



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

The VERCE Science Gateway

(..and data-intensive provenance)

Alessandro Spinuso

spinuso@knmi.nl



“Democratising Computational Seismology Research in Europe”



INGV



Virtual Research Environment

http://portal.verce.eu/home

VERCE
Beta

Welcome Help Pages Support

portal.verce.eu Welcome

The **VERCE** Science Gateway is currently in beta. We are accepting new users in small groups as we roll out new features.

Sing up for an account, you will get notified after our approval.

SIGN UP

Check out our [training pages](#)

IMPORTANT !!

The **NEW** version of the **Forward Modeling** application will be soon available in its **Beta** release.

It includes additional workspaces for **FDSN Data Download**, **Seismic Waveform Processing** and **MISFIT Analysis**. Below a schematic representation of the integrated tool.

To obtain the GRID Certificate which will allow you to join the VERCE Virtual Organisation and use our platform, please follow the [instructions](#)

Need Help? For rapid feedback please contact us at support[AT]verce[DOT]eu. Alternatively you can register to our [support pages](#)

The VERCE Science Gateway has been developed in partnership with the [SCI-BUS](#) project

Virtual Earthquake and seismology
Research Community in Europe e-science environment

Verce - The Science Gateway Too...

Earthquake Ground Motion: Lucca (Pianu...

Virtual Research Environment

http://portal.verce.eu/home

VERCE
Beta

Welcome Help Pages Support

portal.verce.eu Welcome

The **VERCE** Science Gateway is currently in beta. We are accepting new users in small groups as we roll out new features.
Sing up for an account, you will get notified after our approval.

Virtual Earthquake and seismology
Research Community in Europe e-science environment

13 - Verce - The Science Gateway Too...

portal.verce.eu

Reduce the initial burning in HPC seismology
exploiting SaaS on EGI, PRACE and Local resources

FDSN Raw Data

Processed Data and Records in IRDS

To obtain the GRID Certificate which will allow you to join the VERCE Virtual Organisation and use our platform, please follow the [instructions](#)

Need Help? For rapid feedback please contact us at support@verce.eu. Alternatively you can register to our [support pages](#)

The VERCE Science Gateway has been developed in partnership with the [SCI-BUS](#) project





- **Earthquake Simulation: Synthetic Seismograms** for **public** and **custom Earth models** and **Earthquakes** via the execution of **HPC simulation codes** called *solvers* (*SPECFEM3D - SPECFEMGLOBE*)
- **Raw data acquisition & Misfit:** The synthetic data may be compared with real observations stored in distributed archives (FDSN), adopting Data Intensive methods

VERCE Recipe



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

SEISMOLOGICAL INGREDIENTS

MESH

MATERIAL PROPERTIES
(velocity model)

NUMERICAL CODE
(SPECFEM3D)

EVENTS

STATIONS

HPC INGREDIENTS

HPC CONNECTION
HPC CENTER POLICIES
HPC CENTER

SEISMOLOGICAL OUTPUT

SYNTHETIC SEISMOGRAMS

VOLUMETRIC DATA

Shakemap

Movie

SYNT - OBS MISFIT

VERCE Recipe



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

forward-modelling

Setup Results iRods

max extent

Help Layers info

Solver Earthquakes Stations Submit Control

File FDSN

Open

Name: DefaultName

File: Select a file Browse... Upload

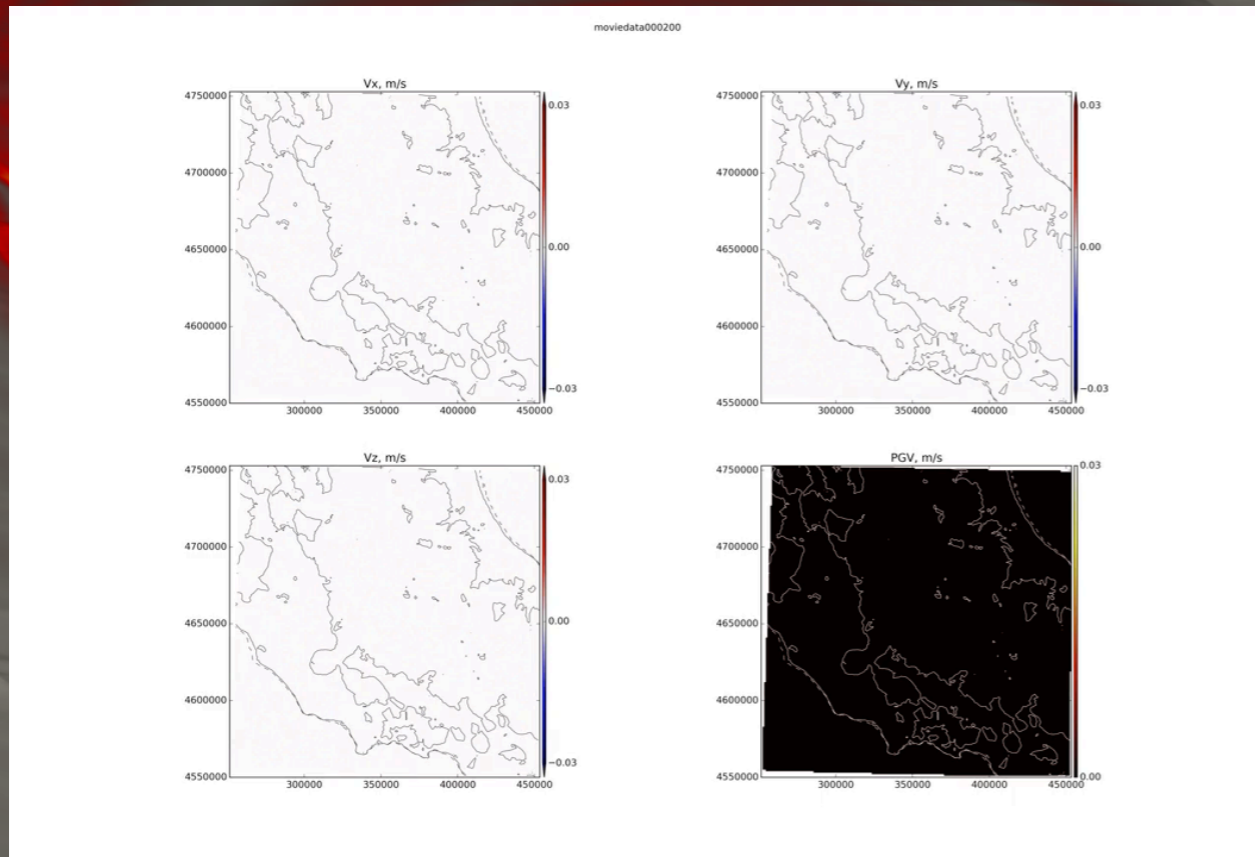
0/15	Desc	Date	Depth	Latitude	Longitude	Magnitude	MT
	LUCCA	2013-06-30T1...	9800.0	44.171	10.2047	4.5	○
					10.2108	4.4	○
					10.135	5.1	○
					10.4543	4.8	○
					9.6703	4.3	○
					10.9502	4.7	○
					10.9795	5.3	○
					10.9663	4.1	○
					11.0657	5.6	○
					11.305	4.2	○
					11.4407	4.9	○
					11.2635	5.8	○
					9.354	4.0	○
					10.009	4.9	○

VERCEDEMO01402044603337.mp4

00:30 -00:29

Powered By Liferay - gUSE

AMATRICE Mw 6.0 24/08/2016



OUTPUT

seismograms, plots, 3D Geometry, Videos, KMZ packages, meshes and models > 10 GB

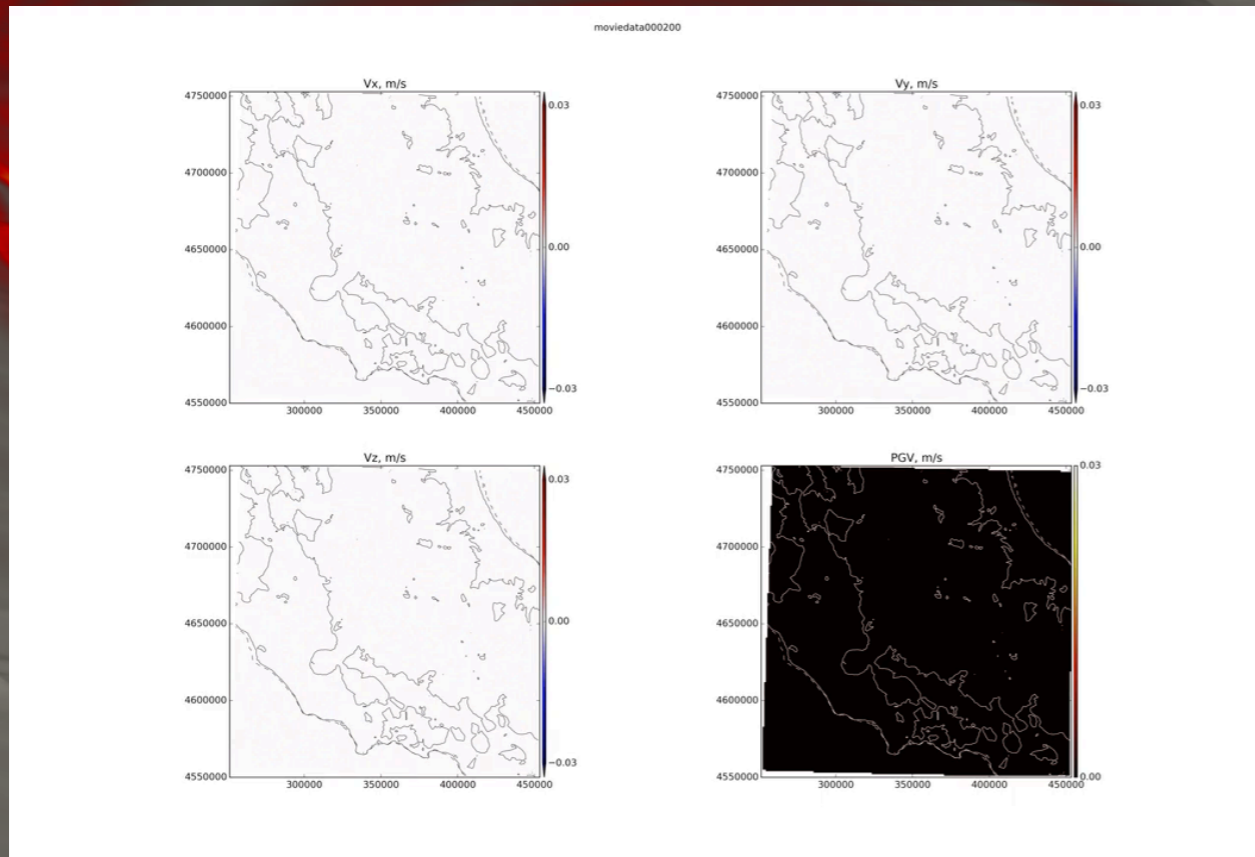
HPC Resources @ Fraunhofer SCAI

VR Products

3D Geometry for CAVE Visualisation experimented at LRZ



AMATRICE Mw 6.0 24/08/2016



OUTPUT

seismograms, plots, 3D Geometry, Videos, KMZ packages, meshes and models > 10 GB

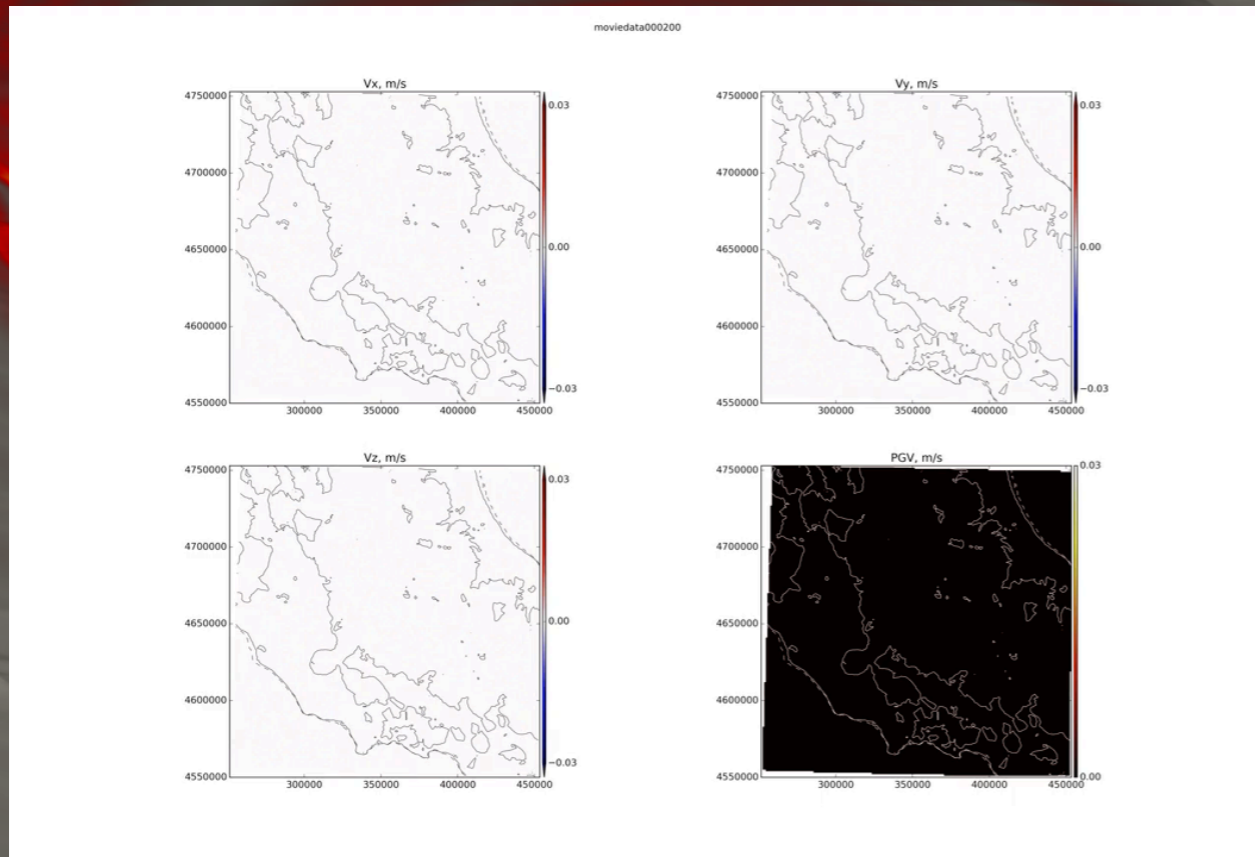
HPC Resources @ Fraunhofer SCAI

VR Products

3D Geometry for CAVE Visualisation experimented at LRZ



AMATRICE Mw 6.0 24/08/2016



OUTPUT

seismograms, plots, 3D Geometry, Videos, KMZ packages, meshes and models > 10 GB

HPC Resources @ Fraunhofer SCAI

VR Products

3D Geometry for CAVE Visualisation experimented at LRZ



User Perspective



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

Earthquake Simulation

Integrated view of community webservices
via the Forward Simulation GUI



User's Actions

1 Select Solver

2 Select Mesh

3 Select Velocity Model

4 Upload Input

Retrieve Input

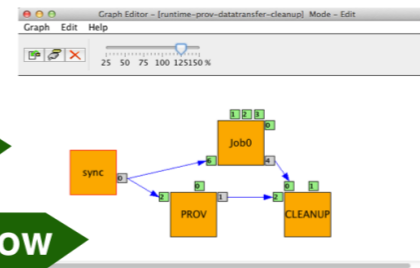
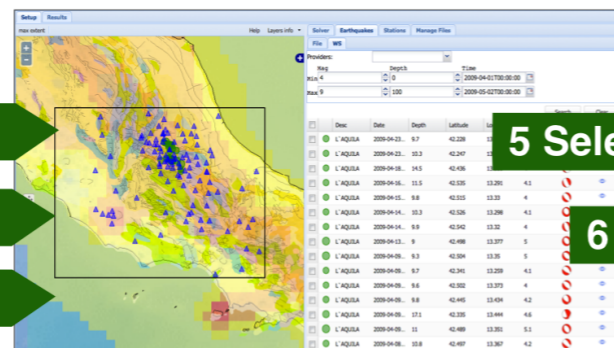
Reuse Input

5 Select Workflow

6 Submit Workflow

7 Access Provenance

Visualise Results



VERCE Science Gateway



Provenance Production



MyProxy
Credential Management Service



Data storage within
the VERCE Federation



Many Tasks
Simulation workflows
in HPC
resources

User Perspective



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

Misfit between SYNTHETICS and DATA

Name	Desc	Workflow	Date
simulation_abruzzo1343000000_1485181669330	15 stazioni	SCAI_mpi_SPECFEM_PRODUCTION__2015-03-03-120220_2015-11-19-151723	23 - 01 - 2017
simulation_abruzzo1343000000_1485180676096	15 stazioni	SCAI_mpi_SPECFEM_PRODUCTION__2015-03-03-120220_2015-11-19-151723	23 - 01 - 2017
simulation_abruzzo1343000000_1485175329249	15 stazioni	SCAI_mpi_SPECFEM_PRODUCTION__2015-03-03-120220_2015-11-19-151723	23 - 01 - 2017
simulation_abruzzo13430000_1485173261691	15 stazioni	SCAI_mpi_SPECFEM_PRODUCTION__2015-03-03-120220_2015-11-19-151723	23 - 01 - 2017
simulation_abruzzo1343000_1485172391804	15 stazioni	SCAI_mpi_SPECFEM_PRODUCTION__2015-03-03-120220_2015-11-19-151723	23 - 01 - 2017
simulation_abruzzo134300_1484940067541	15 stazioni	SCAI_mpi_SPECFEM_PRODUCTION__2015-03-03-120220_2015-11-19-151723	20 - 01 - 2017
simulation_abruzzo134300_1484938577498			
simulation_abruzzo134300_1484938138425			
simulation_abruzzo13430_1484937623558			
simulation_abruzzo13430_1484936993904			
simulation_abruzzo13430_1484936730412			
simulation_abruzzo13333000000000_14849358			
simulation_abruzzo13333000000000_148493546			
simulation_abruzzo1333300000000_1484750697			
simulation_abruzzo13333000000_14829339375			
simulation_abruzzo13333000_1481628715387			
simulation_macerata00_1481563368758			
simulation_abruzzo1333300_1481559849742			
simulation_abruzzo133330_1481300770050	15 stazioni	SCAI_mpi_SPECFEM_PRODUCTION__2015-03-03-120220_2015-11-19-151723	09 - 12 - 2016
simulation_abruzzo133330_1481300521985	15 stazioni	SCAI_mpi_SPECFEM_PRODUCTION__2015-03-03-120220_2015-11-19-151723	09 - 12 - 2016
simulation_macerata00_1481300109953	italia	SCAI_mpi_SPECFEM_PRODUCTION__2015-03-03-120220_2015-11-19-151723	09 - 12 - 2016
simulation_macerata00_1481299879297	italia	SCAI_mpi_SPECFEM_PRODUCTION__2015-03-03-120220_2015-11-19-151723	09 - 12 - 2016
simulation_abruzzo16000000000_1481299507926	15 stazioni	SCAI_mpi_SPECFEM_PRODUCTION__2015-03-03-120220_2015-11-19-151723	09 - 12 - 2016
simulation_abruzzo16000000000_1481299057509	15 stazioni	SCAI_mpi_SPECFEM_PRODUCTION__2015-03-03-120220_2015-11-19-151723	09 - 12 - 2016
simulation_abruzzo1600000000_1481298000065	15 stazioni	SCAI_mpi_SPECFEM_PRODUCTION__2015-03-03-120220_2015-11-19-151723	09 - 12 - 2016
simulation_abruzzo16000000000000_1481119958747	15 stazioni	SCAI_mpi_SPECFEM_PRODUCTION__2015-03-03-120220_2015-11-19-151723	07 - 12 - 2016
simulation_macerata0_1477641667271	italia	SCAI_mpi_SPECFEM_PRODUCTION__2015-03-03-120220_2015-11-19-151723	28 - 10 - 2016

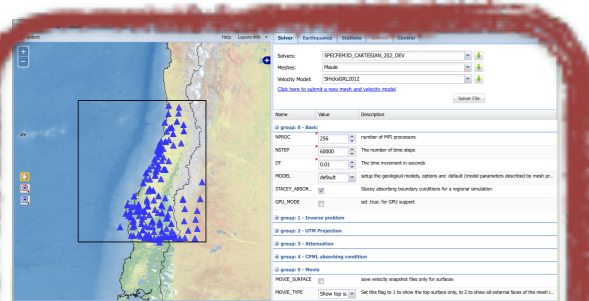
Select Previous Simulation Run for the
parametrisation of RAW Data Caching
Workflow

User Perspective

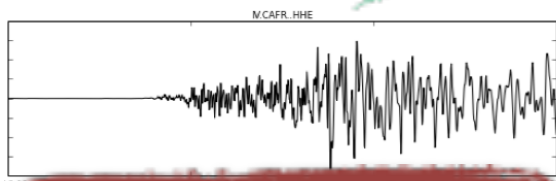
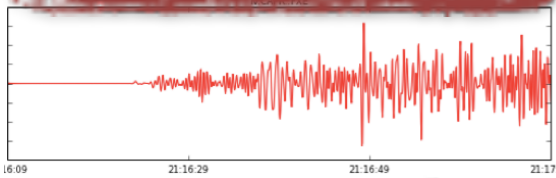


Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

Misfit between SYNTHETICS and DATA

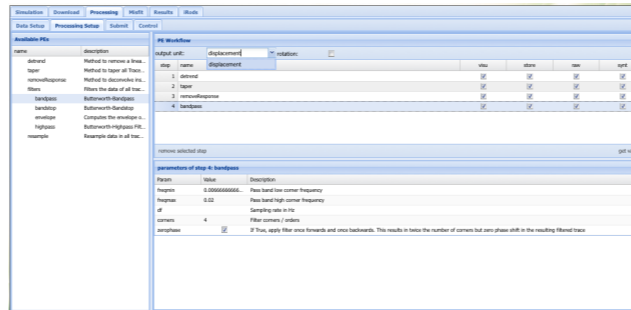


Simulated Synthetics

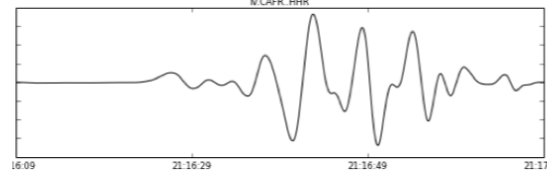
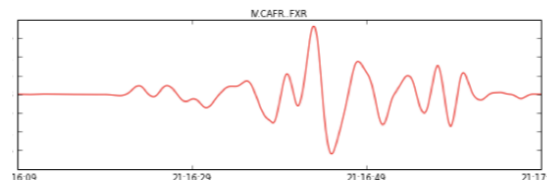


Data Download (FDSN)

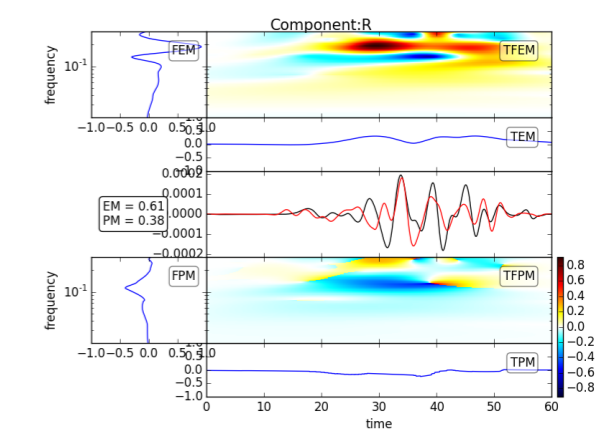
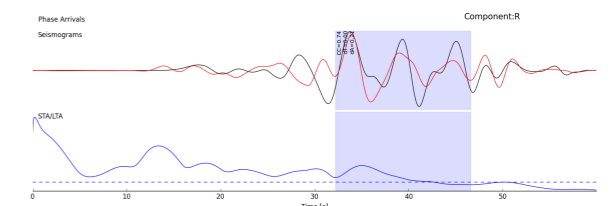
Station	Latitude	Longitude	Depth	Code	Code	Code
...



Data/Synt Processing



Misfit Analysis



dispel
4PY



ObsPy
A Python Framework for Seismology

User Perspective



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

Misfit between SYNTHETICS and DATA

Simulation Runs

FDSN Data Download Runs

Stored in VERCE and Reusable!!

			Date	RunID			
simulation_abruzzo160000_1443707335437	SMU_test_512_30...	9 stazioni	01 - 10 - 2015	download_abruzzo16000_1443537054666...	SCAI_FDSN_data_...	dd	30 - 09 - 2015
simulation_mauleSCAIInomov000_144370...	SCAI_multi_SPECIF...	small256nomov	01 - 10 - 2015	download_abruzzo16000_1443537309549...	SCAI_FDSN_data_...		30 - 09 - 2015
simulation_mauleSCAIInomov000_144370...	SMU_test_512_30...	small256nomov	01 - 10 - 2015	download_abruzzo16000_1443537309549...	SCAI_FDSN_data_...	with visjj	30 - 09 - 2015
simulation_mauleSCAIInomov000_144370...	SMU_test_512_30...	small256nomov	01 - 10 - 2015	download_Frosinone_01429774374753_1...	SCAI_FDSN_data_...	rrr	24 - 09 - 2015
simulation_abruzzo16000_1443537309549	SMU_test_512_30...	9 stazioni	29 - 09 - 2015	download_Frosinone_OLSEN_INF0014297...	SCAI_FDSN_data_...	rrr	24 - 09 - 2015
simulation_abruzzo16000_1443537054666	SMU_test_512_30...	9 stazioni	29 - 09 - 2015	download_abruzzo160_1443105413838_1...	FDSN_data_2015-...	rrr	24 - 09 - 2015

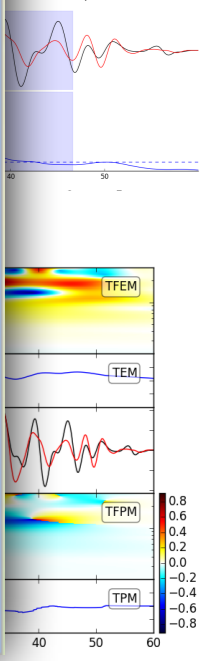
commons stations

<input checked="" type="checkbox"/>	Network	Station	Source
<input checked="" type="checkbox"/>	IV	FIAM	syn, raw
<input checked="" type="checkbox"/>	IV	CAFR	syn, raw
<input checked="" type="checkbox"/>	IV	CESX	syn, raw
<input checked="" type="checkbox"/>	IV	CERA	syn, raw
<input checked="" type="checkbox"/>	IV	GIUL	syn, raw
<input checked="" type="checkbox"/>	IV	FAGN	syn, raw
<input checked="" type="checkbox"/>	IV	CERT	syn, raw
<input type="checkbox"/>	IV	AQT1	syn
<input type="checkbox"/>	IV	CAMP	syn
<input type="checkbox"/>	IV	RNI2	raw

Sensors' data with common properties for Misfit analysis (time-range, event)

is

Component:R

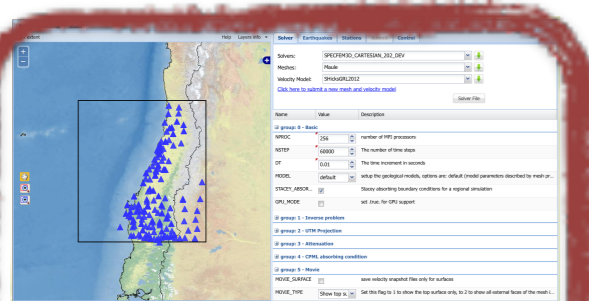


User Perspective

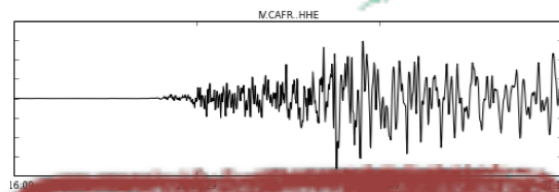
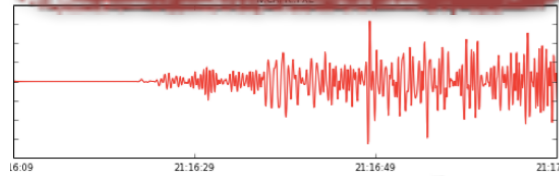


Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

Misfit between SYNTHETICS and DATA

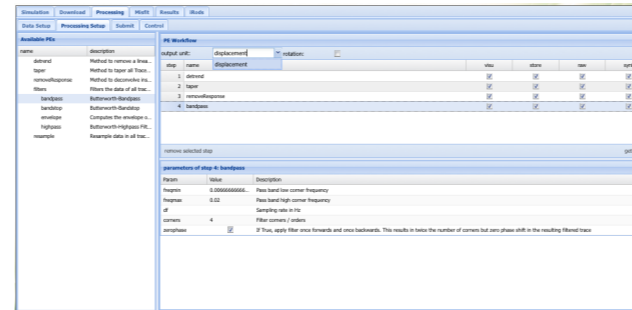


Simulated Synthetics

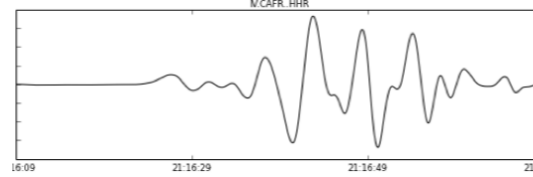
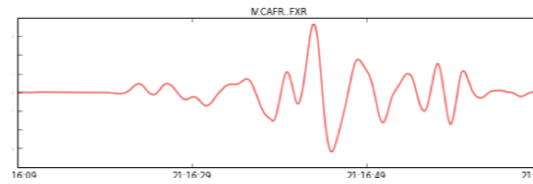


Data Download (FDSN)

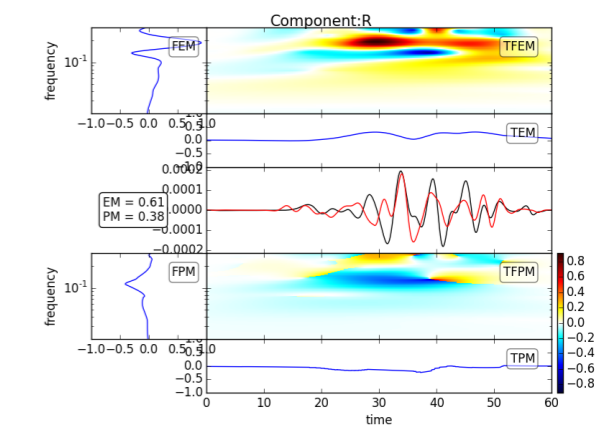
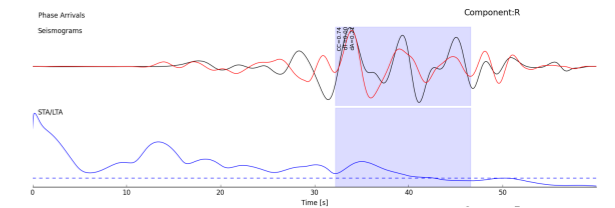
Station	Latitude	Longitude	Depth	Date	Station	Latitude	Longitude	Depth	Date
01	52.13	5.13	0.0	2013-08-04	01	52.13	5.13	0.0	2013-08-04
02	52.13	5.13	0.0	2013-08-04	02	52.13	5.13	0.0	2013-08-04



Data/Synt Processing



Misfit Analysis



dispel
4PY



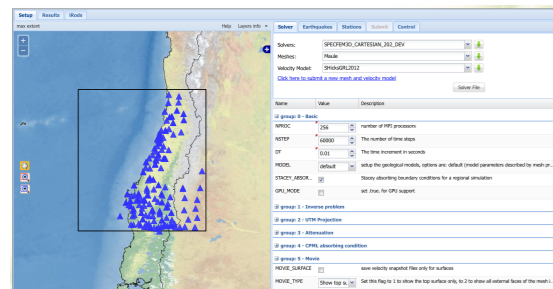
ObsPy
A Python Framework for Seismology

User Perspective

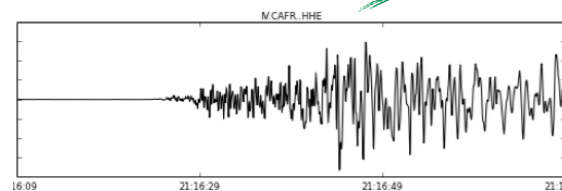
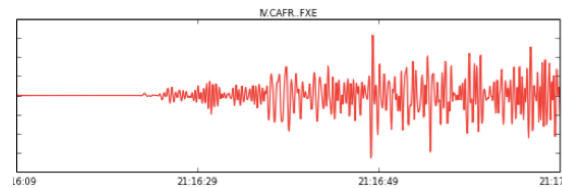


Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

Misfit between SYNTHETICS and DATA

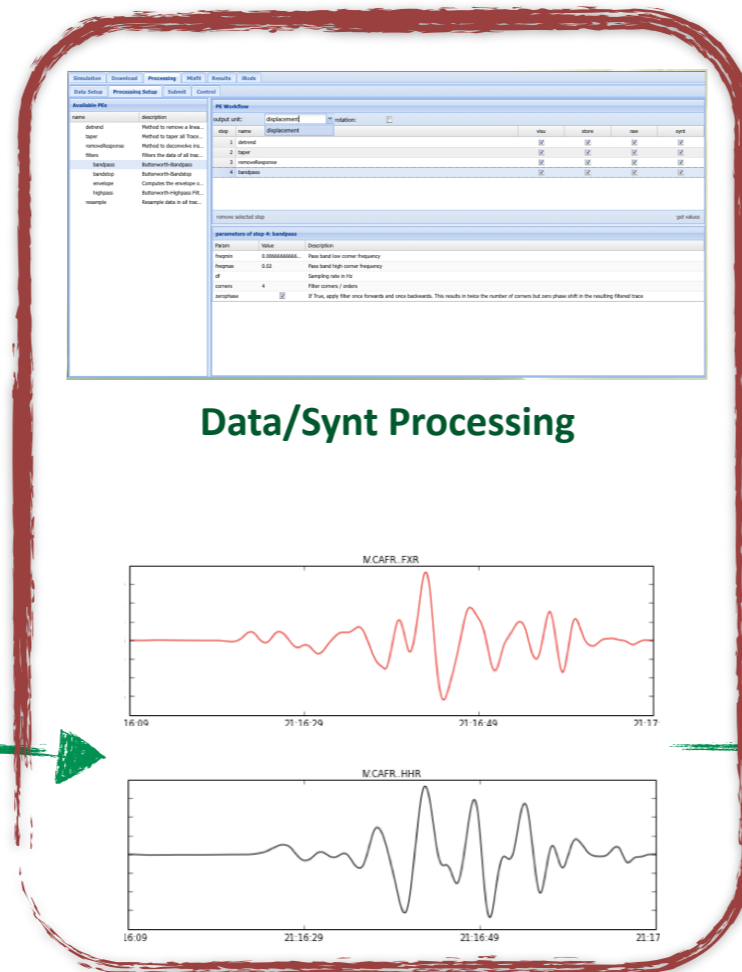


Simulated Synthetics

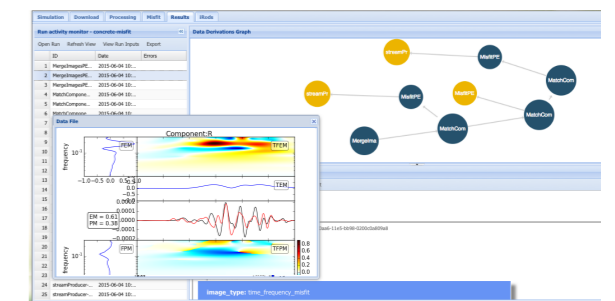


Data Download (FDSN)

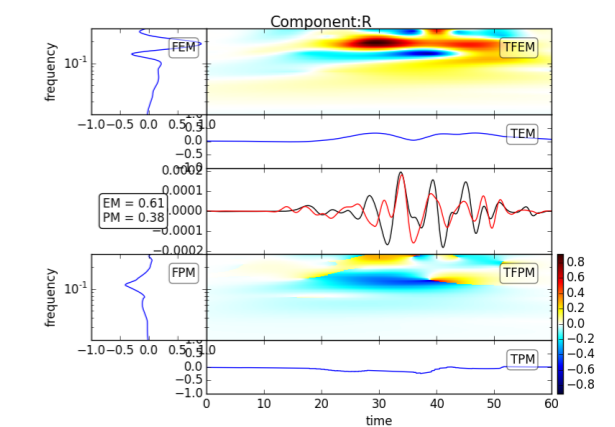
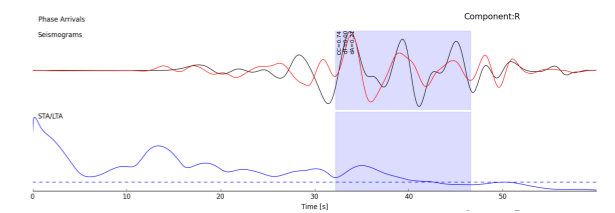
Station	Description	Date	Time
...



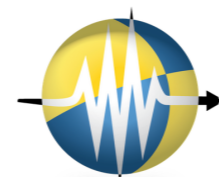
Data/Synt Processing



Misfit Analysis



dispel
4PY



ObsPy
A Python Framework for Seismology

User Perspective



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

Misfit between SYNTHETICS and DATA

The screenshot shows the iRods software interface with several key components highlighted by green callouts:

- Activation / Visualisation / Storage:** A green box at the top right highlights the top navigation bar and a table with columns for 'visu', 'store', 'raw', and 'synt'.
- Pipeline Sequence:** A green box in the center highlights the 'PE Workflow' section, which lists a sequence of processing steps: 1. detrend, 2. taper, 3. removeResponse, and 4. bandpass.
- Toolbox of Processing Elements:** A green box on the left highlights the 'Available PEs' list, which includes elements like detrend, taper, removeResponse, filters, bandpass, bandstop, envelope, highpass, and resample.
- Processing Element's Parameters Setup:** A green box at the bottom highlights the configuration panel for the 'bandpass' element, showing parameters such as pass band corner frequencies, sampling rate, and filter orders.

Additional elements visible in the interface include a 'Simulation' menu, 'Data Setup', 'Processing Setup', 'Submit', and 'Control' buttons. On the right side, there are several plots, including a time-series plot labeled 'Component:R' and a spectrogram labeled 'TFPM'.

4PY



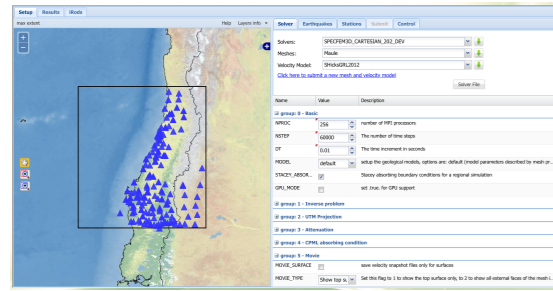
A Python Framework for Seismology

User Perspective

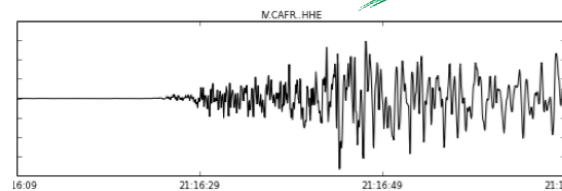
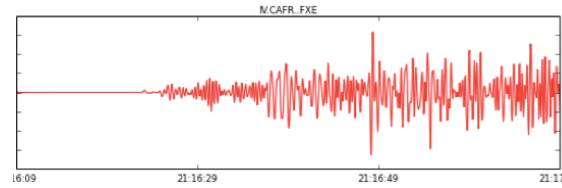


Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

Misfit between SYNTHETICS and DATA

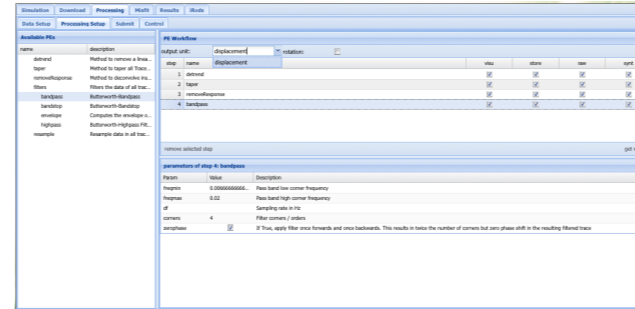


Simulated Synthetics

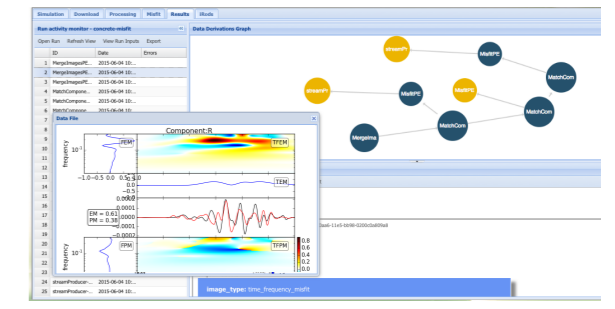
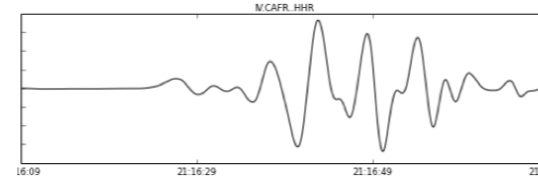
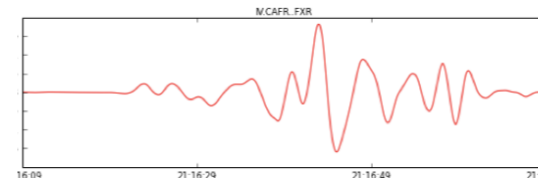


Data Download (FDSN)

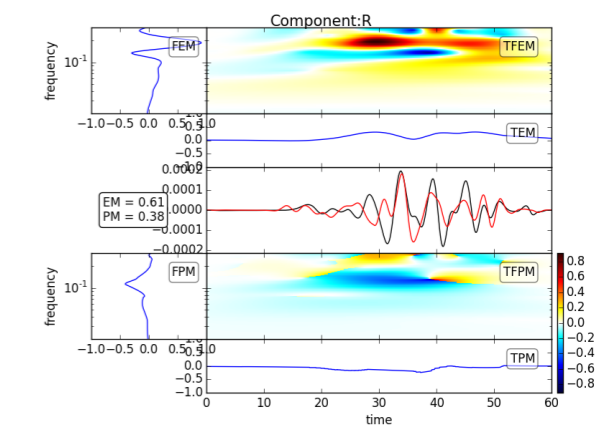
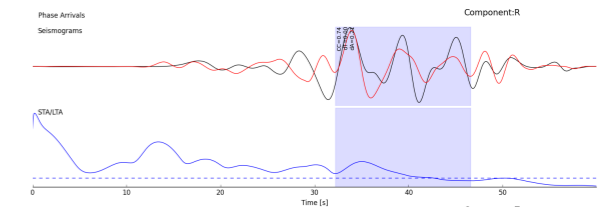
Station	Network	Code	Location	Depth	Instrument	Start	End
M/CAFR_FXE	CAFR	FXE
M/CAFR_HHE	CAFR	HHE



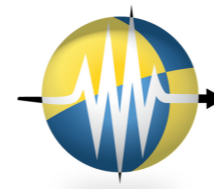
Data/Synt Processing



Misfit Analysis



dispel
4PY



ObsPy
A Python Framework for Seismology

User Perspective



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

Misfit between SYNTHETICS and DATA

The image illustrates a user's perspective on misfit between synthetics and data through a workflow and software interface.

Workflow Diagram (Left): A vertical flowchart showing the data processing pipeline:

- streamProducer116** (output_pyflex, output_time_frequency) feeds into **input PyflexPE117** (window_tapering, image).
- input PyflexPE117** feeds into **input WindowTaperingPE120** (output).
- input WindowTaperingPE120** feeds into **input MisfitPE118** (image).
- input MisfitPE118** feeds into **input MatchComponents119** (output).
- input MatchComponents119** feeds into **input MergeImagesPE121**.

Software Interface (Center): A screenshot of the 'Run activity monitor - concrete-misfit' application. It features a table of runs and a 'Data File' window.

ID	Date	Errors
1	2015-06-04 10:...	
2	2015-06-04 10:...	
3	2015-06-04 10:...	
4	2015-06-04 10:...	
5	2015-06-04 10:...	
6	2015-06-04 10:...	
7	2015-06-04 10:...	
8	2015-06-04 10:...	
9	2015-06-04 10:...	
10	2015-06-04 10:...	
11	2015-06-04 10:...	
12	2015-06-04 10:...	
13	2015-06-04 10:...	
14	2015-06-04 10:...	
15	2015-06-04 10:...	
16	2015-06-04 10:...	
17	2015-06-04 10:...	
18	2015-06-04 10:...	
19	2015-06-04 10:...	
20	2015-06-04 10:...	
21	2015-06-04 10:...	
22	2015-06-04 10:...	
23	2015-06-04 10:...	
24	2015-06-04 10:...	
25	2015-06-04 10:...	

Data File Window (Center): Displays seismic data for 'Component:R' with the following plots:

- TFEM:** Time-Frequency Misfit plot showing a color scale from 0.0 to 0.8.
- TEM:** Time-Error Misfit plot showing a color scale from -1.0 to 1.0.
- TPM:** Time-Phase Misfit plot showing a color scale from -0.8 to 0.8.
- FEM:** Frequency-Error Misfit plot showing a color scale from -1.0 to 1.0.
- FPM:** Frequency-Phase Misfit plot showing a color scale from -0.8 to 0.8.

Parameters shown: EM = 0.61, PM = 0.38.

Data Derivations Graph (Right): A network diagram showing the relationships between processing steps: streamPr, MisfitPE, MatchCom, and Mergolma. A red circle highlights the 'Mergolma' node.

Text and Log (Bottom Right): The text 'Generates W3C-PROV' is displayed above a log entry: '0aa6-11e5-bb98-0200c0a809a8'. Below this is a time-series plot with 'time' on the x-axis (0 to 60) and values on the y-axis ranging from -1.0 to 0.0.

Logos (Bottom): The logos for '4py' and 'Obspy - A Python Framework for Seismology' are visible at the bottom of the image.

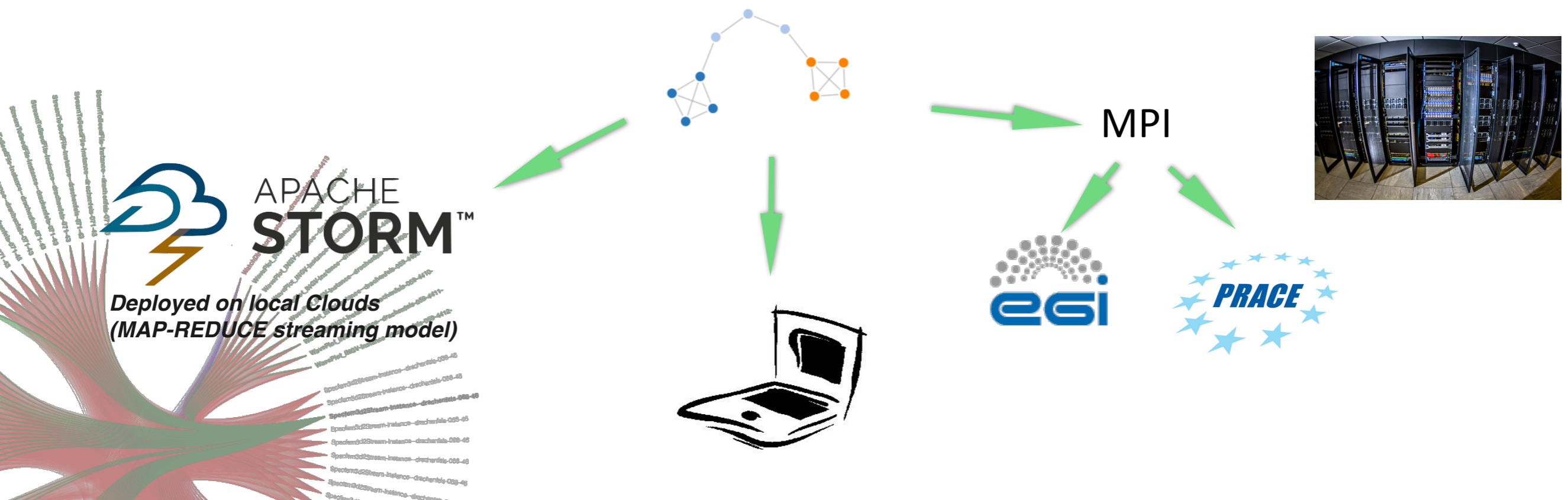


Python library used to describe **abstract workflows** for distributed data-intensive applications.

Support for composition: Single components may be defined by having their own internal workflows.

Workflows described in **dispel4Py** can be automatically executed in numerous parallel environment.

Docker containers available supporting multiple execution environments (MPI, SharedMemory) and the integration with other workflow systems (eg. Pegasus)



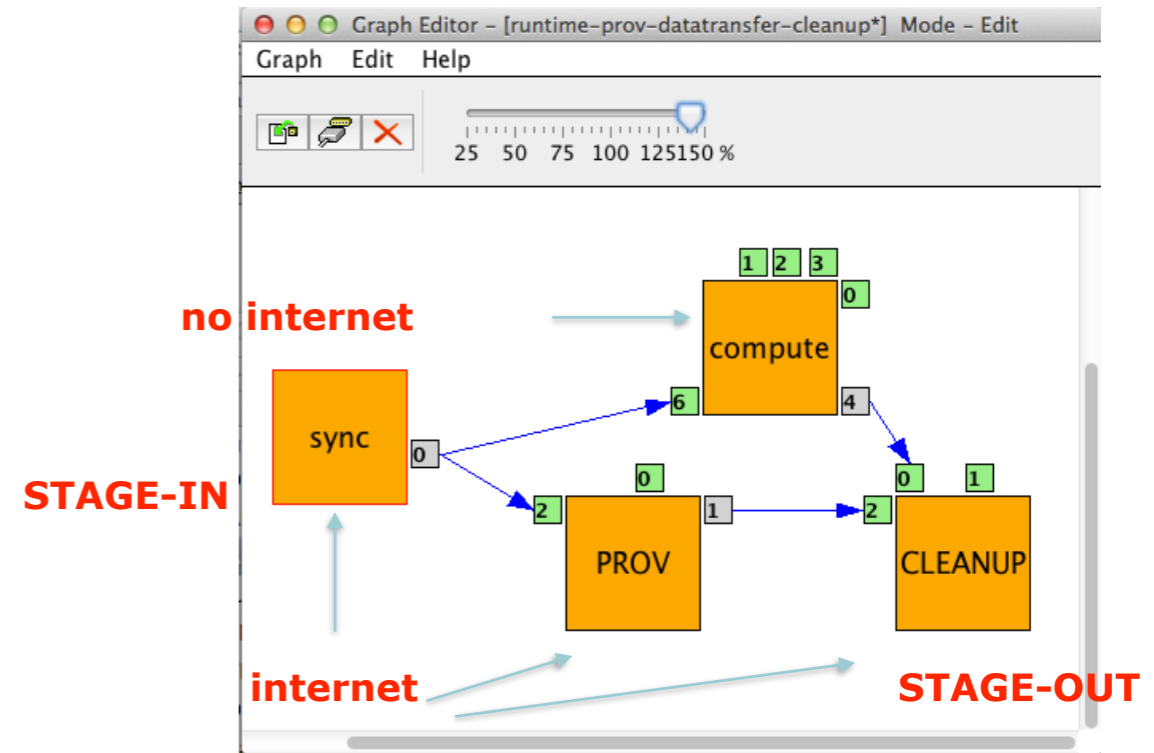
Platform's Control Workflows (WS-PGRADE)

1

HPC Control Workflow

GLOBUS & UNICORE jobs to Grid and HPC Clusters.

Frequent limitations in connectivity require special serial jobs for runtime data and metadata management (Sync, PROV, Cleanup)

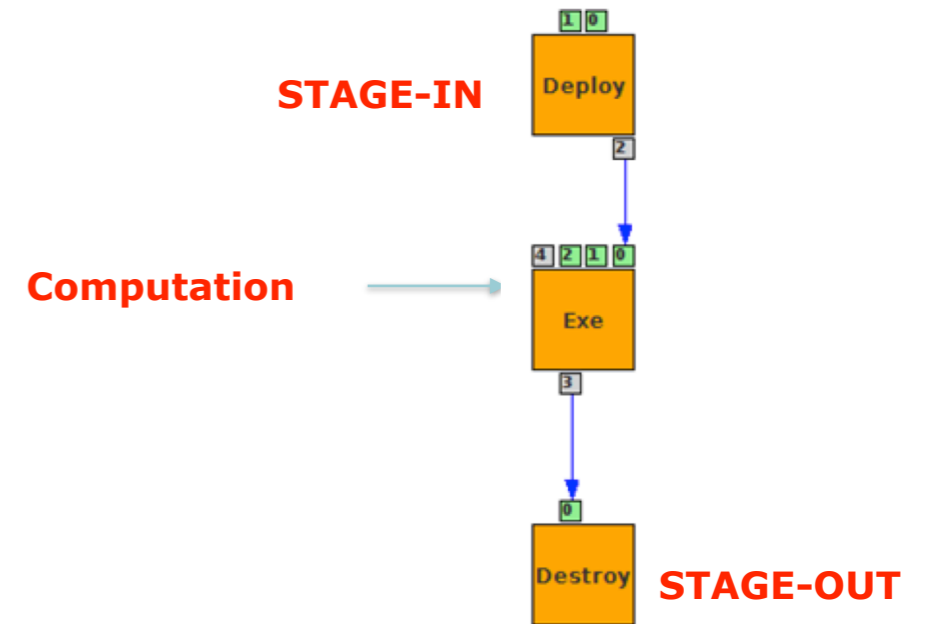


2

Cloud Control Workflow

(Deploy - Compute - Destroy)

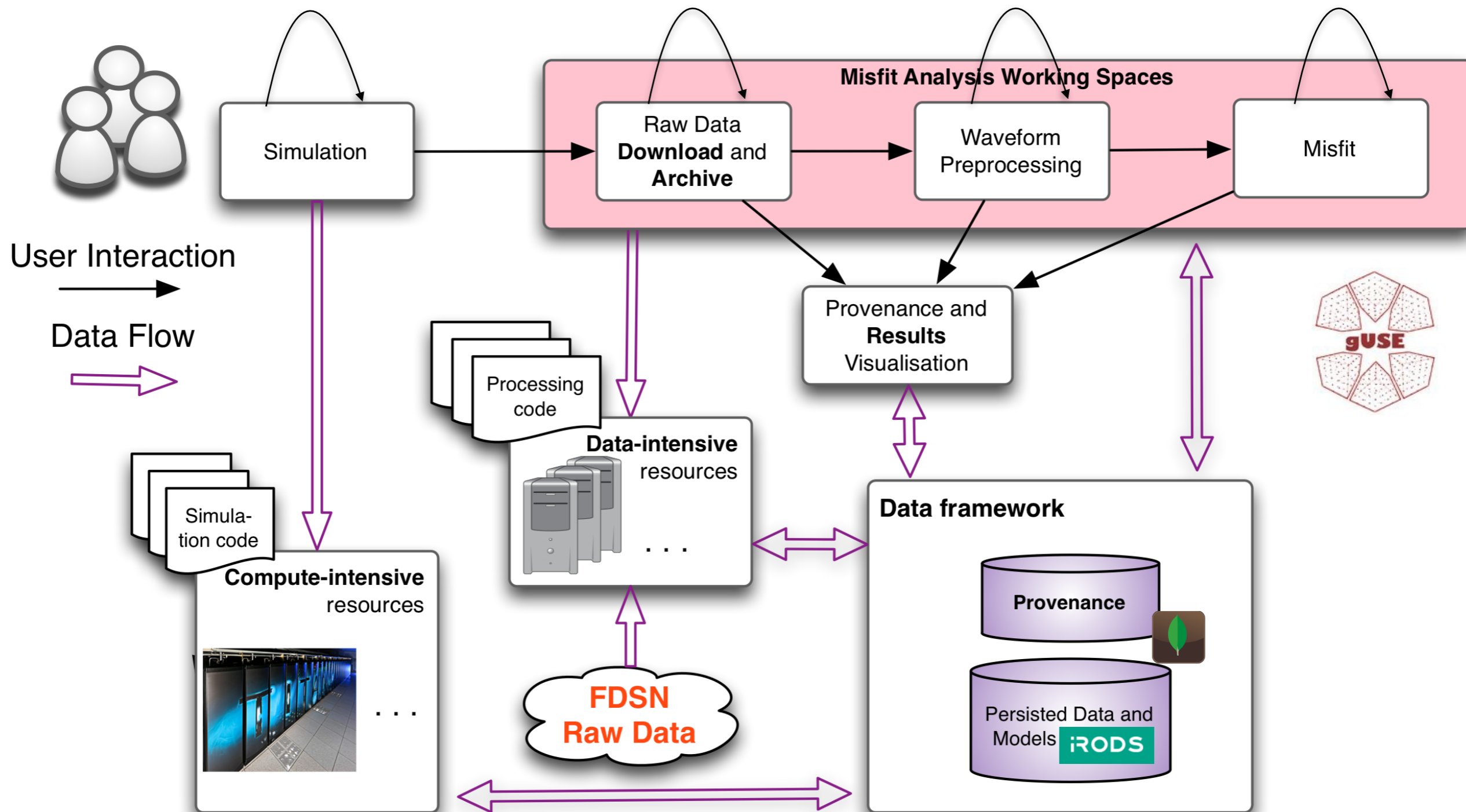
Less limitations in outbound connectivity, easier control workflow.



System & E-Infra Perspective



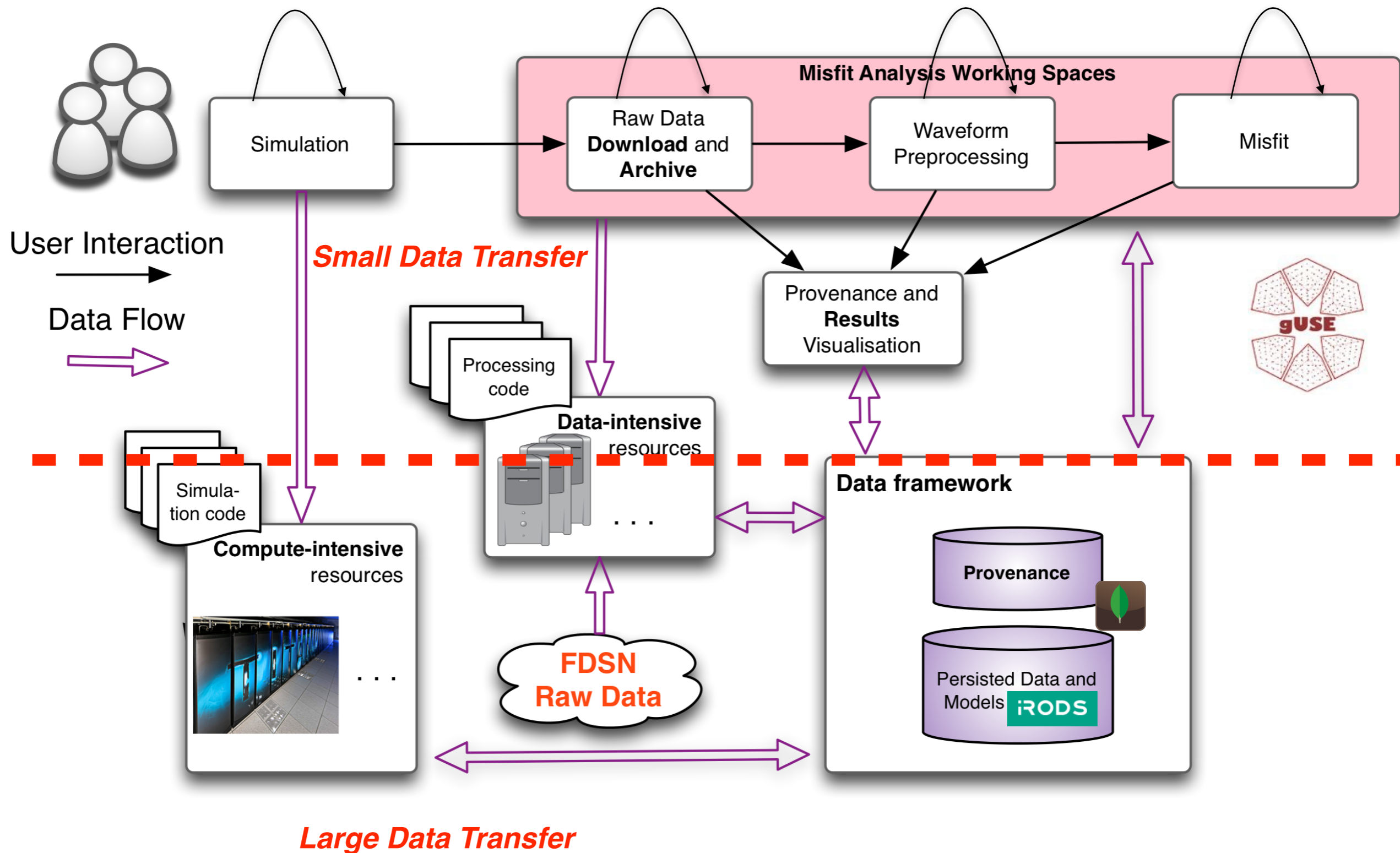
Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu



System & E-Infra Perspective



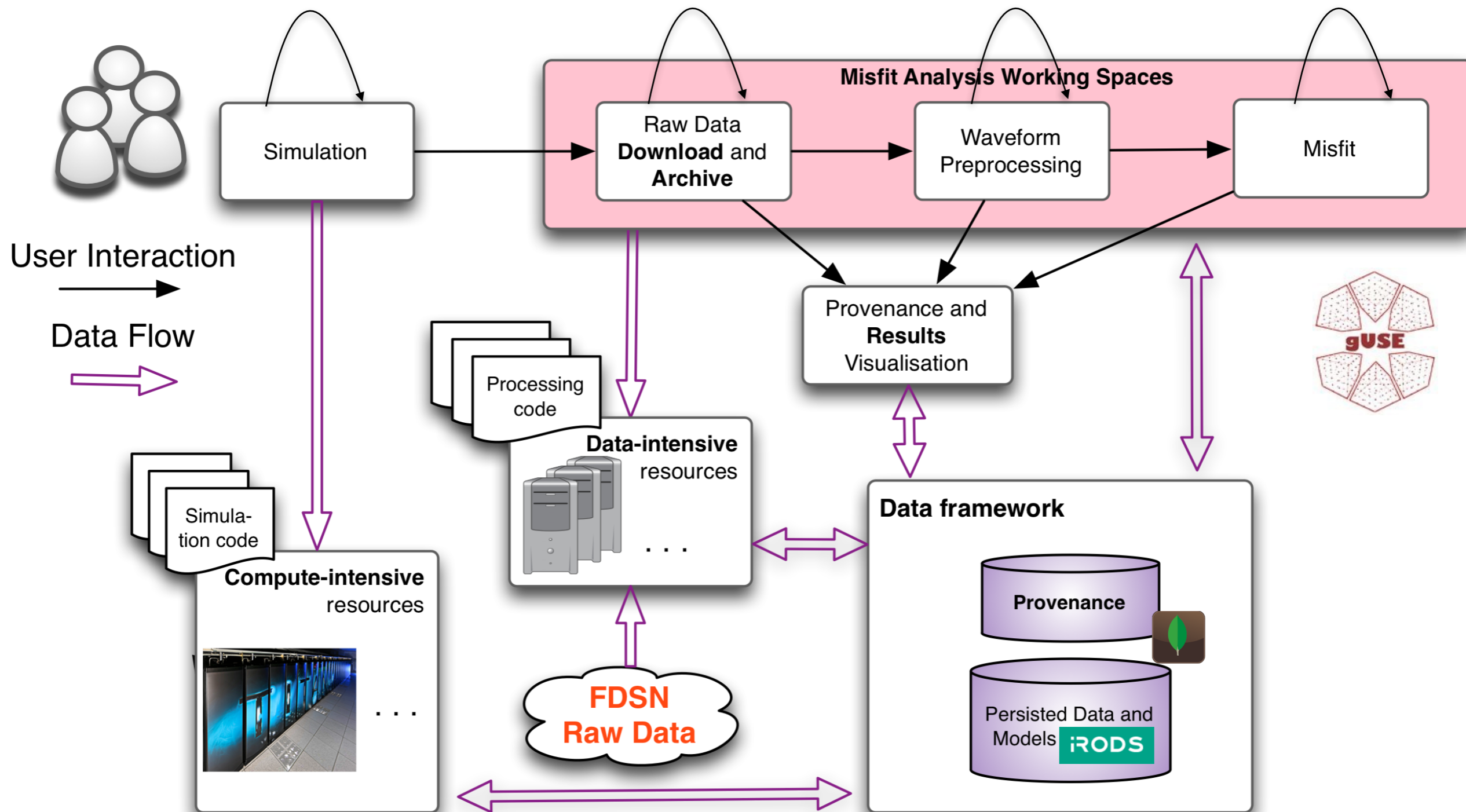
Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu



System & E-Infra Perspective



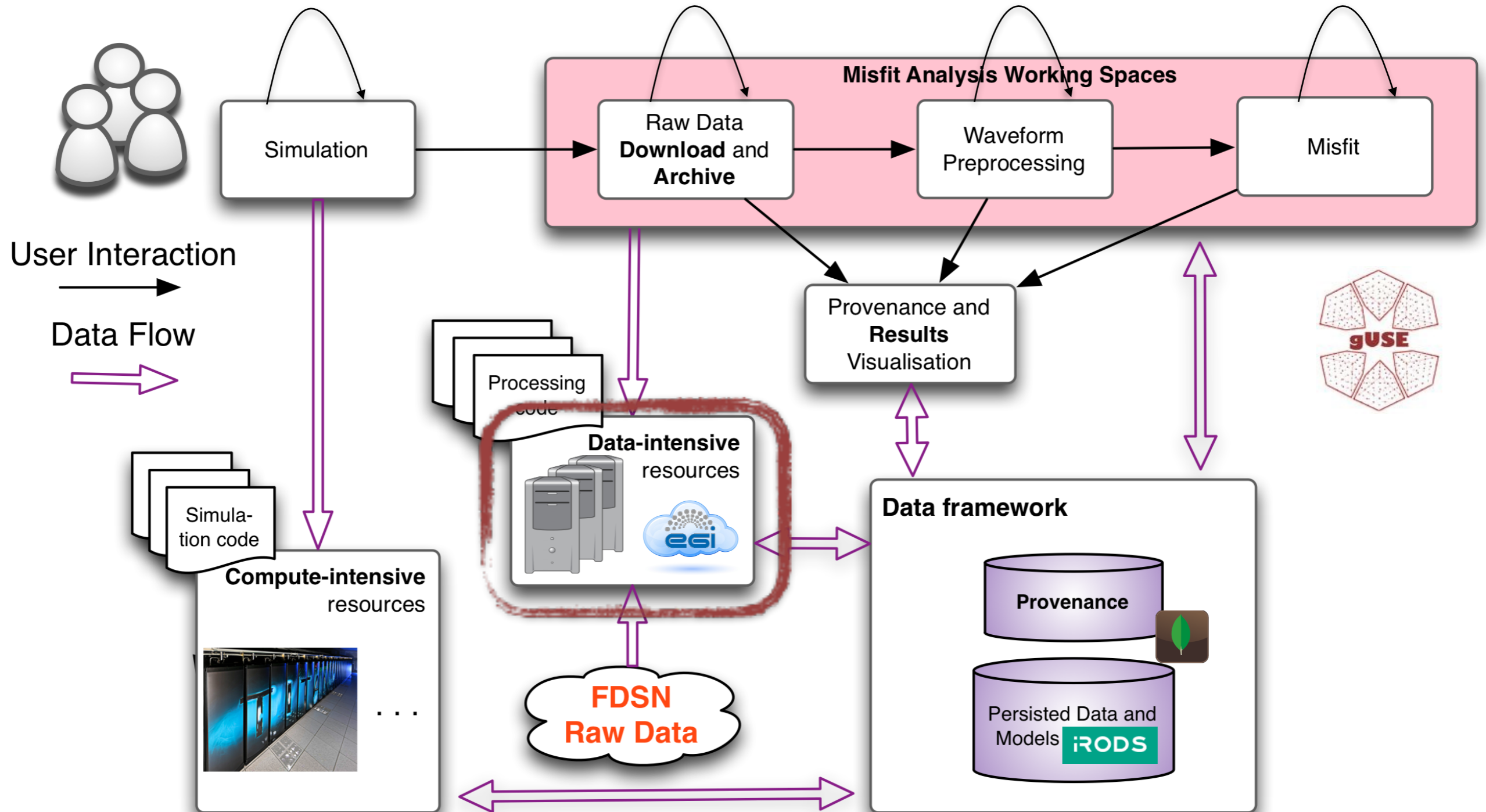
Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu



System & E-Infra Perspective



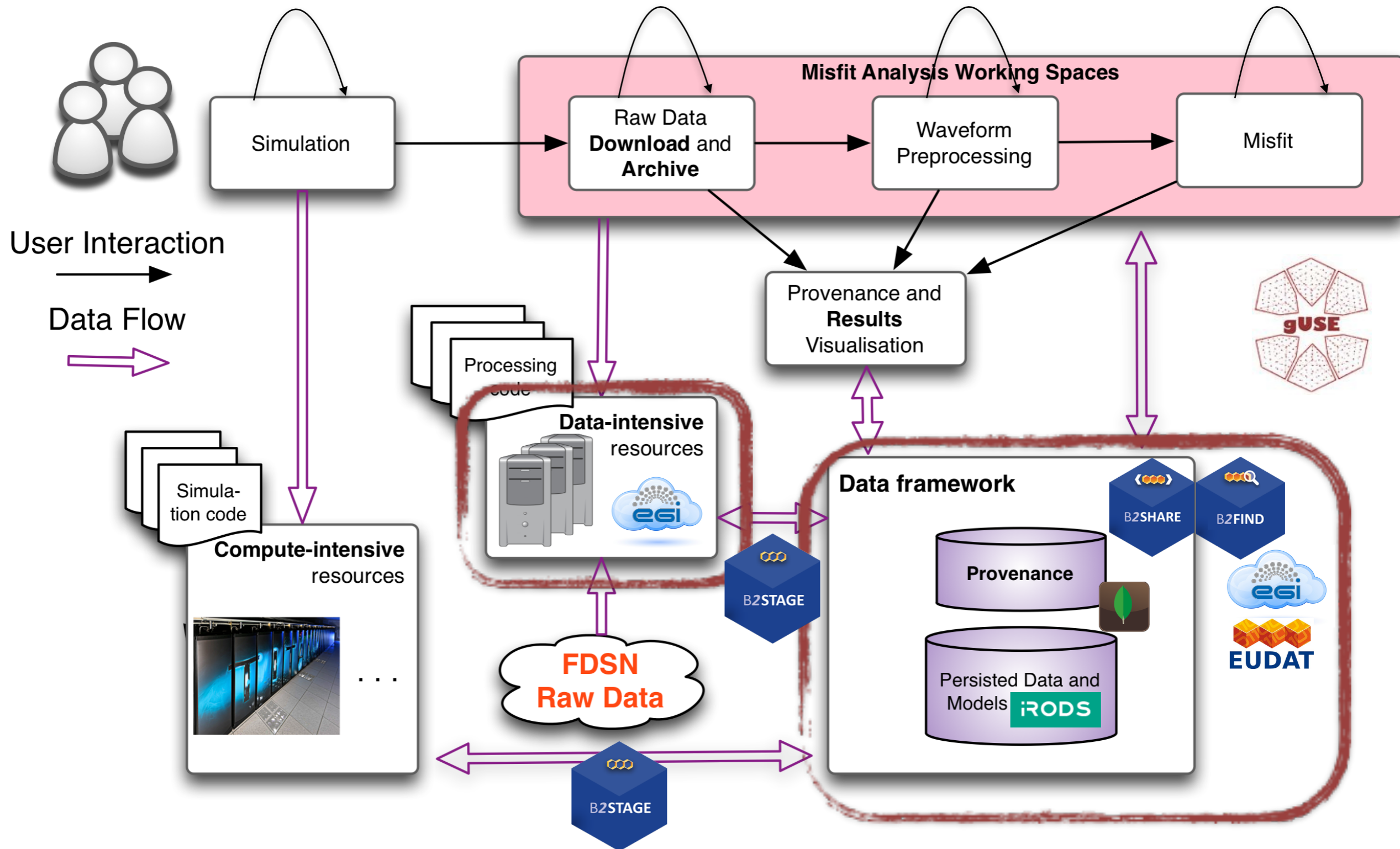
Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu



System & E-Infra Perspective



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu



ACCOUNT creation and GRID Authentication



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

- * Anyone can sign up and browse data, but real use of the scientific tools will require a GRID Certificate
- * Extensive documentation for acquiring certificate
- * Clear procedures for infrastructure access
- * Built in tool for proxy management

How to get your Certificate

Access to PRACE (HPC) and EGI (GRID)

The first step to access to Grid resources *en general* is to obtain adequate authentication mean delivered by an accredited issuer.

For PRACE resources, access to each resource is independent from the others and must be negotiated separately.

For EGI a complementary step is to register to a VO. While passing all these barriers, you should as well ask for an account on a UI

X.509 Certificates

X.509 certificates are requested in order to use many kind of academic computing structures, and mainly Grids such as PRACE and EGI which are both critical components of the VERGE computing platform.

Grid Certificates are delivered on a national basis, be it for individuals or computing resources by Certificates Authorities members of the International Grid Trust Federation, EuGridPMA being it's european side.

Here is the list for countries where VERGE partners are located:

- **France:** <https://igc.services.cnrs.fr/GRID2-FR/>
- **Germany:** http://www.gridka.de/cgi-bin/frame.pl?seite=ca/d_inhalt-en.html or https://pki.pca.dfn.de/grid-root-ca/cgi-bin/pub/pki?RA_ID=101 DFN Guide
- **Italy:** <http://security.fi.infn.it/CA>
- **Netherlands:** <http://ca.dutchgrid.nl>
- **UK:** <http://www.ngs.ac.uk/certoverview>

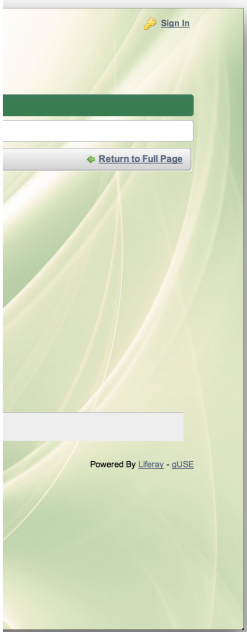
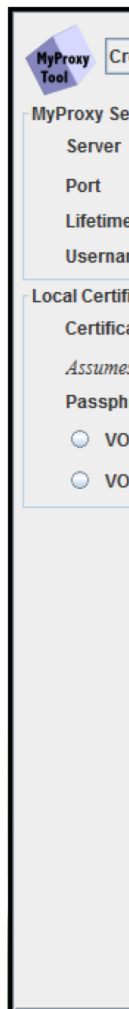
You should ask for a *personal* certificate of at least medium strength (512 bytes or more private key length). Authentication relies on cryptography and trust through a continuous chain of trusted bodies granting that the next to come is trustable as well. You are now part of this chain so please carefully read some [security](#)

ACCOUNT creation and GRID Authentication

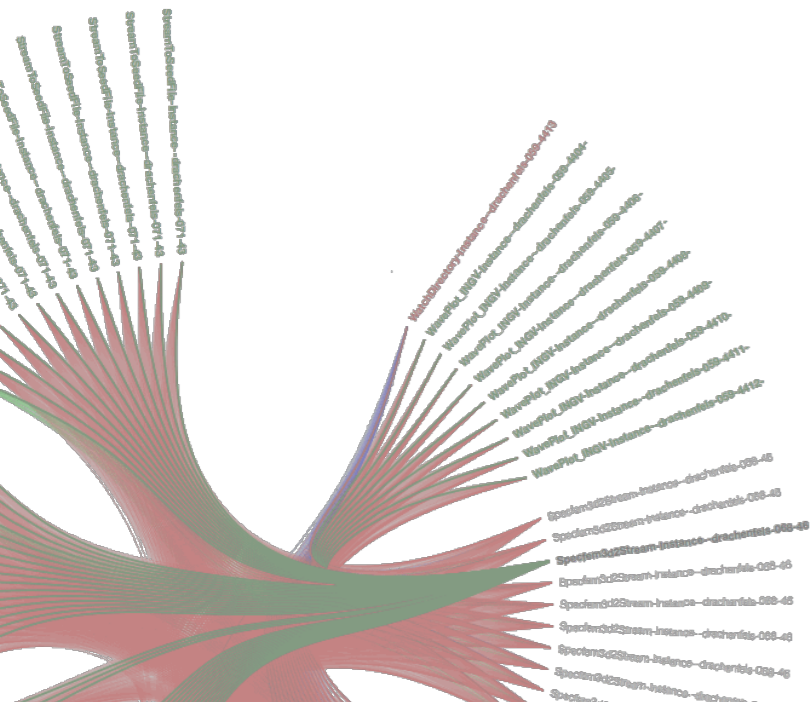
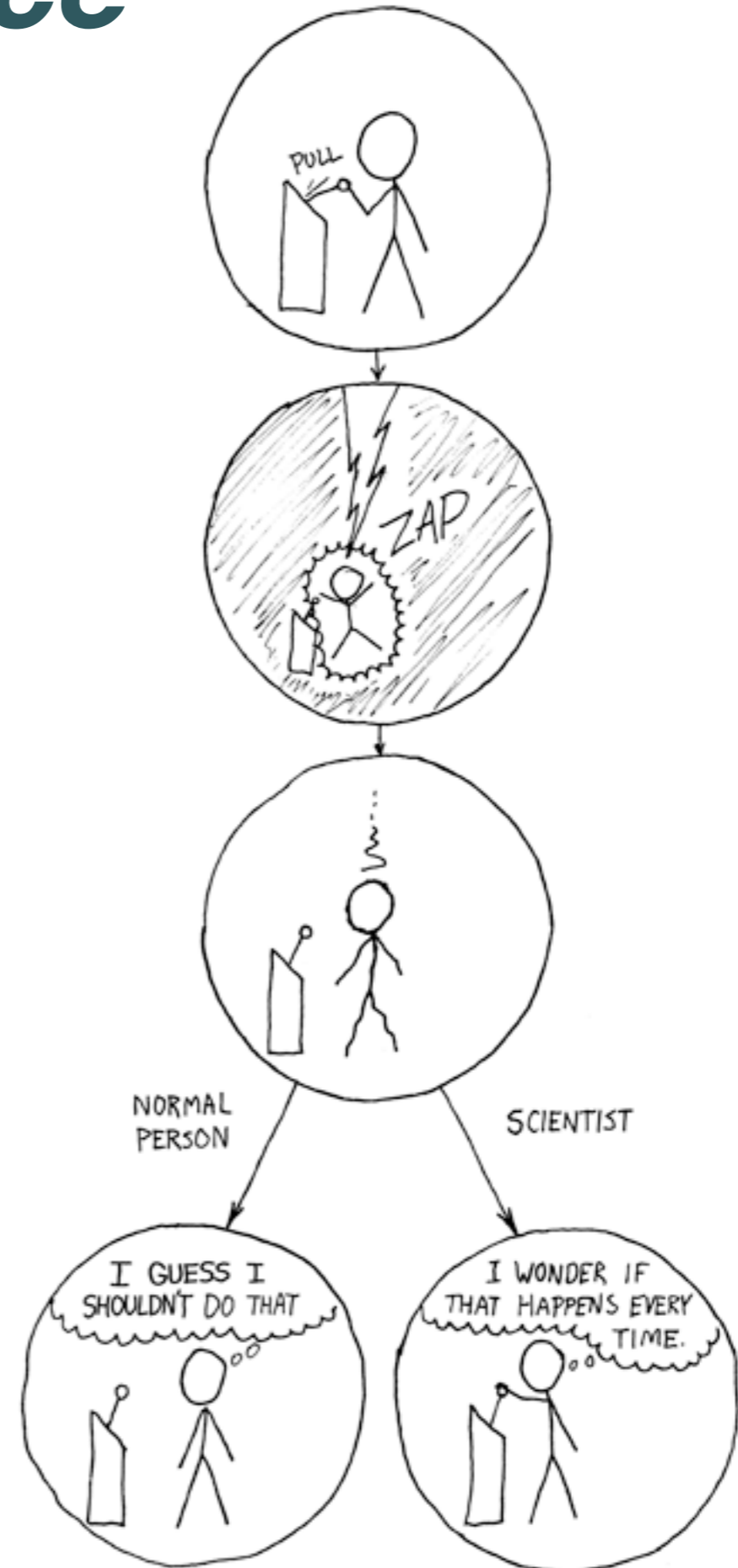


Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

- * Any tools
- * External
- * Clean
- * Built



Reproducible Science

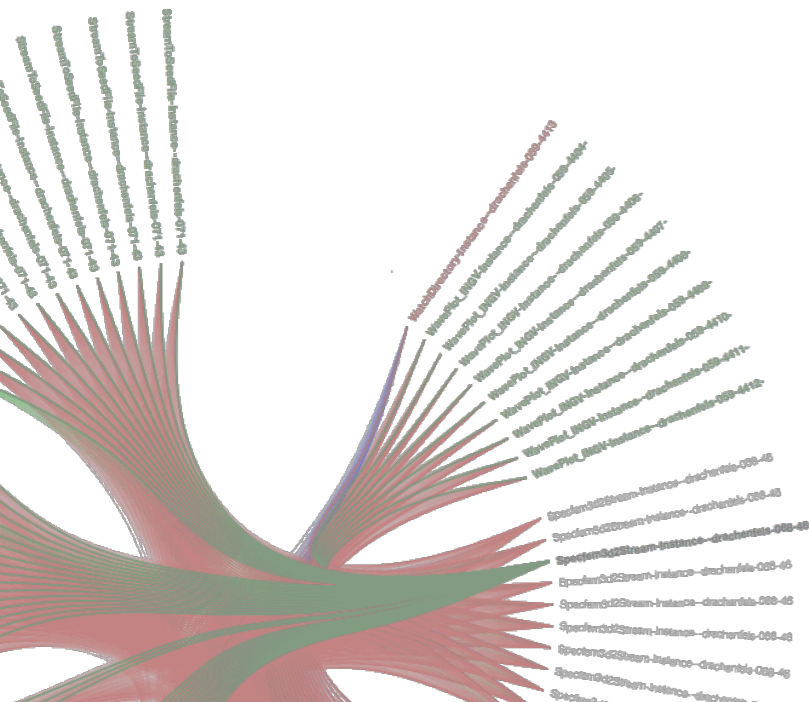
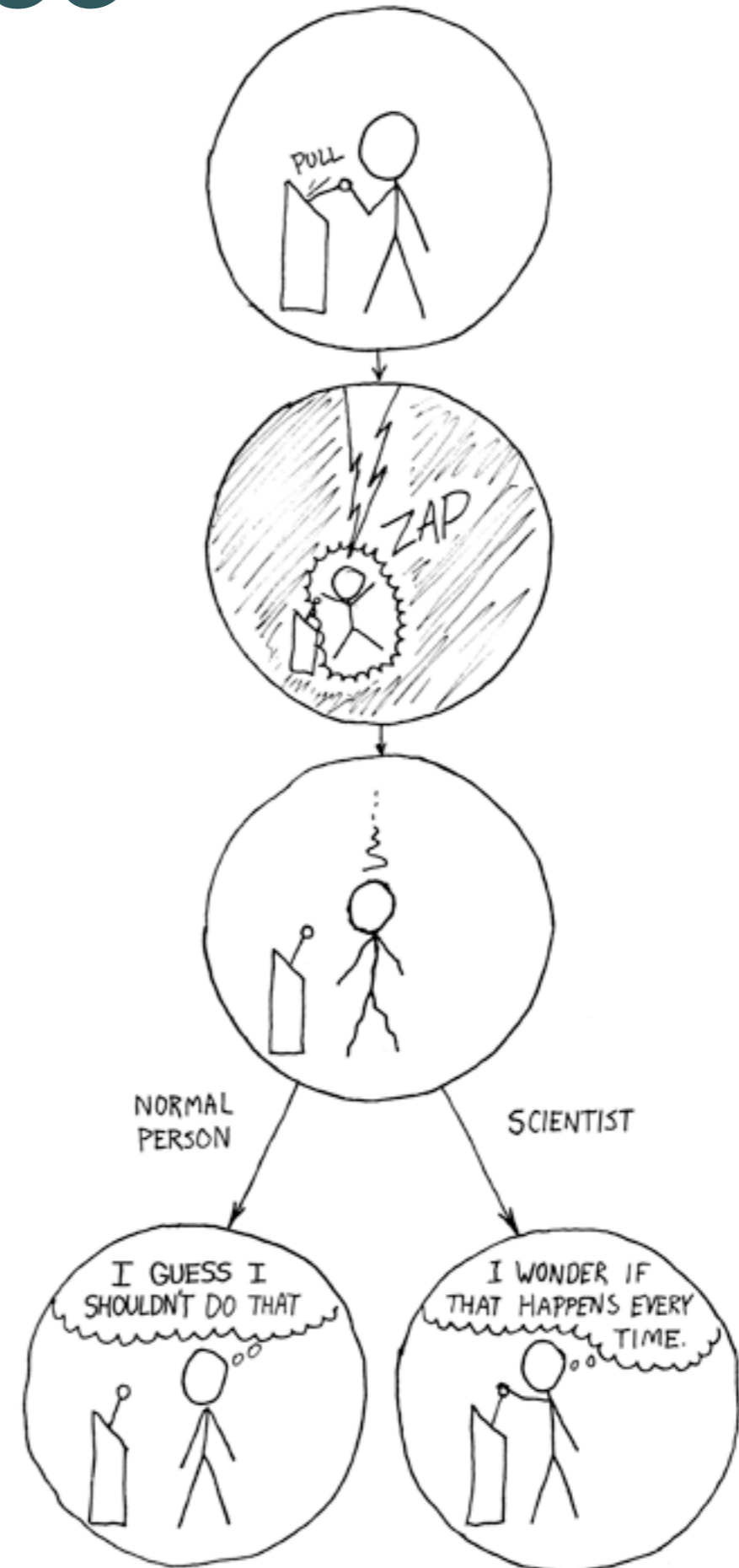


Reproducible Science

What does this suggest?

Scientists aren't normal..

Reproducibility is masochism..



Reproducible Science

What does this suggest?

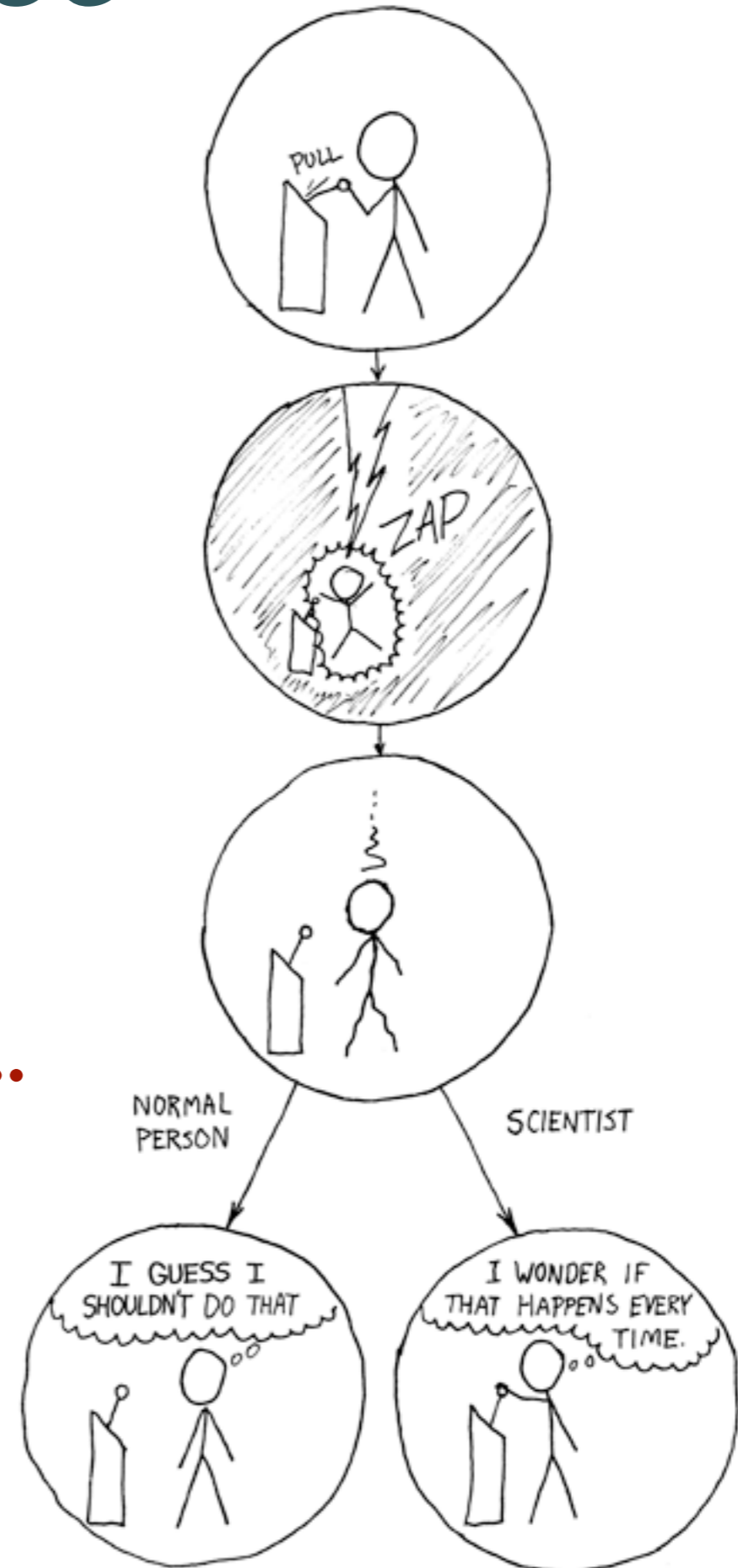
Scientists aren't normal..

Reproducibility is masochism..

Reproducibility is fundamental but difficult to achieve.

Not always convenient/possible..

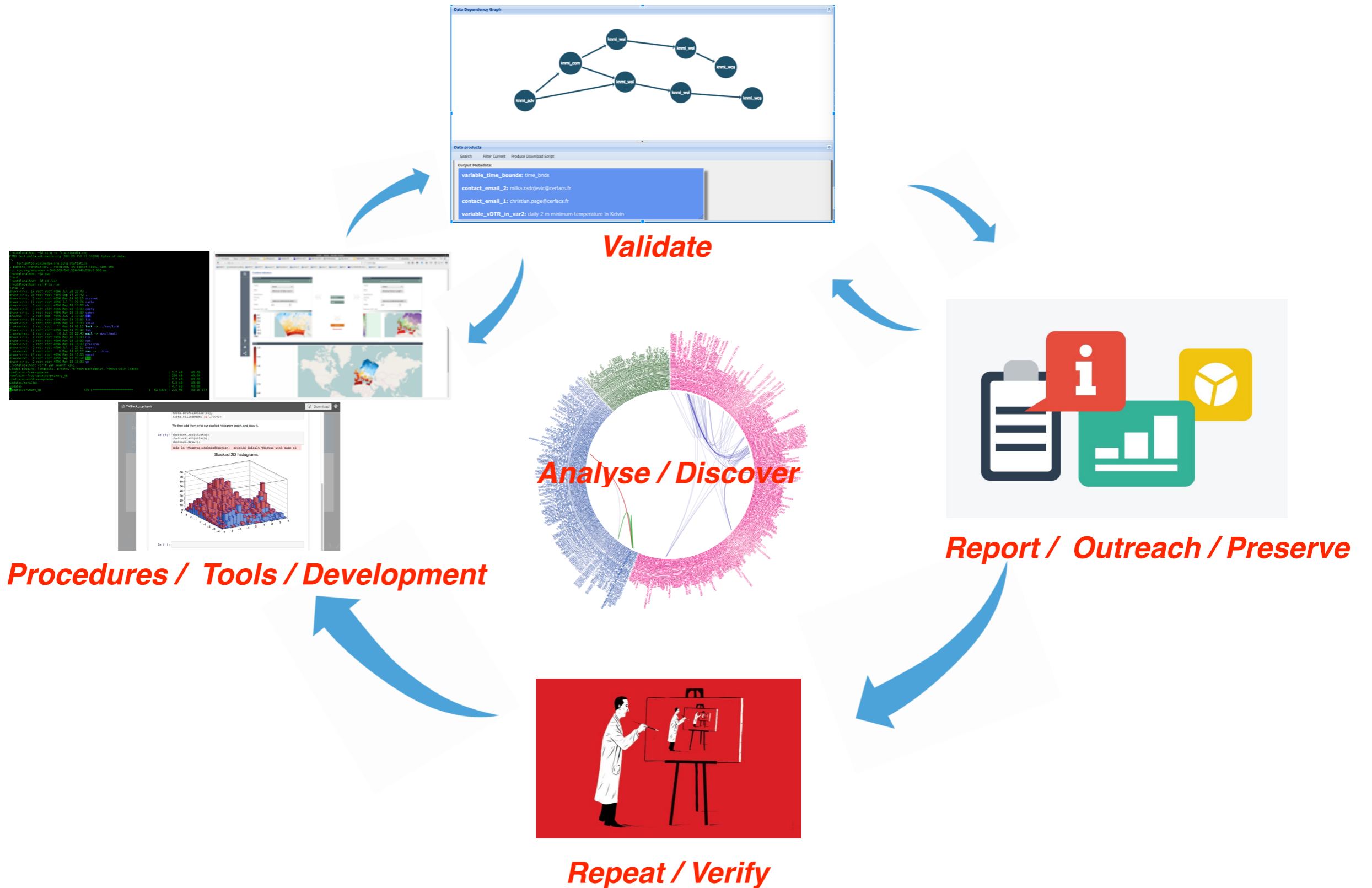
We need more than just rerun



Reproducibility Cycle(s)

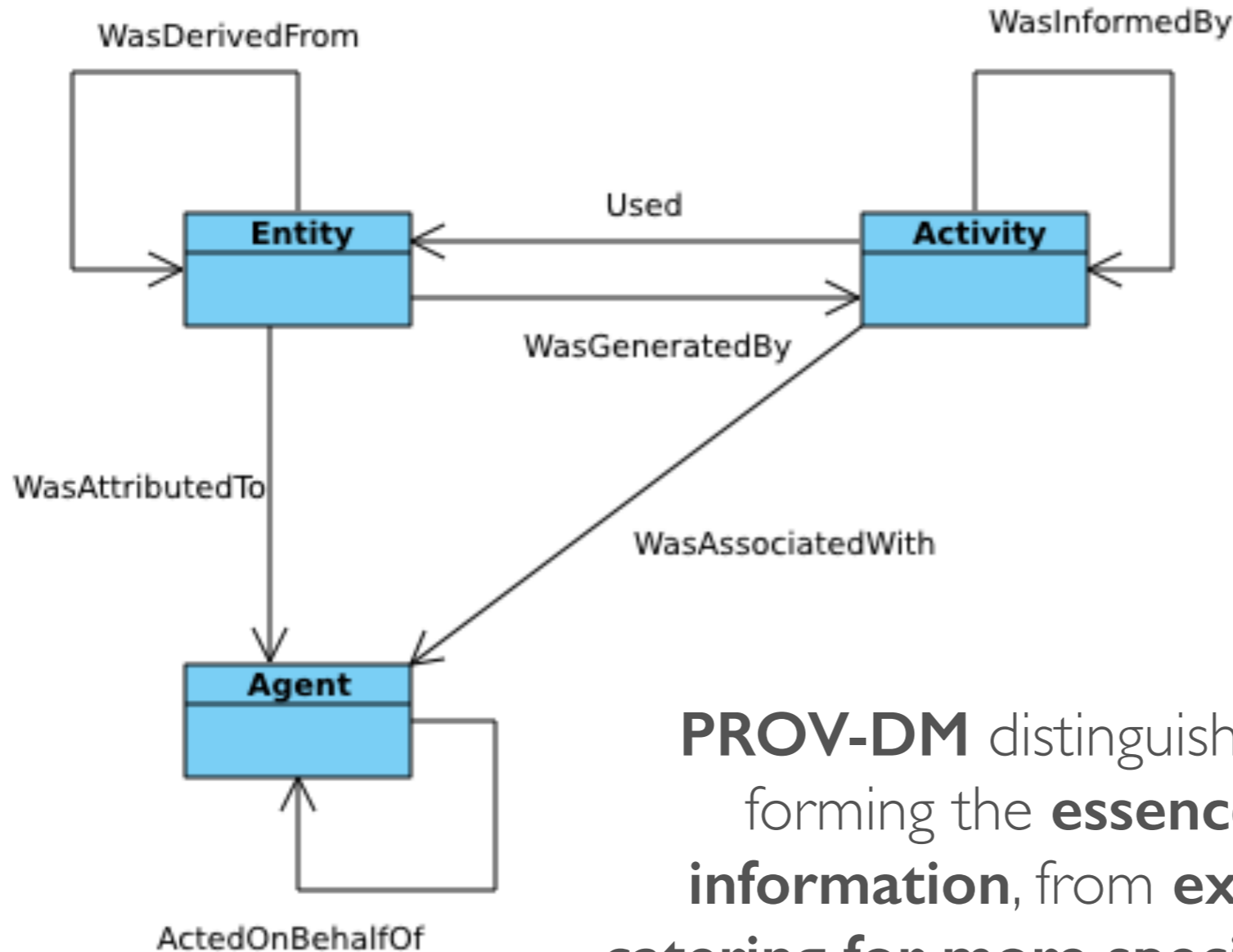


Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu



Provenance Model

Data-Lineage → *Knowledge*



W3C *PROV*

PROV-DM distinguishes **core structures**, forming the **essence of provenance information**, from **extended structures** catering for more specific uses of provenance.

Enactment ← *Machinery & Data*

Challenges

*“Automated system should provide **support** for a **consistent and effective acquisition** of provenance metadata.” [A. Misra] [I. Foster.]*

Expert users are part of the process - configuration and contextualisation:

“How much domain metadata should be contained?”

Scale of the provenance records:

“What level of granularity is needed to describe provenance of complex objects?”

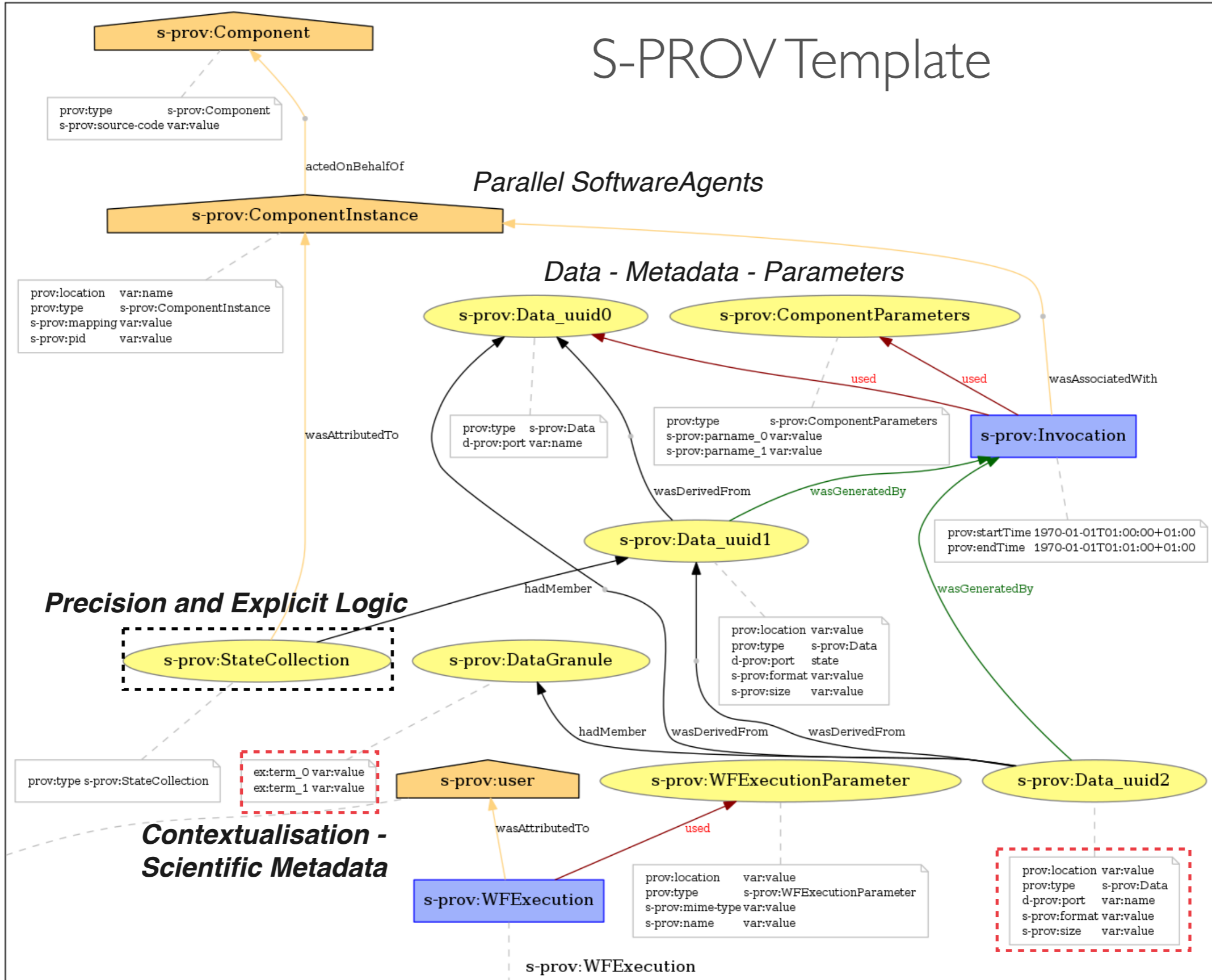
“Manage the scale of the provenance records to be recorded and processed”

Multiple levels of understanding:

“Provenance at different levels of abstraction, extract high-level summaries of provenance from detailed records.”



S-PROV Template



```

prov:type      s-prov:Component
s-prov:source-code var:value
    
```

```

prov:location  var:name
prov:type      s-prov:ComponentInstance
s-prov:mapping var:value
s-prov:pid     var:value
    
```

```

prov:type      s-prov:Data
d-prov:port    var:name
    
```

```

prov:type      s-prov:ComponentParameters
s-prov:parname_0 var:value
s-prov:parname_1 var:value
    
```

```

prov:startTime 1970-01-01T01:00:00+01:00
prov:endTime   1970-01-01T01:01:00+01:00
    
```

```

prov:type s-prov:StateCollection
    
```

```

ex:term_0 var:value
ex:term_1 var:value
    
```

```

prov:location var:value
prov:type     s-prov:Data
d-prov:port  state
s-prov:format var:value
s-prov:size  var:value
    
```

```

foaf:name var:name
    
```

```

prov:location  var:value
prov:type      s-prov:WFExecutionParameter
s-prov:mime-type var:value
s-prov:name    var:value
    
```

```

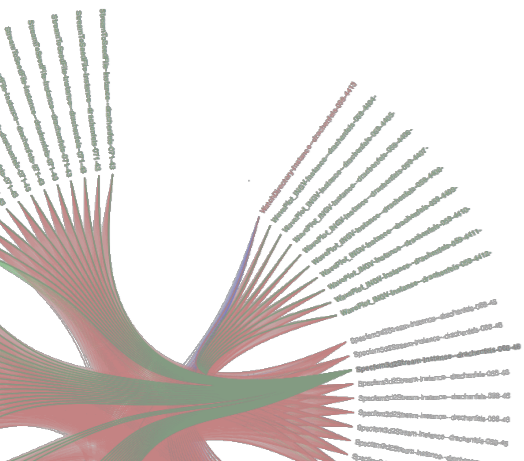
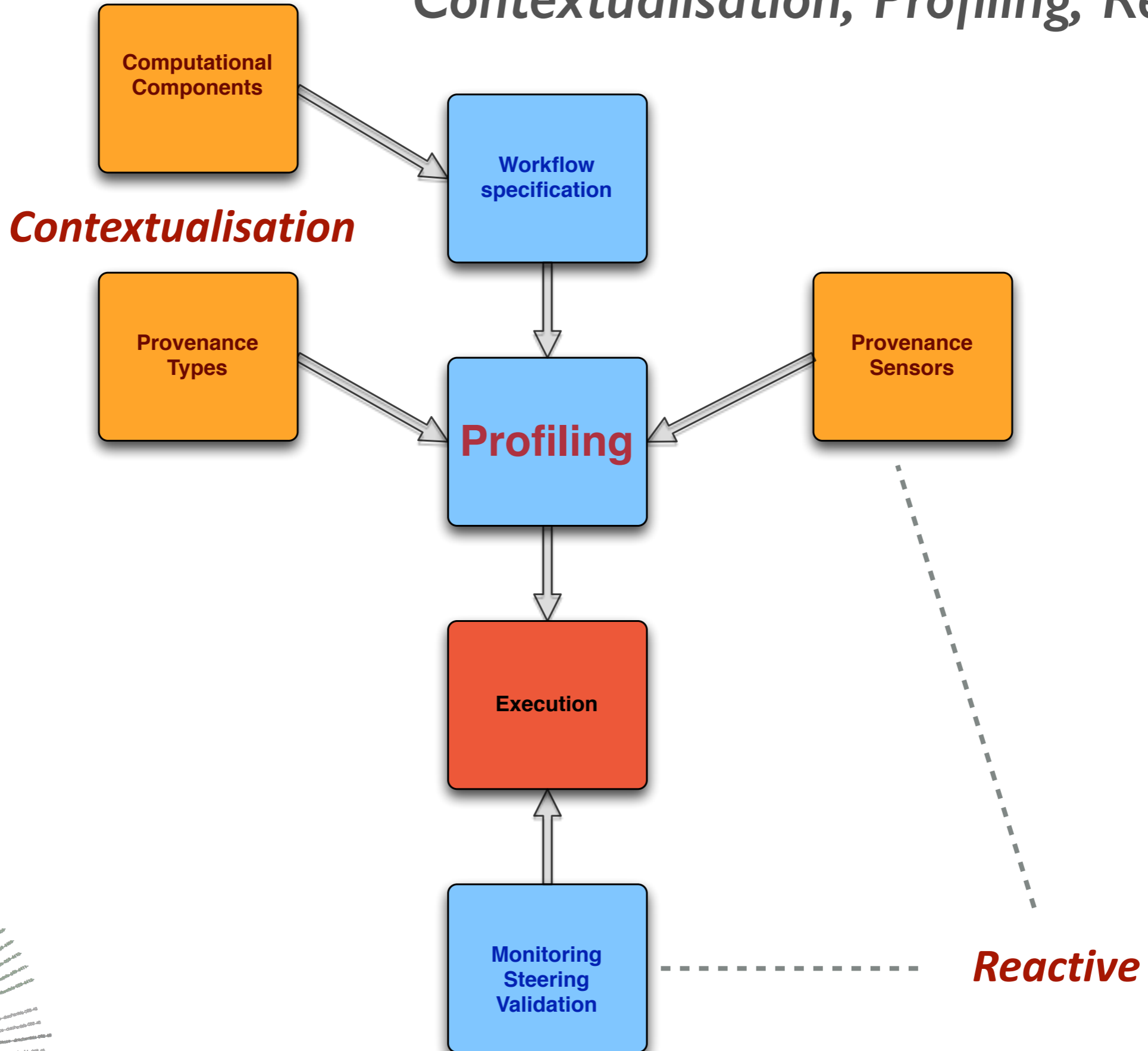
prov:location var:value
prov:type     s-prov:Data
d-prov:port  var:name
s-prov:format var:value
s-prov:size  var:value
    
```

```

prov:startTime 1970-01-01T01:00:00+01:00
prov:endTime   1970-01-01T01:01:00+01:00
s-prov:description var:value
s-prov:mapping    var:value
s-prov:name       var:name
s-prov:runId     var:value
s-prov:tags      var:value
s-prov:type      s-prov:WFExecution
    
```

Active Provenance framework

Contextualisation, Profiling, Reactive



Tools: S-ProvFlow - Reproducibility as a Service

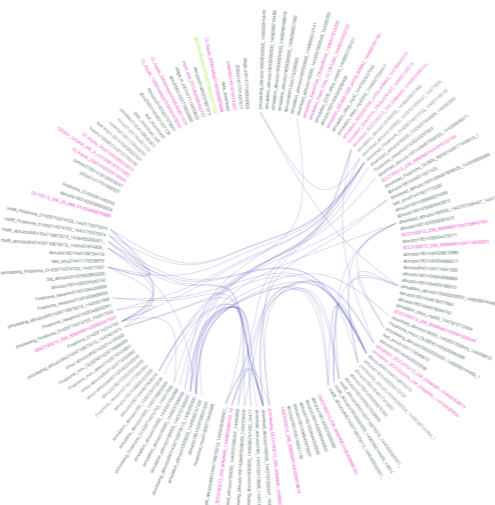
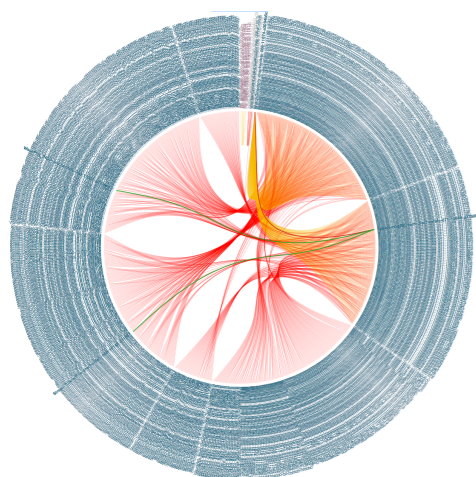
The screenshot displays the S-ProvFlow interface with several key components:

- Run Activity Monitor:** A table listing 25 runs with columns for ID, Date, and Messages. The messages include tasks like 'MergeImagesPE_write_drac...', 'MatchComponents_write_d...', 'PyflexPE_write_drachenfels...', 'streamProducer_write_drac...', 'PE_stagein_write_draco-ext...', and 'StreamMapper_write_draco...'.
- Data Dependency Graph:** A flow diagram showing nodes like 'Mergelma', 'MatchCom', 'PyflexPE', and 'streamPr'. A red arrow points to the 'Mergelma' node, labeled 'Data Resource'. The graph is annotated with 'Stateful invocation' and 'Navigation'.
- Search Data Window:** A dialog box with 'Values' Range and 'Annotations' tabs. It shows search terms 'sampling_rate, station', min values '20,CESX', max values '40,CESX', and mime-type 'image/png'.
- Data products:** A window showing 'Output Metadata' with fields like 'delta_syn: 0.05', 'id_raw: 043c00e4-0197-11e6-ab20-0025907b26', 'station_raw: CESX', and 'prov:type_raw: waveform'.
- Data File:** A window showing 'Link to data files or data images preview....' with two sets of plots for 'Component: E' and 'Component: N', including 'Phase Arrivals Seismograms' and 'STACIA'.

Monitoring:
System and user
Messages

Discovery:
Search on contextual metadata
(within and across runs)

Preview and Download



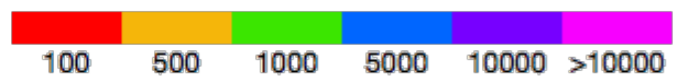
Visual analytics techniques on provenance

Perspectives on:

- data-intensive processes
- users and applications interactions
- data-reuse
- exploitation of resources..

Visual Analytics: Computation Distribution and execution modes

Edges: Data Transfer (bytes)



Nodes: Mappings



drachenfels-060

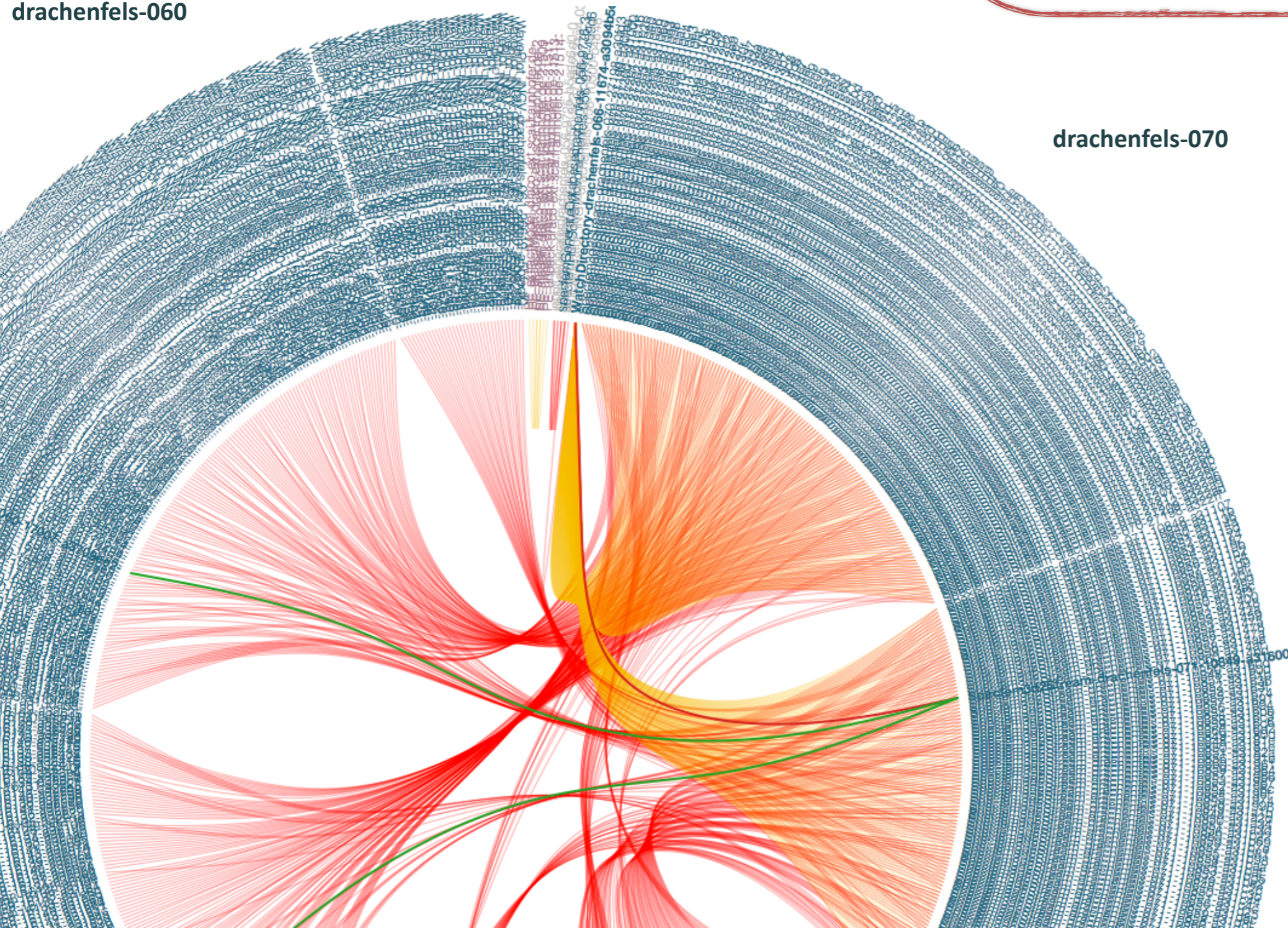
drachenfels-061

drachenfels-login

drachenfels-070

drachenfels-071

Level: iterations
Select Level
Iteration Range: minidx 0 maxidx 10
Grouping: worker
Select Grouping



Visual Analytics: Collaborative Perspectives

Data-reuse across Users

Vertex: Runs with color-coding for Users

Edges: Data exchange between runs

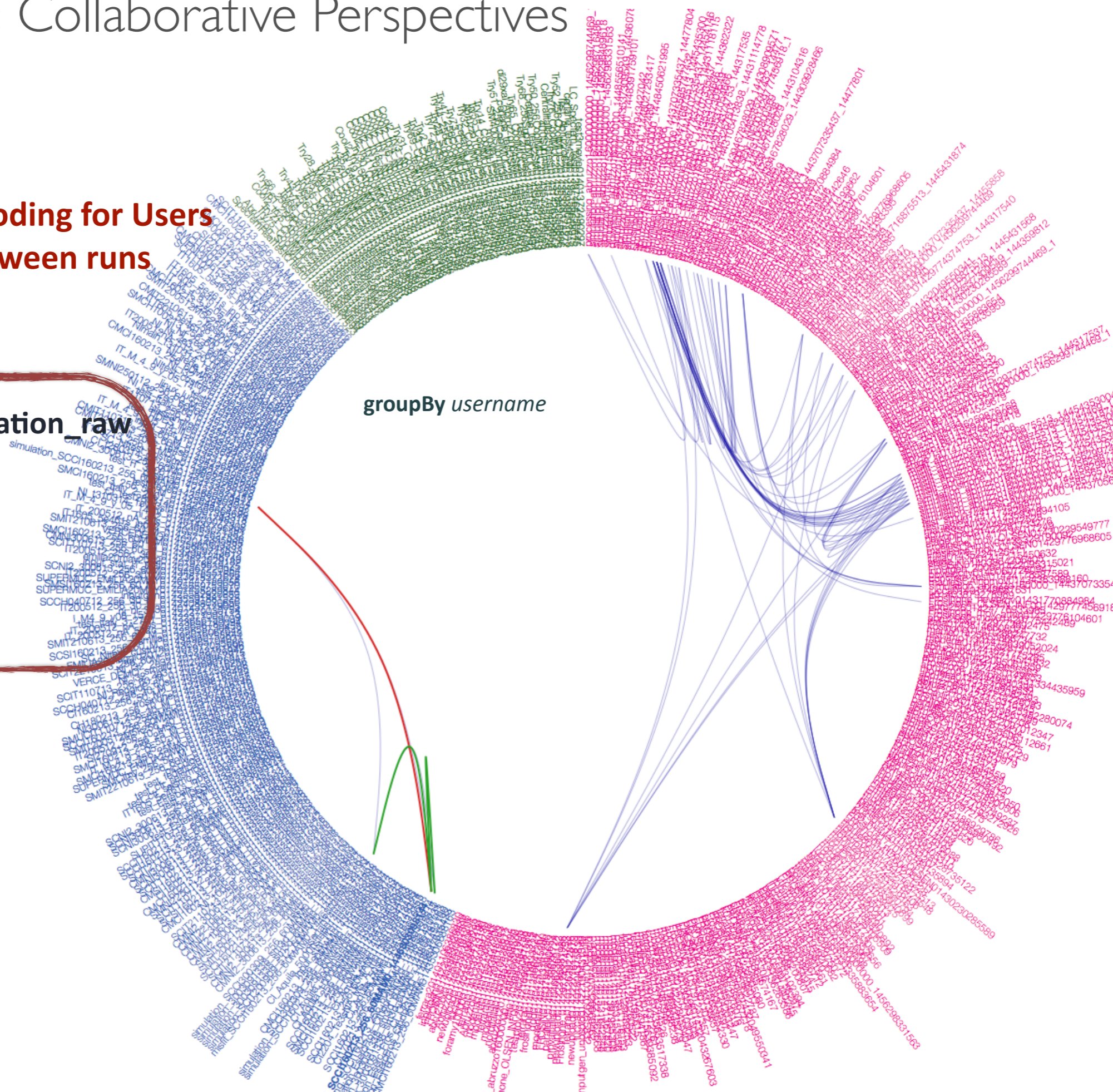
keys:magnitude,station,station_raw

minvalues: 4,CERA,CERA

maxvalues: 6,CERA,CERA

users: aspinuso,fmagnoni

cerlane



Visual Analytics: Infrastructure Perspectives

Data-reuse across Infrastructures

Vertex: Runs with color-coding for Users

Edges: Data exchange between runs

keys:magnitude,station_raw

minvalues: 4,CERA,CERA

maxvalues: 6,CERA,CERA

users: aspinuso,fmagnoni

cerlane

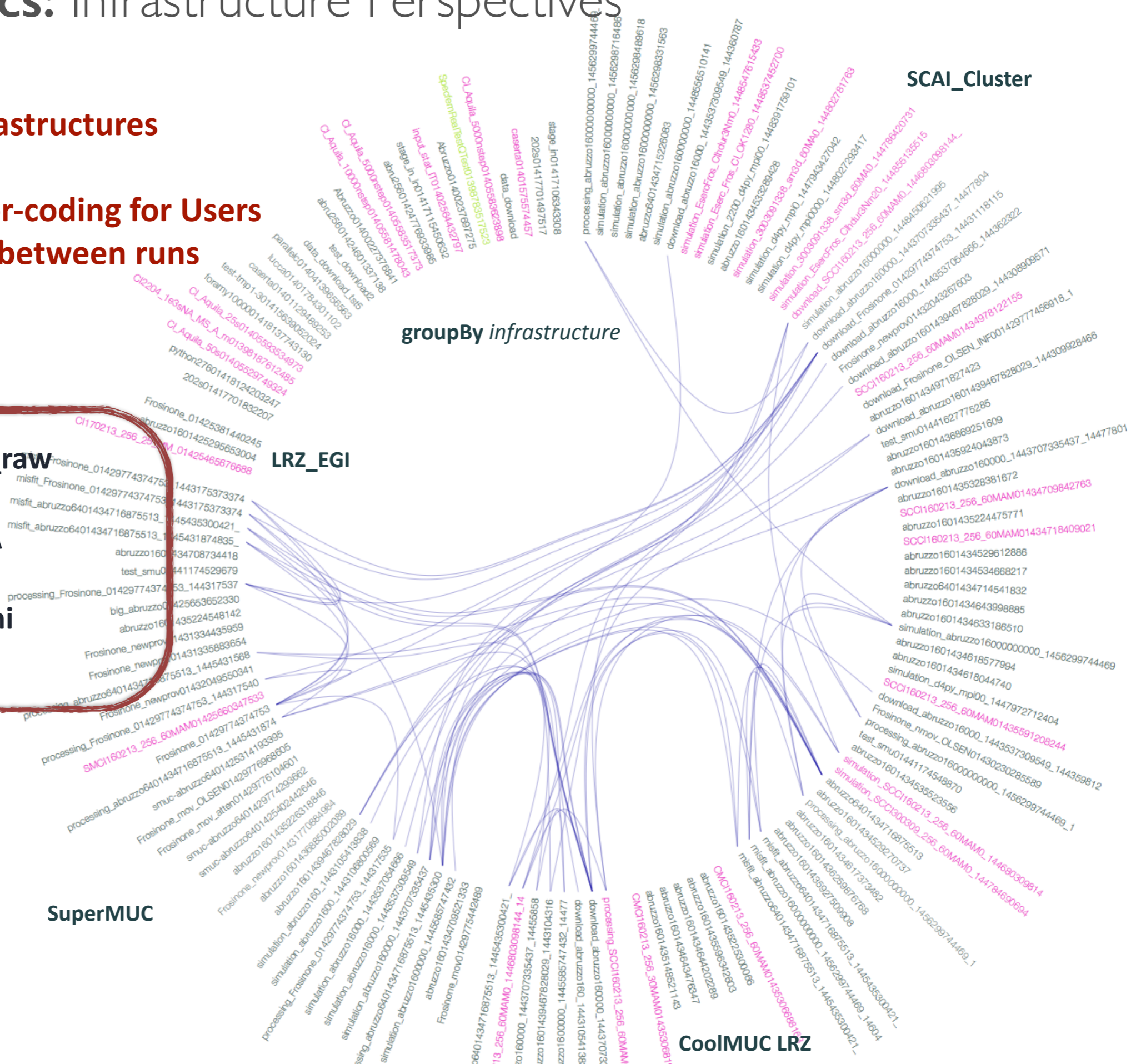
SuperMUC

LRZ_EGI

groupBy infrastructure

SCAI_Cluster

CoolMUC LRZ



2.5 - Explore the resulting provenance with static and interactive visualisations

The W3C-PROV provenance trace for a target data element is visualised.

```
In [14]: import prov
import io
import StringIO
from prov.model import ProvDocument, ProvBundle, ProvException, first, Literal
from prov.dot import prov_to_dot

def provToSvg(xml,output_f):

    xml_doc = StringIO.StringIO()
    xml_doc.write(str(xml))
    xml_doc.seek(0, 0)
    #print xml_doc
    doc=ProvDocument.deserialize(xml_doc,format="xml")
    dot = prov_to_dot(doc)
    return dot.create(format=output_f)

#prov_doc=open(prov).read()

#print prov1

svg_content=provToSvg(prov1,"png")

with open("PROV.png","w+") as text_file:
    text_file.write(str(svg_content))

from IPython.display import Image
Image("PROV.png")

# visualise NetCDF provenance in PNG
```

Out[14]:



1.2 Const

Instantiates t

In [3]:

```
#Initialis
def create
readX
readX.

analyse
analyse
analyse

write>
write>

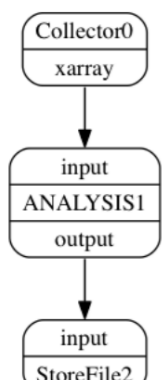
graph
graph.
graph.

return
```

graph = createWorkflowGraph()

```
from dispel4py.visualisation import display
display(graph)
```

SETTING NAME: Read
SETTING NAME: Analysis
SETTING NAME: Write



https://github.com/KNMI/wps_workflow/blob/master/ProvenanceDemo.ipynb

```
<?xml xmlns:prov="http://www.w3.org/ns/prov#" xmlns:s-prov="http://s-prov"
```

```
i.nl-10439-0758e376-d367-11e6-9b5f-901b0e877023">
s376-d367-11e6-9b5f-901b0e877023</s-prov:_id>
```

```
s8e376-d367-11e6-9b5f-901b0e877023</s-prov:runId>
v:startTime>
```

```
wId>
ne>
```

```
396.knmi.nl-10439-0758e376-d367-11e6-9b5f-901b0e877023">
cnmi.nl-10439-083c05d4-d367-11e6-9b5f-901b0e877023">
.nl-10439-083c05d4-d367-11e6-9b5f-901b0e877023"/>
:tor-pc150396.knmi.nl-10439-083b344c-d367-11e6-9b5f-901b0e877023"/>

cnmi.nl-10439-083c1c04-d367-11e6-9b5f-901b0e877023">
.nl-10439-083c1c04-d367-11e6-9b5f-901b0e877023"/>
low-pc150396.knmi.nl-10439-083c1326-d367-11e6-9b5f-901b0e877023"/>
```

```
i="s-prov:wgb_pc150396.knmi.nl-10439-0840b278-d367-11e6-9b5f-901b0e877023">
prov:Data_pc150396.knmi.nl-10439-0840b278-d367-11e6-9b5f-901b0e877023"/>
s-prov:Invocation_Write-pc150396.knmi.nl-10439-083c294c-d367-11e6-9b5f-901b0e877023"/>
```

```
rov:Invocation_collector-pc150396.knmi.nl-10439-083b344c-d367-11e6-9b5f-901b0e877023">
5T16:50:15.456858</prov:startTime>
16:50:15.462283</prov:endTime>
```

```
rov:Invocation_Write-pc150396.knmi.nl-10439-083c294c-d367-11e6-9b5f-901b0e877023">
5T16:50:15.463131</prov:startTime>
16:50:15.493009</prov:endTime>
```

```
rov:Invocation_Workflow-pc150396.knmi.nl-10439-083c1326-d367-11e6-9b5f-901b0e877023">
5T16:50:15.462562</prov:startTime>
16:50:15.462899</prov:endTime>
```

```
v:Parameters_Write-Instance--pc150396.knmi.nl-10439-083b30e6-d367-11e6-9b5f-901b0e877023">
QName">s-prov:Parameters</prov:type>
```

```
v:Data_pc150396.knmi.nl-10439-0840b278-d367-11e6-9b5f-901b0e877023">
xsd:string"></prov:location>
QName">s-prov:Data</prov:type>
-prov:annotations>
ormat>
nl-10439-0840b278-d367-11e6-9b5f-901b0e877023</s-prov:id>
port>
ze>
```