

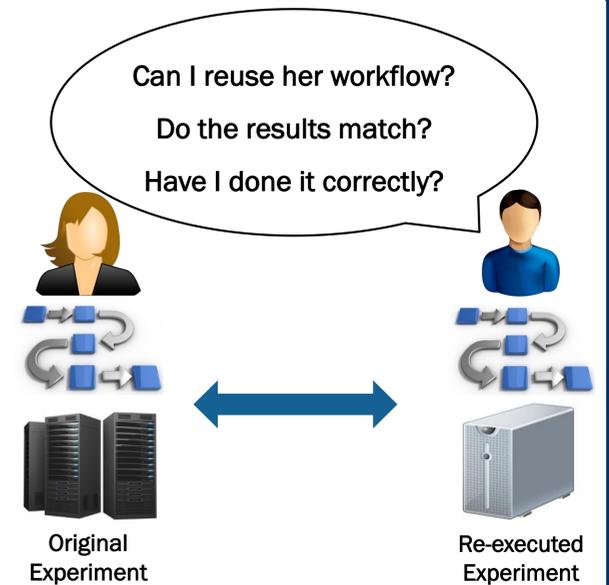
# REPLICABLE SCIENTIFIC EXPERIMENTS

Tomasz Miksa

Vienna University of Technology  
SBA Research, Austria

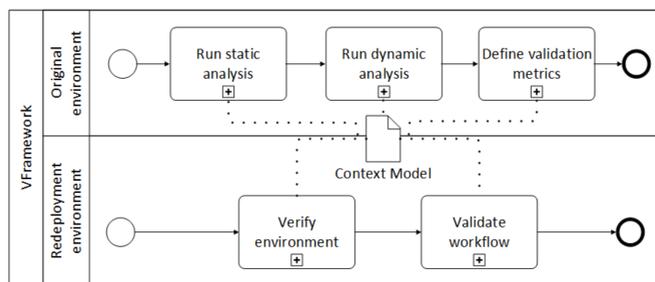
## Problem Description

- e-Science
- Research Infrastructures
- Research requires
  - special tooling and software
  - workflows to
    - capture
    - transform
    - visualise
    - interpret data
- Studies show low reproducibility in
  - medicine
  - economy
  - computer science
- Reproducibility requires
  - well documented research workflows
  - precise information on the experiment's environment



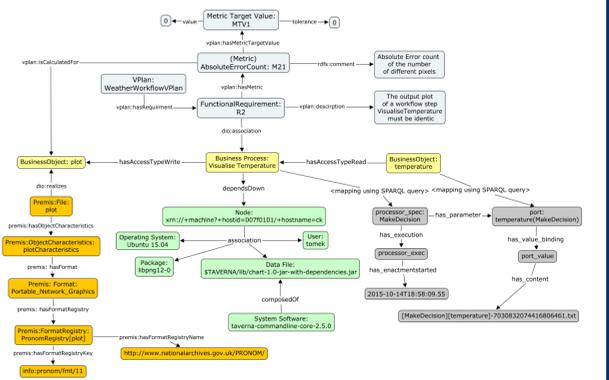
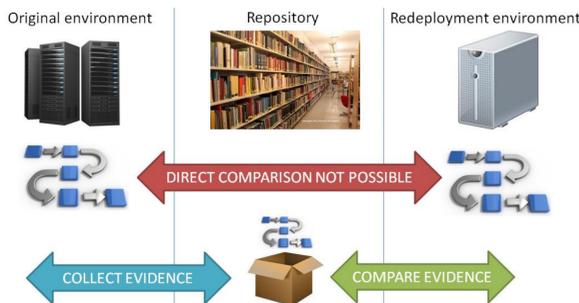
## Methodology

- VFramework**
  - collects evidence about the original execution
  - verifies and validates the re-execution
  - needs no access to both environments at the same time
  - is applicable also in digital preservation settings



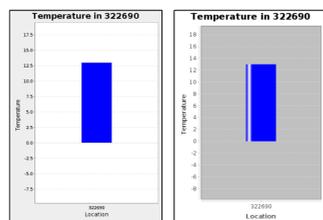
- VFramework steps**
  - Run static analysis
    - describes workflow model
    - defines workflow boundaries
  - Run dynamic analysis
    - monitors workflow execution
    - identifies dependencies (software libraries, scripts, packages, web services, etc.)
    - captures the processed data

- Define validation metrics
  - generates metrics based on file formats
- Context model**
  - OWL ontology
  - stores collected evidence
  - modular architecture
    - domain independent and domain specific ontologies



## Results

- Various workflow representations
  - Taverna workflows
  - Python scripts
  - bash scripts
- VFramework
  - identified all software
  - captured all data
  - identified differences between environments
  - validated results using file format based metrics
- Mock-up of external services can replace deterministic web services for known requests



Validation report for the WeatherExample

Evaluation result: There are 2 not fulfilled metrics. Please see tables below for details.  
Comparison performed using following workflow execution traces  
Original workflow: ID: 378462b-e71c-46f7-87b3-17888e02977  
Timestamp: 2015-10-14 13:58:06 475  
Compared workflow: ID: ea62ba87-9f58-4af9-90c3-e1759570ba80  
Timestamp: 2015-11-13 13:58:56 483

Table 1: Overview of requirements

Requirement	Description	Is Fulfilled
R1	The inputs to the workflow are the same	True
R2	The outputs of the workflow are the same	False
R3	The workflow step ExtractTemperature must have identical outputs	True
R4	The workflow step GetWeatherData must have identical outputs	True
R5	The workflow step MakeDecision must have identical outputs	True
R6	The workflow step ExtractWeatherData must have identical outputs	True
R7	The workflow step VisualizeTemperature must have identical outputs	False
R8	Execution duration of each of the workflow steps shall be similar	True

Table 2: List of requirements and metrics that failed

Req	Sub-req	Sub-req description	Measure point	Metric	Validity
R2	R2.1	The output plot of workflow step VisualizeTemperature must be identical	plot	ImageFingerprintQuality	True
R7	R7.1	The output plot of workflow step VisualizeTemperature must be identical	plot	ImageFingerprintQuality	False
				AbsoluteMetricCount	True
				AbsoluteMetricCount	False

## Future Work

- Actionable Data Management Plans**
  - reduce bureaucracy imposed on researchers
  - describe data, workflows, and environment used in experiments
  - have machine-readable template
  - are semi-automatically generated using tools that
    - import data from various workflow specifications
    - monitor experiment execution
    - analyse provided data samples
  - validate provided information
  - integrate with existing DMP tools (e.g. DMP Online)



Dr. Tomasz Miksa  
TU Wien, SBA Research  
Favoritenstrasse 16, 1040 Wien, Austria  
tmiksa@sba-research.org