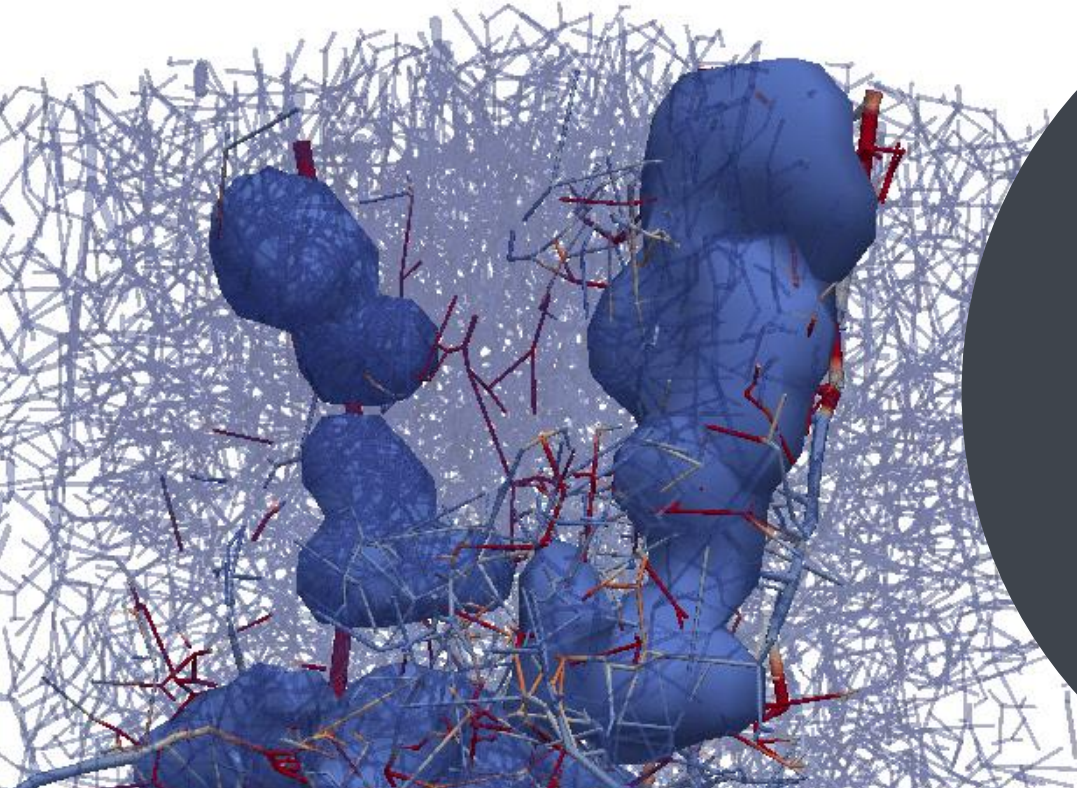


University of Stuttgart
Germany

E-Science-Tage 2019:
Data To Knowledge
27-29 March 2019
Heidelberg



Sustainable infrastructure for the integration of research software in data repositories

using the example of DUNE/DuMu^x

Anett Seeland¹
Timo Koch²
Sibylle Hermann³
Bernd Flemisch²

¹ Technische Informations- und Kommunikationsdienste (TIK), Uni Stuttgart

² Institut für Wasser- und Umweltsystemmodellierung (IWS), Uni Stuttgart

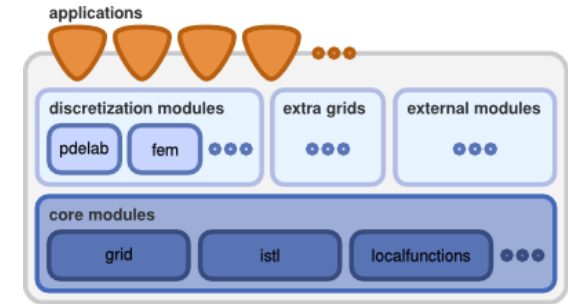
³ Universitätsbibliothek, Uni Stuttgart

Background

DuMu^x and DUNE



- DUNE – a numerical software framework for solving PDEs
 - Developed at over 10 universities in Europe
 - Open-source development model / license
 - Highly *modular*, loosely connected modules
 - Over 20 years of development
 - Highly templated generic C++ code
- DuMu^x – application module, porous media simulator
 - Modular structure
 - Open-source development model / license
 - Ca. 10 years of development



get Dune / Dumux at

<https://www.dune-project.org/>
<http://dumux.org/>

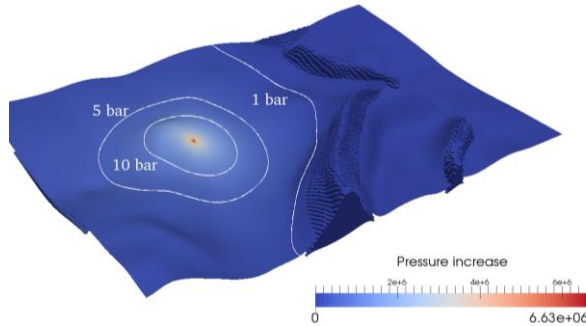
repositories at

<https://gitlab.dune-project.org/groups/core>
<https://git.iws.uni-stuttgart.de/dumux-repositories/>

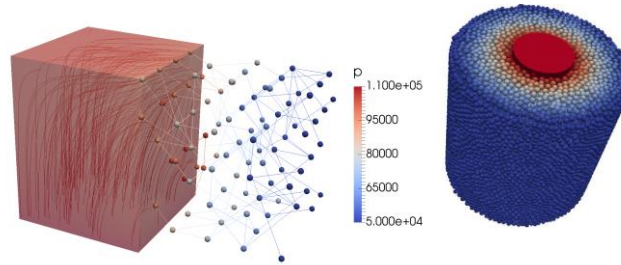
DuMu^X – DUNE for Multi-{Phase, Component, Scale, Physics, ...} flow and transport in porous media



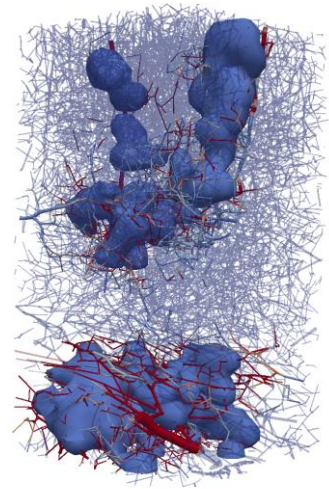
Groundwater contamination
(Alexander Kissinger)



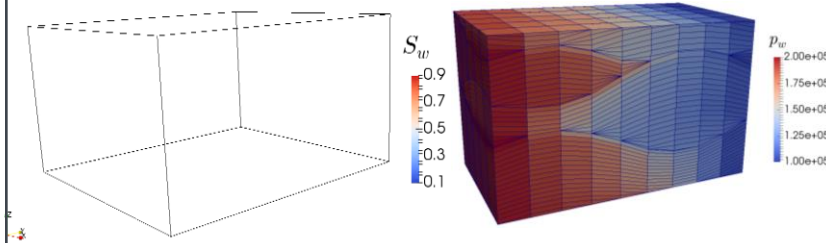
Porennetwork – Darcy coupling
(Kilian Weishaupt, T. K.)



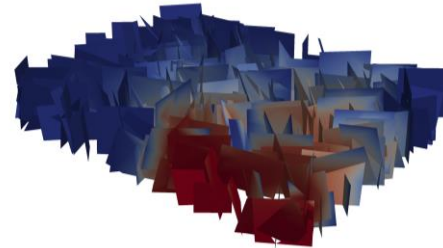
Capillary networks
1D-3D coupling (T. K.)



Two-phase flow on cornerpoint grids
with NL-TPFA (Martin Schneider)



Discrete fracture networks
(Dennis Gläser)



DuMu^x – DUNE for Multi-{Phase, Component, Scale, Physics, ...} flow and transport in porous media



- mainly developed at LH2 (Dept. of Hydromechanics and Modelling of Hydrosystems)
- “...has had [10,647 commits](#) made by [80 contributors](#) representing [182,340 lines of code](#)” (Black Duck Open Hub)
- “...took an estimated [47 years of effort](#) (COCOMO model [person-years]) starting with its [first commit in July, 2010](#) ending with its [most recent commit 2 days](#) ago” (Black Duck Open Hub)
- Over 100 publications with the help of DuMux
 - from these: 25 journal articles, 8 PhD theses without collaborator from LH2
- Since 2015 at LH2 Uni Stuttgart: DuMux-PUB
- Every publication (B/M/PhD theses and journal) gets a public module containing
 - Source code + instructions, Dockerfile, Docker image

DuMu^x – DUNE for Multi-{Phase, Component, Scale, Physics, ...} flow and transport in porous media



Problems

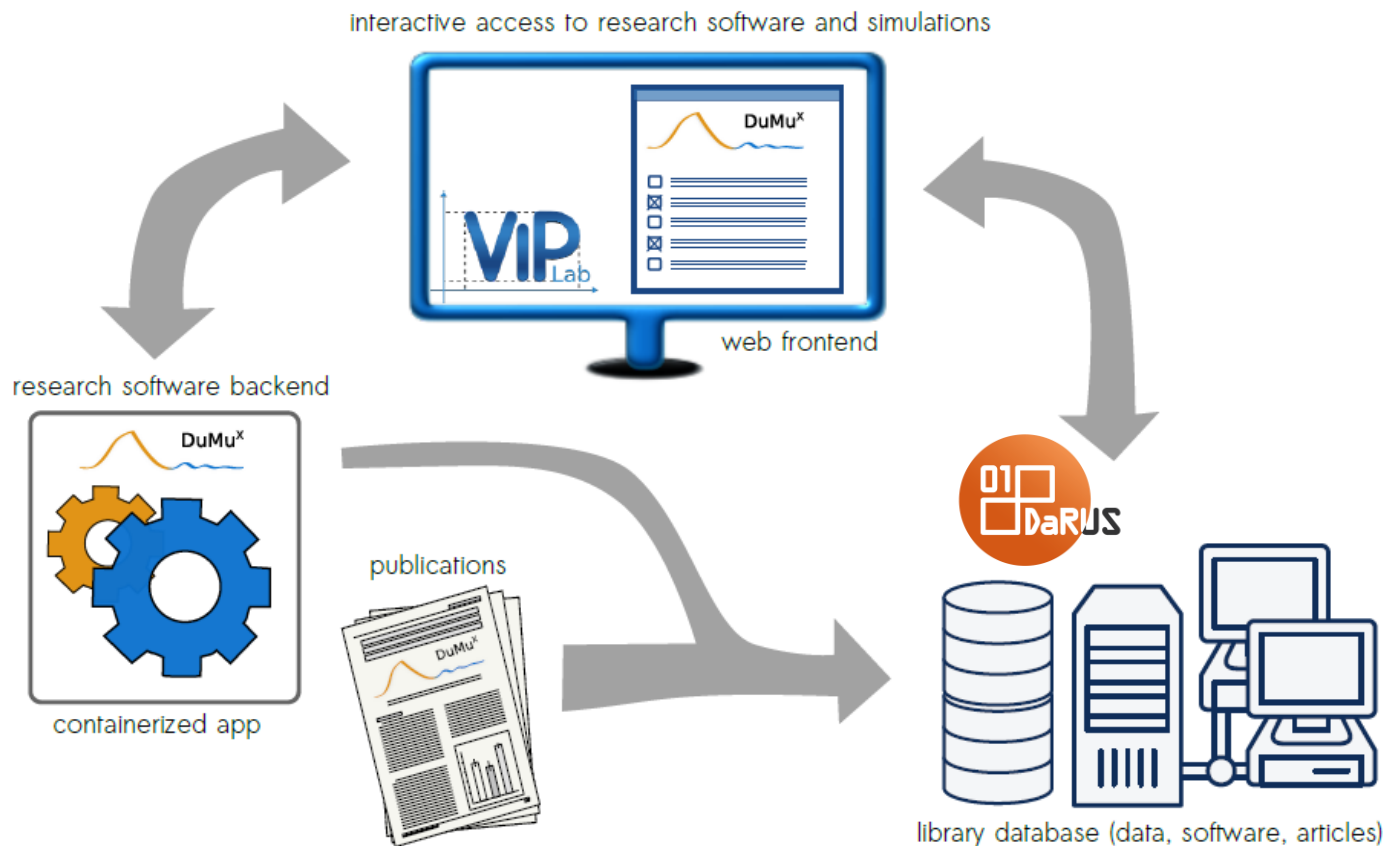
- Generic C++ code with heavy use of templates (**hard for beginners**)
- Many changes **require programming skills** and need recompilation
- **Installation** of software and **dependencies** requires some programming knowledge
- Simulations hard to **reproduce** for a scientific peer without knowledge of DUNE

DFG Project SuSI

Solution strategy

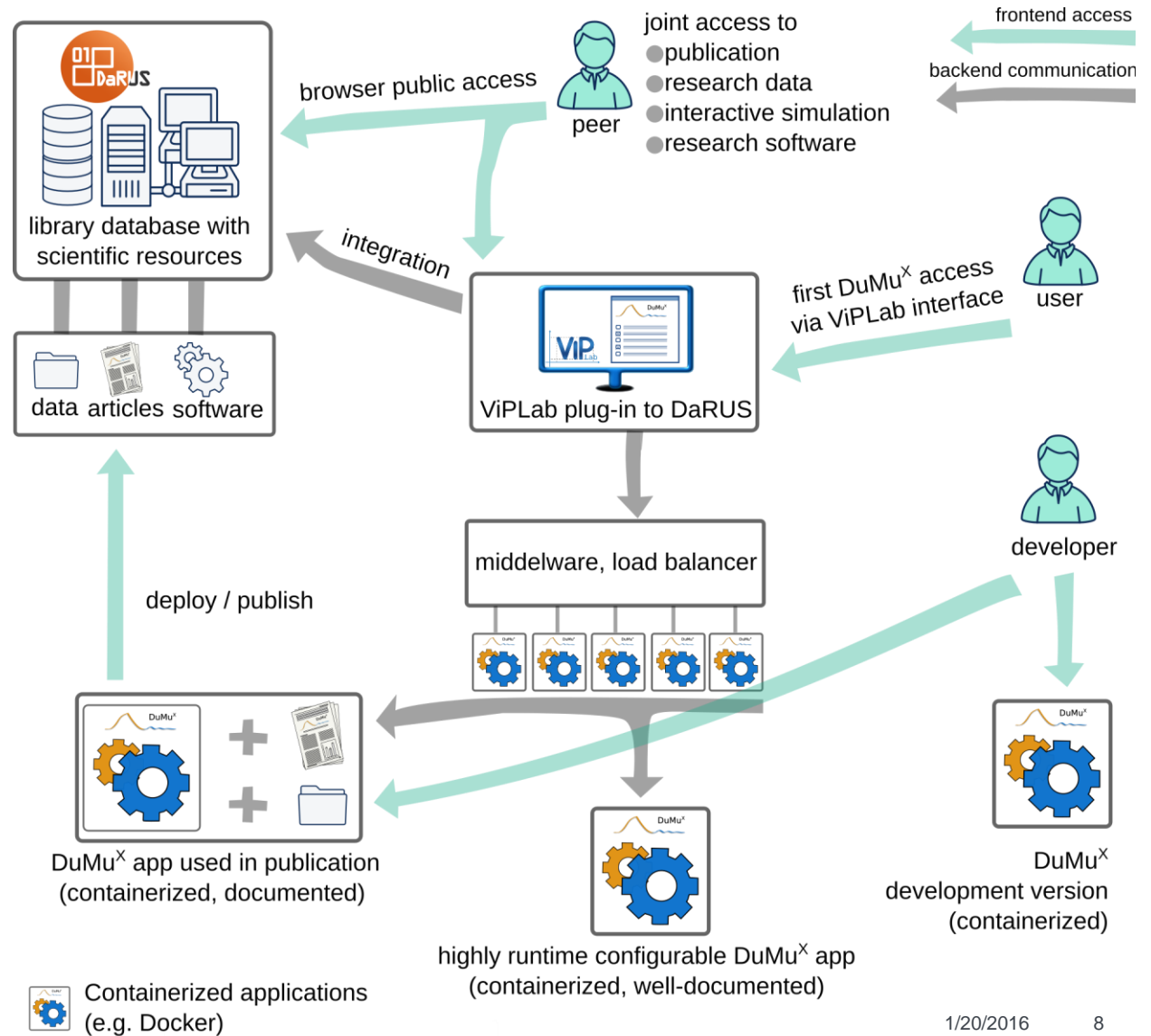
DFG Project SuSI

Since February 2019 (funding within LIS, Wissenschaftl. Literaturversorgung und Informationssysteme)



DFG Project SuSI

Project details



DFG Project SuSI

1

Improve usability of DuMux on the **research software** side
(1.1) Improve source code (1.2) Containerization

2

Web frontend for maximum **accessibility** and (re-)usability (ViPLab)

3

Integral and sustainable concept for **storing data, software**, publication in data repository (DaRUS) with ViPLab plug-in

1 Containerization

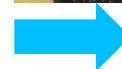
Containerization

- Encapsulates OS, Dependencies and source code + binaries (e.g. Docker container)
- Three modes of data
 - Source code
 - Instruction file (Dockerfile), descriptive installation file with version numbers
 - Binary Image (Docker image), used to launch a container
- Different encapsulated environments for
 - First application with DuMux
 - Developing DuMux
 - Running a demonstrator or an application (associated with publication)
 - Automated testing

2 ViPLab

The Virtual Programming Lab ViPLab ...

- Is a project at TIK since 2009 together with some institutes of the University of Stuttgart
- Aims:
 - not only theoretical tasks but also programming
 - programming tasks for lectures with large numbers of students (e.g. numerics)
- Supported and initialized by students and financed in part by tuition fees
- Making own programming experience is very important - also for the professional training of a Bachelor student
- Motivation: consistent exercise and examination environment
- Is used since 2010 in exercises/ tutorials



Working with computer

ViPLab-DuMuX example for teaching

ViPLab Demo-DuMuX Grundwasser-Simulation

Finish the Test

Homepage Grundwasser

Question 1 of 1
Not answered (editing)

Actions ▾

DuMuX-Grundwasser-Simulator

Source

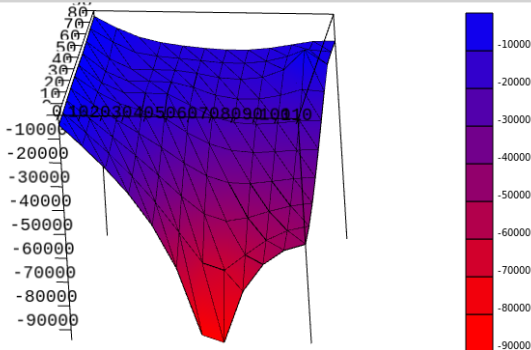
```
1 [TimeManager]
2 TEnd = 1.0
3 DtInitial = 1.0
4
5 [Grid]
6 UpperRight = 120 100 # coordinates of upper right corner of rectangle
7 Cells = 12 10 # number of cells in each coordinate direction
8 Depth = 20 # depth of the domain in z direction (a single cell)
9
10 [BoundaryConditions]
11 # Insert Boundary conditions in the following format:
12 # Divide the boundary in segments (intervals)
13
14 # Start-Coordinate End-Coordinate Type Value
15
16 # Type is 0 for Dirichlet boundary conditions and 1 for Neumann BC
17 # Dirichlet values are given in [m]
18 # Neumann values are given in [m/s]
19 # Example 10 20 1 0 denotes a no-flow BC between 10m and 20m
20
21 Left = "
22 0 100 0 50
23 "
24
25 Right = "
26 "
27
28 Bottom = "
29 60 80 1 -0.5
30 "
31
32 Top = "
33 "
34
35 [SpatialParams]
36 Permeability = 1e-4 # the background permeability k_f [m/s]
37
38 # Position and permeability of lenses
39 # Insert in the following format:
40
41 # Left Right Bottom Top Permeability
42
43 Lenses = "
44 20 50 50 80 1e-2
45 "
46
47 [Source]
48 # Source terms at specific points
49 # Insert in the following format:
50 # X-Coordinate Y-Coordinate Flux(in m³/s)
51
52 Sources = "
53 100 70 62.5
54 "
```

Elliot Carver: Mr. Jones, are we ready to release our new software?
Jones: Yes, sir. As requested, it's full of bugs, which means people will be forced to upgrade for years.
- James Bond, Tomorrow Never Dies

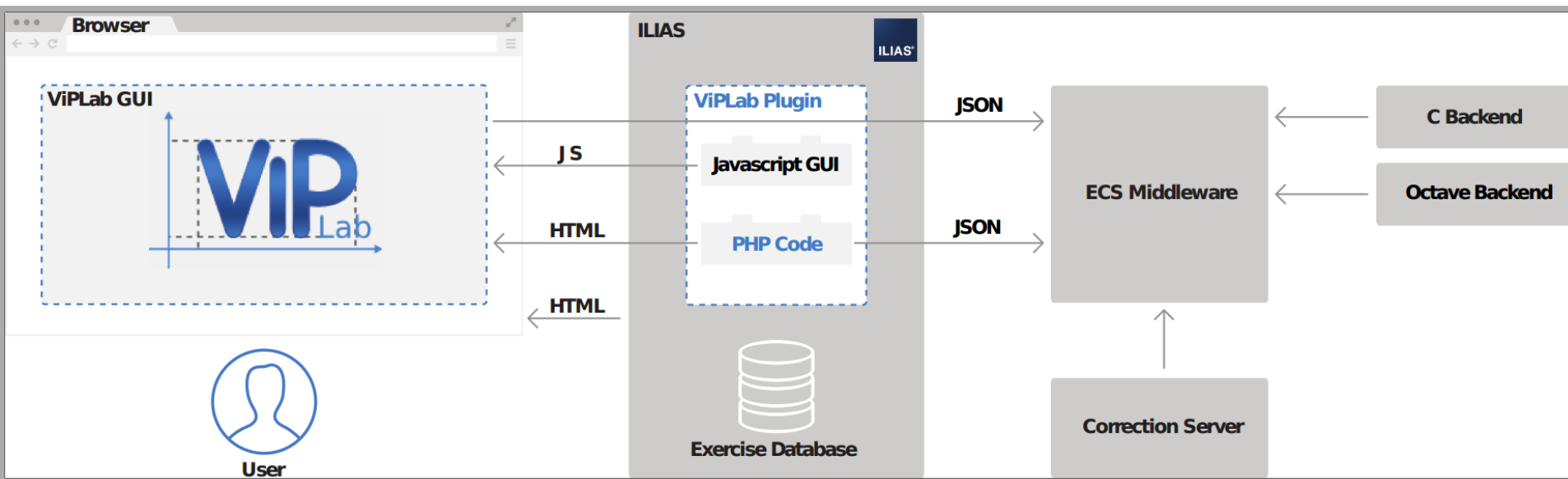
Reading parameters from file dumux-in.
Initializing problem 'groundwater'
Initialization took 0.0006015 seconds on 1 processes.
The cumulative CPU time was 0.0006261 seconds.
Writing result file for current time step

x	y	h	v_x	v_y
5	5	-4157	0.08511	-0.0009625
15	5	-1.276e+04	0.08924	-0.00317
25	5	-2.201e+04	0.09898	-0.006567
35	5	-3.256e+04	0.118	-0.01241
45	5	-4.56e+04	0.1541	-0.02373

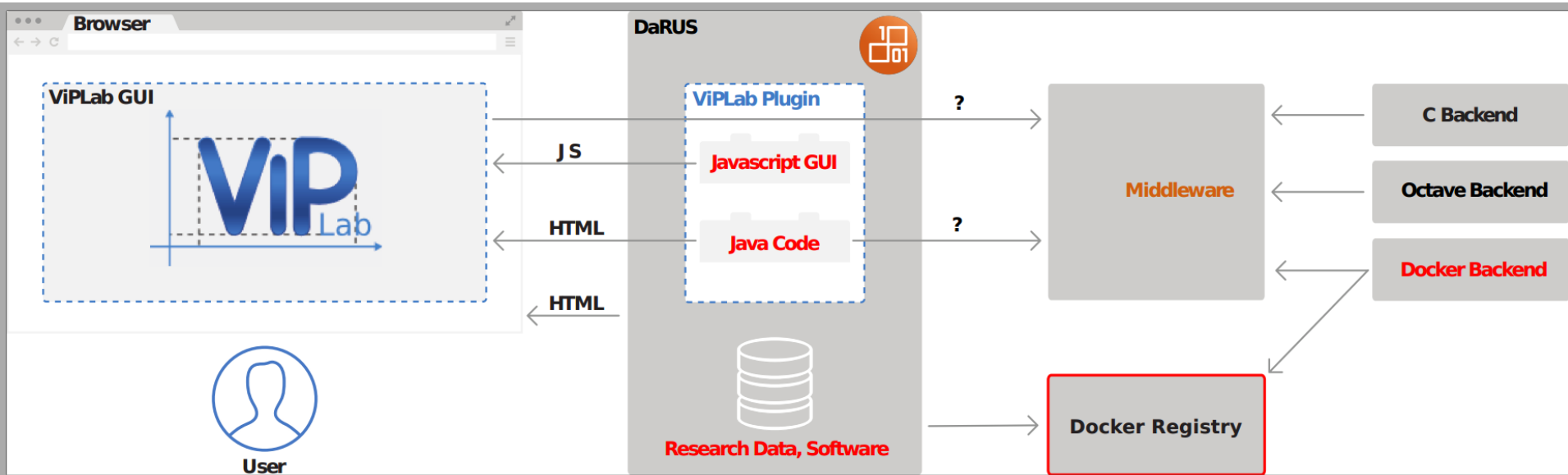
Interpolated 3D Plot



ViPLab-Architecture Changes



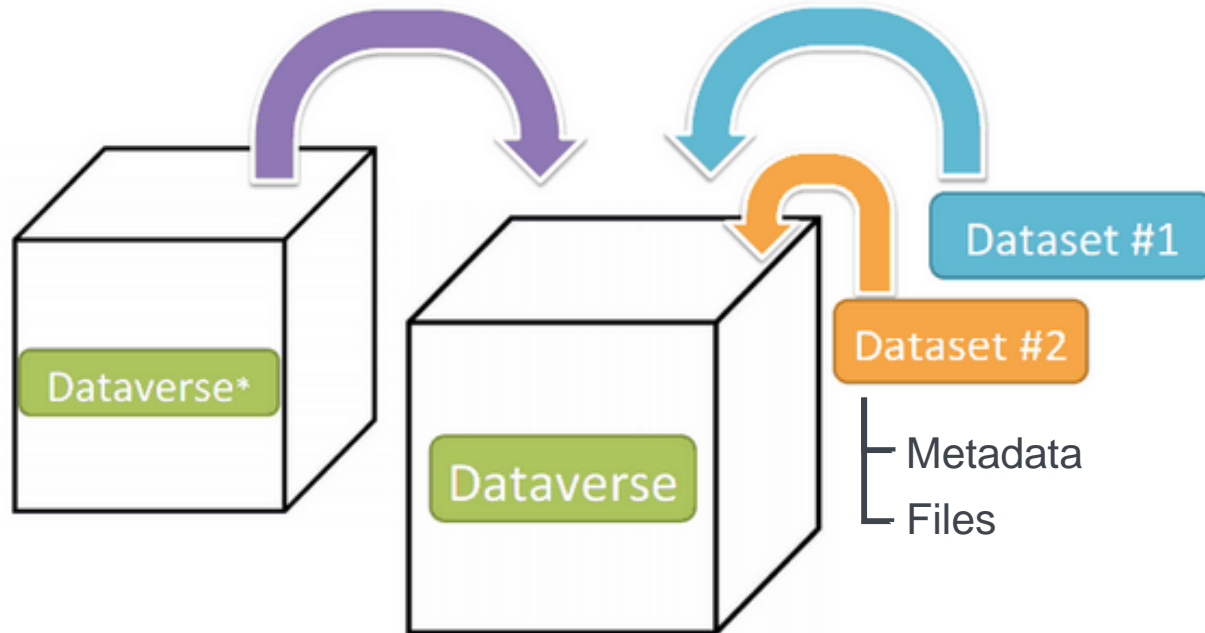
ViPLab-Architecture Changes



3 DaRUS

DaRUS

Data Repository of the University of Stuttgart



Based on Dataverse:

- Open source research data repository software
- Repository hosts multiple virtual archives called Dataverses

Software Dataset

- Archivability of software
- Persistent identifier
- Versioning
- Metadata: CodeMeta, Schema.org
 - Support for automatic metadata extraction

Datanverse

Search - User Guide Support Log In

Dataset Metadata ▾

Software Description ▲

Software Title Software associated with publication "High-resolution behavioral mapping of electric fishes in Amazonian habitats"

License MIT license (MIT)

Description "Fishtracker" is the two-step algorithm described in Madhav, Jayakumar et. al, 2017 to track multiple freely moving weakly electric fish using measurements from a grid of electrodes in a known configuration (see "2_Fishtracker"). Also included is code used to generate the data and results figures from the Madhav, Jayakumar et. al., 2017 (see "3_PaperFigs") and code for analyzing laboratory data (see "4_AnalysisCommon").

Artifact Type Source Code

Date Published 2017-04

Code Repository Link <https://github.com/mmadhav/fishtracker>

Programming Language MATLAB

Function Data collection / organization; Visualization

Interaction Method Graphical; Programmatic

Software Contributors Madhav, Manu Developer
Jayakumar, RaviKrishnan Developer

Dependencies Software Package or Platform - MATLAB v. R2016a
Library / Module - CircStat2012a, Hungarian, MagnetGInput, SCIRTS, addpath, recuse, datastructure, dispense, distinguishable_colors, export_fig, findobj, ginputa, matlab, ndhfilter, parfor, progress, plot, release, progressbar, savefast, serialization, subplotplot. Refer to README in 2_Fishtracker for more information.

Other related software Interoperable Software - File 3_PaperFigs (download available from this project): code used to generate the data and results figures from the Madhav, Jayakumar et. al., 2017.
Interoperable Software - File 4_AnalysisCommon (download available from this project): code for analyzing laboratory data. analysisData.m from each lab data folder included in this project uses code from this folder to initially process the raw data.

- Use Plugin Interface of Dataverse for ViPLab
- Adapt “Explore” feature of Dataverse for Docker Containers

Metrics

0 Downloads

Contact

Share

Publish

Edit

Susl Docker Explore

Draft

Unpublished

Seeland, Anett, 2019, "Susl Docker Explore", <https://doi.org/10.5072/darus-187>, DaRUS, DRAFT VERSION

Cite Dataset

Learn about Data Citation Standards

Description

This is a test on how the explore button of ViPLab could be integrated into Dataverse.

Subject

Computer and Information Science

Related Publication

Related Publication with DOI

Files

Metadata

Terms

Versions

Search this dataset...

Find

1 to 2 of 2 Files

Upload Files

Edit Files

Download

Dockerfile

Plain Text - 2.6 KB - Mar 25, 2019 - 0 Downloads

MD5: 93707ec5a1e3c7852b5fdefc75e51d0d

Docker spec file

Docker

Explore

Download

pubtable_koch2017a

Plain Text - 1.4 KB - Mar 25, 2019 - 0 Downloads

MD5: 671df10e1ef96310d82e51d82fd7069e

Bash script to sync Docker image

Docker

Download

DFG Project SuSI

Discussion

Summary / Outlook

- Improve the usability and sustainability of the research software DuMu^x
 - Archivability
 - Linking to publication, reproducibility
- With the help of the infrastructure services ViPLab (TIK) and DaRUS (UB)
 - Solution strategy independent of software

Next steps:

- May '19: Kick-Off Workshop with DuMu^x user community
- Summer '19: DaRUS production launch

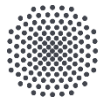
Summary / Outlook

- Improve the usability and sustainability of the research software DuMu^x
 - Archivability
 - Linking to publication, reproducibility
- With the help of the infrastructure services ViPLab (TIK) and DaRUS (UB)
 - Solution strategy independent of software

Next steps:

- May '19: Kick-Off Workshop with DuMu^x user community
- Summer '19: DaRUS production launch

Questions? Comments?



University of Stuttgart
Germany

Thank you!



e-mail

phone +49 (0) 711 685-

fax +49 (0) 711 685-

University of Stuttgart

DuMu^x – DUNE for Multi-{Phase, Component, Scale, Physics, ...} flow and transport in porous media



Typical applications

- Input Data → 1-100 MB
- Compile time → up to 2 min
- Runtime → from seconds to days (on cluster)
- Output → 10 MB to 10 GB

DuMu^x – DUNE for Multi-{Phase, Component, Scale, Physics, ...} flow and transport in porous media



Quality assurance:

- Version control with git
- Issue tracker, mailing lists
- Transparent development → GitLab, (GitHub, Bitbucket,...)
- Code review process, partly automated → merge requests, commit hooks, and
- Automated testing and benchmarks
- Documentation (code, models, assumptions, parameters with literature)
- User meetings / courses
- Write maintainable, reusable, (generic) code

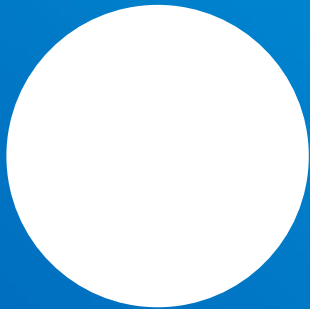
Related Work

- Commercial services → costs, privacy
 - Shiny App
 - Code Ocean
- (Jupyter)-Notebooks → not for complex software
 - Binderhub
- Docker for Software storage → no execution environment
 - Docker Hub (Zenodo for DOI)



University of Stuttgart
Germany

Thank you!



e-mail

phone +49 (0) 711 685-

fax +49 (0) 711 685-

University of Stuttgart