Project EnhanceR:
Enabling Open Science, Shareable, reproducible research containers

Sergio Maffioletti
EnhanceR project director
https://www.EnhanceR.ch

EnhanceR: enhancing Research through IT expertise

- A federation of Research IT specialists in Switzerland
- Offer consulting and technical services to research groups
- Grow provider capacities and capabilities
- Knowledge transfer to researchers
- Work with other national services and networks
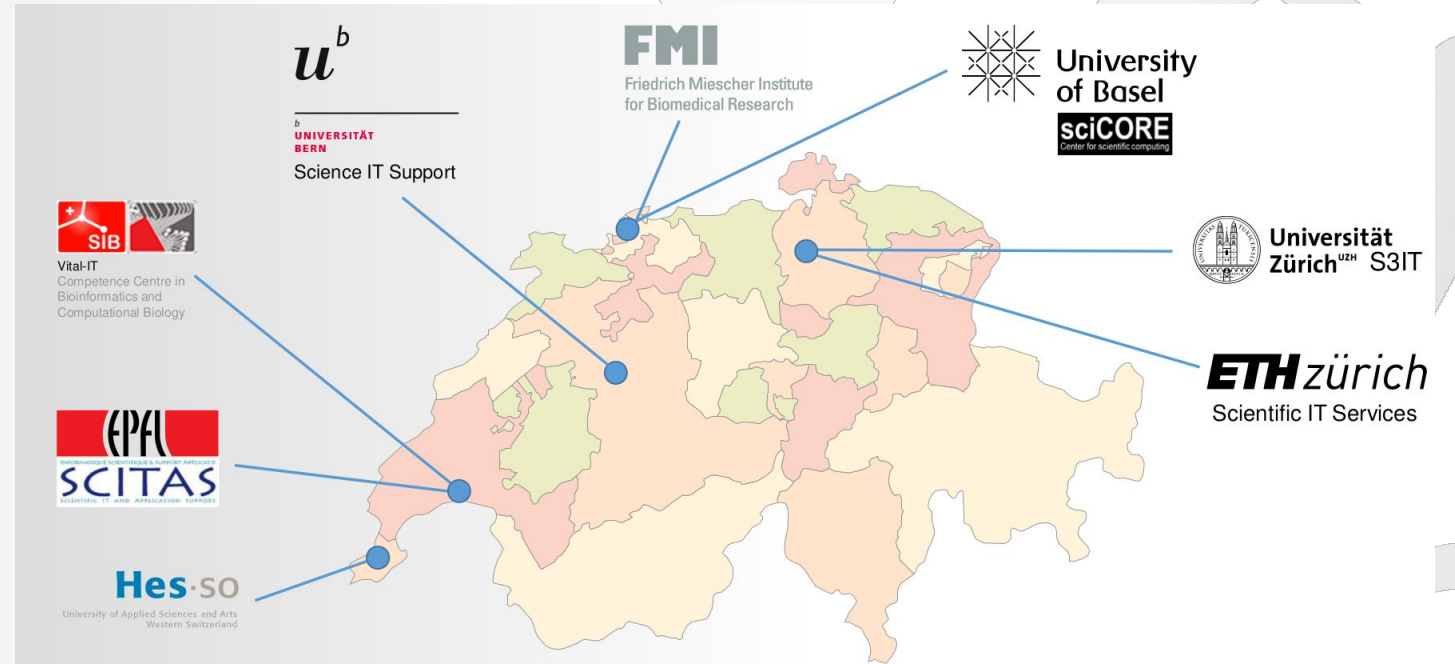
Open Science: principles

- ## Shift to 'data culture'
  - Build idea into process, approaches, dissemination and work
  - Data as the initial output of research. Papers, insights and discovery as the second stage output.
- ## Openness by default
  - Require openness unless there is a clear case for privacy
  - Open source software, open formats, open repositories
- ## FAIR
  - Use as underlying principle and practical guide
- ## Building skills
  - All activities include training and knowledge transfer
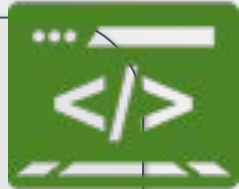 - Explicit goal of raising community capability

Supporting open science

Know-How

- Consulting researchers
- Best practices
- Identify upcoming challenges

Technology

- New tools and services
- Continuum of infrastructure: Desktop, Cluster, Cloud, HPC
- Data analysis in containers

Engage with end-users and communities

Education

- National Training program
- DMP templates & support
- Legal advice and information

Policies

- Align with funding agencies
- Business models & services for sustainability
- Governance challenges in federated system

EnhanceR | Enhancing Research through IT Expertise

https://www.enhancer.ch

Develop container-based data analytics,

deploy and validate on the supporting infrastructures.

Driven by personalised medicine initiative: SPHN

- Cutting edge science using private medical data
- Strong need for collaboration and portability of data AND analysis
- Bring computation to data

https://www.enhancer.ch/casestudies

Focus on Reproducible research: Data Analytics

Technical challenges

- Heterogeneous infrastructure and different technologies

Policy/Human challenges

- Establish new practises for researchers to develop their data analytics and validate the reproducibility and interoperability

https://www.enhancer.ch/casestudies
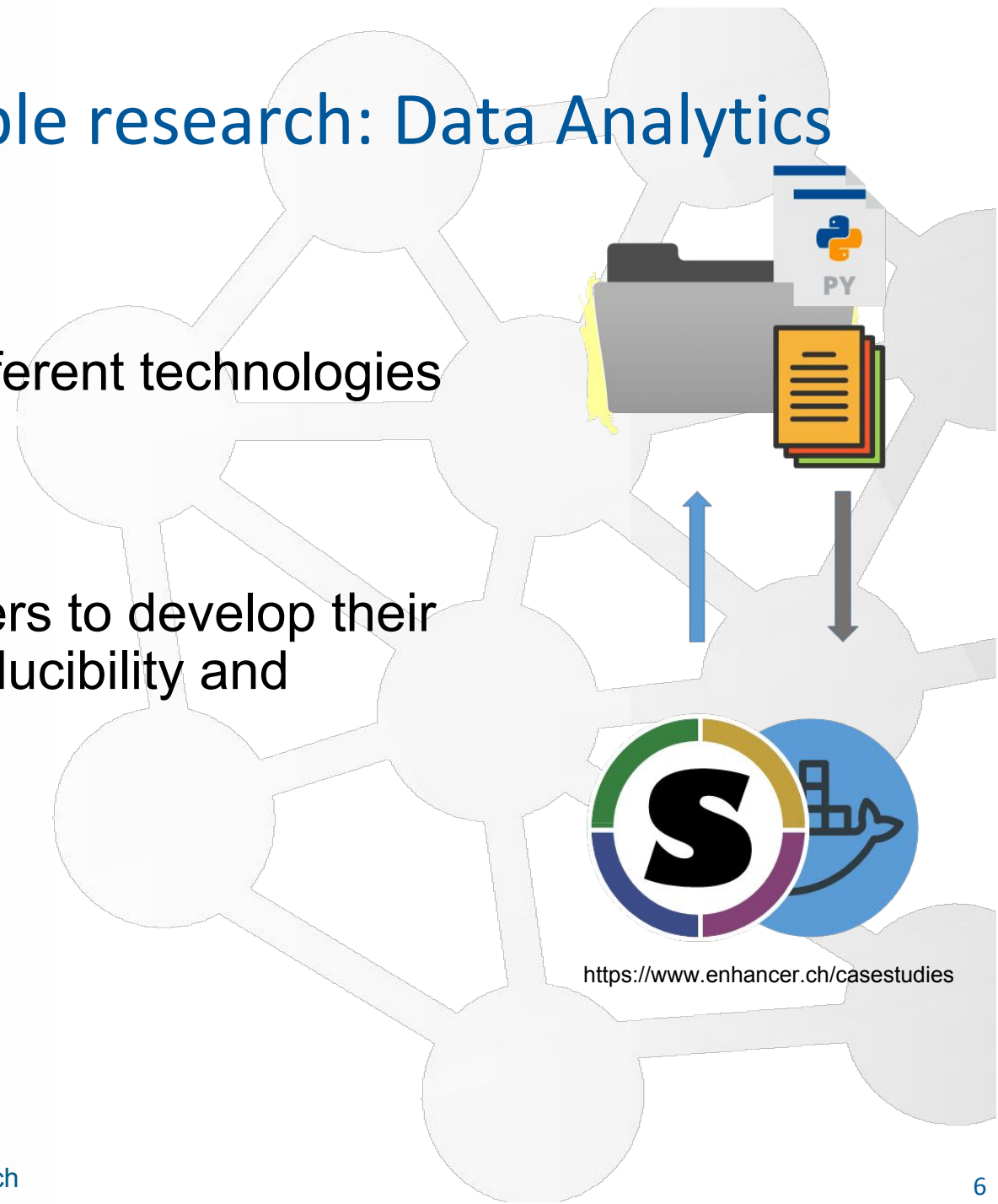
Applications and data analytics in containers

```
1  # docker build --rm -t beast .
2
3  FROM ubuntu:16.04
4  MAINTAINER Sergio Maffioletti - S3IT: service and Support for Science IT - University of Zurich, Switzerland.
5  LABEL description="Run BEAST2"
6
7  ARG beast_version
8  ARG babel_version
9  ARG beast_packages_folder_name
10 RUN apt-get update -y
11 RUN apt-get install -y autoconf build-essential unzip wget git openjdk-8-jdk libt
12 # Install Beagle
13 RUN git clone --depth=1 https://github.com/beagle-dev/beagle-lib.git
14 WORKDIR /beagle-lib
15 RUN ./autogen.sh
```


DOCKERFILE
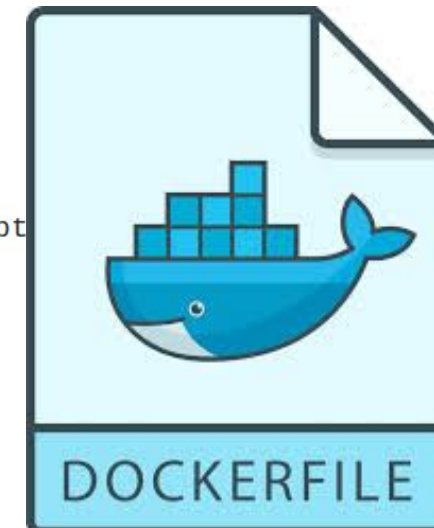
docker build

Docker Image

Singularity wrapper to build, run and test pipelines

data-analysis.img

description.yaml
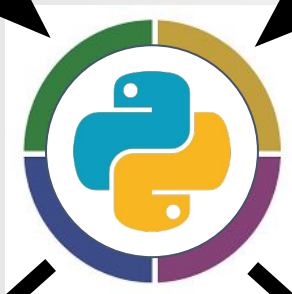
```yaml
name: hello_world_pipeline
version: 1
author: Balazs Laurenczy
author_org: ETHZ
substitutions:
  name: "hello_world_pipeline"
binds:
  - "/usr/lib64:/usr/lib64"
build:
  type: pull
  source: docker://blaurenczy/hello_world:latest
run:
  commands:
    - "{exec} python3 test.py > {name}.out 2> {name}.err"
test:
  validate_commands:
    - "[[ \"$(md5sum {name}.out | cut -f1 -d ' ')\" \
      == \"06c19b37d27dfda293492a4459fe3bc3\" ]] && echo 0"
```

singularity-pipeline

BUILD   **RUN**   **VALIDATE**

https://github.com/kav2k/singularity-pipeline

Focus on Reproducible research

- <u>Docker/Singularity</u> containers

  - contains full execution environment AND data analytics workflows.

- <u>Documented</u> usecases

  - what's for, what goes in, what comes out

- Container's <u>validator</u>

  - reference dataset + expected results

- <u>Registry</u> of supported containers and certified providers (soon available)

https://www.enhancer.ch/casestudies
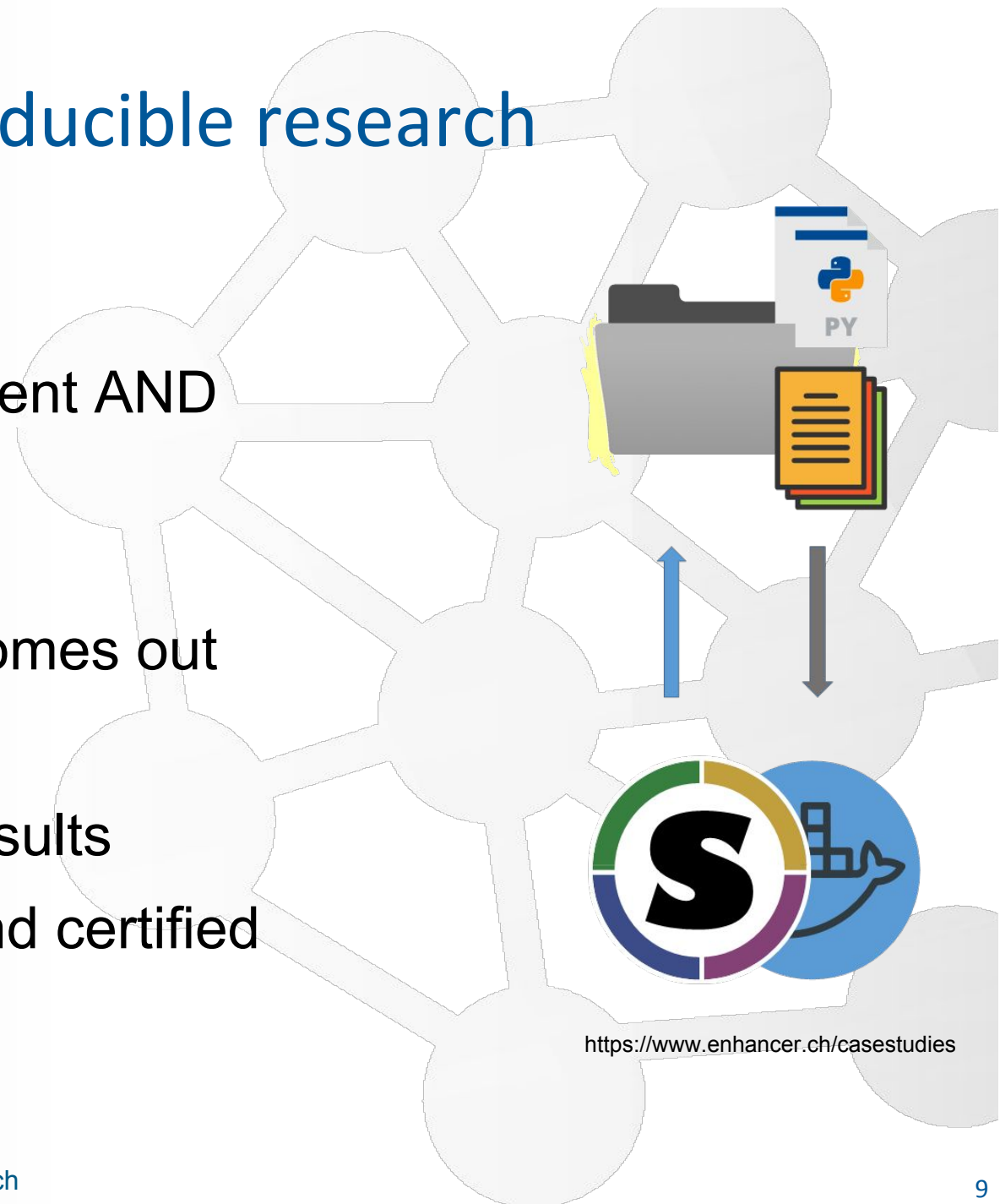
How can we help each other ?

- Use existing supported portable/reproducible containers for your research
- Turn your usecase in a portable/reproducible one
- Shape the EnhanceR OpenScience roadmap

Read more about our activities: https://www.EnhanceR.ch

Contact us: hello@EnhanceR.ch

Follow us: https://twitter.com/eScienceCH