



### **DARE:** sustained power from abstraction and mapping

### at Creating platform-driven elnfrastructure innovation on EOSC

### **NCSR Demokritos campus**

Malcolm Atkinson, University of Edinburgh

10 July 2019



















# Outline

- Vision and Goals
- Key insights
- Architecture
- Achievements and Work in Progress

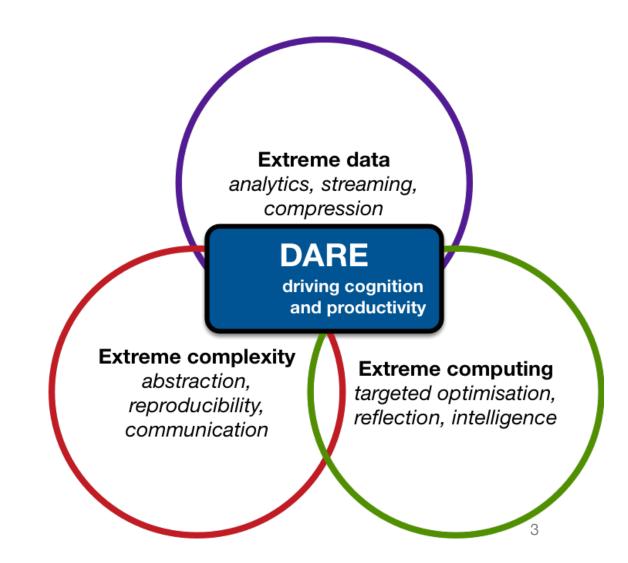




- Growth unlimited and sustained
  - Data
  - Computation
  - Complexity

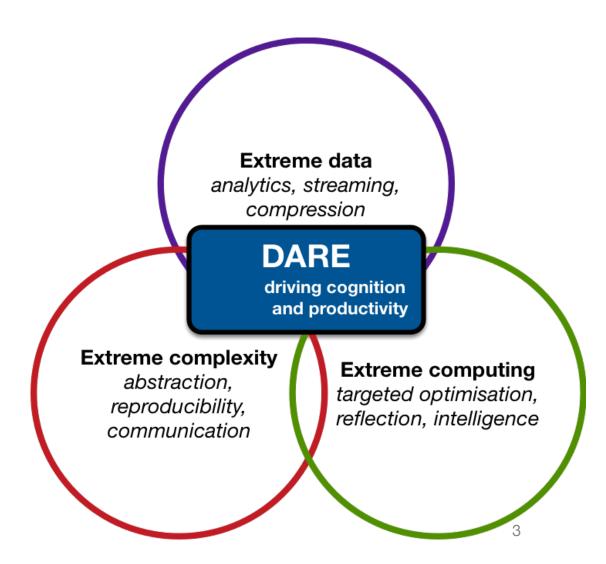


- Growth unlimited and sustained
  - Data
  - Computation
  - Complexity
- Challenges demanding and urgent





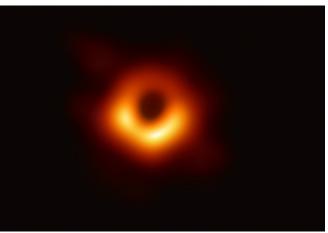
- Growth unlimited and sustained
  - Data
  - Computation
  - Complexity
- Challenges demanding and urgent
- Conserving scarce resources
  - Human intellectual effort
  - Willingness to collaborate
  - Energy / GHG emissions



#### Science and collaborations

#### **"Black Hole Image Makes History; NASA Telescopes Coordinated Observations"**

A black hole and its shadow have been captured in an image for the first time, a historic feat by an international network of radio telescopes called the Event Horizon Telescope.





"Science is powerful"

#### "Collaboration is the key to cancer research"

To fight the disease effectively, researchers from across the scientific spectrum and beyond must join forces.

2018, Neal Savage

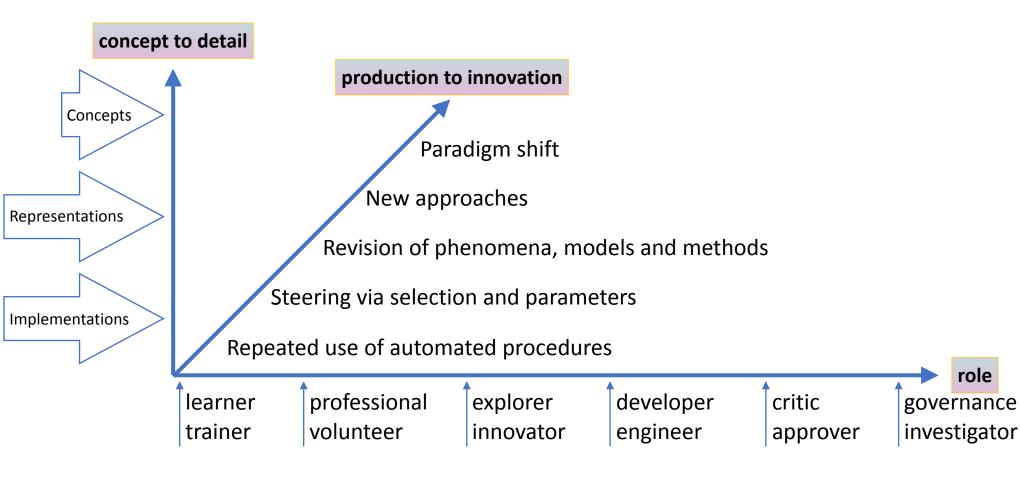


Science can solve some of the world's most complex problems, but only when the best ideas are brought together. #TogetherScienceCan

Acknowledge Luca Trani, KNMI & EPOS

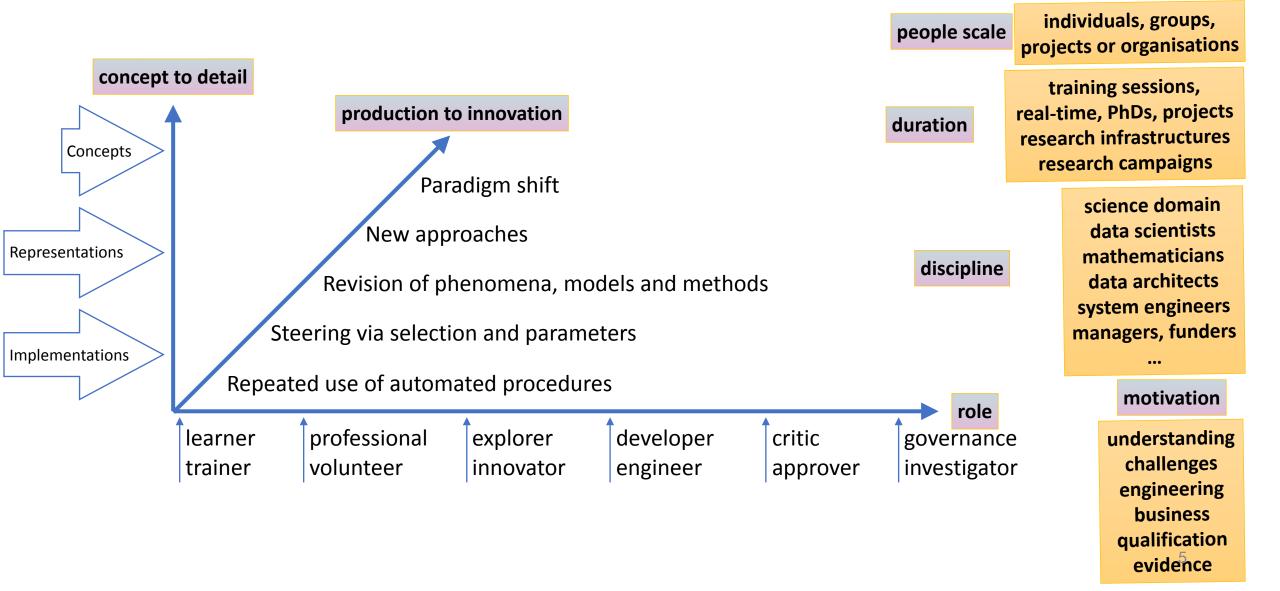


# Respond to community complexity



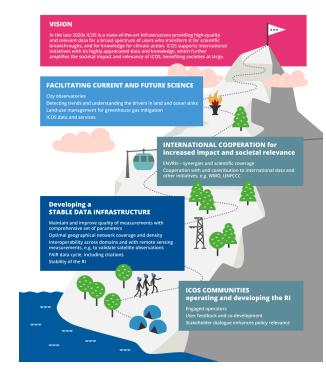


# Respond to community complexity



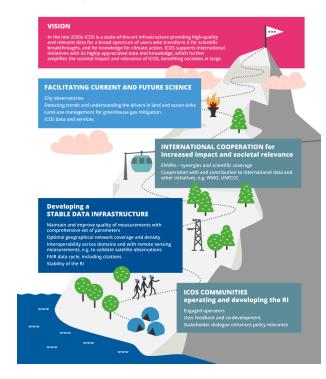






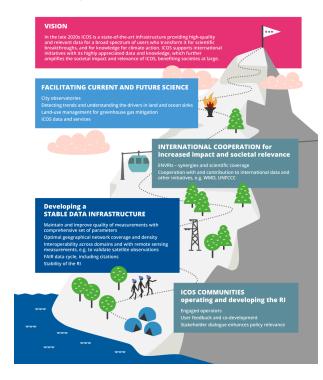


- Deliver stability
  - established practices unchanged
  - refined methods unchanged
  - negotiated agreements unchanged



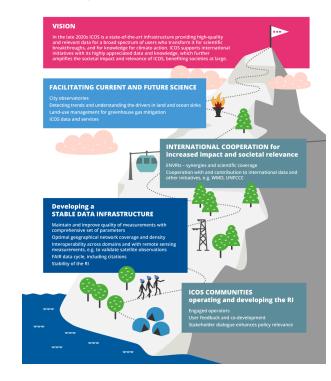


- Deliver stability
  - established practices unchanged
  - refined methods unchanged
  - negotiated agreements unchanged
- New capabilities



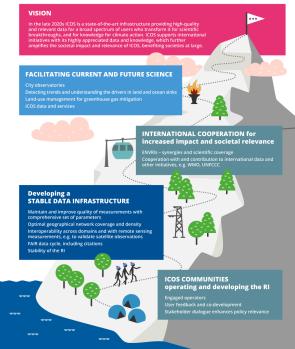


- Deliver stability
  - established practices unchanged
  - refined methods unchanged
  - negotiated agreements unchanged
- New capabilities
- Improved productivity





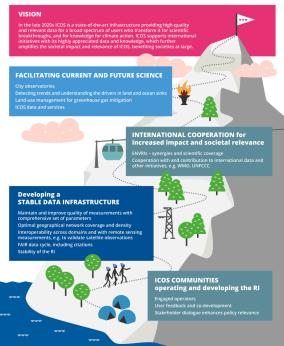
- Deliver stability
  - established practices unchanged
  - refined methods unchanged
  - negotiated agreements unchanged
- New capabilities
- Improved productivity
- Agile response to new challenges and opportunities





- Deliver stability
  - established practices unchanged
  - refined methods unchanged
  - negotiated agreements unchanged
- New capabilities
- Improved productivity
- Agile response to new challenges and opportunities
- Riding the technology wave







### Intention



|<sup>3</sup>



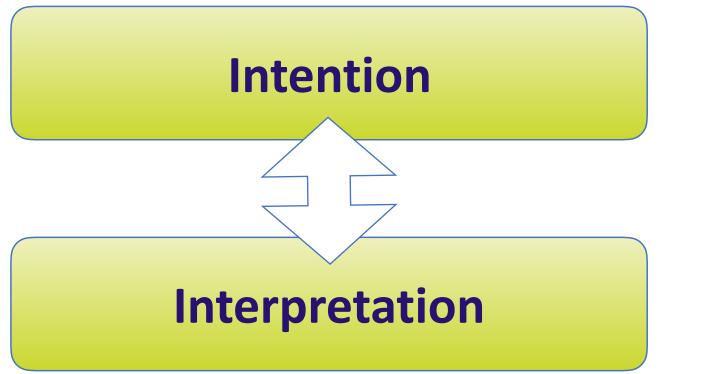
### Intention

refined precise honoured persistently domain focused shaped by campaigns



<mark>-</mark>3



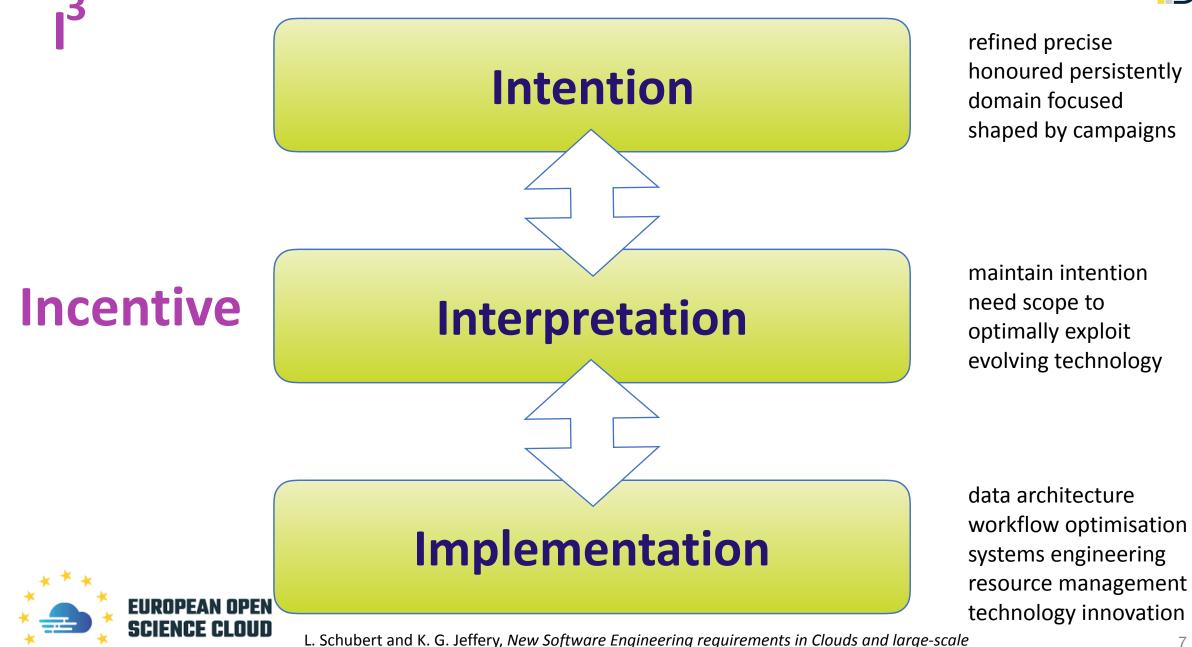


refined precise honoured persistently domain focused shaped by campaigns

maintain intention need scope to optimally exploit evolving technology







systems, IEEE Cloud Computing 2 (1) 48–58, 2015.



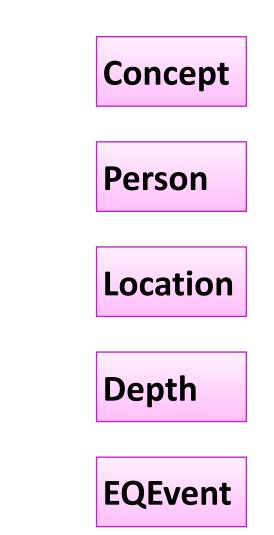
## Outline

- Vision and Goals
- Key insights
- Architecture
- Achievements and Work in Progress



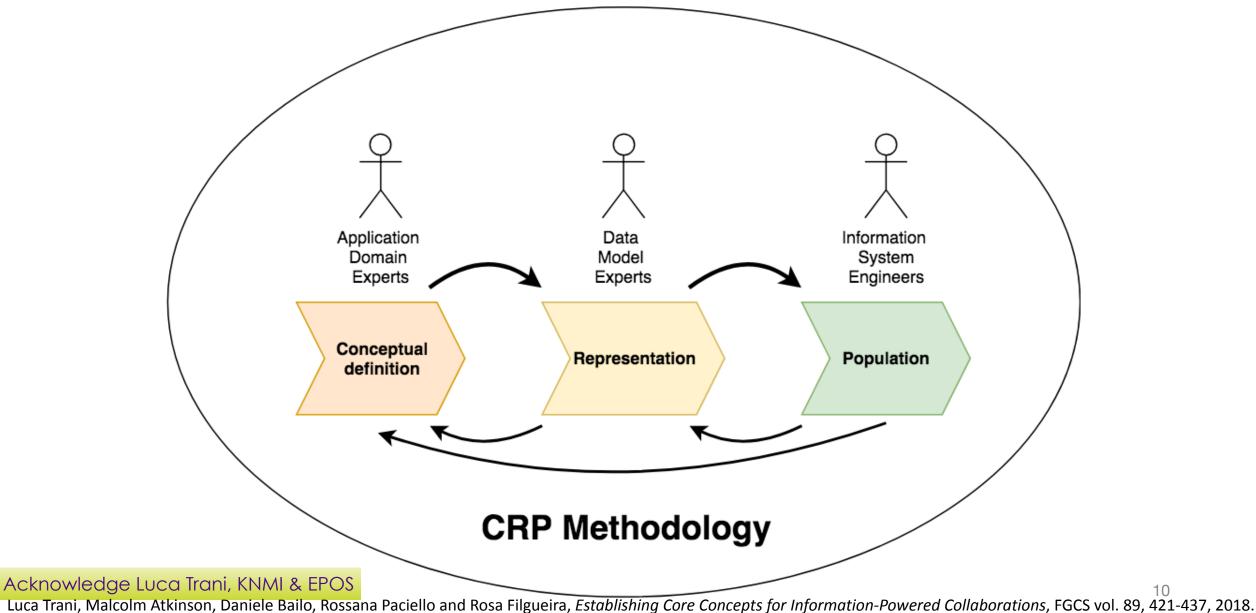
### Concepts

- what a communication is about
- precise terminology
- precise properties
- precise relationships
- Representations
  - how they are communicated
- Populations
  - the instances in play





## Clarifying the concepts in use



10



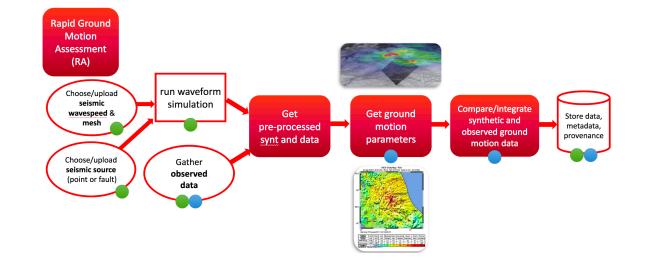
#### Methods

- purpose and processes
- conceptual inputs
- controls, parameters and active provenance
- conceptual results delivered
- Notations
  - creation and refinement
  - avoid distracting detail and lock-ins
  - workflows as requests for action
- Workloads
  - testing & validation
  - steered production
  - automated routine



#### Methods

- purpose and processes
- conceptual inputs
- controls, parameters and active provenance
- conceptual results delivered
- Notations
  - creation and refinement
  - avoid distracting detail and lock-ins
  - workflows as requests for action
- Workloads
  - testing & validation
  - steered production
  - automated routine





Rapid Ground

#### Methods

- purpose and processes
- conceptual inputs
- aued multiplaisation for capabilities and cost saving immediate local action for testing controls, parameters and active provenance
- conceptual results delivered
- Notations
  - creation and refinement
  - avoid distracting detail and lock-ins
  - workflows as requests for action
- Workloads
  - testing & validation
  - steered production
  - automated r





### • Data

- primary, intermediate, evidential
- from human input, instruments, observations, simulations and analyses
- often standardised imperfectly
- every possible diversity
- Organisation and structure
  - data models, metadata, storage systems, databases, document, latent
  - optimised encodings
- Populations
  - many, many instances growing in diversity, number and size



### • Data

- Organisation and structure
- Populations



Collections	Concept { }
<ul> <li>users create, build, manage and use them</li> <li>software works with them</li> </ul>	Person
<ul> <li>Representation</li> <li>sets, bags, tuples, trees, sequences, time series, …</li> </ul>	Location { }
<ul> <li>optimised encodings, scientific databases</li> <li>Populations</li> </ul>	Depth []
<ul> <li>many instances growing in diversity, number &amp; size</li> </ul>	EQEvent { }



### Collections

- Unstantialisation for engineering and cost saving users create, build, manage and use them
- software works with them
- Representation
- Local control for user engagement sets, bags, tuples, trees, sequer
  - optimised encodings, sei
- Populations
  - many insta





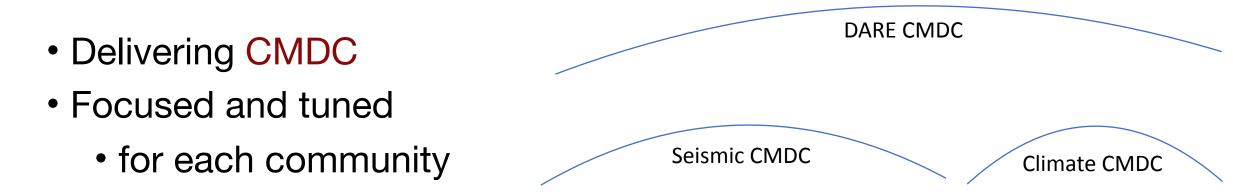




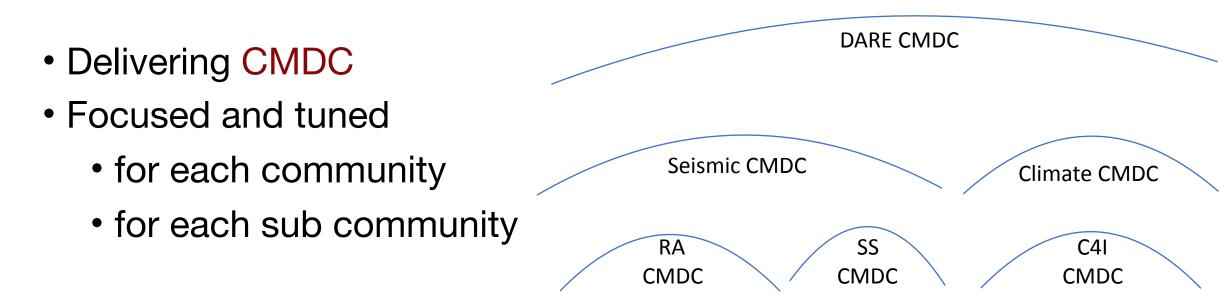






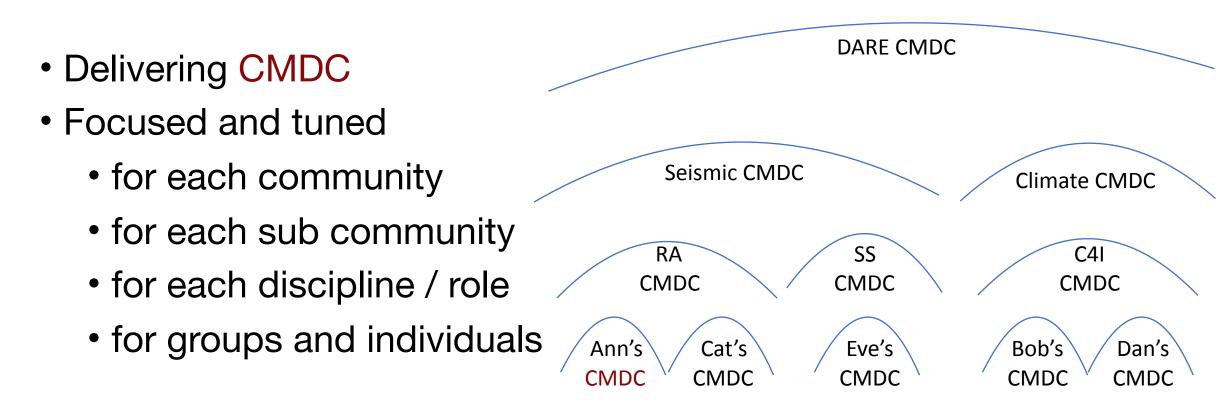








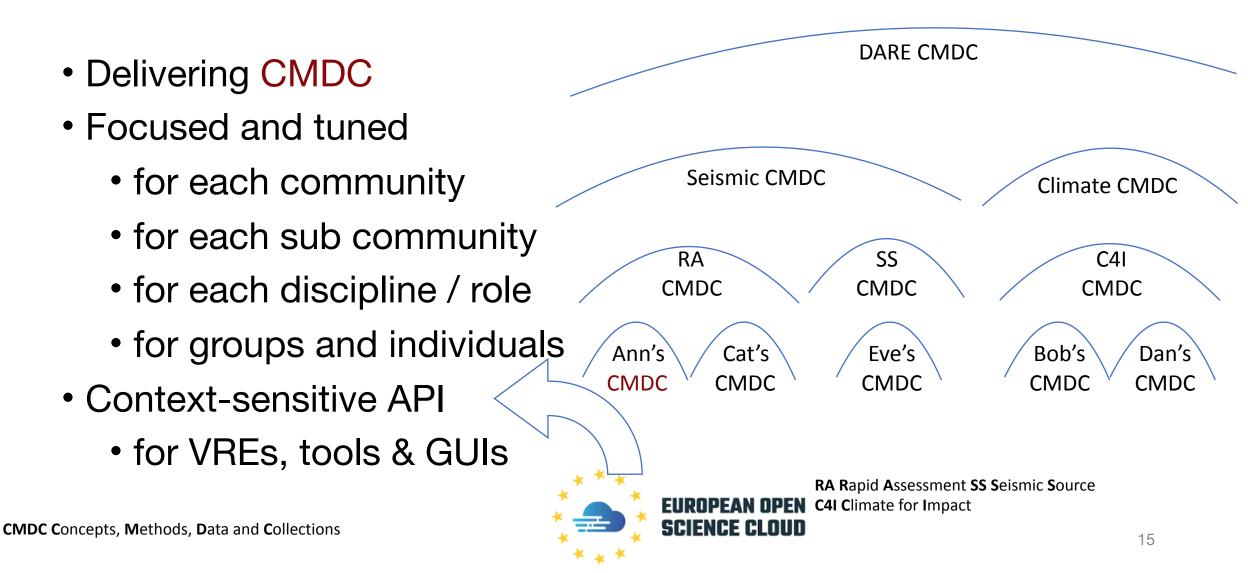
### Specialised but related work contexts



RA Rapid Assessment SS Seismic Source C4I Climate for Impact



## Specialised but related work contexts

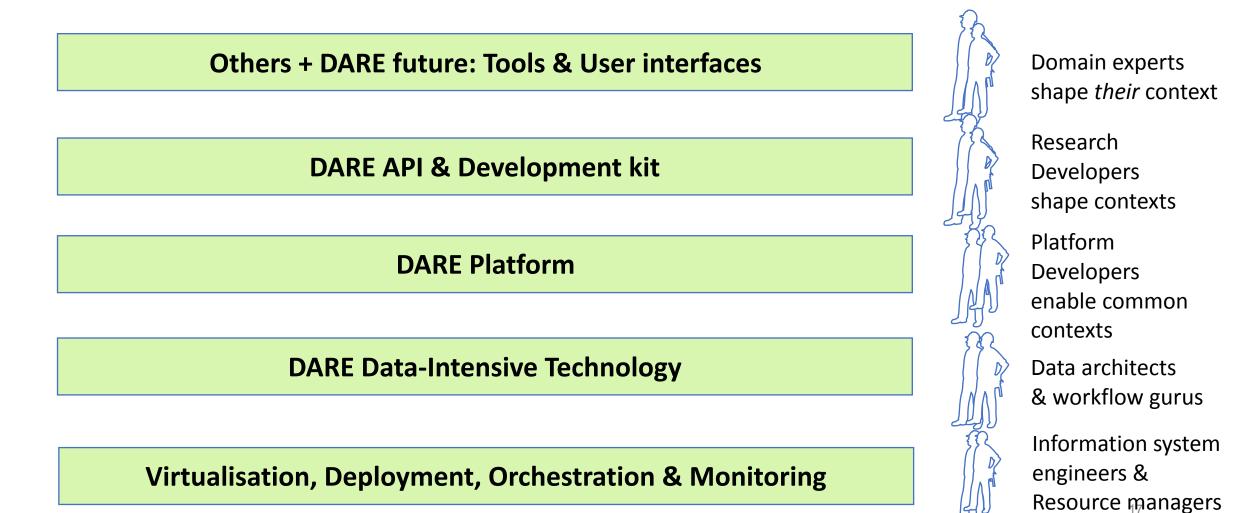




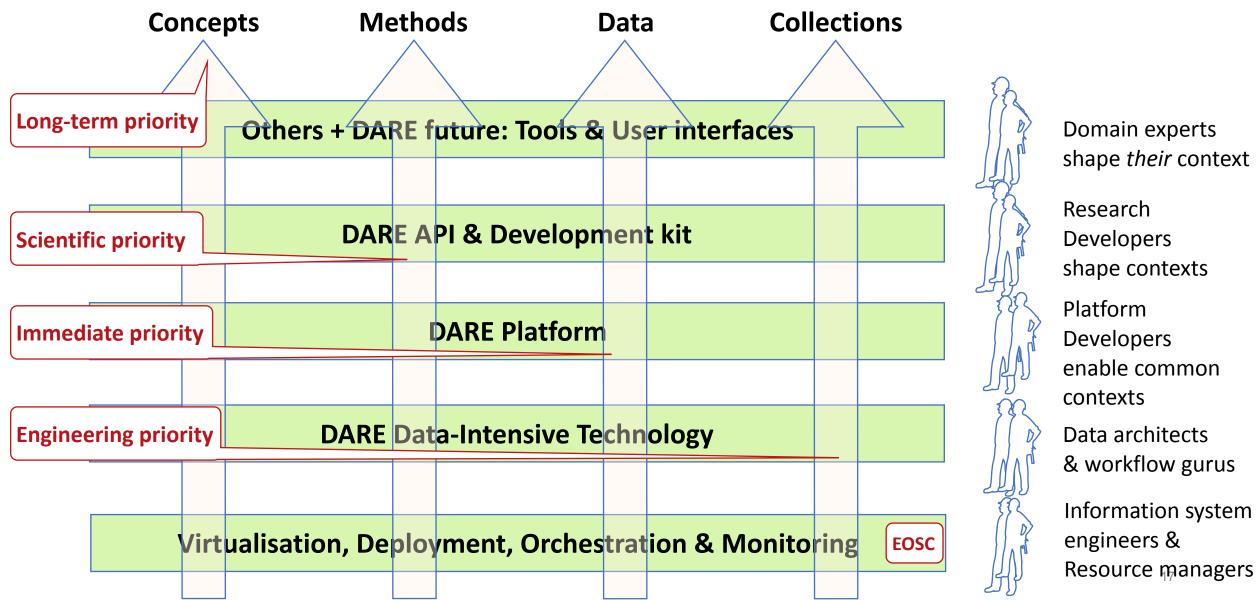
## Outline

- Vision and Goals
- Key insights
- Architecture
- Achievements and Work in Progress

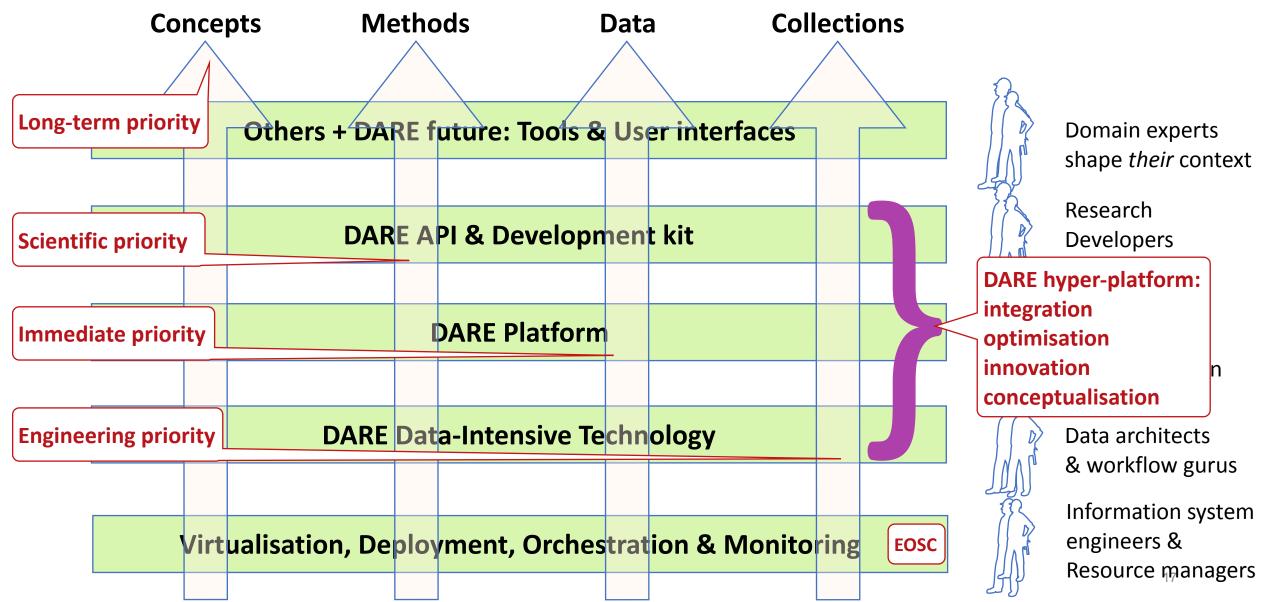
# DARE users create, use, refine and discard



#### DARE users create, use, refine and discard

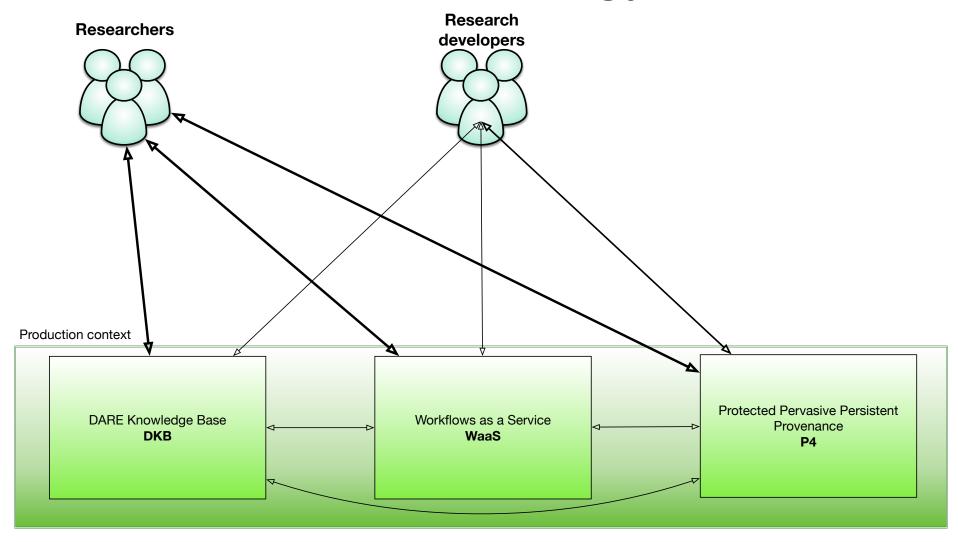


#### DARE users create, use, refine and discard



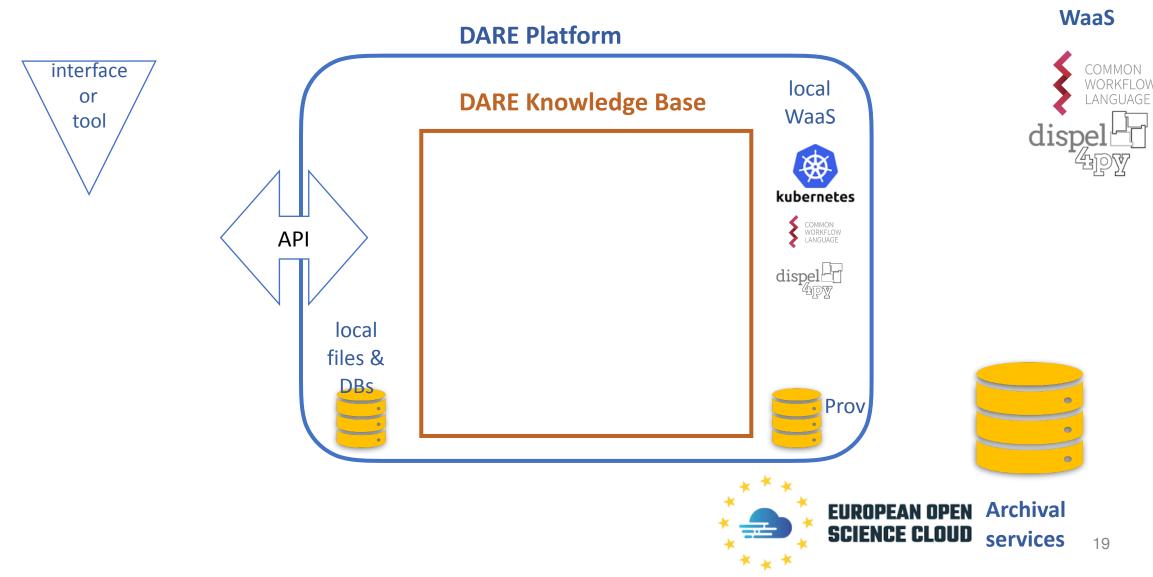


### The three DARE technology pillars



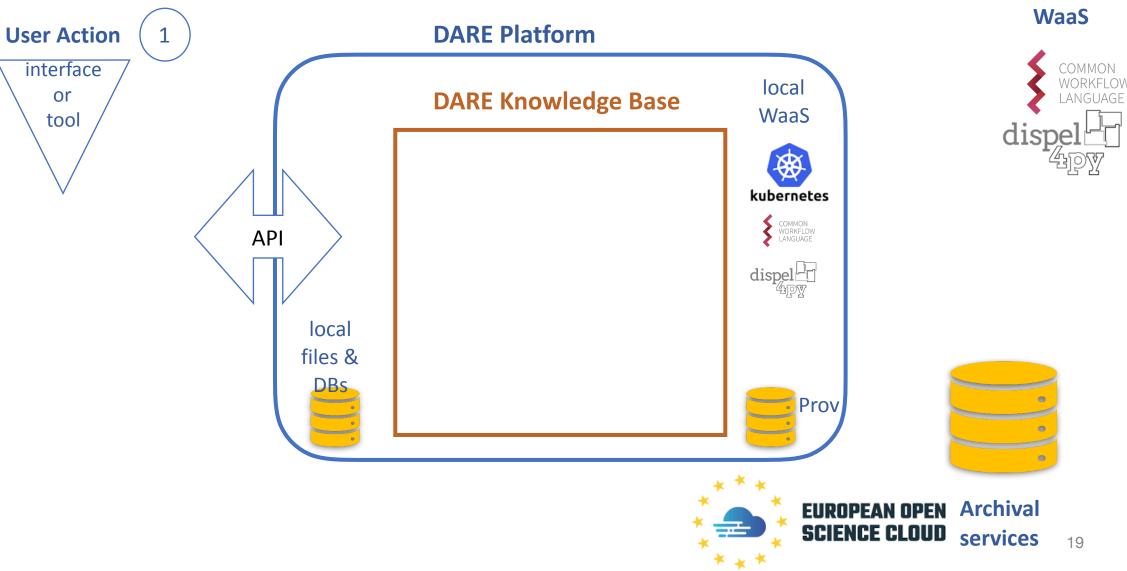


**External services** 



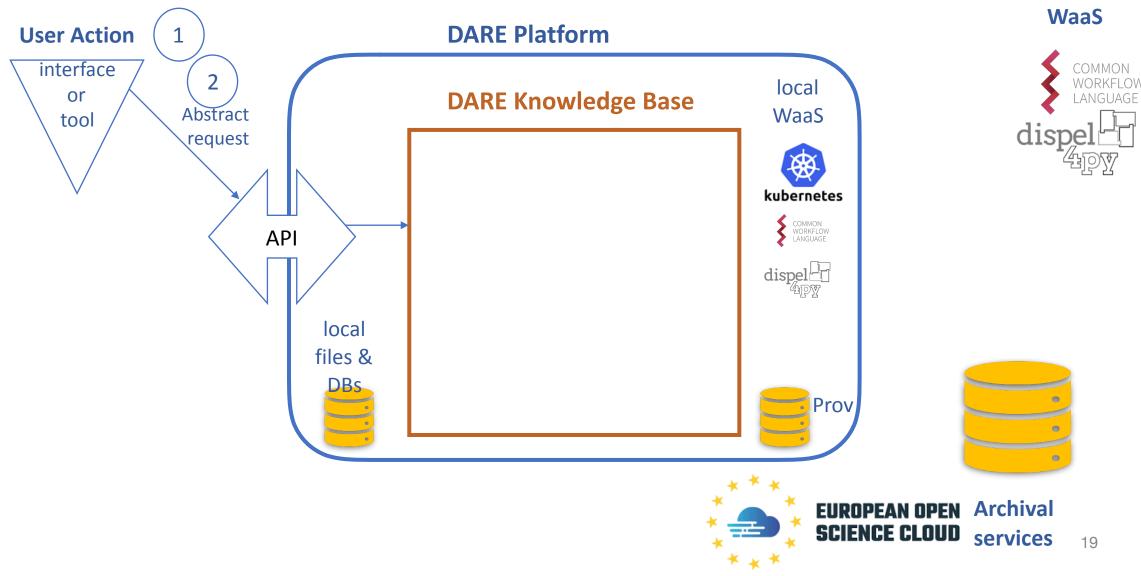


**External services** 





**External services** 



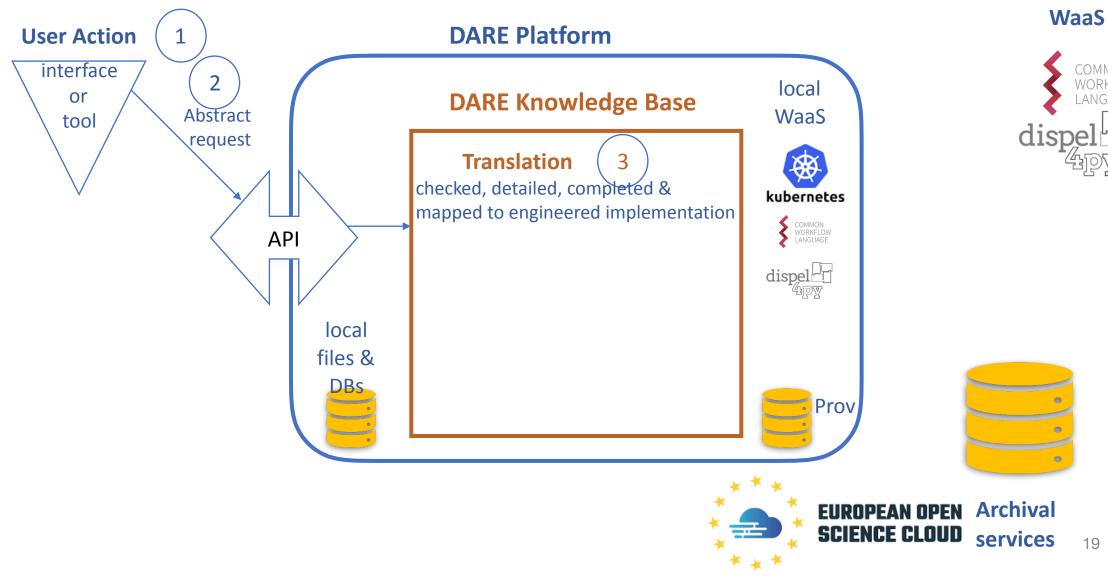


COMMON

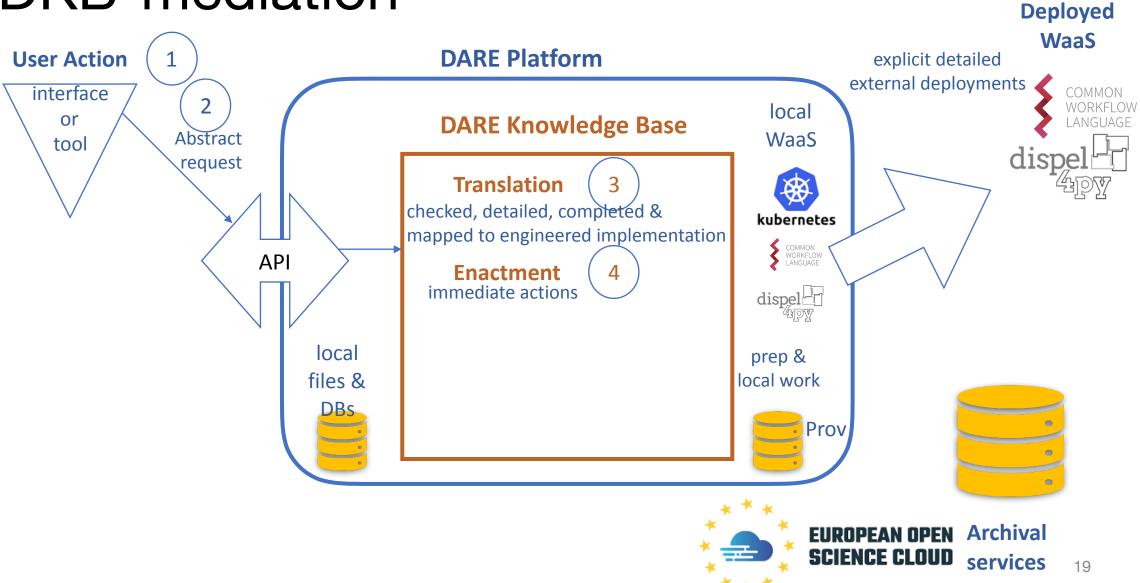
19

WORKFLOW

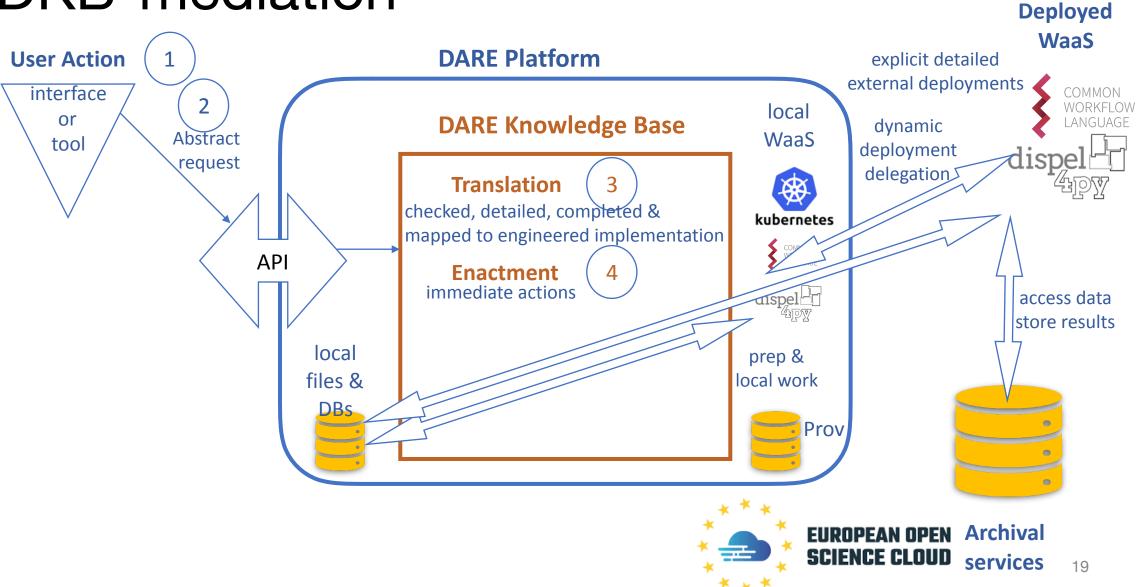
**External services** 



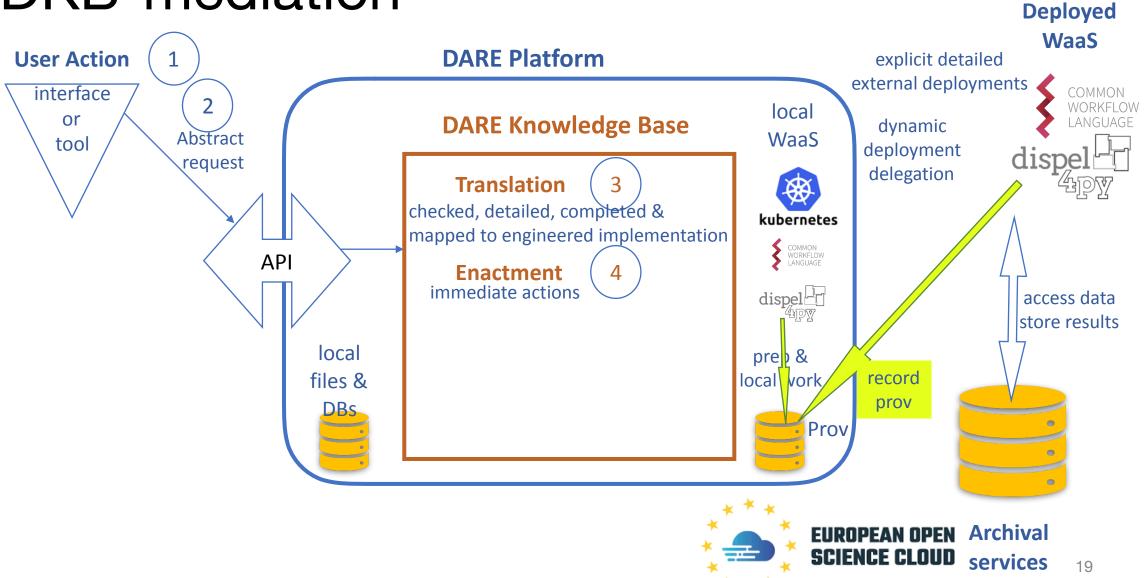




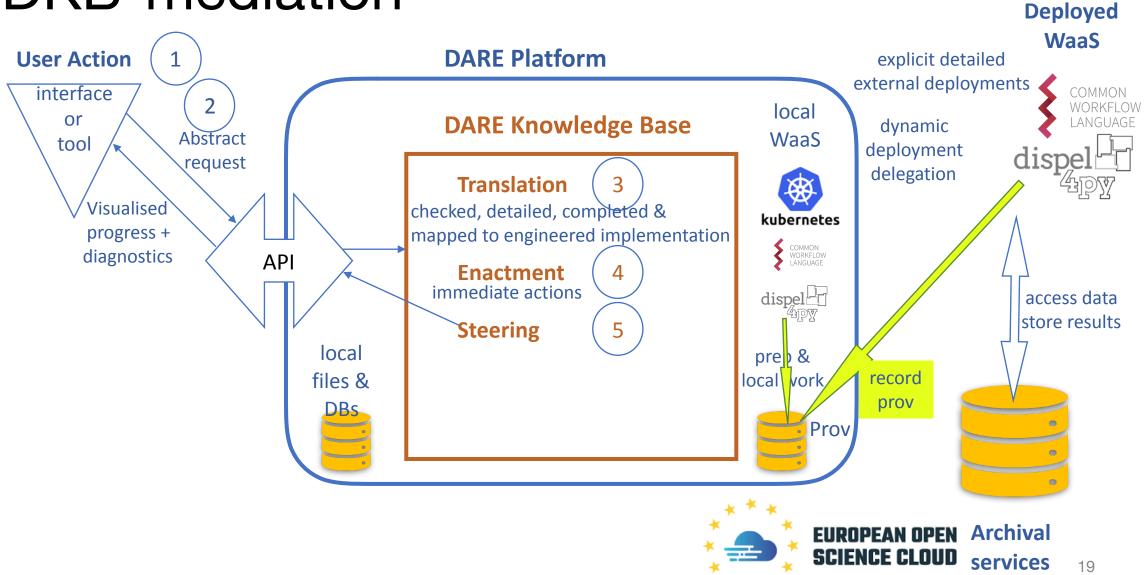














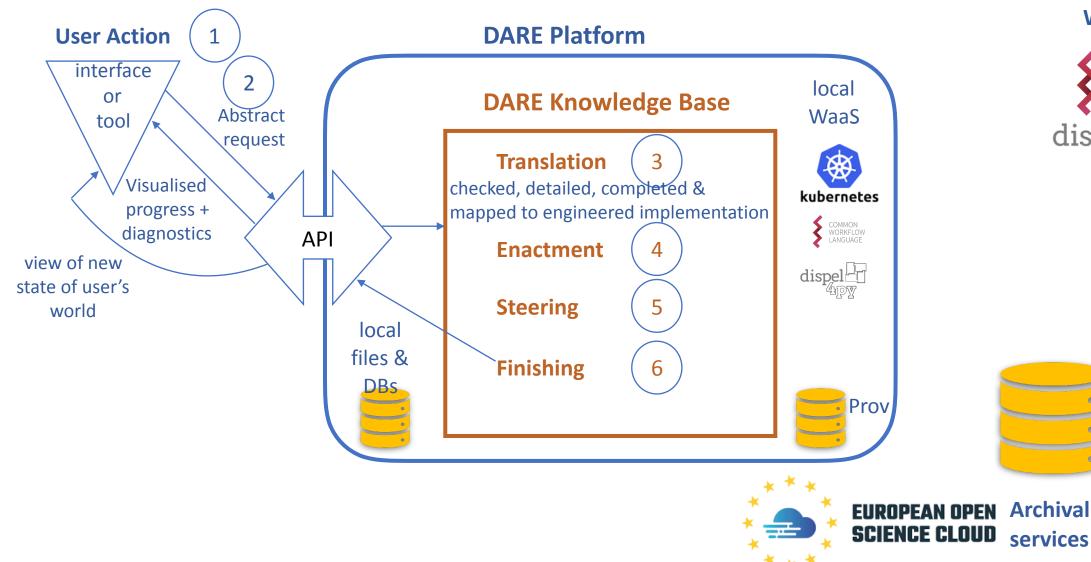
WaaS

COMMON

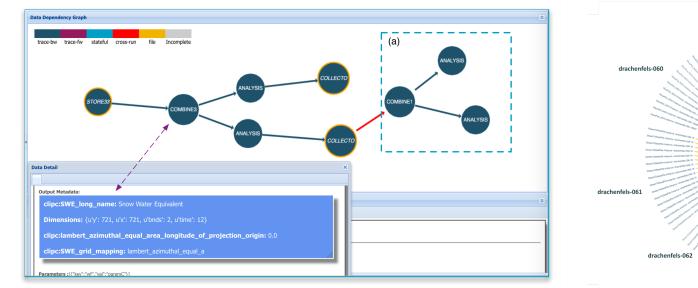
19

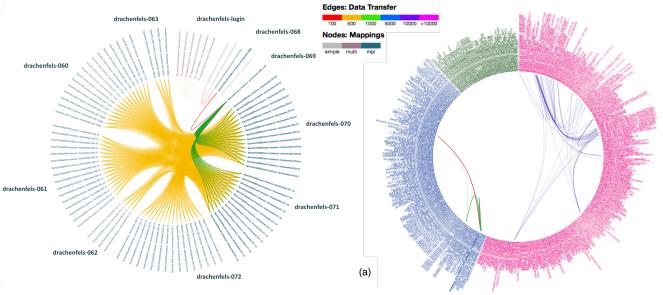
WORKFLOW

**External services** 



#### **Provenance Tools**





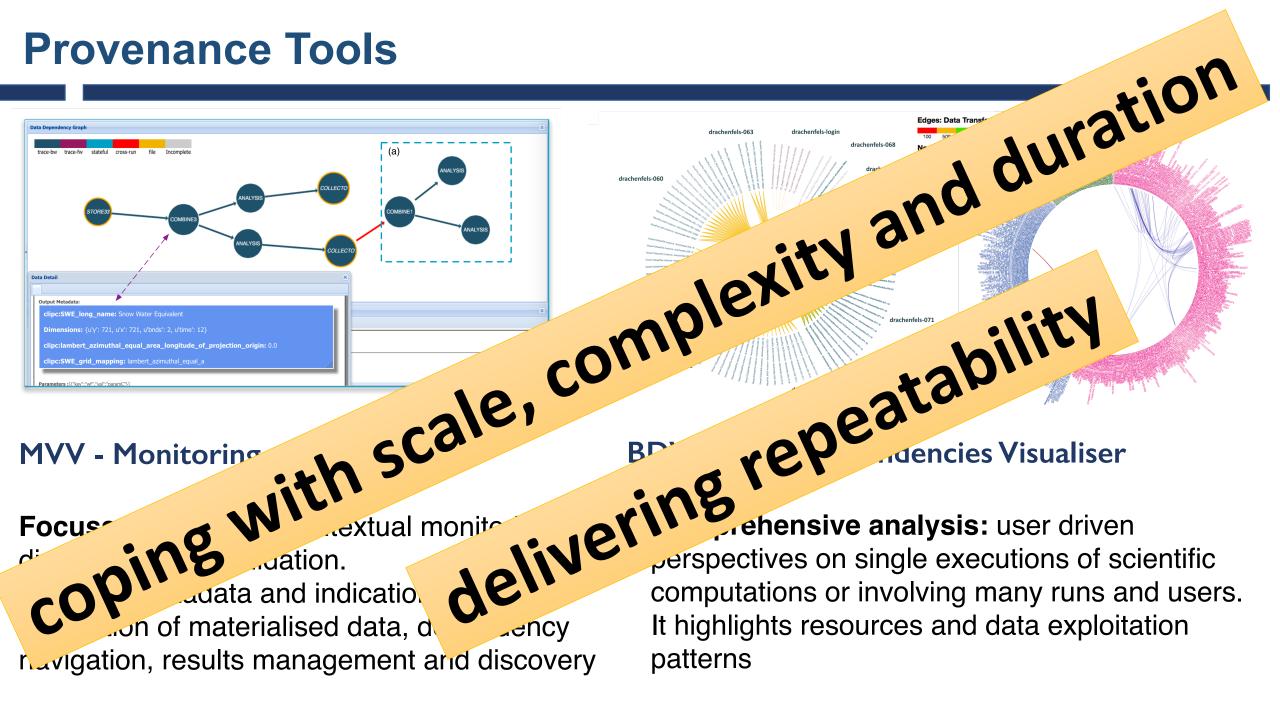
#### **MVV - Monitoring and Validation**

**Focussed analysis:** contextual monitoring, diagnostics and validation. Provides metadata and indication on the production of materialised data, dependency navigation, results management and discovery

#### **BDV** — Bulk Dependencies Visualiser

**Comprehensive analysis:** user driven perspectives on single executions of scientific computations or involving many runs and users. It highlights resources and data exploitation patterns

#### **Provenance Tools**





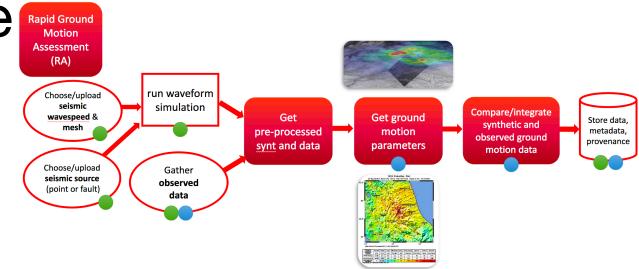
## Outline

- Vision and Goals
- Key insights
- Architecture
- Achievements and Work in Progress

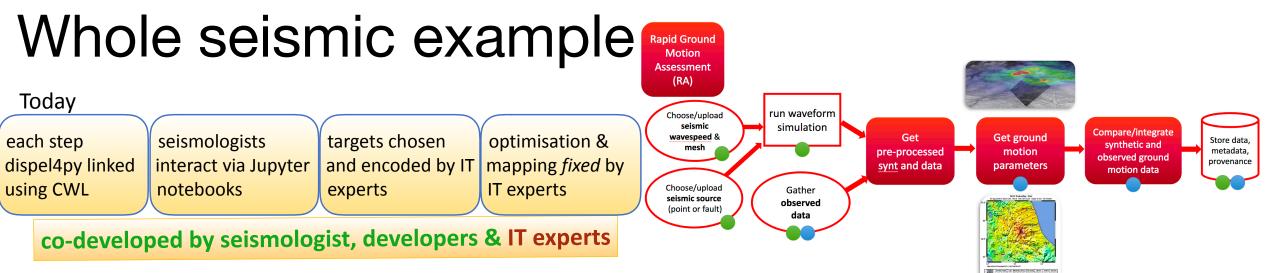


### Whole seismic example

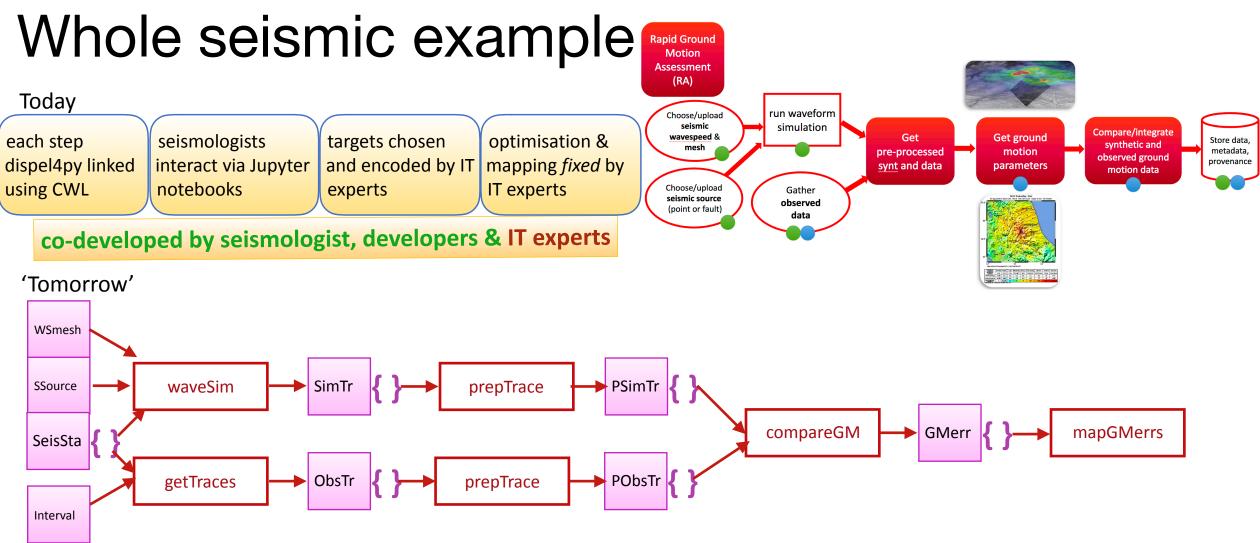
Today



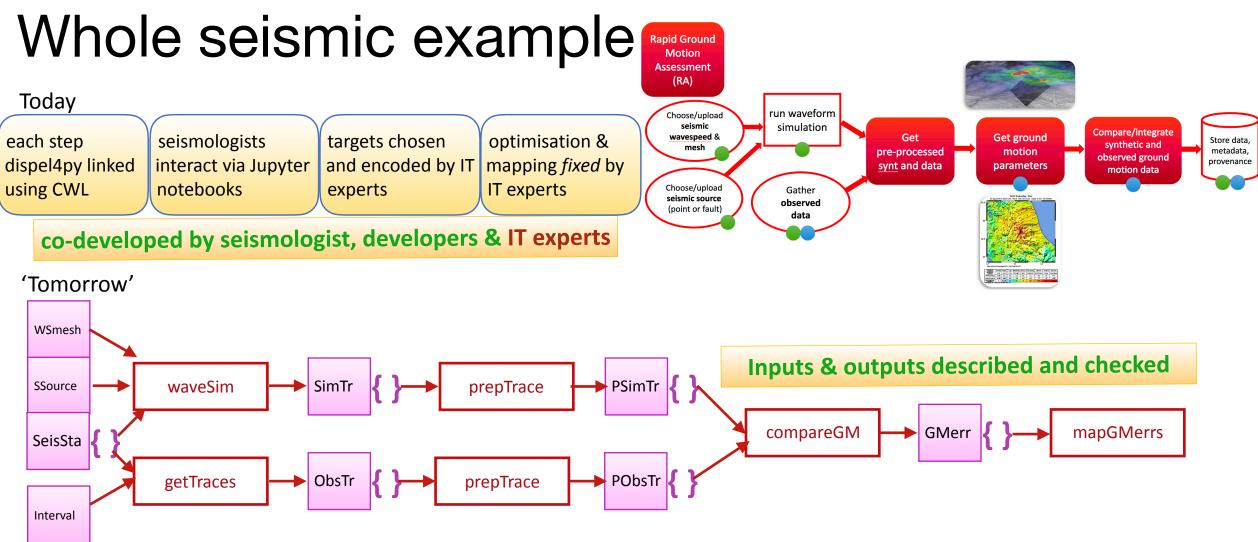




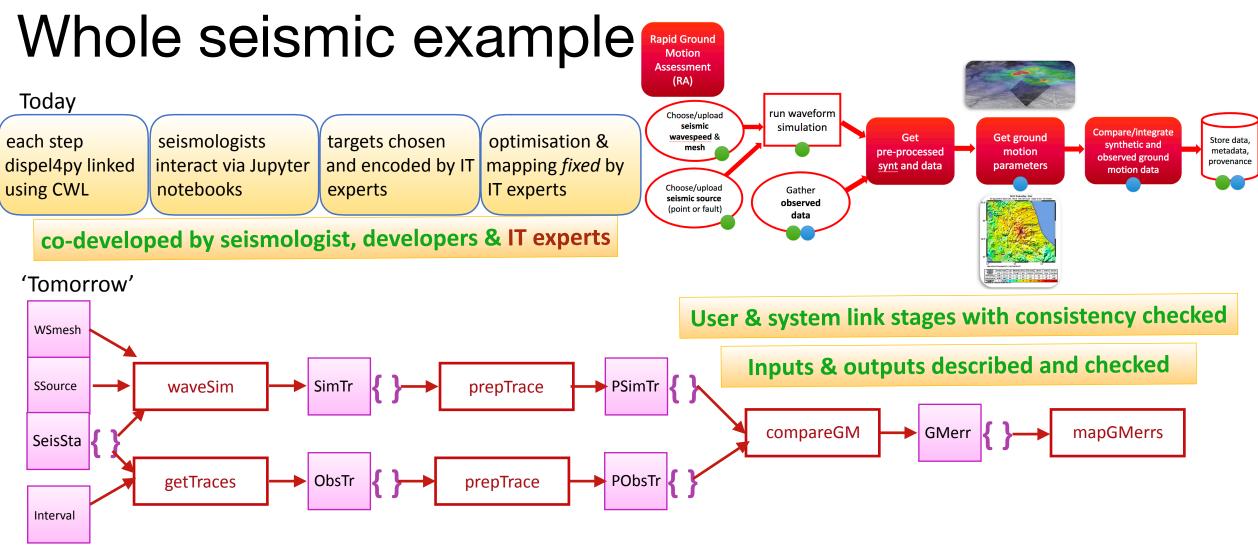




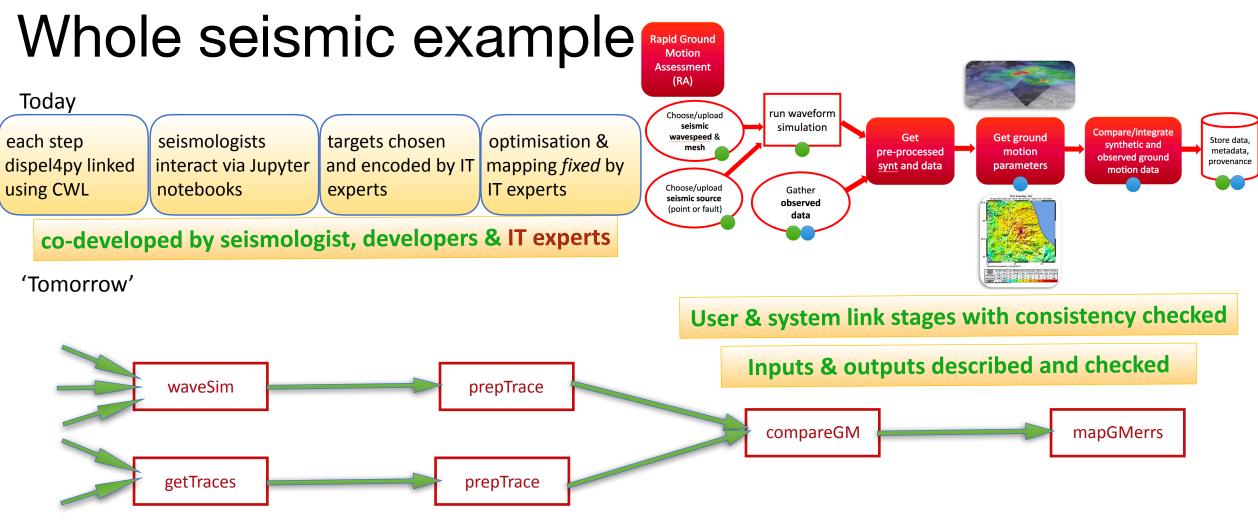




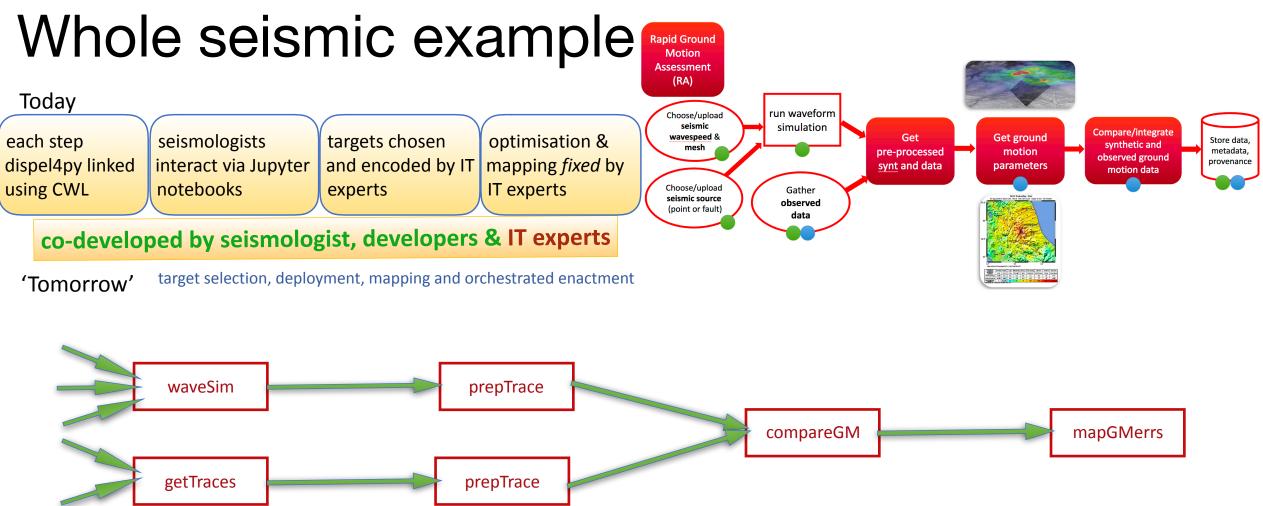




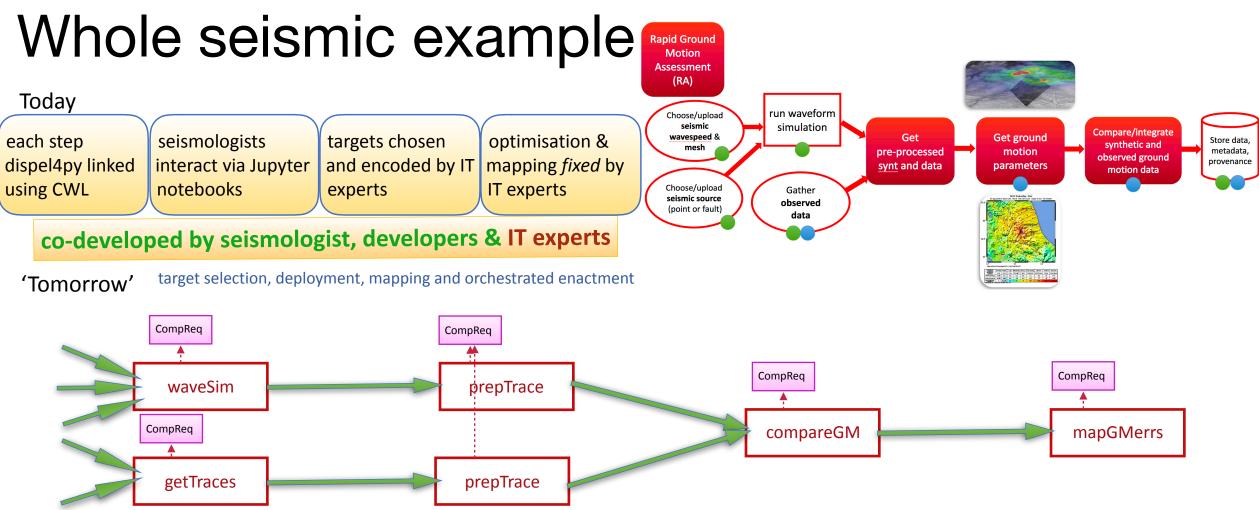




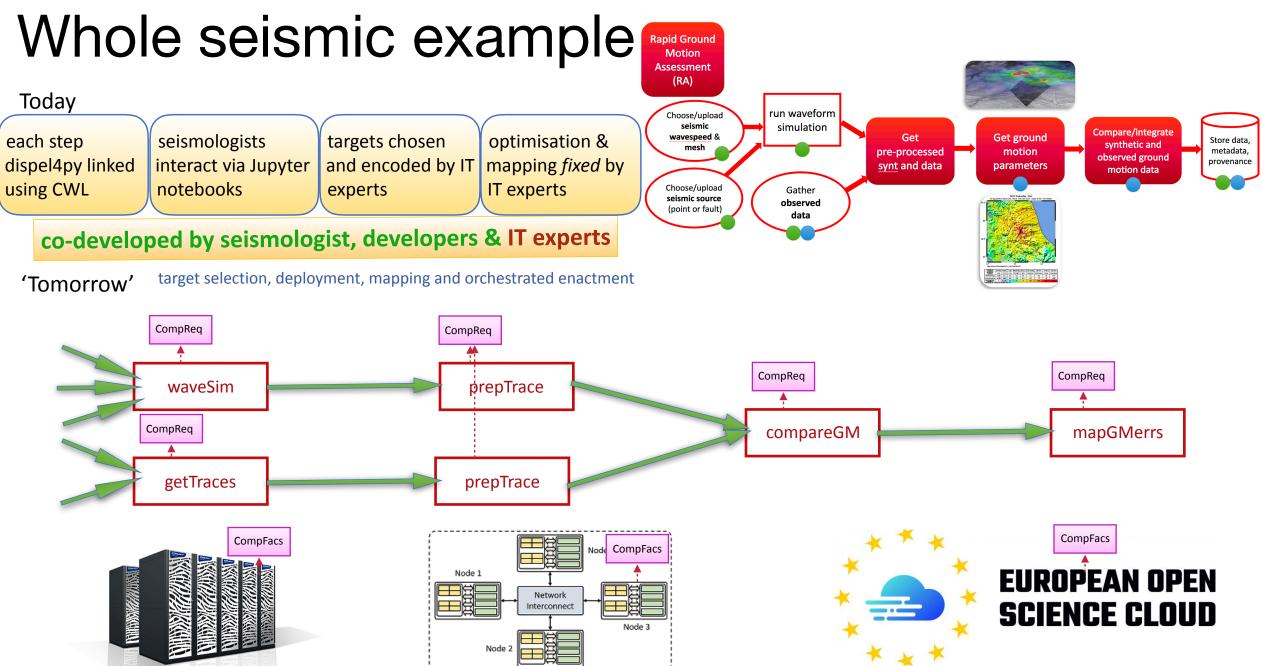




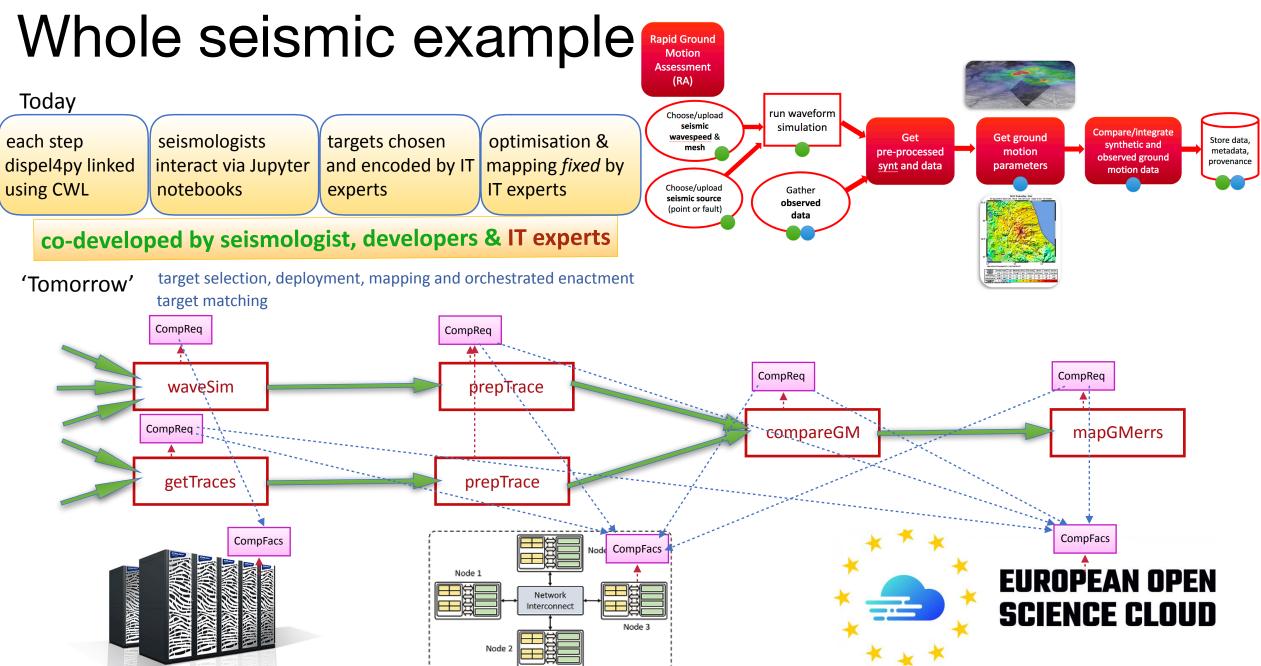




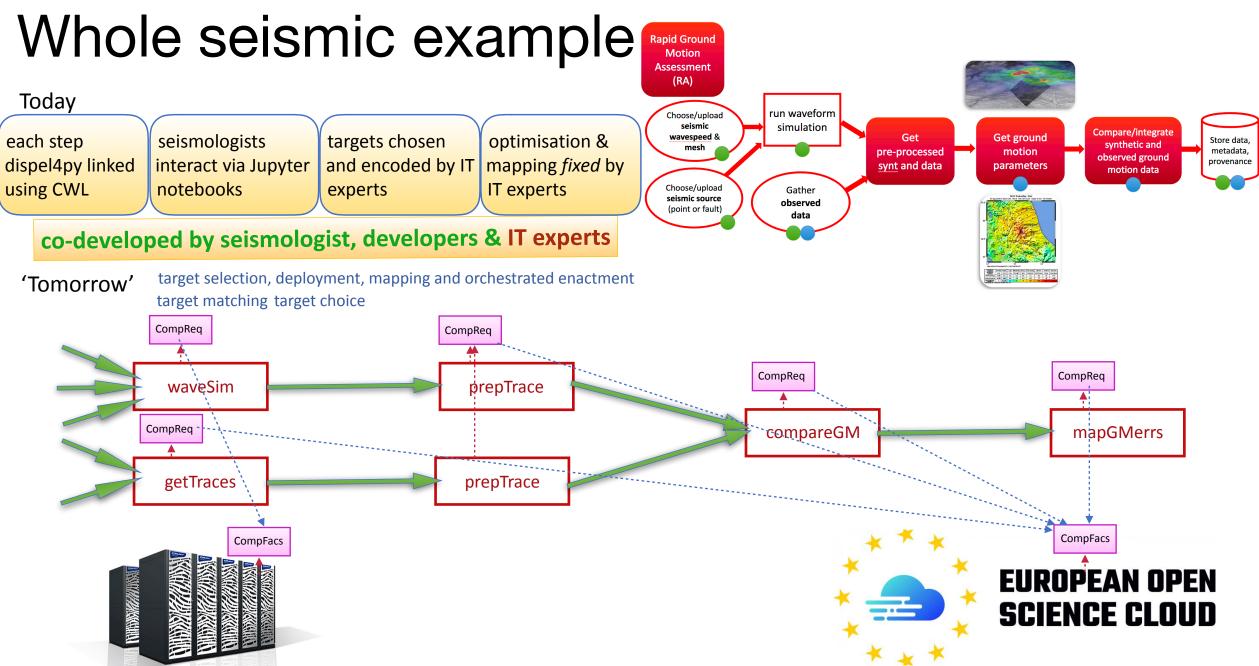




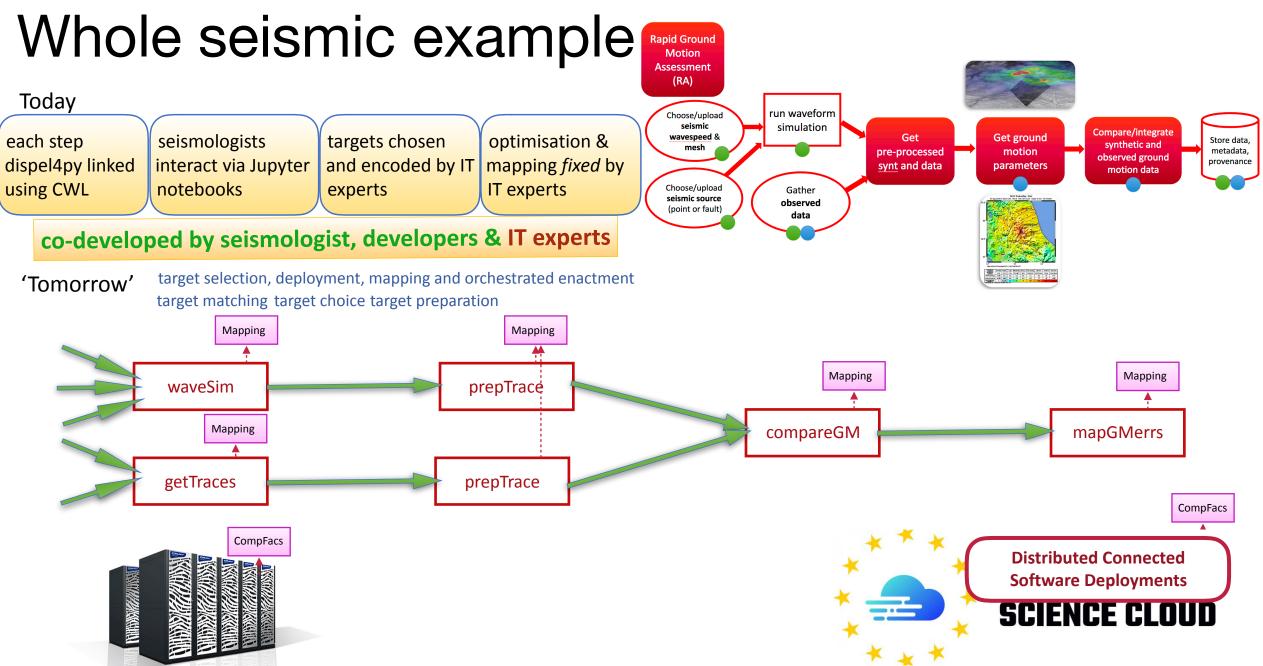




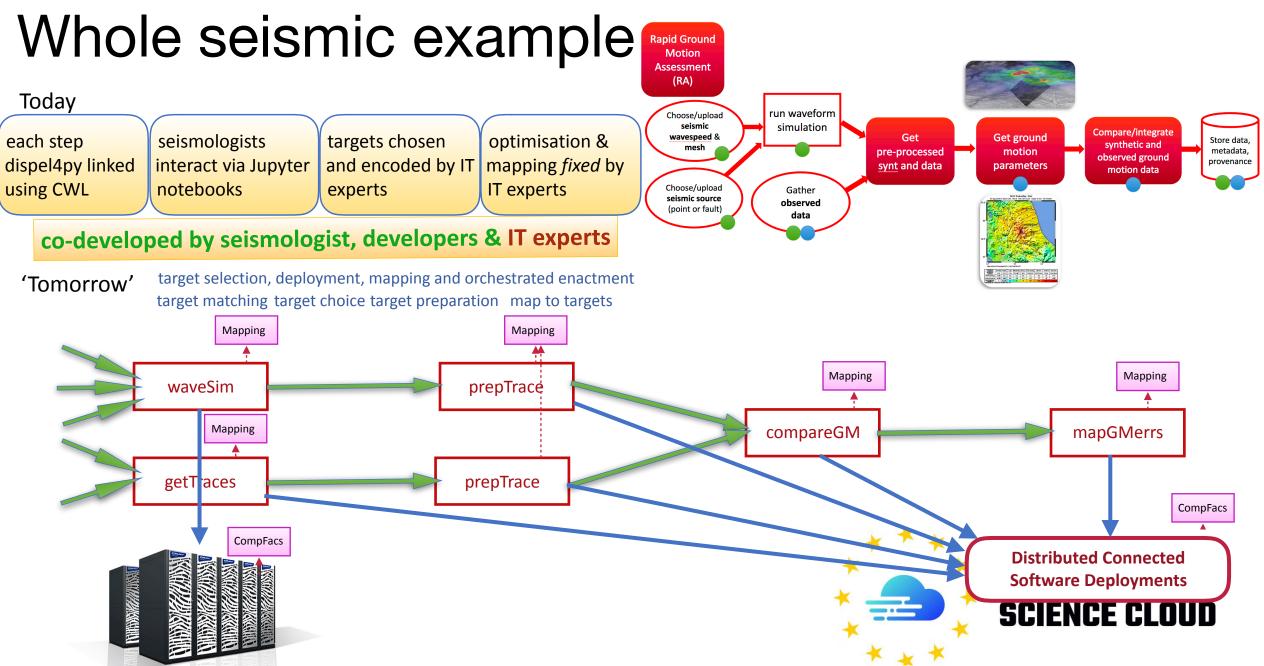






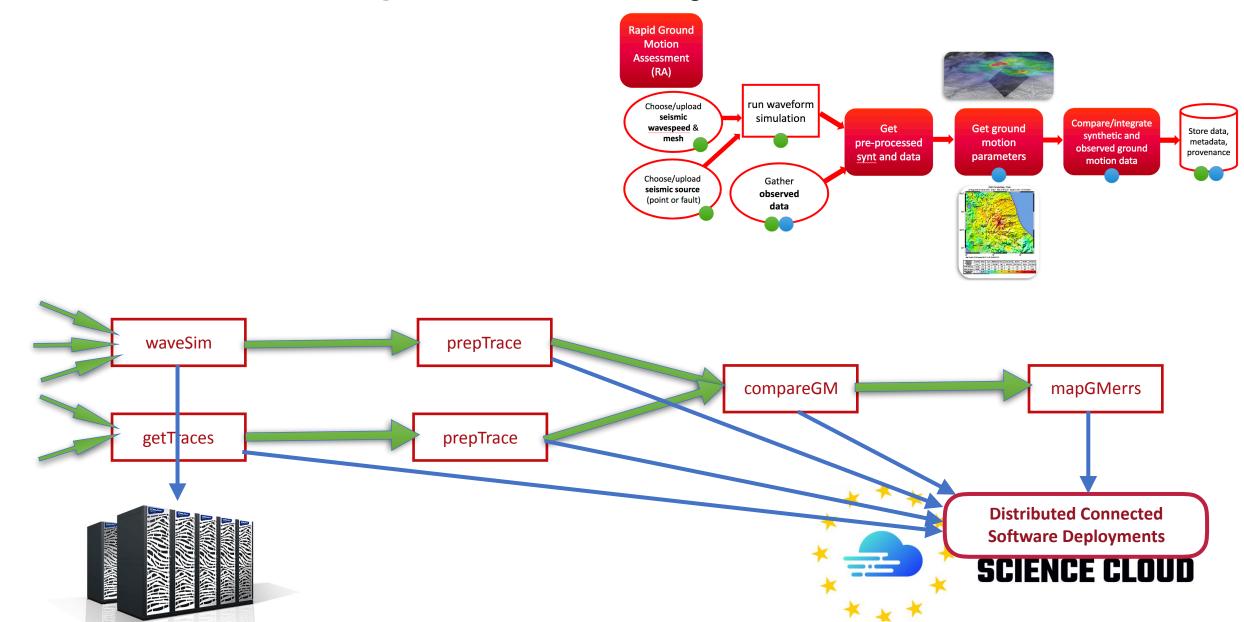






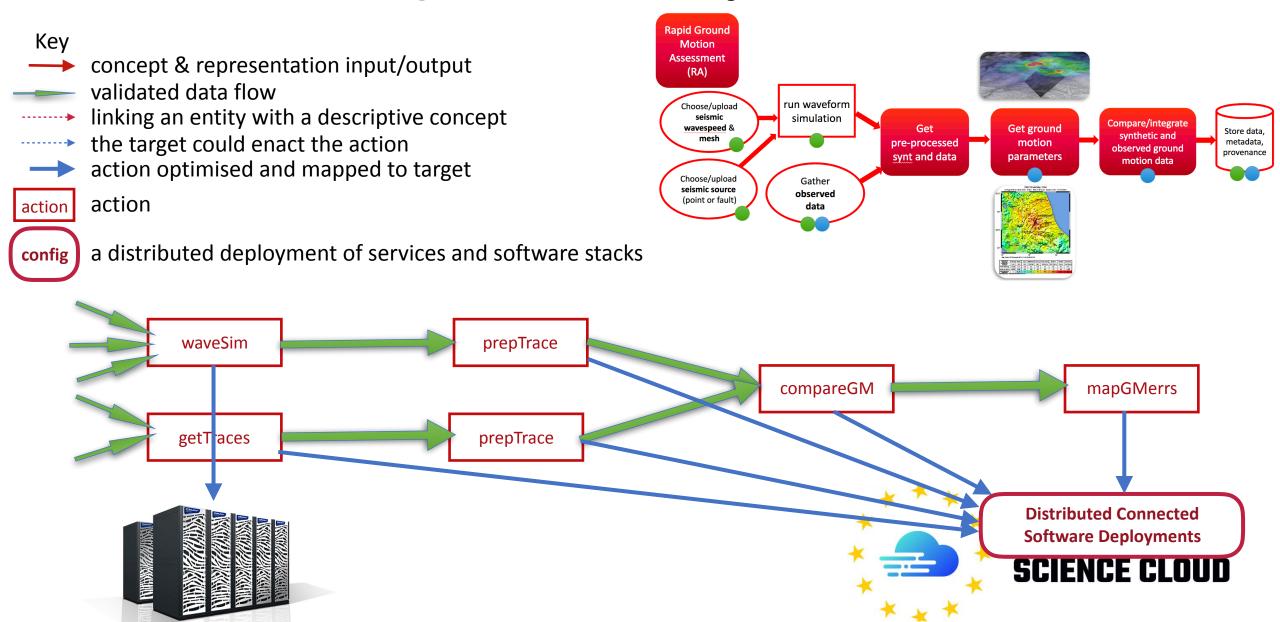


# Seismic example summary



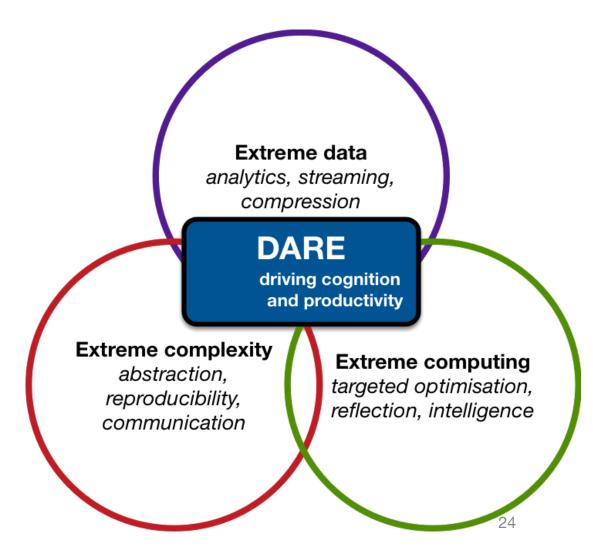


# Seismic example summary





### Summary to take home

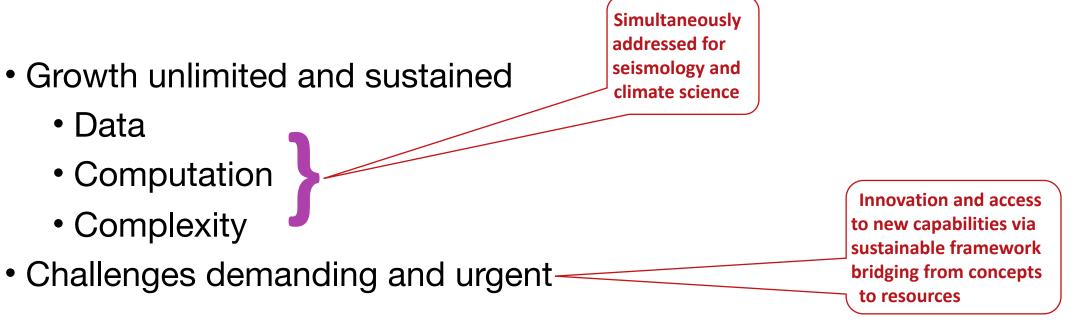




#### Summary to take home Simultaneously addressed for seismology and Growth unlimited and sustained climate science • Data Computation Extreme data analytics, streaming, Complexity compression DARE driving cognition and productivity **Extreme complexity Extreme computing** abstraction, targeted optimisation, reproducibility, reflection, intelligence communication



# Summary to take home





# Summary to take home

- Growth unlimited and sustained
  - Data
  - Computation
  - Complexity
- Challenges demanding and urgent-
- Conserving scarce resources
  - Human intellectual effort
  - Willingness to collaborate
  - Energy / GHG emissions

Simultaneously addressed for seismology and climate science

> Innovation and access to new capabilities via sustainable framework bridging from concepts to resources

Delivered via the CMDC quartet, work contexts, provenancedriven tools and optimisation



**DARE advances:** 

#### Summary to take home new API for developers dispel4py + CWL workflows Simultaneously SPECFEM3D on demand addressed for **Jupyter notebook sessions** seismology and Growth unlimited and sustained **Pervasive provenance + tools** climate science imminent: Data conceptualisation Computation Innovation and access Complexity to new capabilities via sustainable framework Challenges demanding and urgent bridging from concepts to resources Conserving scarce resources **Delivered via the** Human intellectual effort **CMDC** quartet, work contexts, provenancedriven tools and • Willingness to collaborate optimisation • Energy / GHG emissions



# Thank you



Luca Trani, Malcolm Atkinson, Daniele Bailo, Rossana Paciello and Rosa Filgueira, *Establishing Core Concepts for Information-Powered Collaborations*, FGCS vol. 89, 421-437, 2018. Malcolm Atkinson, Rosa Filgueira, Iraklis Klampanos, Antonis Koulourikos, Amrey Krause, Federica Magnoni, Christian Pagé, Andreas Rietbrock and Alessandro Spinuso *Comprehensible control for researchers and developers facing data challenges*, to appear in proceedings of IEEE eScience, 2019; <u>https://drive.google.com/open?id=1JT6RdaFV0pTQ8aWx-8LY8etXlvd5mRag</u>.

project-dare.eu



# Thank you constinues to the second se

L. Schubert, K. G. Jeffery, New Software Engineering requirements in Clouds and large-scale systems, IEEE Cloud Computing 2 (1) 48–58, 2015.

Luca Trani, Malcolm Atkinson, Daniele Bailo, Rossana Paciello and Rosa Filgueira, *Establishing Core Concepts for Information-Powered Collaborations*, FGCS vol. 89, 421-437, 2018. Malcolm Atkinson, Rosa Filgueira, Iraklis Klampanos, Antonis Koulourikos, Amrey Krause, Federica Magnoni, Christian Pagé, Andreas Rietbrock and Alessandro Spinuso *Comprehensible control for researchers and developers facing data challenges*, to appear in proceedings of IEEE eScience, 2019; <u>https://drive.google.com/open?id=1JT6RdaFV0pTQ8aWx-8LY8etXlvd5mRag</u>.

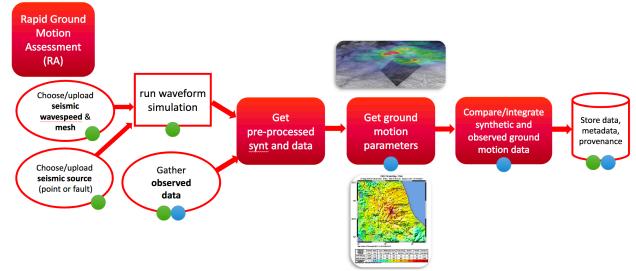
project-dare.eu



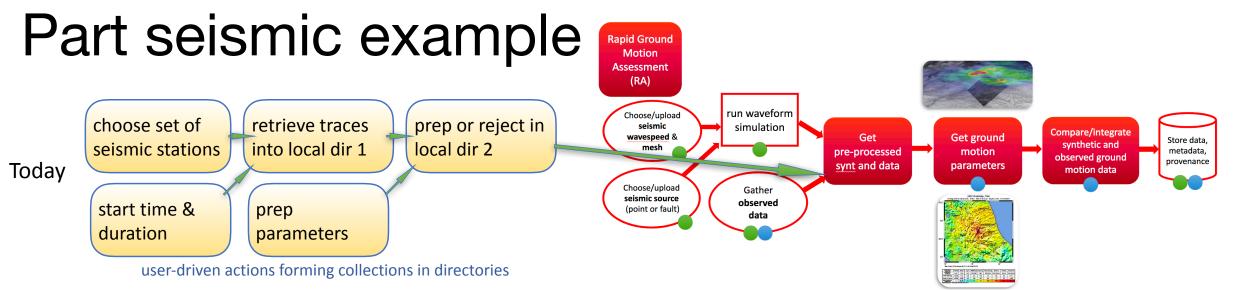




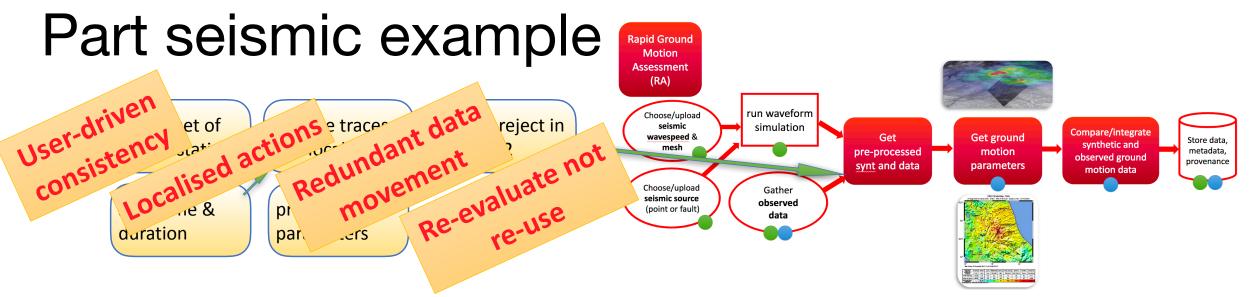
#### Part seismic example



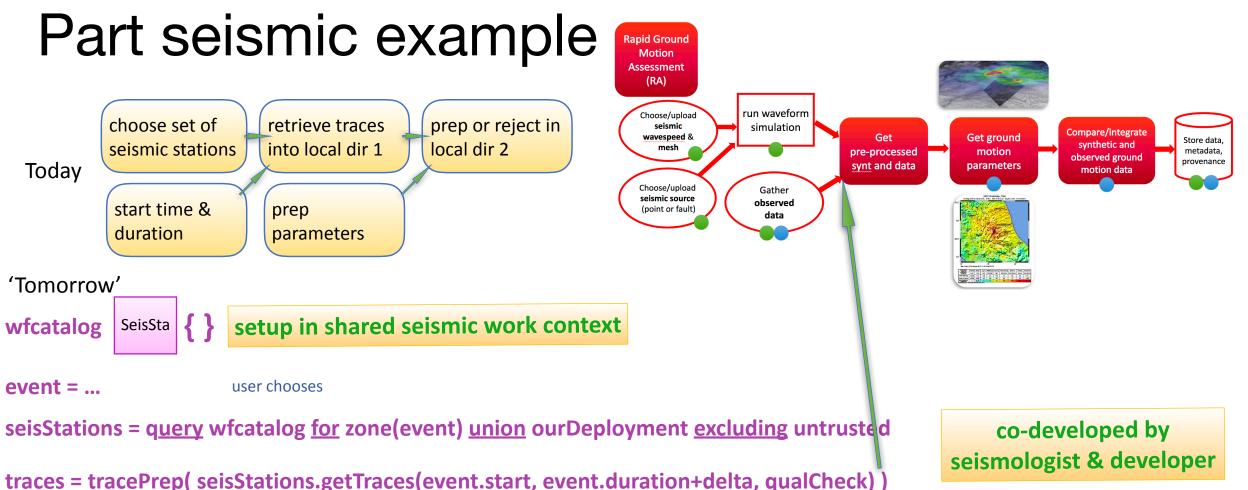




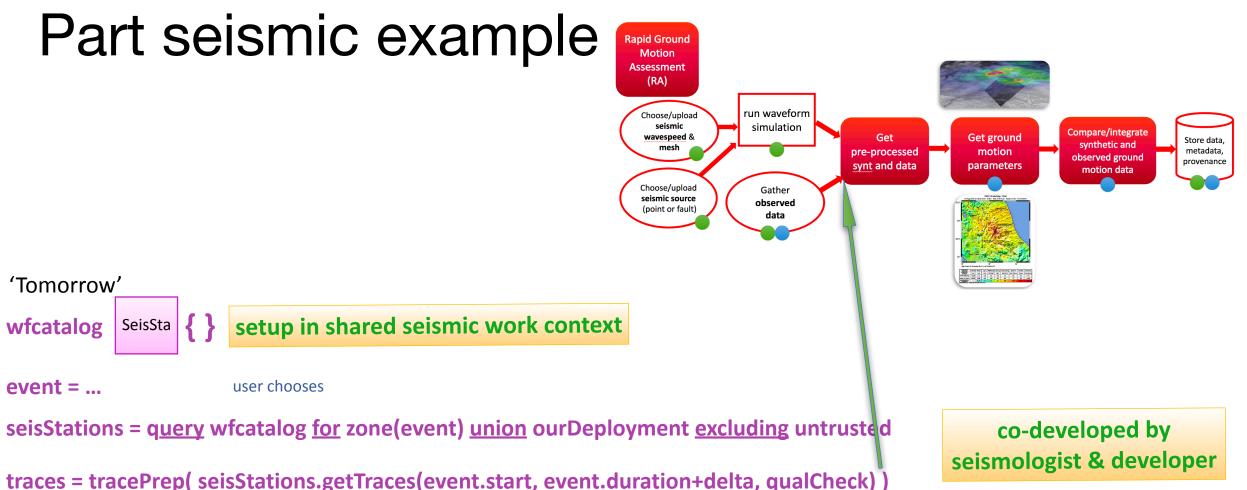




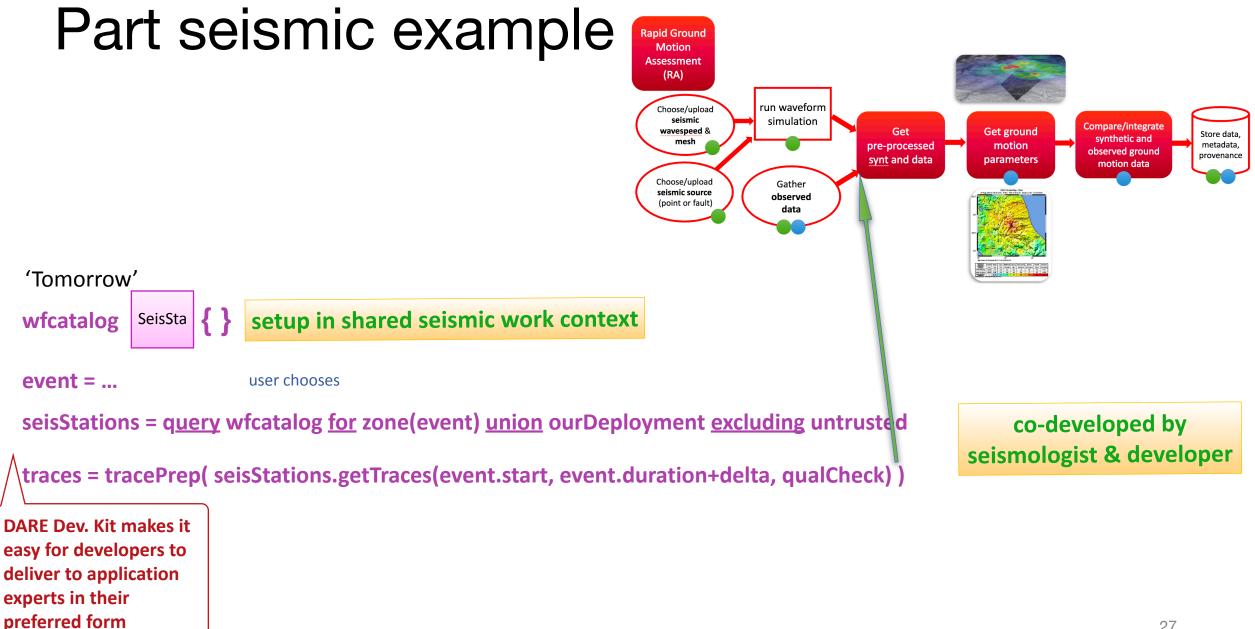




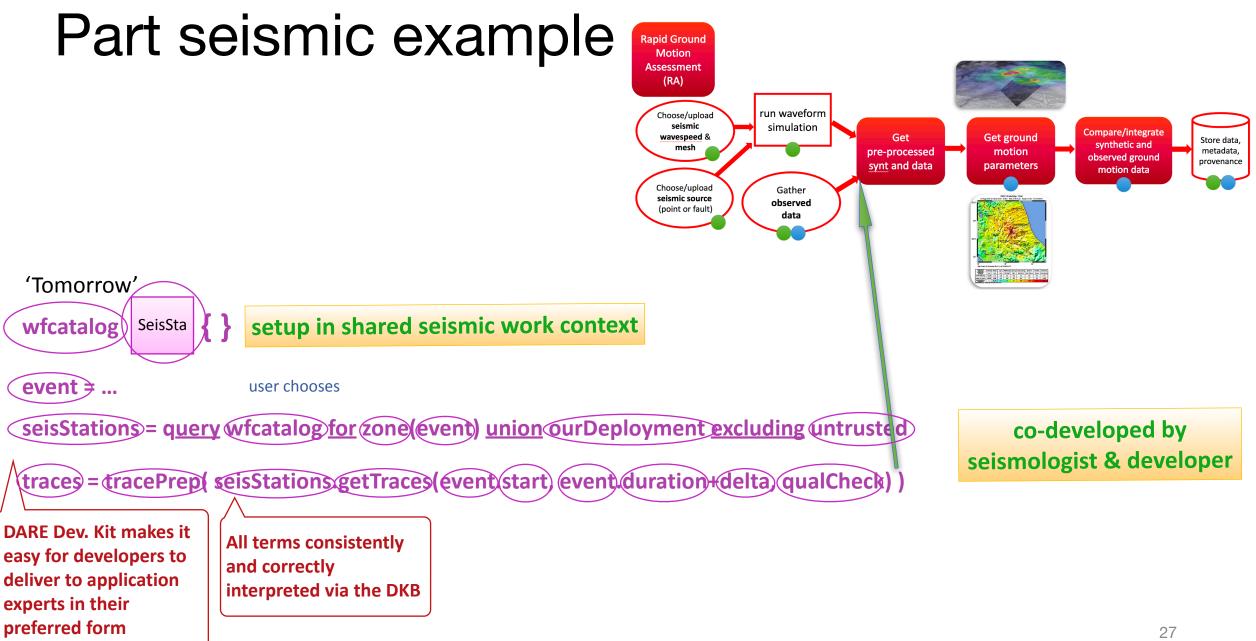




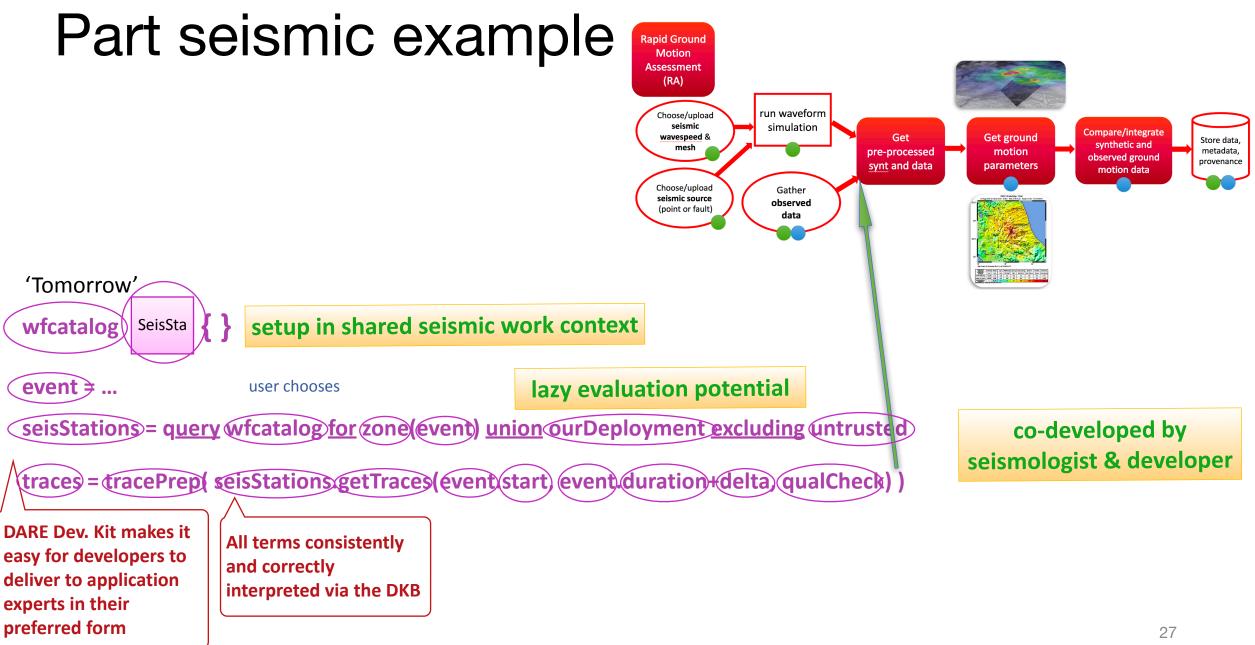




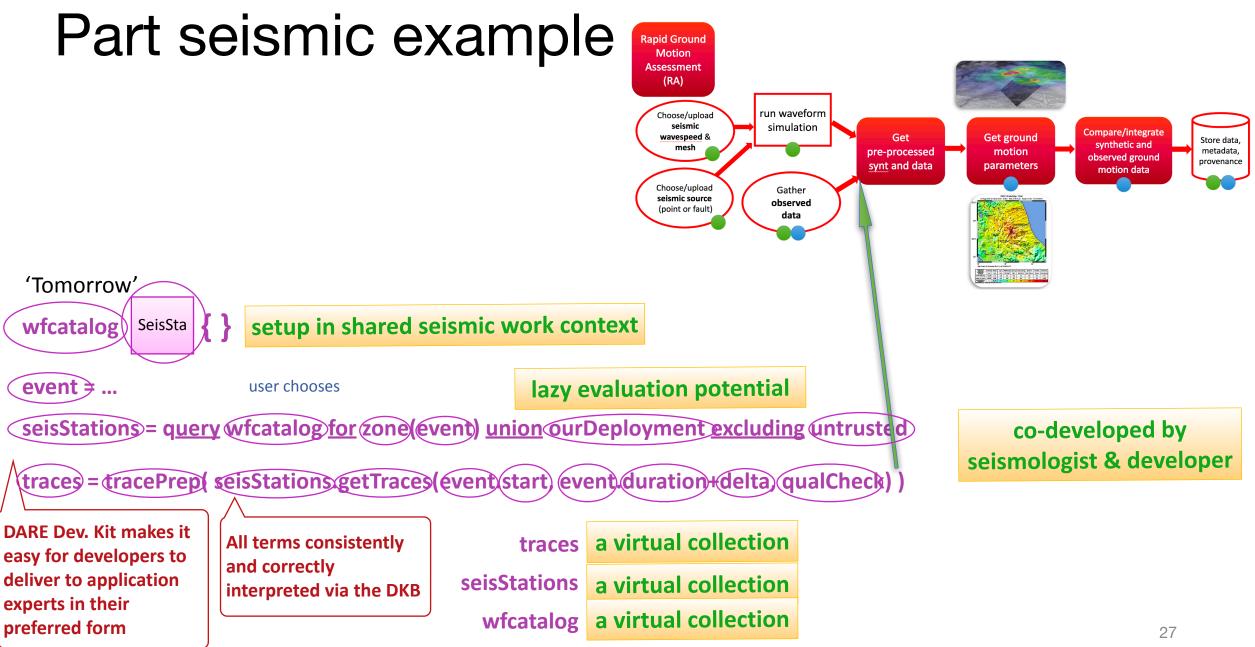




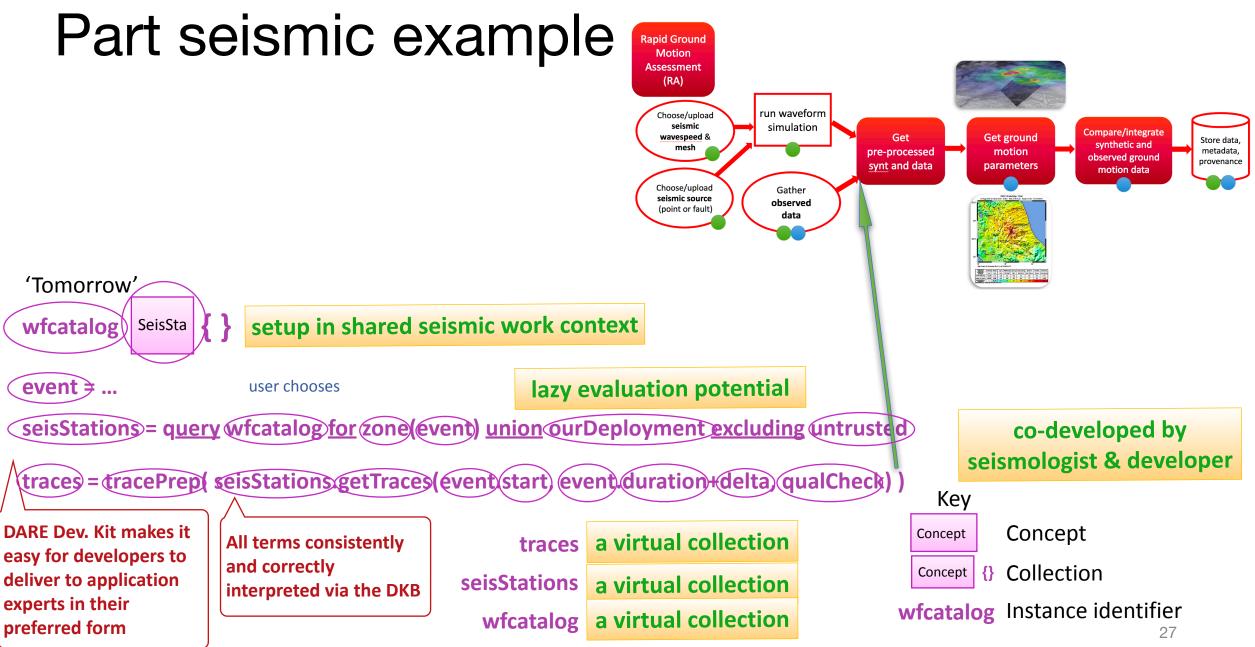
















- Concepts researchers require
  - tailored views
  - controlled incremental development
  - controlled sharing
  - detail only when wanted
  - trustworthy persistent interpretation



- Concepts researchers require
  - tailored views
  - controlled incremental development
  - controlled sharing
  - detail only when wanted
  - trustworthy persistent interpretation
- e-Infrastructure Concepts engineers require
  - bridge researchers' worlds to engineering
  - common sustainable framework
  - to tailor and automatically populate views
  - controlled incremental development
  - managed releases
  - visual steered grouped summaries
  - affordable optimised interpretation



- Concepts researchers require
  - tailored views
  - controlled incremental development
  - controlled sharing
  - detail only when wanted
  - trustworthy persistent interpretation
- e-Infrastructure Concepts engineers require
  - bridge researchers' worlds to engineering
  - common sustainable framework
  - to tailor and automatically populate views
  - controlled incremental development
  - managed releases
  - visual steered grouped summaries
  - affordable optimised interpretation

#### DARE Knowledge Base





- Concepts researchers require
  - tailored views
  - controlled incremental development
  - controlled sharing
  - detail only when wanted
  - trustworthy persistent interpretation
- e-Infrastructure Concepts engineers require
  - bridge researchers' worlds to engineering
  - common sustainable framework
  - to tailor and automatically populate views
  - controlled incremental development
  - managed releases
  - visual steered grouped summaries
  - affordable optimised interpretation

# **DARE Knowledge Base Time-stamped sequence of entries**



- Concepts researchers require
  - tailored views
  - controlled incremental development
  - controlled sharing
  - detail only when wanted
  - trustworthy persistent interpretation
- e-Infrastructure Concepts engineers require
  - bridge researchers' worlds to engineering
  - common sustainable framework
  - to tailor and automatically populate views
  - controlled incremental development
  - managed releases
  - visual steered grouped summaries
  - affordable optimised interpretation

#### DARE Knowledge Base

Time-stamped sequence of entries

Local tokens for external entities



- Concepts researchers require
  - tailored views
  - controlled incremental development
  - controlled sharing
  - detail only when wanted
  - trustworthy persistent interpretation
- e-Infrastructure Concepts engineers require
  - bridge researchers' worlds to engineering
  - common sustainable framework
  - to tailor and automatically populate views
  - controlled incremental development
  - managed releases
  - visual steered grouped summaries
  - affordable optimised interpretation

#### DARE Knowledge Base

Time-stamped sequence of entries

Local tokens for external entities

Local representations for internal entities



- Concepts researchers require
  - tailored views
  - controlled incremental development
  - controlled sharing
  - detail only when wanted
  - trustworthy persistent interpretation
- e-Infrastructure Concepts engineers require
  - bridge researchers' worlds to engineering
  - common sustainable framework
  - to tailor and automatically populate views
  - controlled incremental development
  - managed releases
  - visual steered grouped summaries
  - affordable optimised interpretation

#### DARE Knowledge Base

Time-stamped sequence of entries

Local tokens for external entities

Local representations for internal entities

**Concepts, Methods, Data and Collections** 



- Concepts researchers require
  - tailored views
  - controlled incremental development
  - controlled sharing
  - detail only when wanted
  - trustworthy persistent interpretation
- e-Infrastructure Concepts engineers require
  - bridge researchers' worlds to engineering
  - common sustainable framework
  - to tailor and automatically populate views
  - controlled incremental development
  - managed releases
  - visual steered grouped summaries
  - affordable optimised interpretation

#### DARE Knowledge Base

Time-stamped sequence of entries

Local tokens for external entities

Local representations for internal entities

**Concepts, Methods, Data and Collections** 

Dynamic populations: create, update, use, discard



- Concepts researchers require
  - tailored views
  - controlled incremental development
  - controlled sharing
  - detail only when wanted
  - trustworthy persistent interpretation
- e-Infrastructure Concepts engineers require
  - bridge researchers' worlds to engineering
  - common sustainable framework
  - to tailor and automatically populate views
  - controlled incremental development
  - managed releases
  - visual steered grouped summaries
  - affordable optimised interpretation

#### DARE Knowledge Base

Time-stamped sequence of entries

Local tokens for external entities

Local representations for internal entities

**Concepts, Methods, Data and Collections** 

Dynamic populations: create, update, use, discard

Virtual: local control external action + DKB delta



- Concepts researchers require
  - tailored views
  - controlled incremental development
  - controlled sharing
  - detail only when wanted
  - trustworthy persistent interpretation
- e-Infrastructure Concepts engineers require
  - bridge researchers' worlds to engineering
  - common sustainable framework
  - to tailor and automatically populate views
  - controlled incremental development
  - managed releases
  - visual steered grouped summaries
  - affordable optimised interpretation

#### DARE Knowledge Base

Time-stamped sequence of entries

Local tokens for external entities

Local representations for internal entities

**Concepts, Methods, Data and Collections** 

Dynamic populations: create, update, use, discard

Virtual: local control external action + DKB delta

Tailoring: import (view of) context, import & align bundles + new entities



- Concepts researchers require
  - tailored views
  - controlled incremental development
  - controlled sharing
  - detail only when wanted
  - trustworthy persistent interpretation
- e-Infrastructure Concepts engineers require
  - bridge researchers' worlds to engineering
  - common sustainable framework
  - to tailor and automatically populate views
  - controlled incremental development
  - managed releases
  - visual steered grouped summaries
  - affordable optimised interpretation

#### DARE Knowledge Base

**Time-stamped sequence of entries** 

Local tokens for external entities

Local representations for internal entities

**Concepts, Methods, Data and Collections** 

Dynamic populations: create, update, use, discard

Virtual: local control external action + DKB delta

Tailoring: import (view of) context, import & align bundles + new entities

Innovation: local + promotion <u>or</u> branch & import of releases



- Concepts researchers require
  - tailored views
  - controlled incremental development
  - controlled sharing
  - detail only when wanted
  - trustworthy persistent interpretation
- e-Infrastructure Concepts engineers require
  - bridge researchers' worlds to engineering
  - common sustainable framework
  - to tailor and automatically populate views
  - controlled incremental development
  - managed releases
  - visual steered grouped summaries
  - affordable optimised interpretation

#### DARE Knowledge Base

**Time-stamped sequence of entries** 

Local tokens for external entities

Local representations for internal entities

**Concepts, Methods, Data and Collections** 

Dynamic populations: create, update, use, discard

Virtual: local control external action + DKB delta

Tailoring: import (view of) context, import & align bundles + new entities

Innovation: local + promotion <u>or</u> branch & import of releases

One DKB per DARE platform deployment updated platform core releases