

Where's the data?

Data storage at the university's compute center

jan.leendertse@rz.uni-freiburg.de



Agenda

Solutions at the university of Freiburg

Storage planning as process

bwSFS as platform for an UFR ecosphere

Storage solutions

Storage system	Provider	System	Self-Service
Homedirectory	RZ	Mounting as drive	√
Group share (Isilon)	RZ	Classic file storage	
Group share (bwSFS)	RZ	Hot/warm research data	
bwSync&Share	KIT	Cloud storage (Nextcloud)	√
FreiData	UFR	Publication of research data	√
OpenCloud (in discussion)	RZ	Flexible data handling	
TSM tape library	RZ	Cold research data	

Data lifecycle as planning scheme

Phases in lifecycle are connected

Data changes through lifecycle

Access rights changes between phases

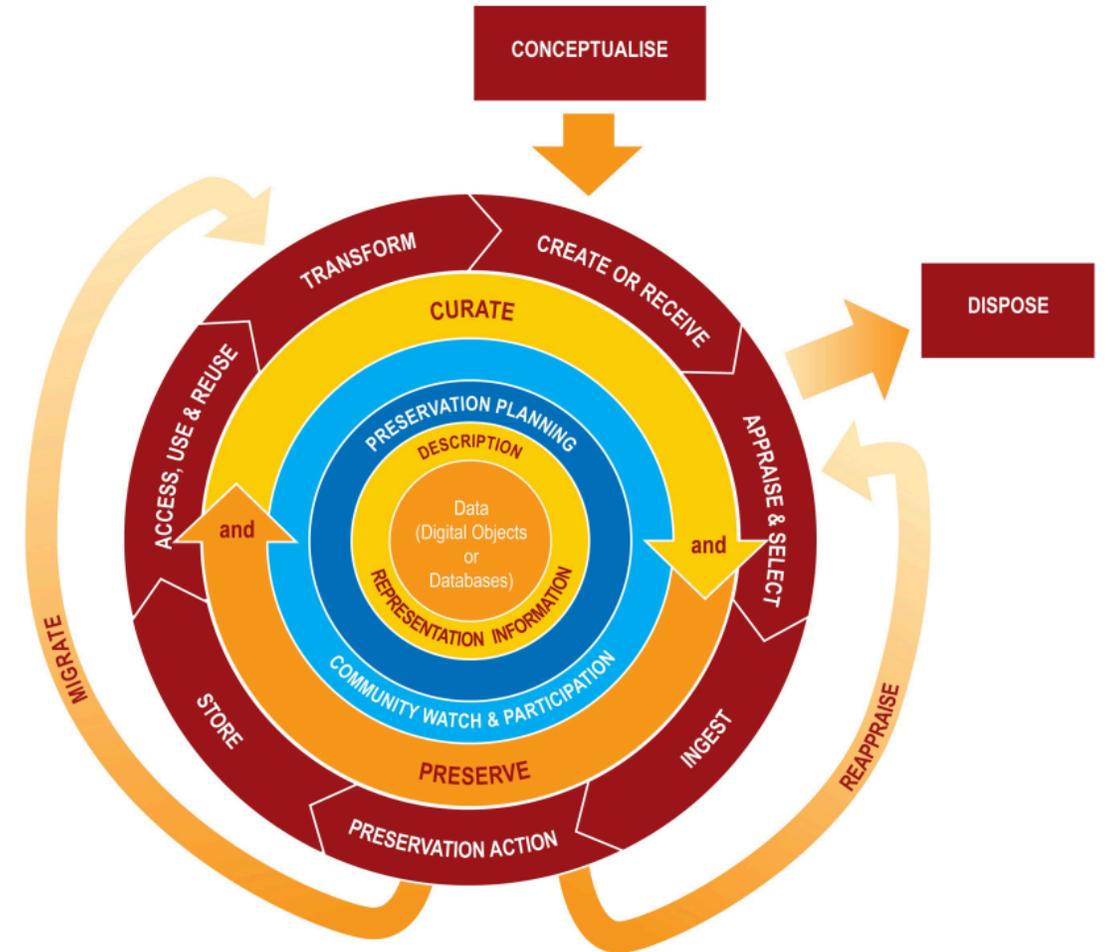


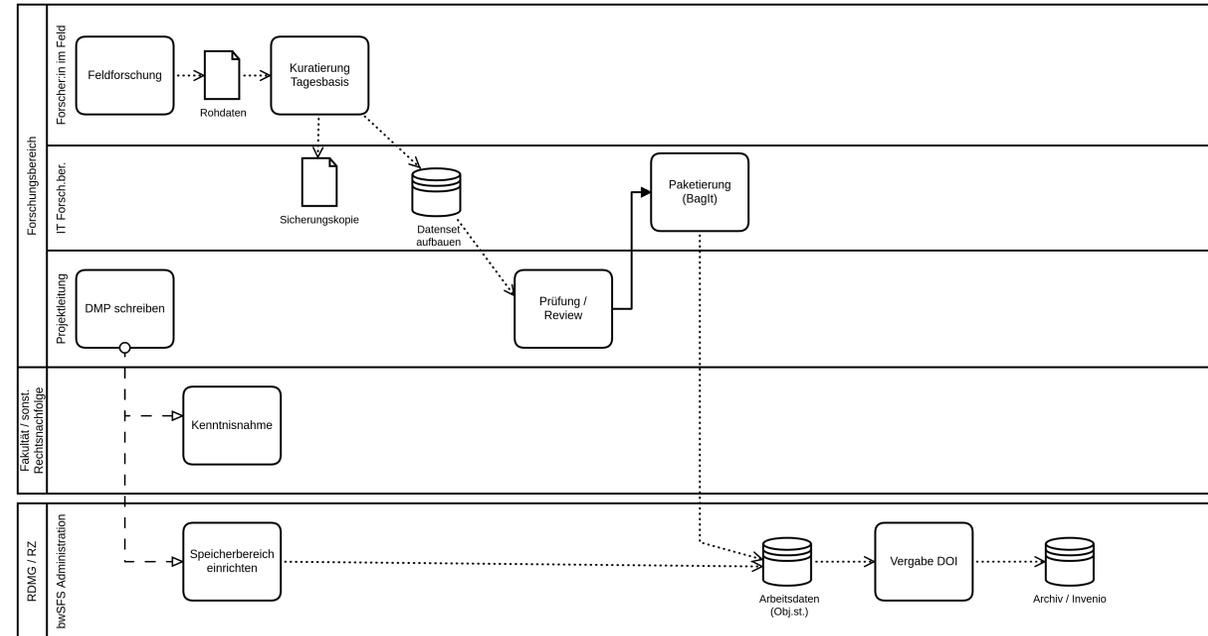
Figure 1. The DCC Curation Lifecycle Model.

Example from archaeology

Process oriented graphics in BPMN

Data produced on external camp sites

Connectivity between storage systems part of planning



Yet another lifecycle model

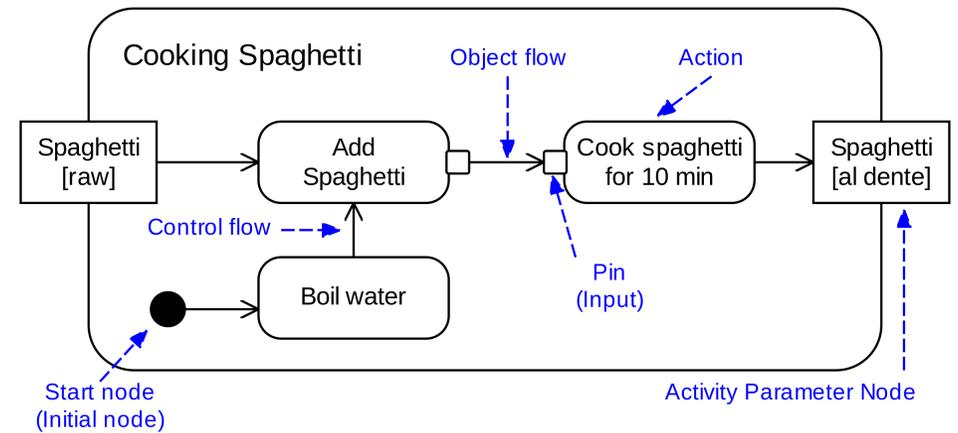
See <https://cdf.uni-freiburg.de/en/rdm-catalogue/>

Data management is a continuous process, and storage planning is an (important) detail



Process modelling

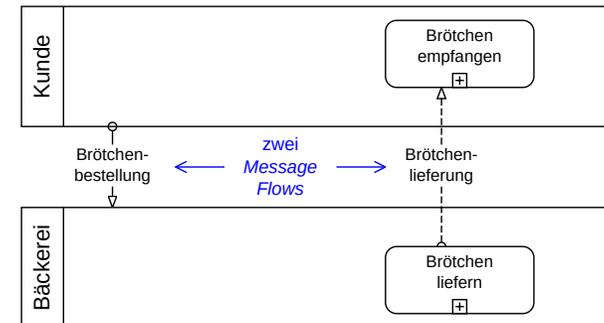
UML



Another notation tool

Business process notation

http://www.bpmb.de/images/BPMN2_0_Poster_DE.pdf



Template with service layers

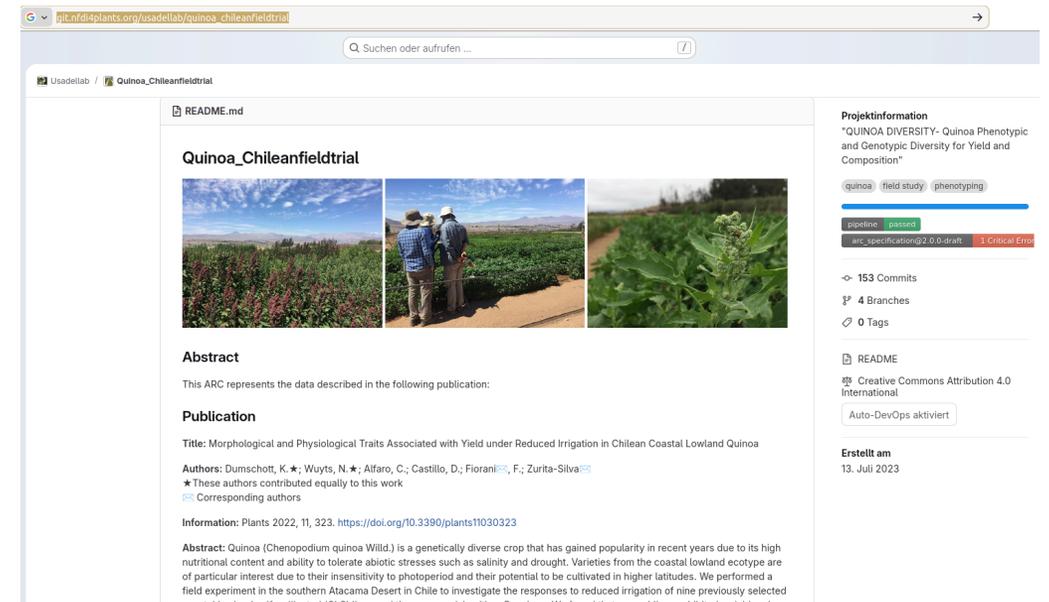
Phases in linear and with layer switching

		Plan	Collect	Process	Analyse	Preserve	Share	Reuse
Service Layers	Work group							
	University							
	N.N. (External Prov.)							

Workflow for datasets

Transparent repository

https://git.nfdi4plants.org/usadellab/quinoa_chileanfieldtrial



The screenshot shows a GitHub repository page for 'Quinoa_Chileanfieldtrial' under the 'usadellab' organization. The repository name is 'Quinoa_Chileanfieldtrial'. The page displays a README.md file with the following content:

Quinoa_Chileanfieldtrial



Abstract

This ARC represents the data described in the following publication:

Publication

Title: Morphological and Physiological Traits Associated with Yield under Reduced Irrigation in Chilean Coastal Lowland Quinoa

Authors: Dumschott, K. ★; Wuyts, N. ★; Alfaro, C.; Castillo, D.; Fiorani, F.; Zurita-Silva

★ These authors contributed equally to this work

Corresponding authors

Information: Plants 2022, 11, 323. <https://doi.org/10.3390/plants11030323>

Abstract: Quinoa (Chenopodium quinoa Willd.) is a genetically diverse crop that has gained popularity in recent years due to its high nutritional content and ability to tolerate abiotic stresses such as salinity and drought. Varieties from the coastal lowland ecotype are of particular interest due to their insensitivity to photoperiod and their potential to be cultivated in higher latitudes. We performed a field experiment in the southern Atacama Desert in Chile to investigate the responses to reduced irrigation of nine previously selected

Project information: "QUINOA DIVERSITY- Quinoa Phenotypic and Genotypic Diversity for Yield and Composition"

quinoa field study phenotyping

pipeline passed

arc_specific/arc2.0.0-draft 1 Critical Error

153 Commits

4 Branches

0 Tags

README

Creative Commons Attribution 4.0 International

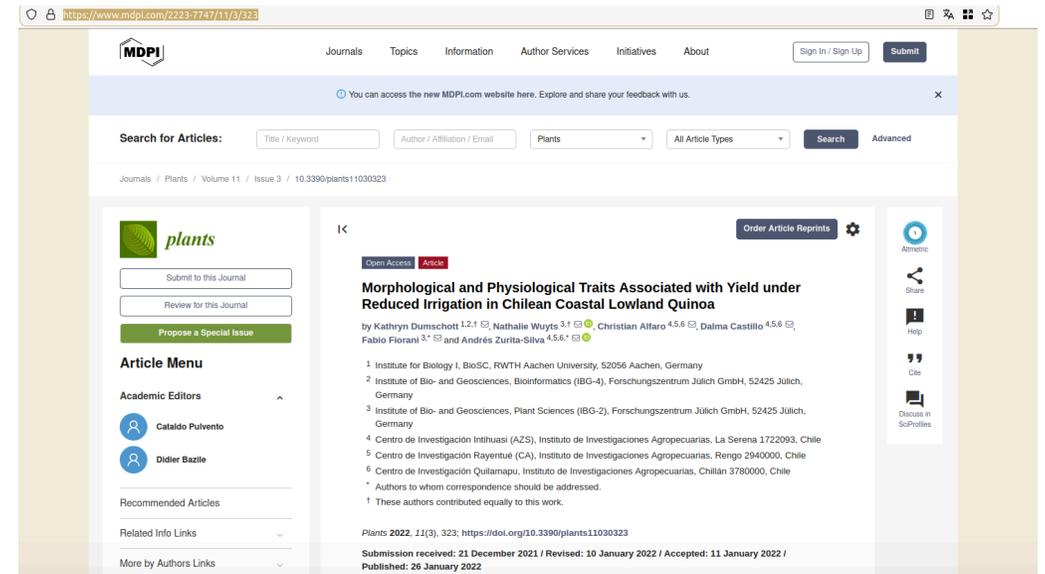
Auto-DevOps aktiviert

Erstellt am 13. Juli 2023

Data publication

Depending on automation citable
publication on prepared publication
platform

External platform



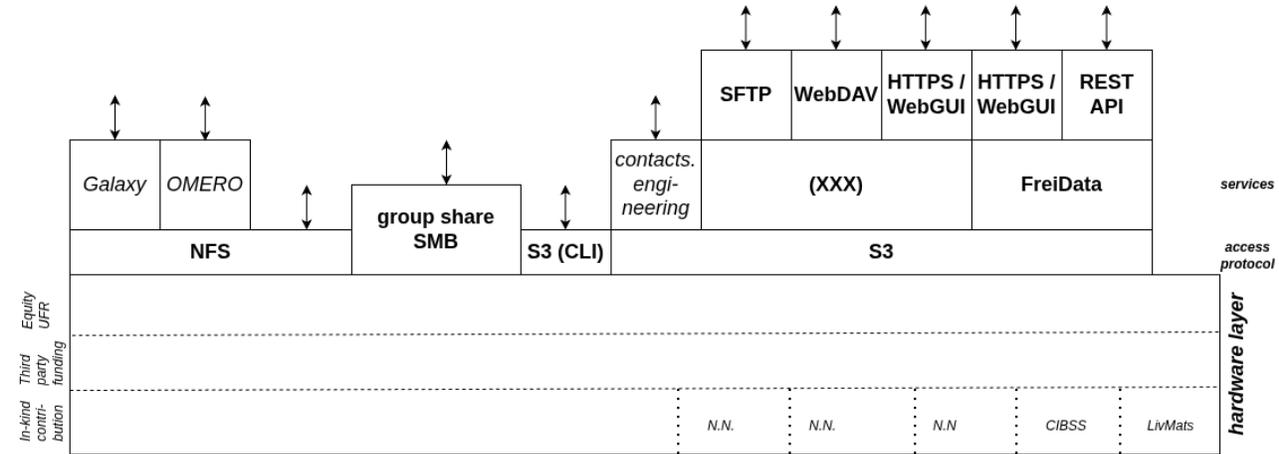
How to contribute

bwSFS is a storage cluster with additional service layers

Contributions open on several layers

Use-Cases

- Databases for Imaging
- Data-Sharing
- Publications
- Hierarchical file structures for curating data
- ...



Open questions and projects

- Offering different storage access protocols
- Implementation of ACL (access control lists) in responsibility of the research groups itself
- visualization of common data structures
- Visualization Fair-Use and paid for

Provider / Contributor

- bold** → computer center
- italic* → others
- bold/italic*** → mixed

Legend

- ↕ Data flow
- Fair-Use
- Paid-Extra for significant higher needs