



Why is it hard to help researchers exploit their data?

Malcolm Atkinson

Data-Intensive Research Group

University of Edinburgh



UvA, Amsterdam, 11 June 2015

Outline

- Are researchers different?
- Data-Intensive thinking
- Extra Pressures
- VERCE
- Data-Intensive methods
 - Principles
 - Strategy
 - Implementation

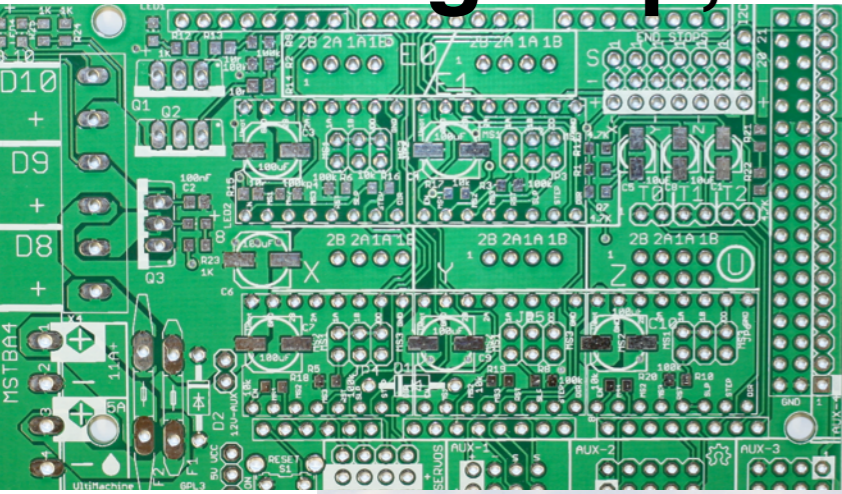


Cornish Coast Path

Researchers different?

Rainbow group, Cambridge

DCS, Edinburgh



1970

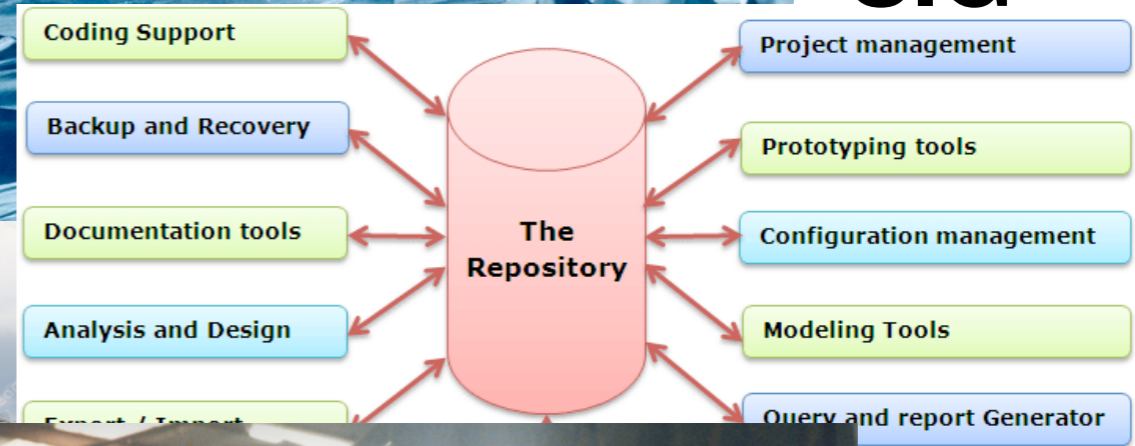
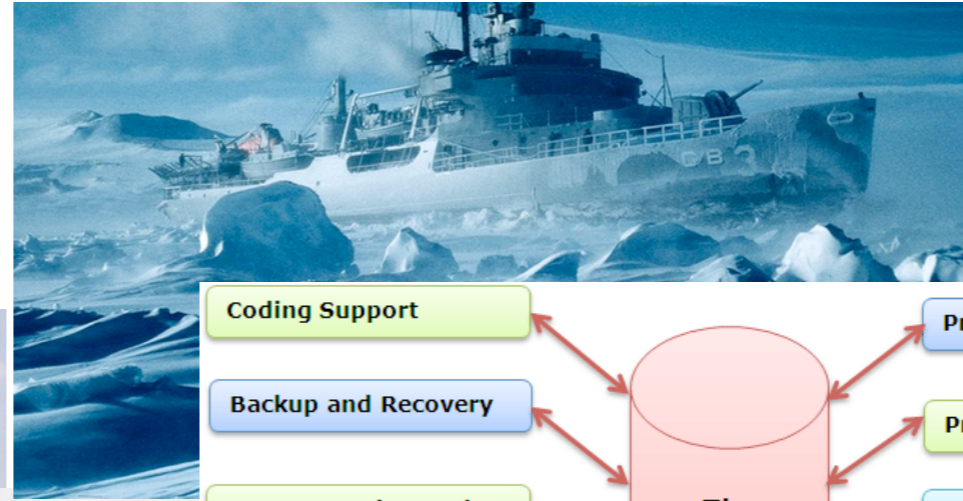
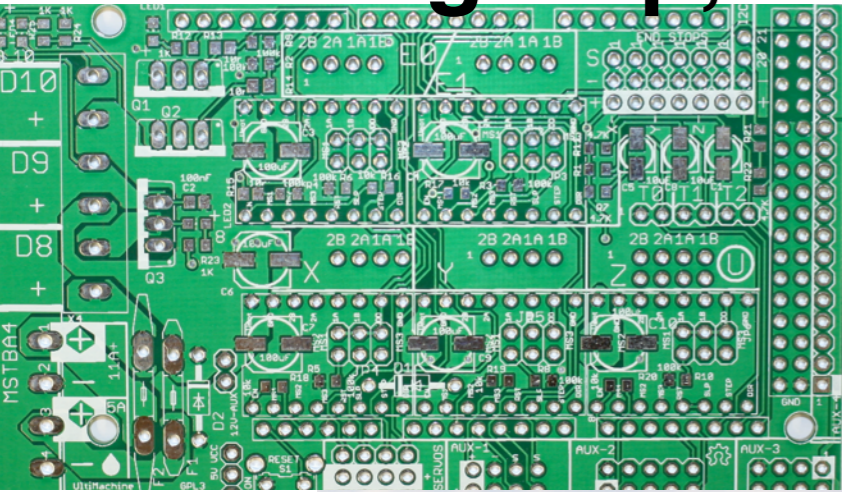
1980



Rainbow group, Cambridge

DCS, Edinburgh

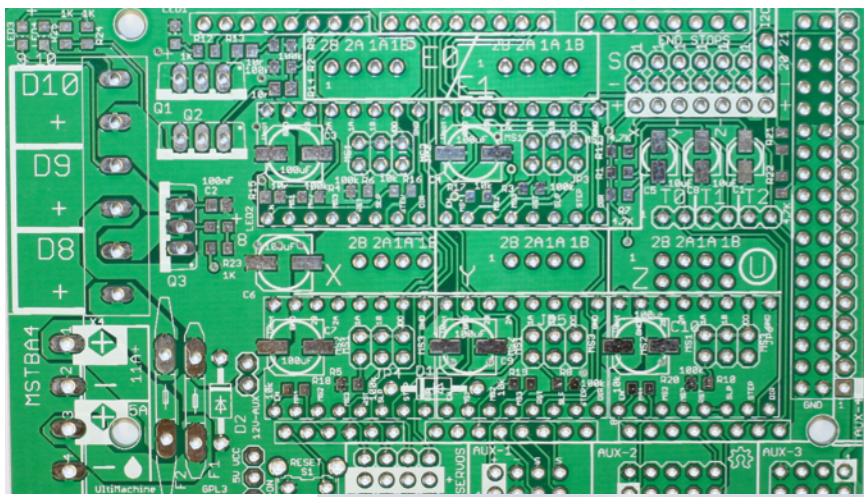
BCS
CAD
SIG



1970

1980





Individuals are gateway to success
Data *always* **distributed**
Data "**owned**" by *independent* "orgs"
Data **complex**: no single model works
Interconnected but imprecisely
Multiple uses / *all* "mission critical"
Conservatism dominates
Change is continuous



1970

1980

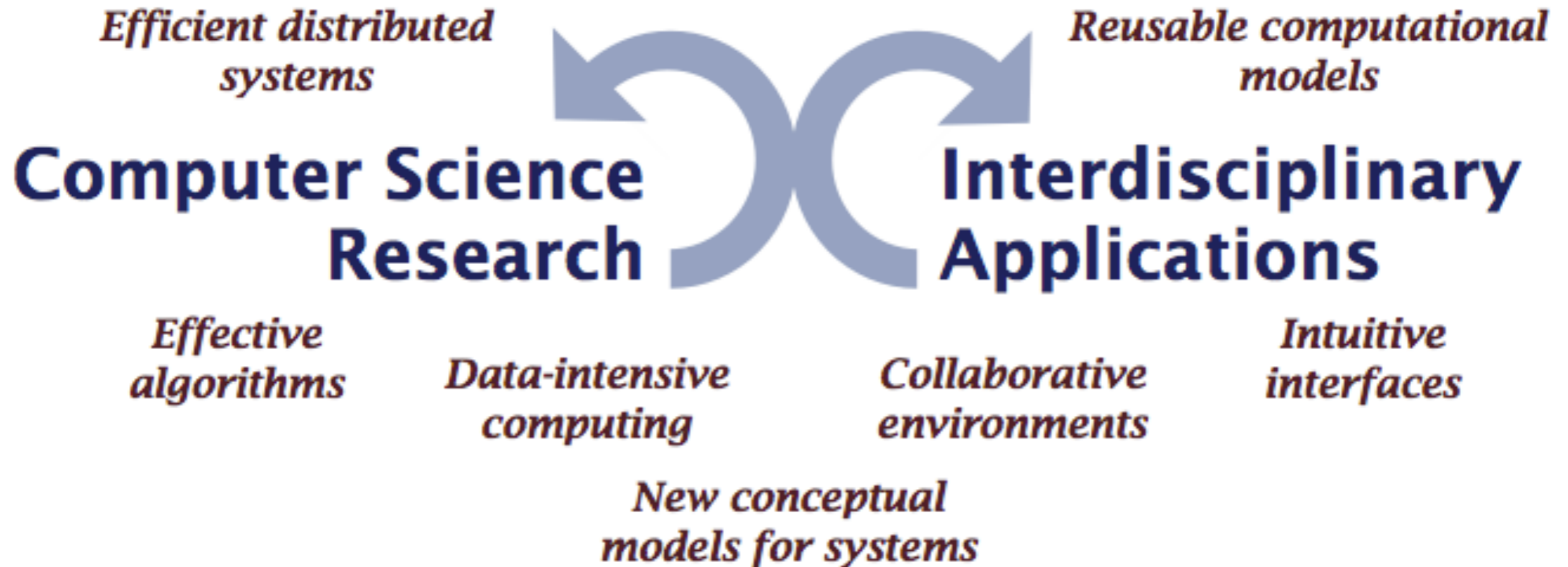


Data-Intensive Thinking

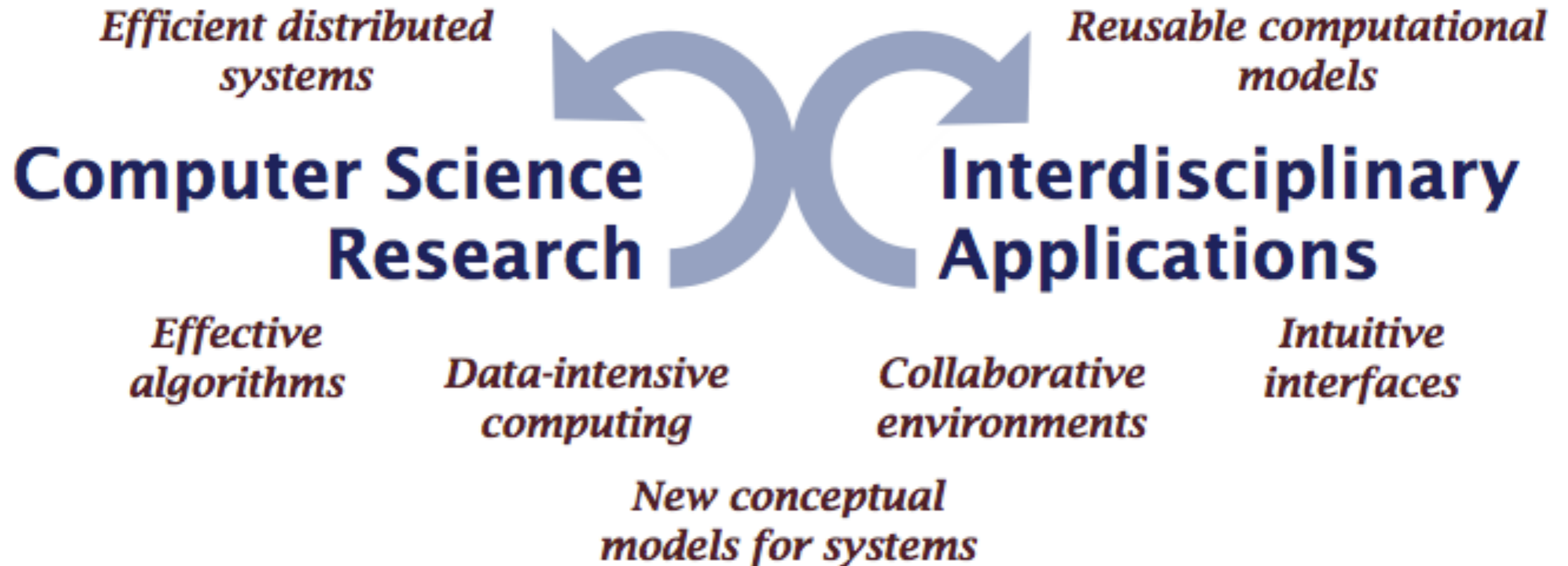


Data-Intensive Thinking

Two rapidly changing worlds

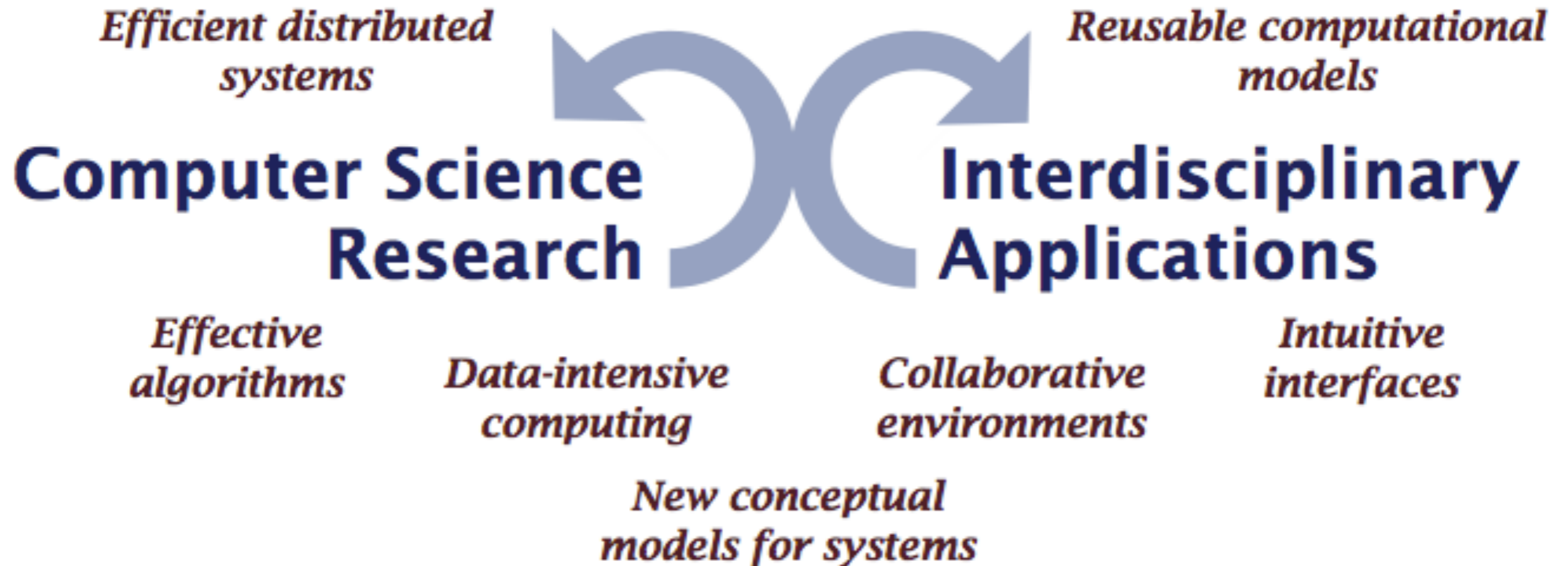


Two rapidly changing worlds



Research is motivated by change and enables change

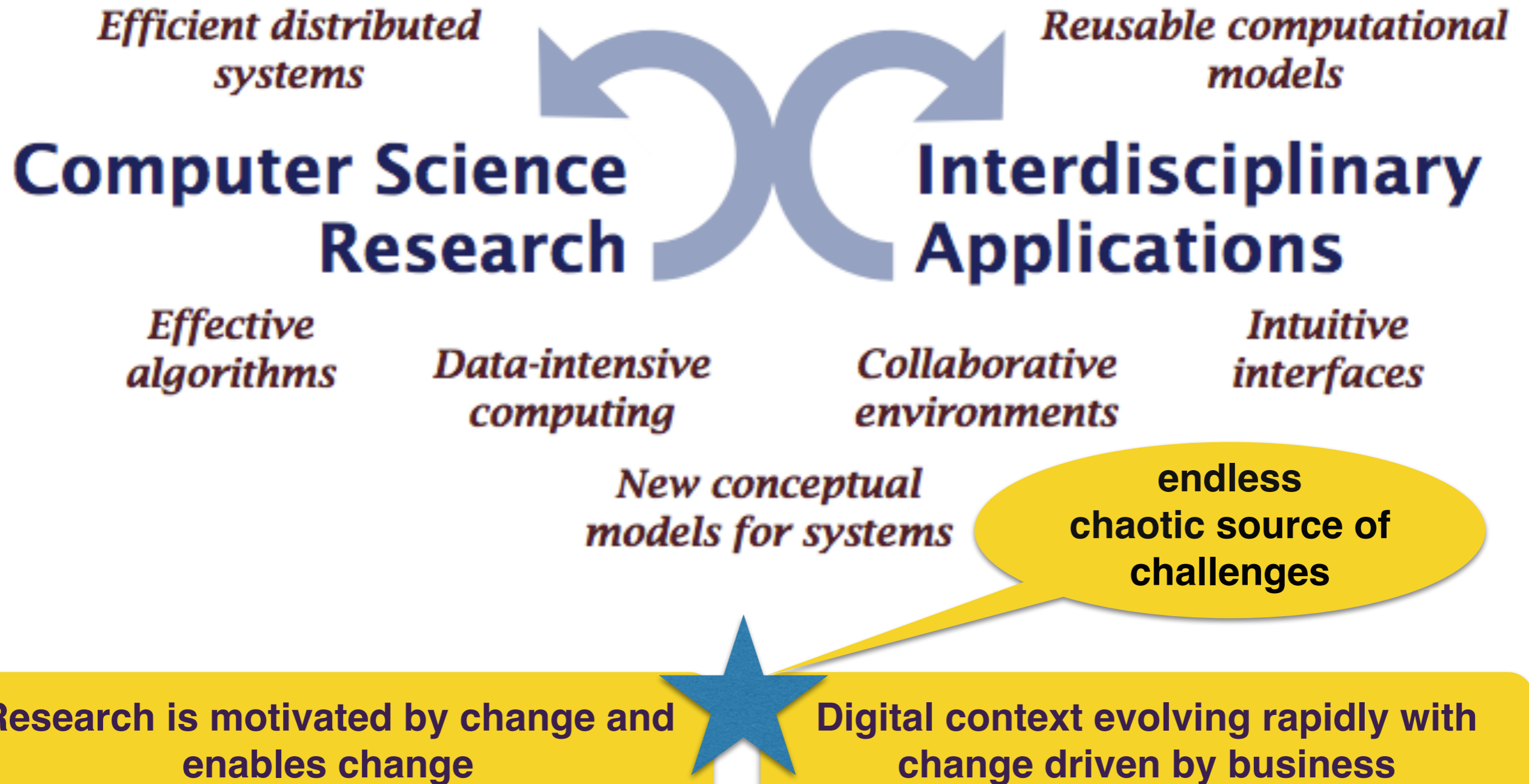
Two rapidly changing worlds



Research is motivated by change and enables change

Digital context evolving rapidly with change driven by business

Two rapidly changing worlds



Wiley Series on Parallel and Distributed Computing

Albert Y. Zomaya, Series Editor

The DATA Bonanza

Improving Knowledge Discovery
in Science, Engineering, and Business

EDITED BY
Malcolm Atkinson
Rob Baxter
Peter Brezany
Oscar Corcho
Michelle Galea
Mark Parsons
David Snelling
Jano van Hemert

IEEE

IEEE
computer
society

WILEY

Admire Project

- Model for Data Driven
 - science & research
 - engineering
 - business
- Abstraction
 - technical detail
- Longevity
 - as digital context evolves

Wiley Series on Parallel and Distributed Computing

Albert Y. Zomaya, Series Editor

The DATA Bonanza

Improving Knowledge Discovery
in Science, Engineering, and Business

EDITED BY
Malcolm Atkinson
Rob Baxter
Peter Brezany
Oscar Corcho
Michelle Galea
Mark Parsons
David Snelling
Jano van Hemert

IEEE

IEEE
computer
society

WILEY

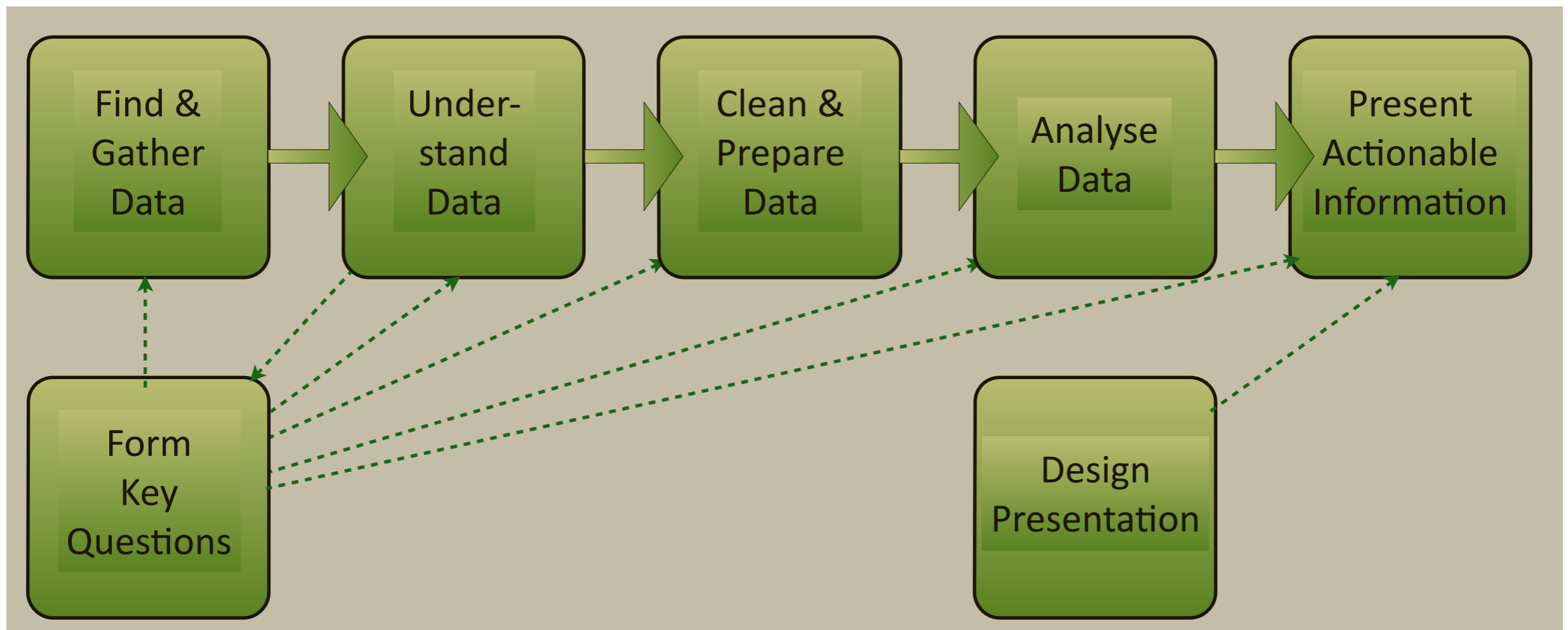
Admire Project

- Model for Data Driven
 - science & research
 - engineering
 - business
- Abstraction
 - technical detail
- Longevity
 - as digital context evolves

Free <http://onlinelibrary.wiley.com/book/10.1002/9781118540343>

Three Groups of Experts

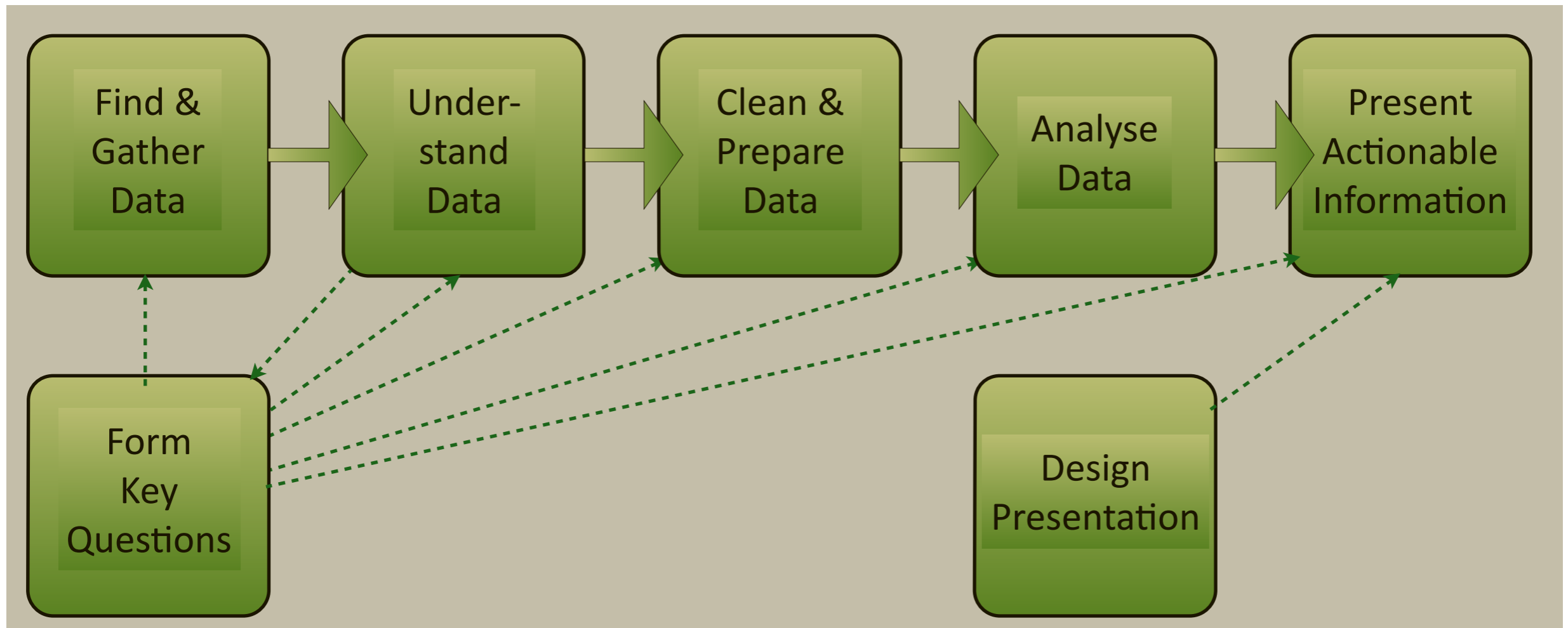
- **Domain expert**
- **Data-analysis experts**
- **Data-intensive engineers**



Three Groups of Experts

- Domain expert
- Data-analysis experts
- Data-intensive engineers

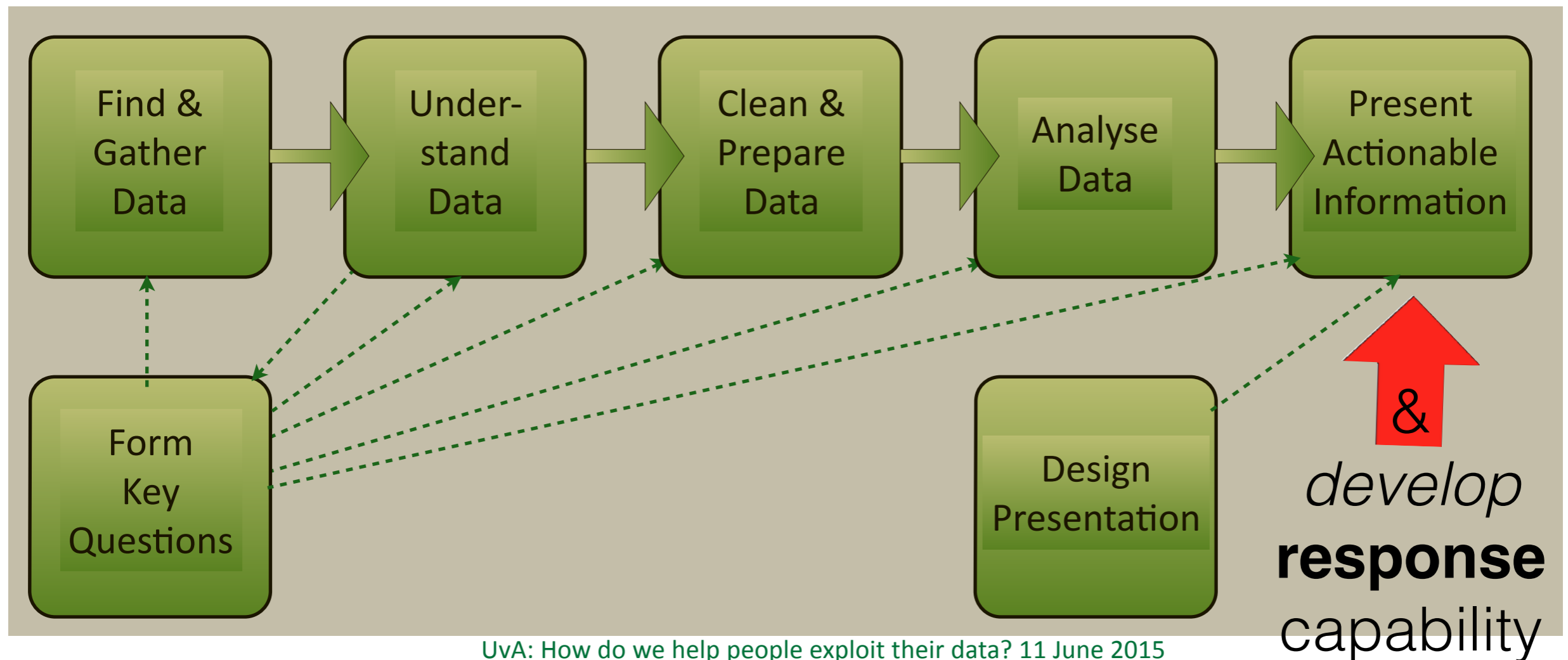
} Working together



Three Groups of Experts

- Domain expert
- Data-analysis experts
- Data-intensive engineers

} Working together



Jim Gray's legacy
• The Fourth Paradigm

Domain Experts

Jim Gray's valediction
*"May all your problems
be technical ones"*

- **Individuals**

- > 90% doing day job delivering services & building the evidence base
- <10% innovating: setting new goals & creating new methods
- Big variation in ITC knowledge
- different subdomains & different targets / changing
- in groups, in projects, in organisations
 - cooperating, competing / allying & pulling in different directions
- in organisational, in national & global cultures and communities
- strongly held preferences for computer interaction

- **Key primary issues**

- *Formulating & refining* scientific methods - **Empower the scientists to do this themselves**
- *Integrating* stages from *different* specialities - **Compose methods without understanding detail**
- Drawing on packaged techniques from other viewpoints - **Well-defined boundaries and semantics**
- *Demonstrable correctness* a **HUGE** challenge
- *Sustained value* as the digital context evolves another **HUGE** challenge

Jim Gray's legacy
• The Fourth Paradigm

Domain Experts

Jim Gray's valediction
*"May all your problems
be technical ones"*

- **Individuals**

- > 90% doing day job delivering services & building the evidence base
- <10% innovating: setting new goals & creating new methods
- Big variation in ITC knowledge
- different subdomains & different targets / changing
- in groups, in projects, in organisations
 - cooperating, competing / allying & pulling in different directions
- in organisational, in national & global cultures and communities
- strongly held preferences for computer interaction

- **Key primary issues**

- *Formulating & refining* scientific methods - **Empower the scientists to do this themselves**
- *Integrating* stages from *different* specialities - **Compose methods without understanding detail**
- Drawing on packaged techniques from other viewpoints - **Well-defined boundaries and semantics**
- *Demonstrable correctness* a **HUGE** challenge
- *Sustained value* as the digital context evolves another **HUGE** challenge

} **abstraction**

Jim Gray's legacy
• The Fourth Paradigm

Domain Experts

Jim Gray's valediction
“May all your problems
be technical ones”

- **Individuals**

- > 90% doing day job delivering services & building the evidence base
- <10% innovating: setting new goals & creating new methods
- Big variation in ITC knowledge
- different subdomains & different targets / changing
- in groups, in projects, in organisations
 - cooperating, competing / allying & pulling in different directions
- in organisational, in national & global cultures and communities
- strongly held preferences for computer interaction

- **Key primary issues**

- *Formulating & refining* scientific methods - **Empower the scientists to do this themselves**
- *Integrating* stages from *different* specialities - **Compose methods without understanding detail**
- Drawing on packaged techniques from other viewpoints - **Well-defined boundaries and semantics**
- *Demonstrable correctness* a **HUGE** challenge
- *Sustained value* as the digital context evolves another **HUGE** challenge

} **abstraction**

Individuals are gateway to success
Data *always* **distributed**
Data “**owned**” by *independent* “orgs”
Data **complex**: no single model works
Interconnected but imprecisely
Multiple uses / *all* “mission critical”
Conservatism dominates
Change is always happening
Rival tribes
Alliances: strategic or tactical

Jim Gray's legacy
• The Fourth Paradigm

Data-Analysis Experts

Jim Gray's valediction
*"May all your problems
be technical ones"*

- **Individuals**

- sub-specialists from mathematics and statistics to application-specific data-analysis
- trade-offs between data/computational cost and reliability and certainty
- favourite problem-solving environments
- different subdomains & different targets / changing
- in groups, in projects, in organisations
 - cooperating, competing / allying & pulling in different directions
- in organisational, in national & global cultures and communities
- strongly held preferences for computer interaction

- **Key primary issues**

- **Correctness** proven / tested ; clarity about scope of applicability / safety
- **Usability** how easily can the domain specialists grasp how to use a technique
- **Support** how much effort is there to sustain the technique and help get it used appropriately
- **Credit and blame** how do we attribute these fairly
- **Sustainability** dependencies and infrastructure independence
- **Relationship** with data-intensive engineering

Jim Gray's legacy
• The Fourth Paradigm

Data-Analysis Experts

Jim Gray's valediction
“May all your problems
be technical ones”

• **Individuals**

- sub-specialists from mathematics and statistics to application-specific data
- trade-offs between data/computational cost and reliability and certainty
- favourite problem-solving environments
- different subdomains & different targets / changing
- in groups, in projects, in organisations
 - cooperating, competing / allying & pulling in different directions
- in organisational, in national & global cultures and communities
- strongly held preferences for computer interaction
- **Key primary issues**
 - **Correctness** proven / tested ; clarity about scope of applicability / safety
 - **Usability** how easily can the domain specialists grasp how to use a technique
 - **Support** how much effort is there to sustain the technique and help get it used appropriately
 - **Credit and blame** how do we attribute these fairly
 - **Sustainability** dependencies and infrastructure independence
 - **Relationship** with data-intensive engineering

Individuals are gateway to success

Data *always* **distributed**

Data “**owned**” by *independent* “orgs”

Data **complex**: no single model works

Interconnected but imprecisely

Multiple uses / *all* “mission critical”

Conservatism dominates

Change is always happening

Rival tribes

Alliances: strategic or tactical

Data-Intensive Engineering

- **Individuals**

- sub-specialists: data storage, data transport, data bases, data curation, ..., computation, software & hardware architectures,..., requirements capture, ..., human-machine interaction, ...
- software communities, language communities, development models, ...
- from demon coders to formalisation experts
- in groups, in projects, in organisations
 - cooperating, competing / allying & pulling in different directions
- in organisational, in national & global cultures and communities
- strongly held preferences for interacting with computational systems

- **Key primary issues**

- **mapping** to existing and changing distributed computing platforms
- **exploiting** systems, architectures and components near optimally
- **Less energy** consumption
- **Sustainability**, how long can the investment survive?
- **Correctness** in the presence of diverse users and diverse infrastructures
- **Support** enabling users of all kinds and colleagues to use what they build

Data-Intensive Engineering

- **Individuals**

- sub-specialists: data storage, data transport, data bases, data curation, ..., computation, ..., human-machine interaction, ...
- software communities, language communities, development models, ...
- from demon coders to formalisation experts
- in groups, in projects, in organisations
 - cooperating, competing / allying & pulling in different directions
- in organisational, in national & global cultures and communities
- strongly held preferences for interacting with computational systems

- **Key primary issues**

- **mapping** to existing and changing distributed computing platforms
- **exploiting** systems, architectures and components near optimally
- **Less energy** consumption
- **Sustainability**, how long can the investment survive?
- **Correctness** in the presence of diverse users and diverse infrastructures
- **Support** enabling users of all kinds and colleagues to use what they build

Individuals are gateway to success

Data *always* **distributed**

Data “**owned**” by *independent* “orgs”

Data **complex**: no single model works

Interconnected but imprecisely

Multiple uses / *all* “mission critical”

Conservatism dominates

Change is always happening

Rival tribes

Alliances: strategic or tactical

Common issues

- **Diversity & distribution**
- **Composability** - integrating experts' insights & effort
- **Longevity** - making intellectual investment worthwhile
- **Correctness** - science, evidence & information trustworthy
- **Scalability** - in multiple dimensions on diverse & mobile DCI
- **Extensibility** - never finished *always* more is **needed**
- **Balance** - avoiding change >90% + Innovators <10%
- **Individuals**, groups, organisations, projects, communities

Extra pressures

Data growth
Expectation growth
User numbers & diversity growth

Data Growth

- Storage capacity
 - Kryder's law Faster than Moore
- Instrumentation
 - Resolution increasing
 - Speed increasing > Moore²
 - Cost / energy dropping
- Automation > Moore
 - laboratories, observatories, businesses, households, ...
- Simulations & Analyses
- Policy & standards > Moore

Data Growth

- Storage capacity
 - Kryder's law
- Instrumentation
 - Resolution increasing
 - Speed increasing
 - Cost / energy dropping

Faster than Moore

> Moore²

> Moore

- Automation

- I, observatories, busi

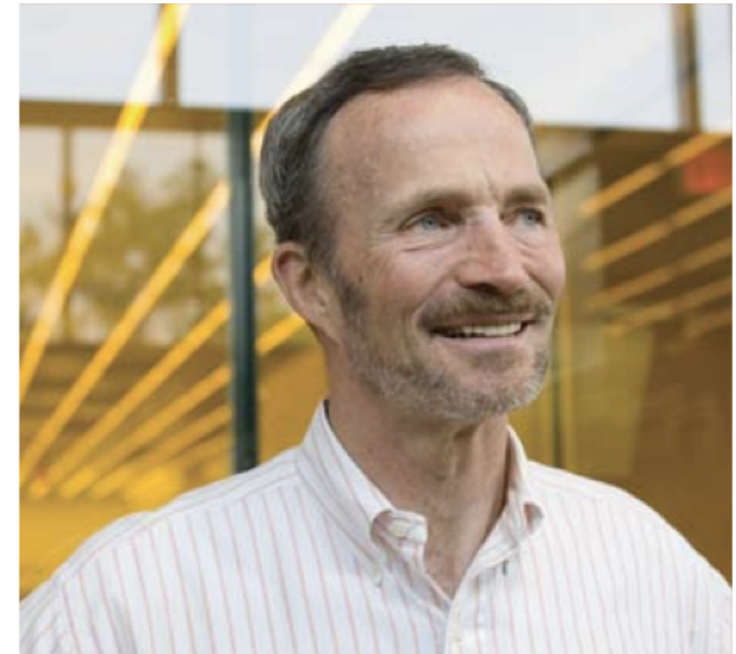
- Simulation & Analyses

> Moore

- Power

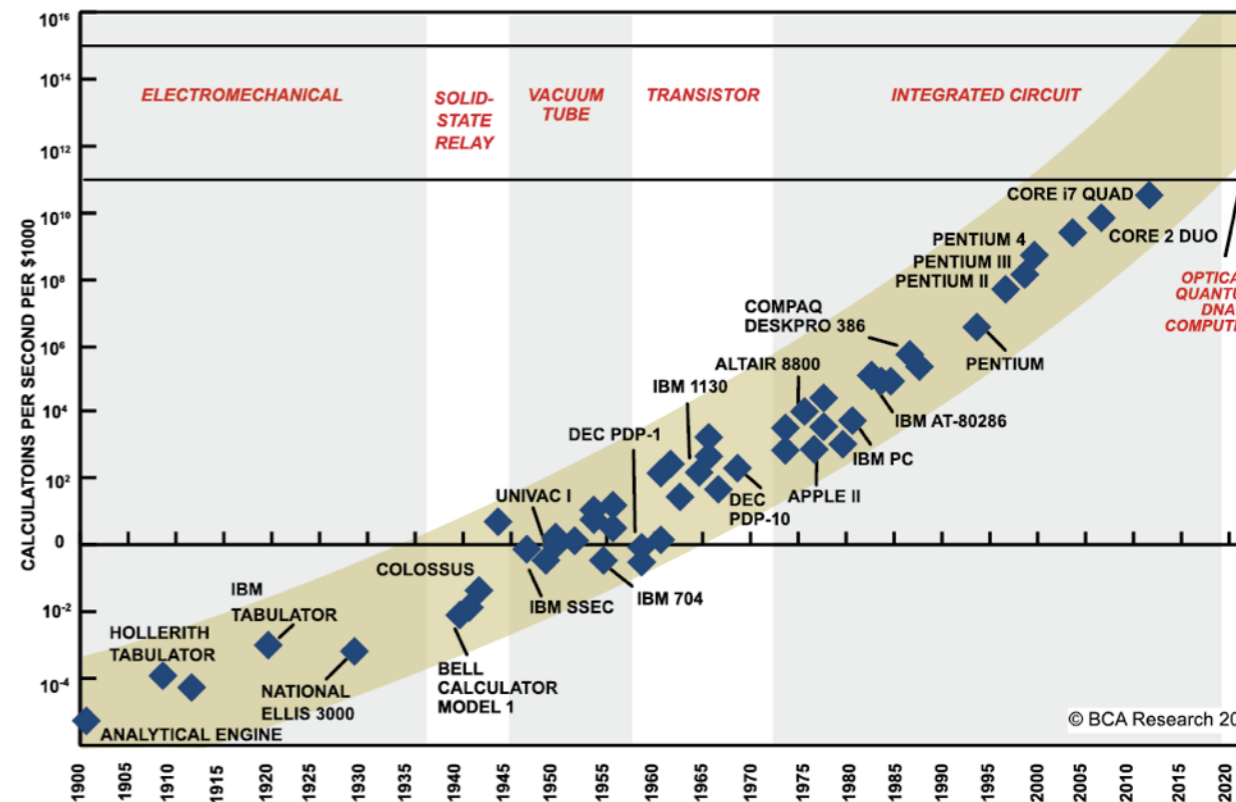


lp people exploit



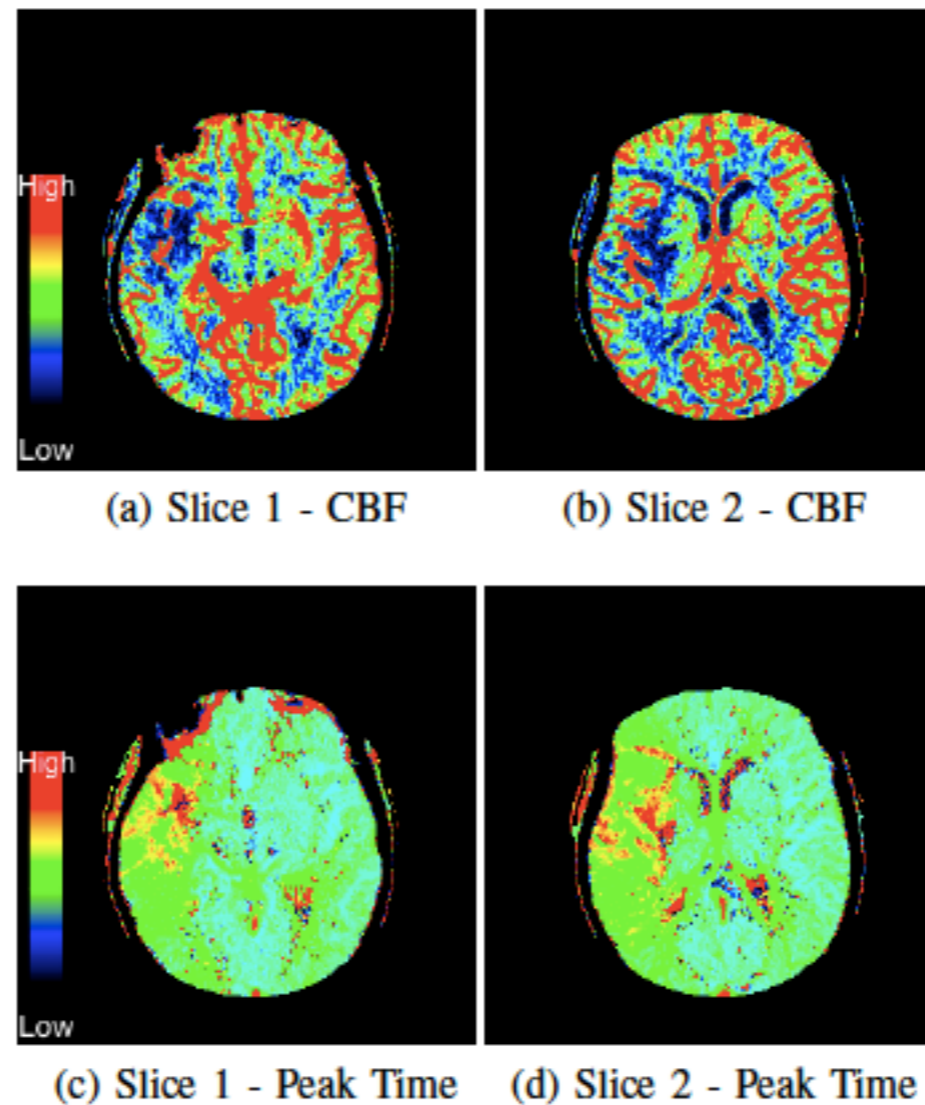
MARK KRYDER: STORING DATA

Leading the way in boosting hard-disk capacity, which has enabled



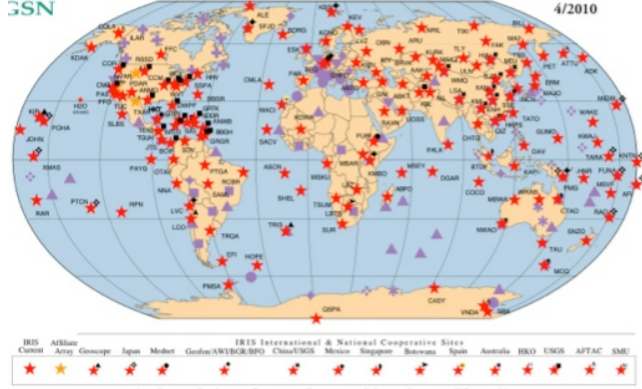
Projects & Alliances

Brain image processing



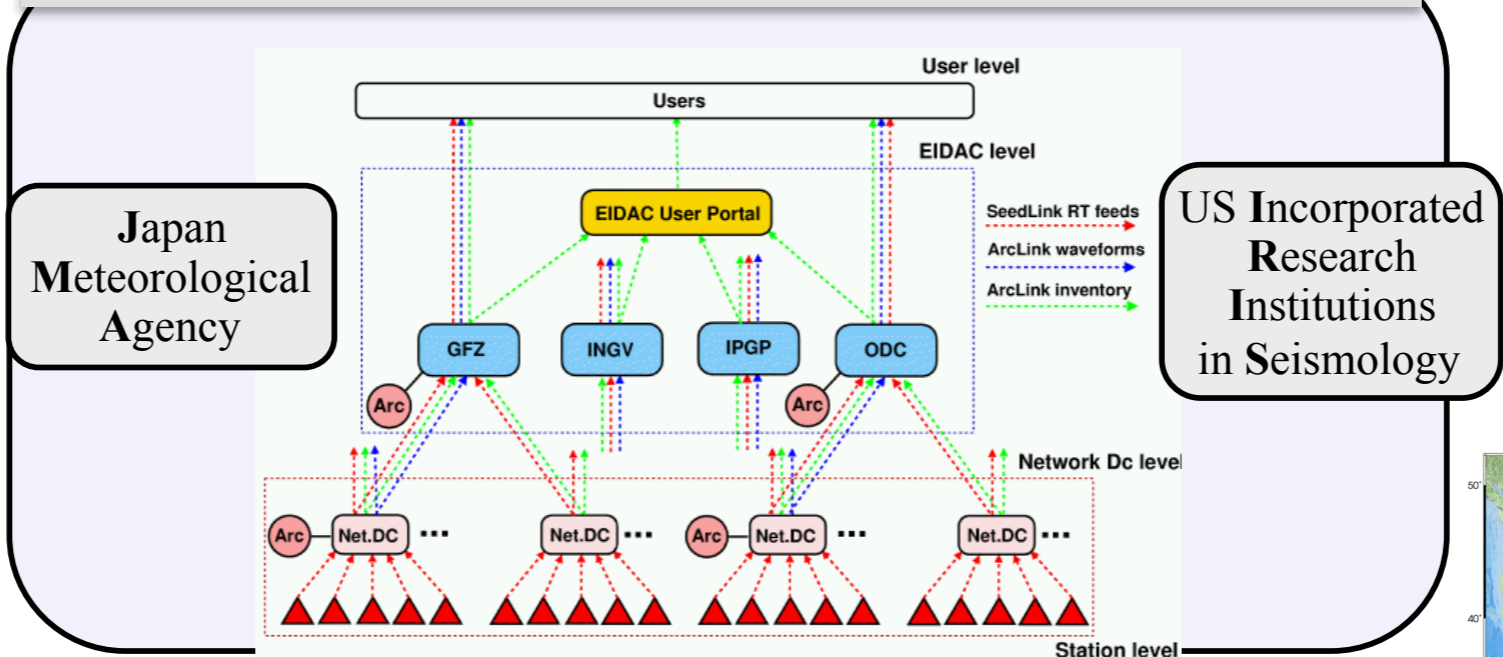
Lesion Area Detection Using Source Image
Correlation Coefficient for CT Perfusion Imaging

Fan Zhu David Rodriguez Gonzalez Trevor Carpenter Malcolm Atkinson Joanna Wardlaw



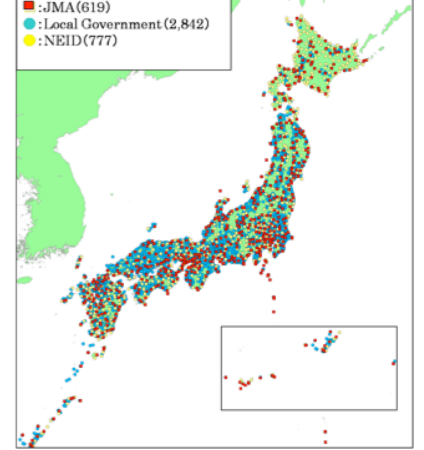
FDSN Global array

EUROPEAN INTEGRATED DATA ARCHIVES (EIDA)



Japan Meteorological Agency

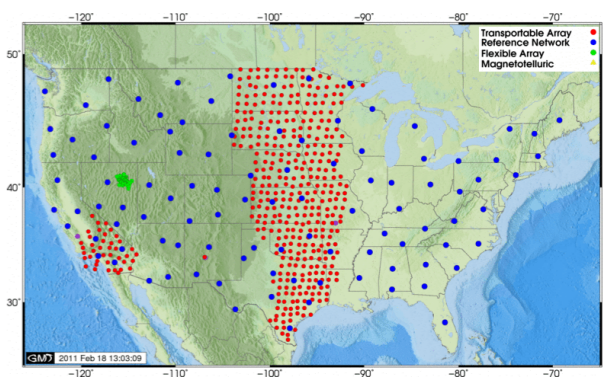
US Incorporated Research Institutions in Seismology



Japan array



European array



US array

Data Intensive Research

- Visualization
- Data analysis / Data mining
- Simulation, inversion, HR imaging

VERCE

e-Science environment for data intensive research based on an extensive service-oriented architecture

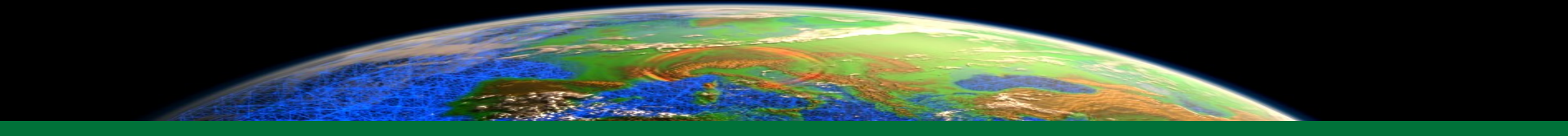
HPC/GRID Infrastructures



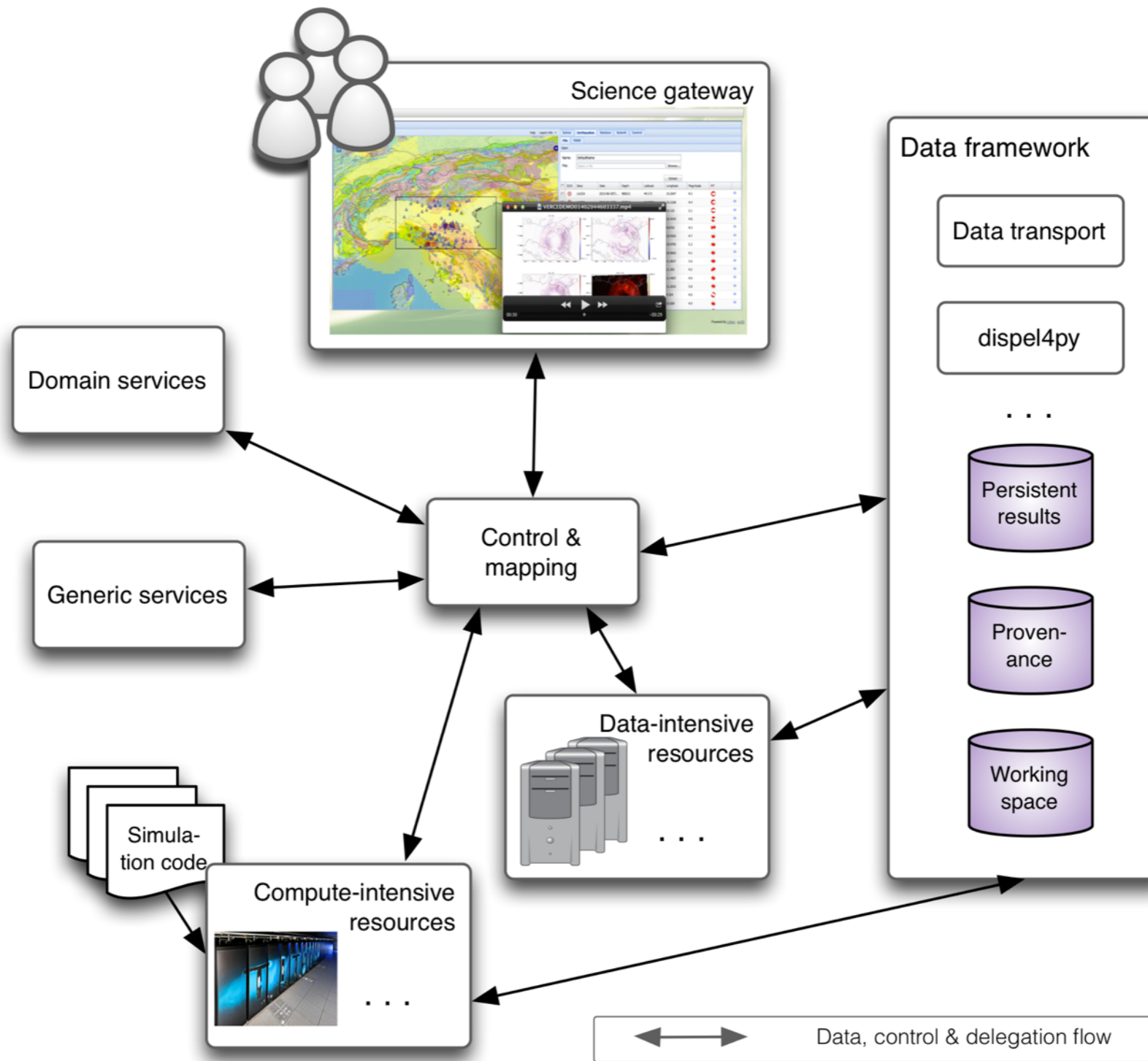
Earth's interior imaging and dynamics: noise correlation, waveform analysis

Natural hazards: new tools for monitoring earthquakes, volcanoes, and tsunami

Interaction of solid Earth with Ocean and Atmosphere: environment, climate changes

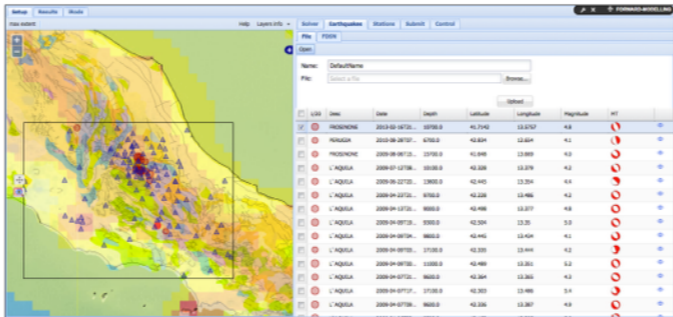


VERCE architecture



2015 “VERCE delivers a productive e-Science environment for seismology research”

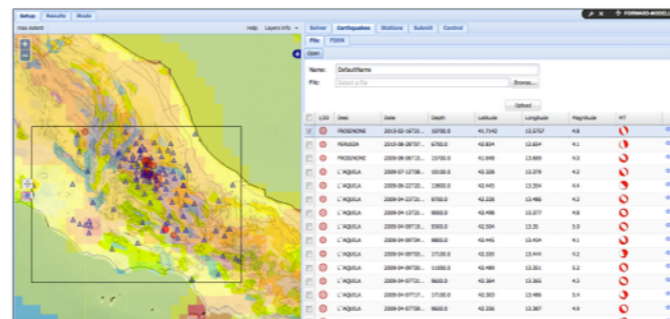
VERCE Platform, Components Interaction



*Science Gateway
Community Applications*



VERCE Platform, Components Interaction



Science Gateway
Community Applications

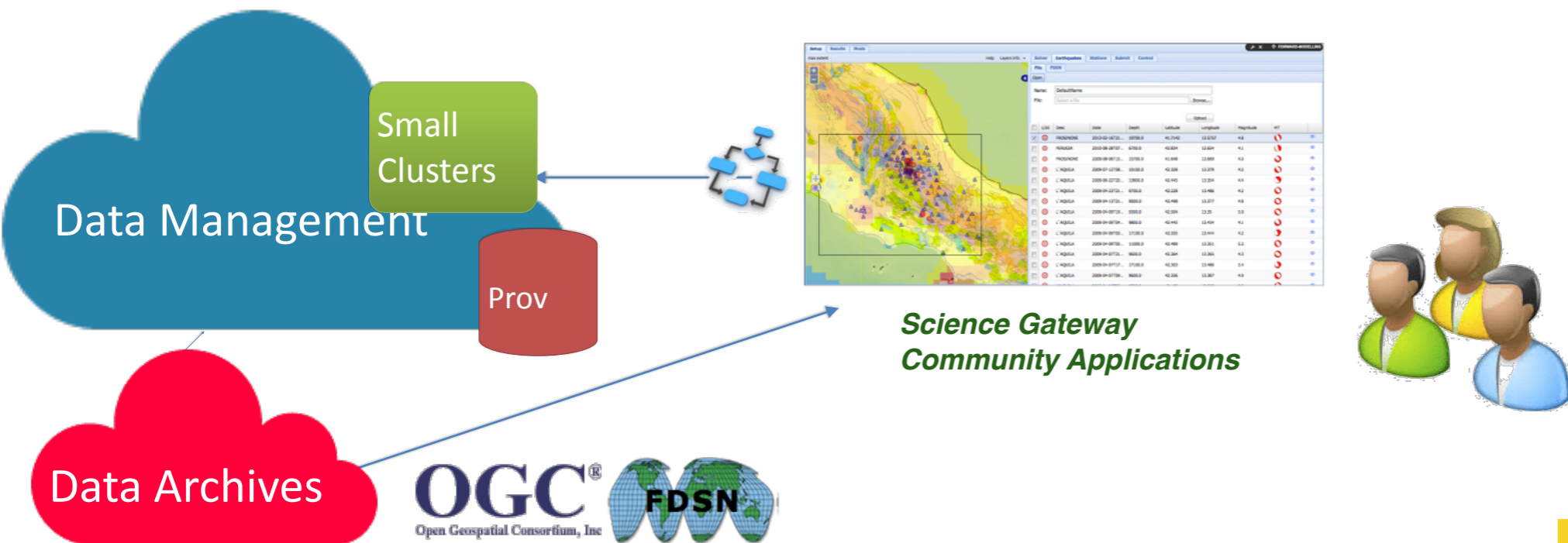


VERCE Platform, Components Interaction



1 - Raw data acquisition

3 - MISFIT

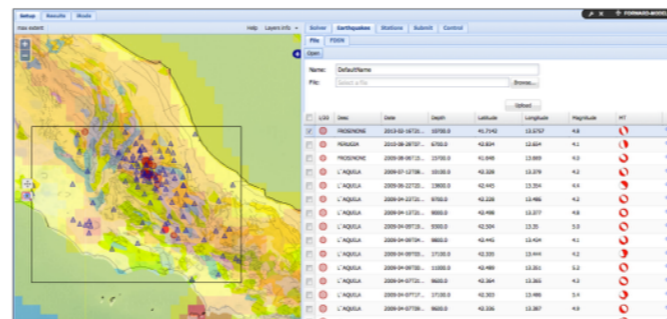


VERCE Platform, Components Interaction

2 - HPC Simulation
(model stage-in)



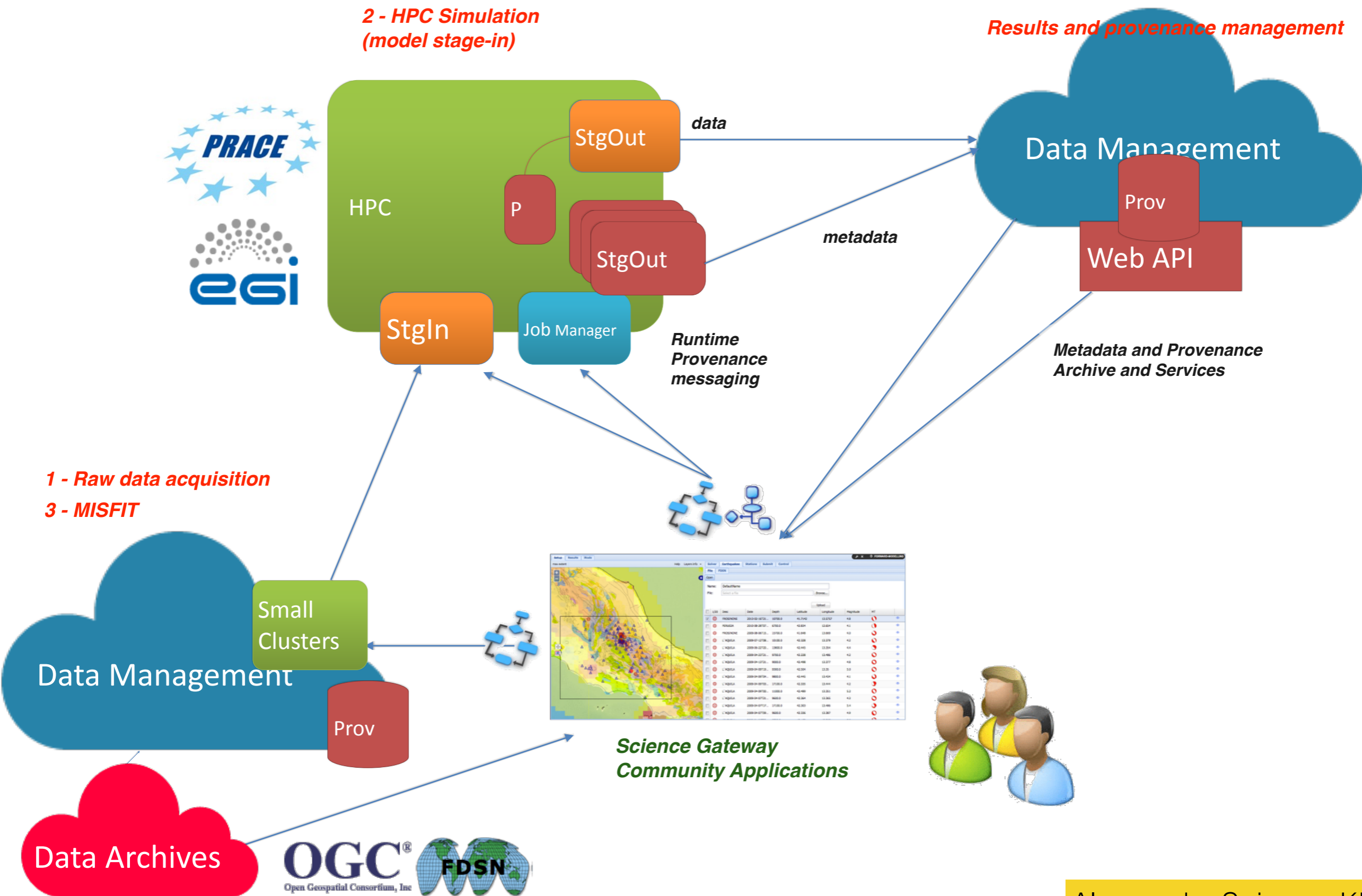
1 - Raw data acquisition
3 - MISFIT



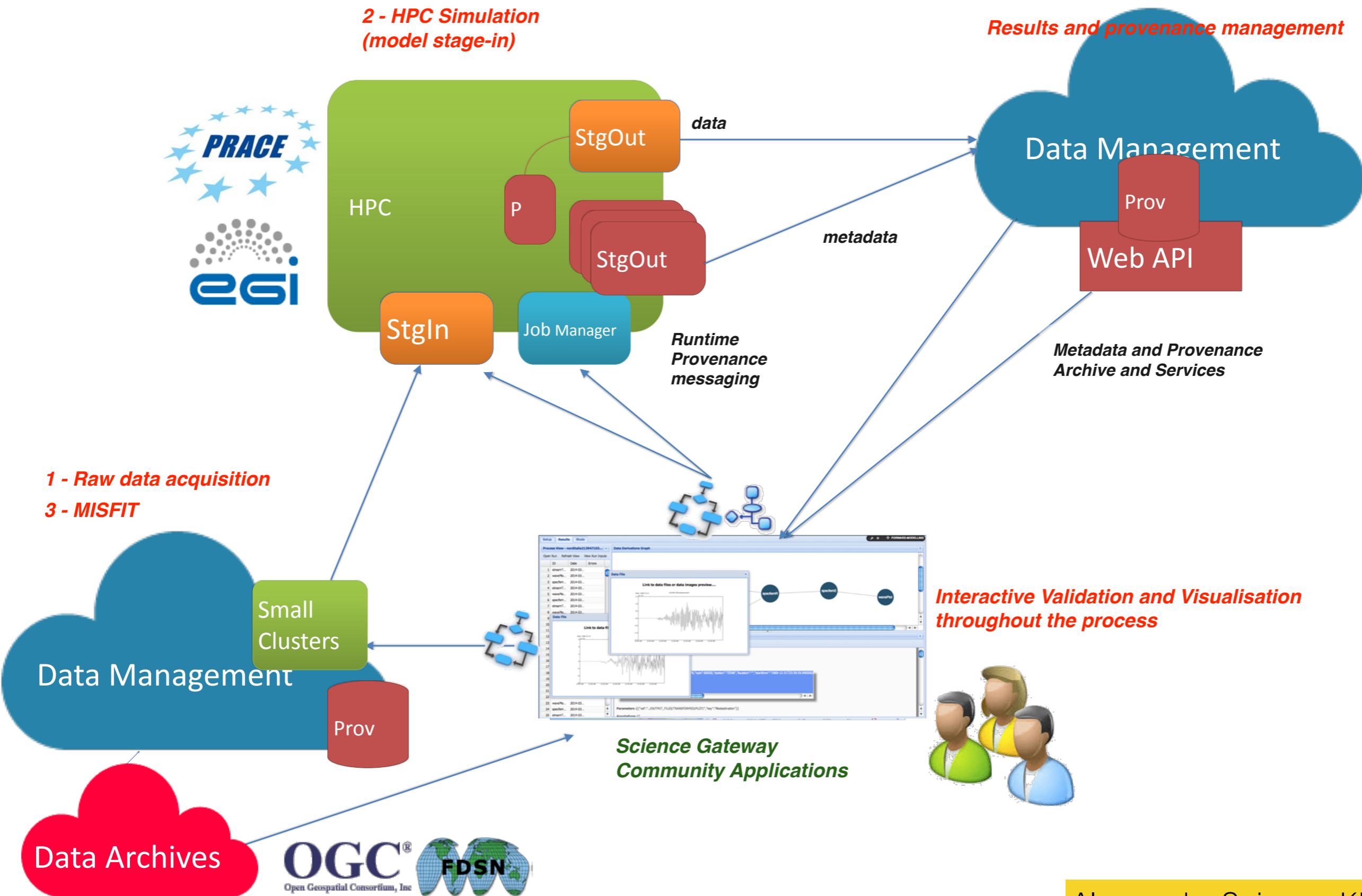
Science Gateway
Community Applications



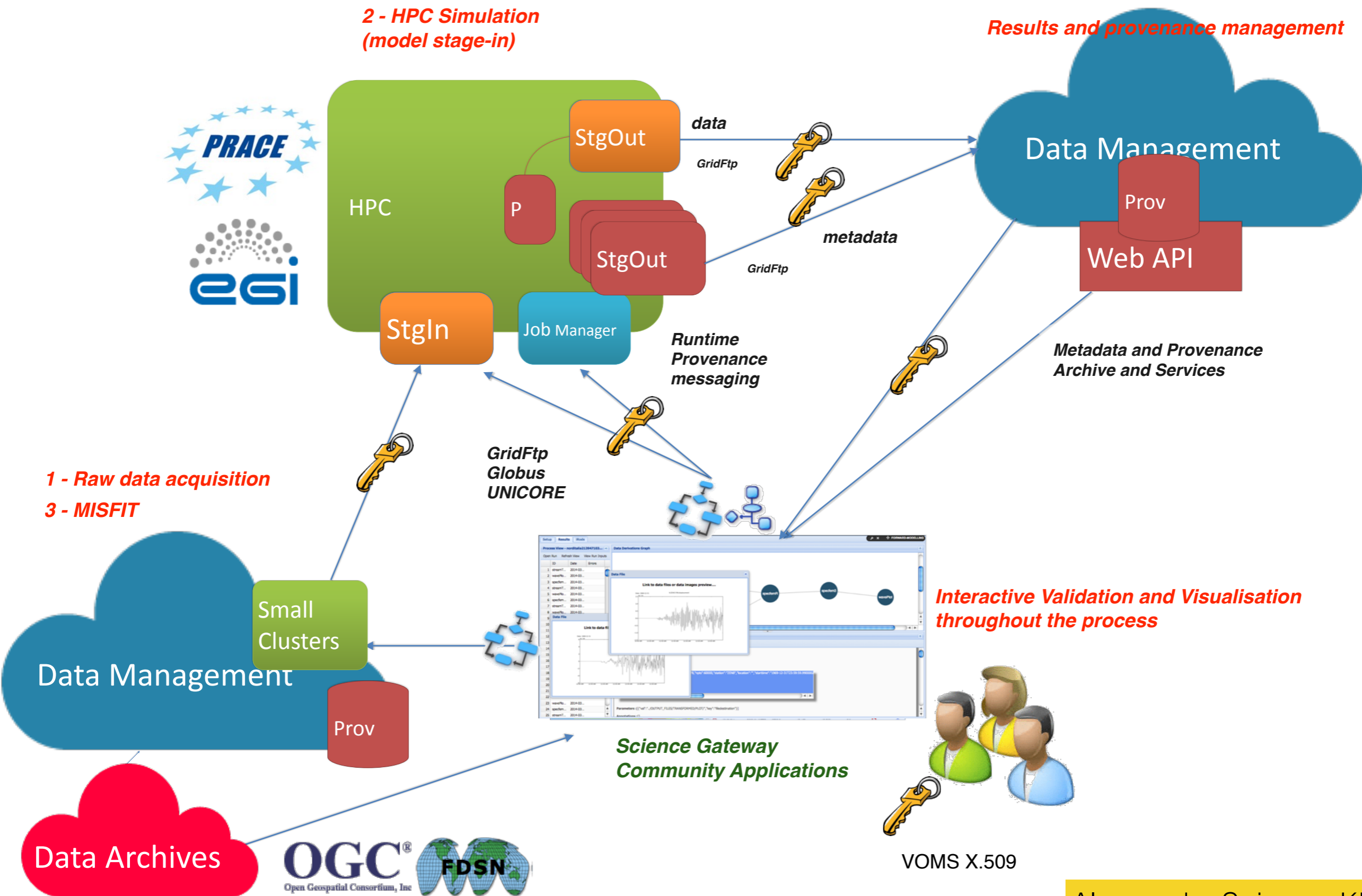
VERCE Platform, Components Interaction



VERCE Platform, Components Interaction



VERCE Platform, Components Interaction





Virtual Earthquake and Seismology Research Community in Europe

Virtual Environment for of **Earthquakes Simulations** and
evaluation of **Earth Models**

<http://portal.verce.eu>

Combined access to **computing infrastructures** (EGI, PRACE, Local Clusters), for
development and execution of large **HPC** computations

Access and use of **European data archives** and services adopting International
standards (FDSN, GCMT, OneGeology, EFEHR, QuakeML)

**Adoption of Workflow Technologies, Data Management and Provenance
System**

Setup Results iRods

max extent

Solver Earthquakes Stations Submit Control

File FDSN

Open

Name: DefaultName

File: Browse...

Upload

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

VERCEDEMO01402044603337.mp4

Powered By [Liferay](#) - [gUSE](#)

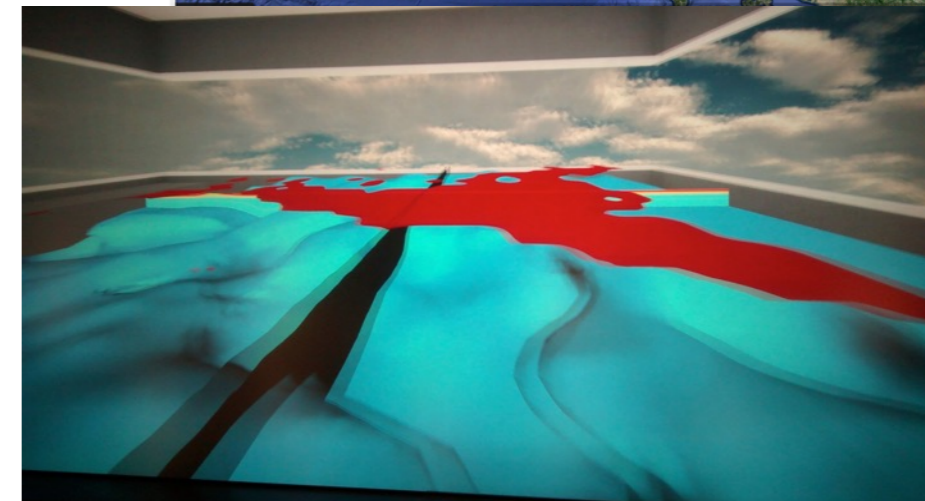
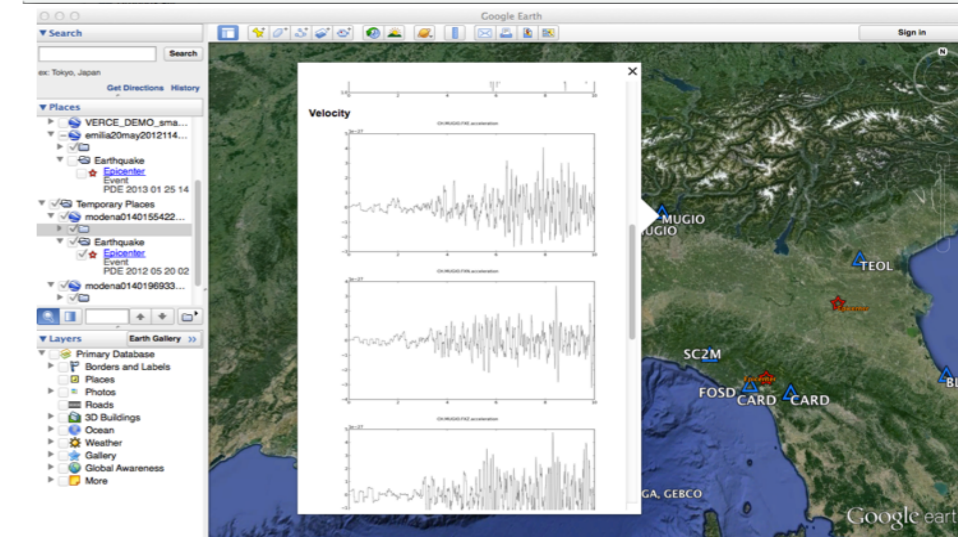
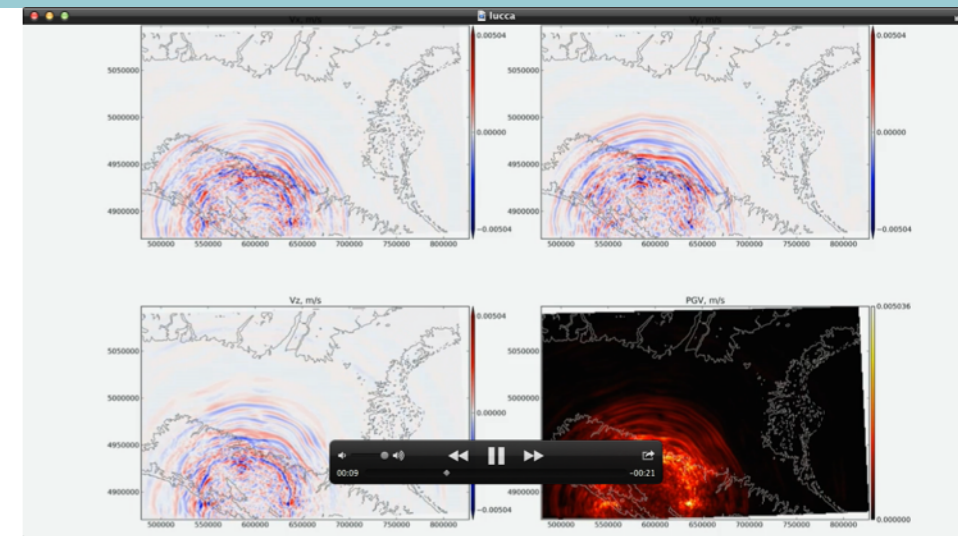


Earthquake Simulation and Misfit Calculation

Data overview:

Synthetic Data: seismograms, plots, 3D Geometry, Videos, KMZ packages, **meshes and models.**
(100 stations = 900 products and metadata)
6-10 GB for a SPECFEM3D simulation on 1000 cores

Raw Data: on demand **access** and **staging** of **observational data** from **EIDA:** Earthquake Metadata, Sensors Metadata, waveform on regional scale.
(At the moment all via the FDSN WEB API)



Human issues!

<http://www.nature.com/news/2011/110914/full/477264a.html>

Nature 14 Sept. 2011

The screenshot shows the top of a Nature website article. The header includes the 'nature' logo and navigation tabs for 'nature news home', 'news archive', 'specials', 'opinion', 'features', 'news blog', and 'nat'. Below the header, there is a 'comments on this story' section, a 'Published online' date of 14 September 2011, and a 'News Feature' label. The main title is 'Scientists on trial: At fault?' with a sub-headline: 'In 2009, an earthquake devastated the Italian city of L'Aquila and killed more than 300 people. Now, scientists are on trial for manslaughter.' The author is Stephen S. Hall. On the left sidebar, there are sections for 'stories by subject' (Earth Sciences, Environmental Science, Policy) and 'stories by keywords' (L'Aquila, Earthquake, Seismology, Law, Italy, Risk communication). At the bottom of the sidebar, there is a 'this article elsewhere' section with links to Connotea, Digg, Facebook, Newsvine, and Del.icio.us.

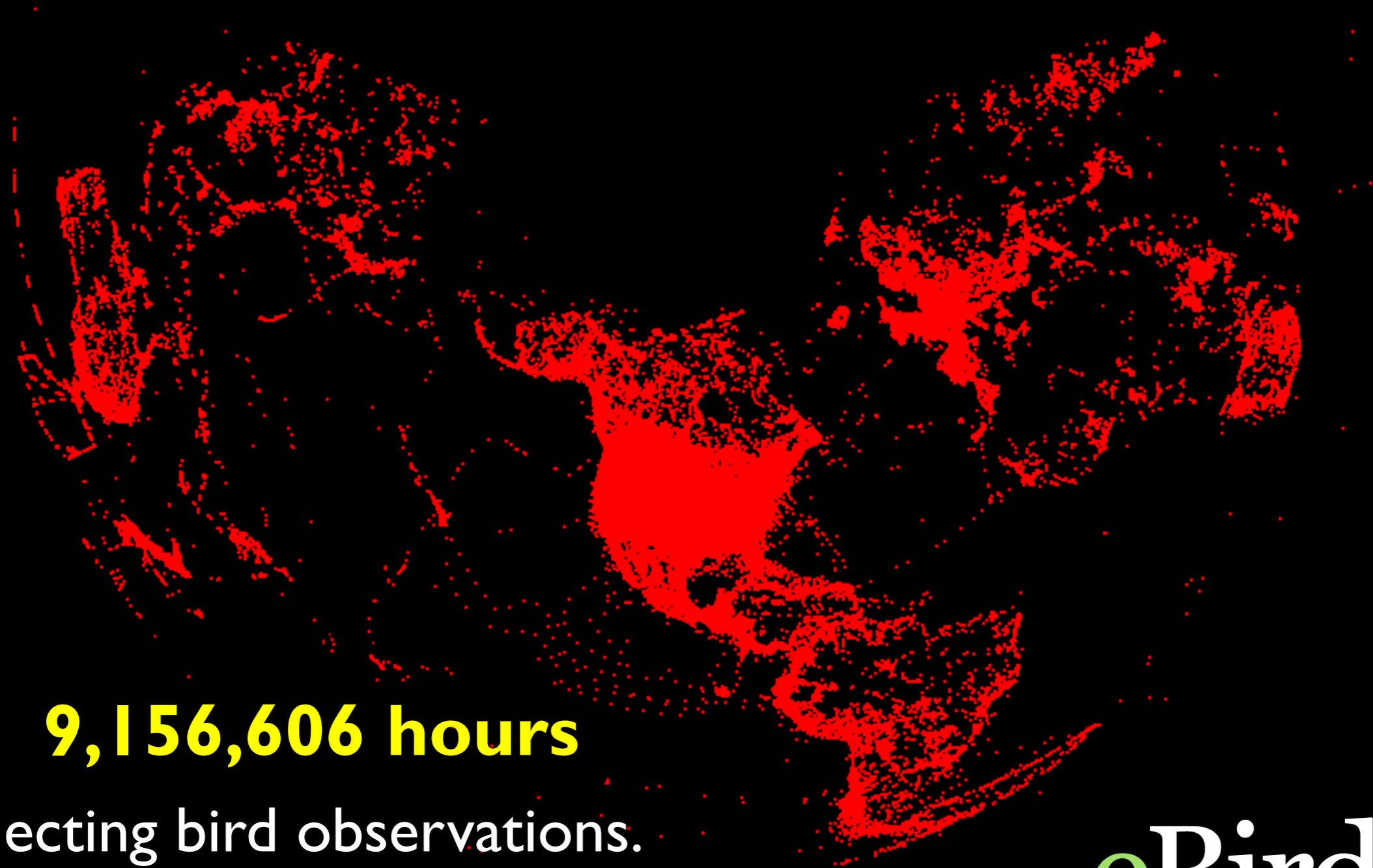
In a trial set to begin next week, an Italian judge will decide whether the symbolic death of L'Aquila — and, more specifically, the earthquake-related deaths of dozens of citizens included in the lawsuit including Vittorini's wife and daughter — constituted a crime due to the negligence of six leading Italian scientists and one government official, who have been charged with manslaughter in connection with the case.

When the charges were first aired in June 2010 by public prosecutor Fabio Picuti, the case was likened to a frivolous attempt by overzealous local prosecutors to make scapegoats out of some of Italy's most respected geophysicists: Enzo Boschi, then-president of Italy's National Institute of Geophysics and Volcanology (INGV) in Rome; Franco Barberi, at the University of 'Rome Tre'; Mauro Dolce, head of the seismic-risk office at the national Department of Civil Protection in Rome; Claudio Eva, from the University of Genova; Giulio Selvaggi, director of the INGV's National Earthquake Centre in Rome; and Gian Michele Calvi, president of the European Centre for Training and Research in Earthquake Engineering in Pavia; as well as government official Bernardo De Bernardinis, then vice-director of the Department of Civil Protection. According to an open letter to the president of Italy, Giorgio Napolitano, signed by more than 5,000



A. Nusca/Polaris/eyevine

From when he was a young boy growing up in a house on Via Antinori in the medieval heart of this earthquake-prone Italian city, Vincenzo Vittorini remembers the ritual whenever the

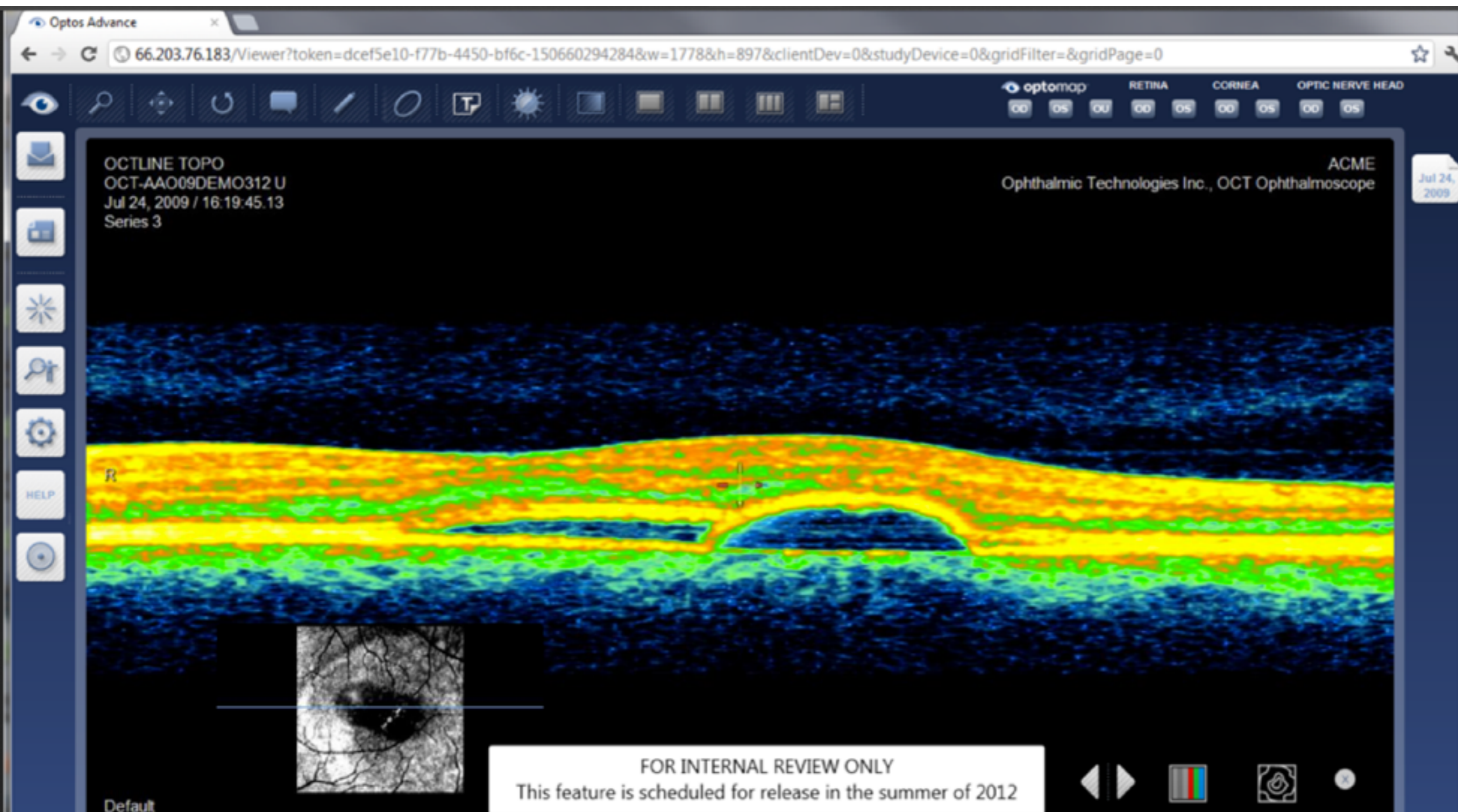


9,156,606 hours

collecting bird observations.

eBird

Optos retinopathy diagnosis



Workflows as a DI strategy

What is a workflow?

A **composition** of steps

to make a data-handling + data-analysis+
simulation journey

Many ways of forming steps

Require **good libraries** of ready made steps

Learn to add your own

Many ways of combining steps

Many computing environments

Recursive — a journey can be a step in another journey

Not necessarily using the same workflow system

What is a workflow?

A **composition** of steps

to make a data-handling + data-analysis+
simulation journey

Many ways of forming steps

Require **good libraries** of ready made steps

Learn to add your own

Many ways of combining steps

Many computing environments

Recursive — a journey can be a step in another journey

Not necessarily using the same workflow system

Common problems

Technology details included

Target details

Composition mono-lingual

Why use a workflow?

Rapid prototyping and experiment
Saving labour and repeated drudgery
Reducing error rates

**Empower the
*domain experts***

Saving you from doing your own housekeeping

Returning resources such as file space

Gathering all your results

Acceleration & scalability due to workflow optimisation,
e.g. parallelisation

Sharing & getting credit for methods

Incrementally improving methods

Combining methods developed by different experts

There are many workflow languages - why invent **dispel**?

Raising the level of discourse

Removing much technology specific information - technology changes
Relieving users from concerns about optimisation

Improving the logical description

Streams of data with auto-iteration over data units
Multiple streams in & multiple streams out
Behaviour, data interpretation & data representation

Covering existing models

Distributed query
 Optimisation based on avoiding IO & characterising operators
Real-time processing
Task-based batch processing

Achieving scalability

After dispel why invent dispel4py?

Re-engineering dispel in **Python**

user demand

python ubiquity

richness of python libraries

potential for open source collaboration increase

rapid prototyping

smooth path from development to production

Light weight composition

Reducing boundary crossings

Mappings to multiple data-intensive infrastructures

workflows run *unchanged* on each target

Concepts

Processing elements

Composed by data streams

Distributed control and termination

Types of PEs

Type	Inputs	Outputs	When to use it
GenericPE	n inputs	n outputs	many inputs and/or many outputs
IterativePE	1 input named 'input'	1 output named 'output'	processing one data block and producing one in each iteration
Consumer PE	1 input named 'input'	no output	only implement <code>_process</code> method
ProducerPE	0 input	1 output named 'output'	could be the first PE. when it does not any inputs and generates 1 output
Simple FunctionPE	1 input named 'input'	1 output named 'output'	only implement <code>_process</code> method
create_iterative_chain	1 input named 'input'	1 output named 'output'	pipeline of functions processing sequentially; creates a composite PE

Types of PEs

Type	Inputs	Outputs	When to use it
GenericPE	n inputs	n outputs	many inputs and/or many outputs
IterativePE	1 input named 'input'	1 output named 'output'	processing one data block and producing one in each iteration
Consumer PE	1 input named 'input'	no output	only implement <code>_process</code> method
ProducerPE	0 input	1 output named 'output'	could be the first PE. when it does not any inputs and generates 1 output
Simple FunctionPE	1 input named 'input'	1 output named 'output'	only implement <code>_process</code> method
create_iterative_chain	1 input named 'input'	1 output named 'output'	pipeline of functions processing sequentially; creates a composite PE

Types of PEs

Type	Inputs	Outputs	When to use it
GenericPE	n inputs	n outputs	many inputs and/or many outputs
IterativePE	1 input named 'input'	1 output named 'output'	processing one data block and producing one in each iteration
Consumer PE	1 input named 'input'	no output	only implement <code>_process</code> method
ProducerPE	0 input	1 output named 'output'	could be the first PE when it does not any inputs and generates 1 output
Simple FunctionPE	1 input named 'input'	1 output named 'output'	only implement <code>_process</code> method
create_iterative_chain	1 input named 'input'	1 output named 'output'	pipeline of functions processing sequentially; creates a composite PE

Empowering Domain Experts

Types of PEs

Type	Inputs	Outputs	When to use it
GenericPE	n inputs	n outputs	many inputs and/or many outputs
IterativePE	1 input named 'input'	1 output named 'output'	processing one data block and producing one in each iteration
Consumer PE	1 input named 'input'	no output	only implement <code>_process</code> method
ProducerPE	0 inputs	1 output named 'output'	could be the first PE, when it does not any inputs and generates 1 output
Simple FunctionPE	1 input named 'input'	1 output named 'output'	only implement <code>_process</code> method
create_iterative_chain	1 input named 'input'	1 output named 'output'	pipeline of functions processing sequentially; creates a composite PE

Data-Intensive Engineering

Empowering Domain Experts

Agile decision factories as a DI strategy

Forums to agree or disagree

- Standards
 - too many
 - too few
 - wrong scope
 - rarely in play
 - inhibit and enable progress

Forums to agree or disagree

- **Standards**
 - too many
 - too few
 - wrong scope
 - rarely in play
 - inhibit and enable progress

Context for fast decisions

Place to bring issues and gather experts
Open informed discussion intense
Conclusions of common path or difference rapid
New topics easily introduced
Global community aware of discussions
Welcome those with issues
and those with solutions

Summary and Conclusions

Exploit the DATA Bonanza

Exploit the DATA Bonanza

- **Educate** all about how to use data
 - The **three categories of expert**
 - Data literate managers, governmental officials & ...
 - A data savvy public
 - High quality opinions and decisions based on data

Exploit the DATA Bonanza

- **Educate** all about how to use data
 - The **three categories of expert**
 - Data literate managers, governmental officials & ...
 - A data savvy public
 - High quality opinions and decisions based on data
- **Long-term** development of leadership
 - curated data
 - expert teams
 - good engineering needed for *usability* and scalability

Exploit the DATA Bonanza

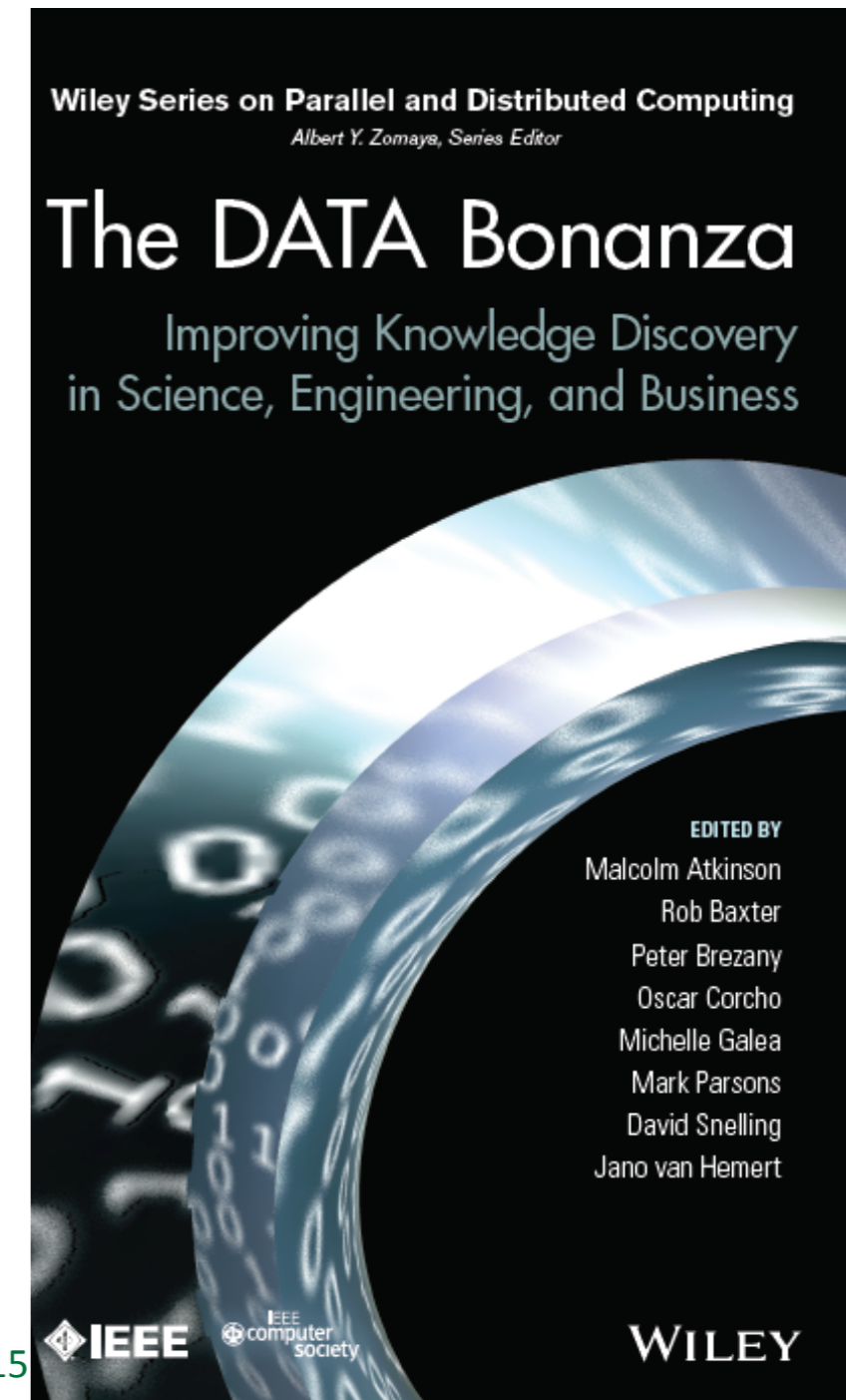
- **Educate** all about how to use data
 - The **three categories of expert**
 - Data literate managers, governmental officials & ...
 - A data savvy public
 - High quality opinions and decisions based on data
- **Long-term** development of leadership
 - curated data
 - expert teams
 - good engineering needed for *usability* and scalability
- **Balanced investment**
 - from collection to “*final mile*” of information delivery

Exploit the DATA Bonanza

- **Educate** all about how to use data
 - The *three categories of expert*
 - Data literate managers, governmental officials & ...
 - A data savvy public
 - High quality opinions and decisions based on data
- **Long-term** development of leadership
 - curated data
 - expert teams
 - good engineering needed for *usability* and scalability
- **Balanced investment**
 - from collection to “*final mile*” of information delivery
- **Open data and processes**
 - encouraging scrutiny, challenge & contribution

Exploit the DATA Bonanza

- **Educate** all about how to use data
 - The **three categories of expert**
 - Data literate managers, governmental officials & ...
 - A data savvy public
 - High quality opinions and decisions based on data
- **Long-term** development of leadership
 - curated data
 - expert teams
 - good engineering needed for *usability* and scalability
- **Balanced investment**
 - from collection to “*final mile*” of information delivery
- **Open data and processes**
 - encouraging scrutiny, challenge & contribution



Why is it hard to help people use data?

- Data represents the world
 - One model unachievable
 - Different models for different niches
 - Varying in precision, frames of reference and terms
- Every user, Every group, Every community, Every organisation
 - is different
 - and *changes*
 - overlapping views and interests limited in scope and time
- Survival in the digital ecosystem
 - depends on capacity to adapt to change
 - diversity
 - relationship with others in your niche

Why is it hard to help people use data?

- Data represents the world
 - One model unachievable
 - Different models for different niches
 - Varying in precision, frames of reference and terms
- Every user, Every group, Every community, Every organisation
 - is different
 - and *changes*
 - overlapping views and interests limited in scope and time
- Survival in the digital ecosystem
 - depends on capacity to adapt to change
 - diversity
 - relationship with others in your niche

**Data is as important
and as difficult as**

**Build on philosophy as
well as mathematics**