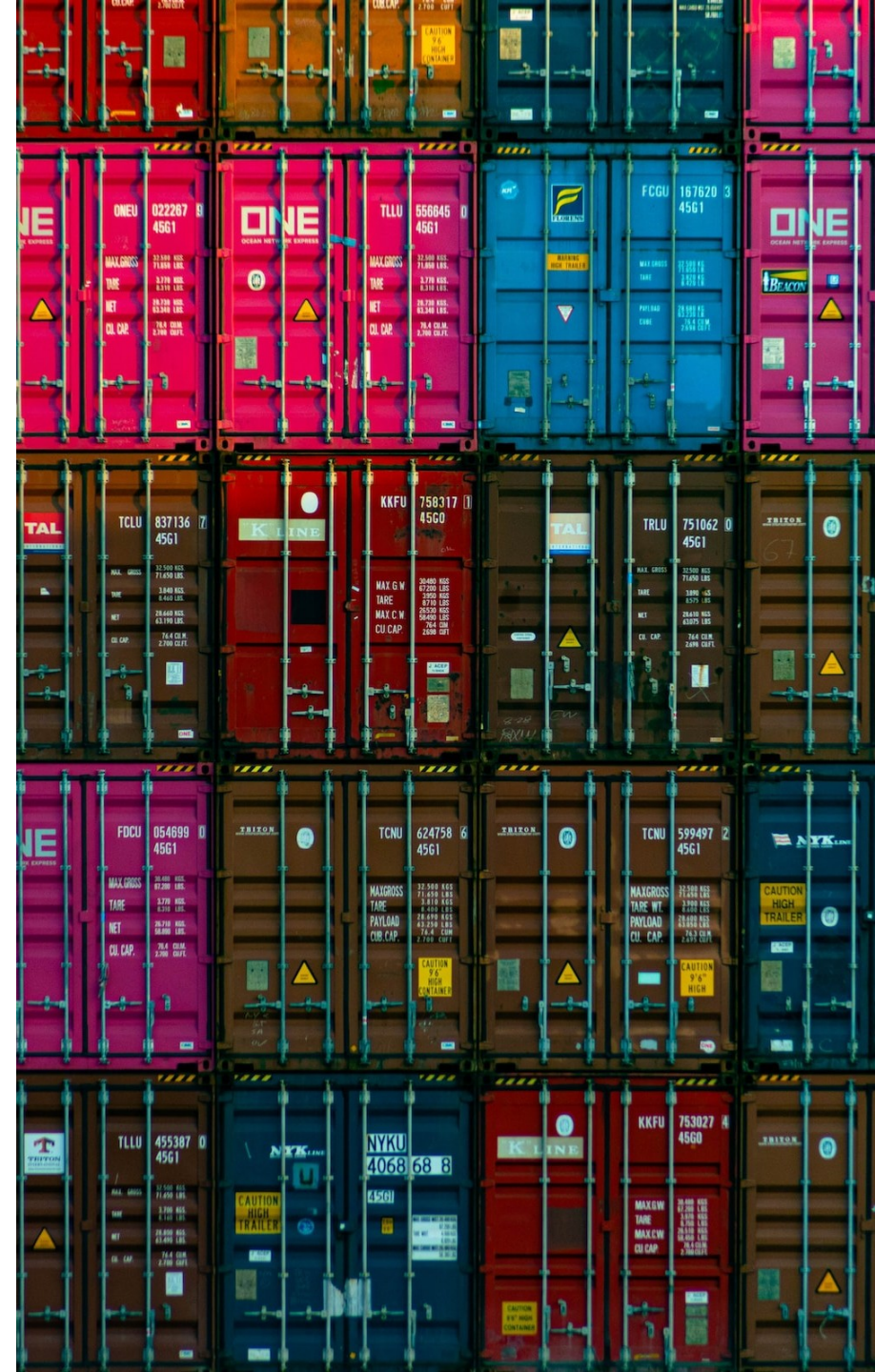
$ iexit

Utrecht University

Sharing science,
shaping tomorrow