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
Researchers different?
Rainbow group, Cambridge
DCS, Edinburgh
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
Data-Intensive Thinking
Data-Intensive Thinking
Two rapidly changing worlds

Computer Science Research

- Efficient distributed systems
- Effective algorithms
- Data-intensive computing

Interdisciplinary Applications

- Reusable computational models
- Collaborative environments
- Intuitive interfaces
- New conceptual models for systems
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

Research is motivated by change and enables change

Digital context evolving rapidly with change driven by business

Endless chaotic source of challenges
Admire Project

• Model for Data Driven
  – science & research
  – engineering
  – business

• Abstraction
  – technical detail

• Longevity
  – as digital context evolves
Admire Project

• Model for Data Driven
  – science & research
  – engineering
  – business

• Abstraction
  – technical detail

• Longevity
  – as digital context evolves

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

- Find & Gather Data
- Understand Data
- Clean & Prepare Data
- Analyse Data
- Present Actionable Information

Form Key Questions

Design Presentation

& develop response capability
Domain Experts

• 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
**Domain Experts**

- **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
**Domain Experts**

- **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

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

**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

**abstraction**
• **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 eInfrastructure independence
  - **Relationship** with data-intensive engineering
Data-Analysis Experts

- **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 eInfrastructure independence
  - **Relationship** with data-intensive engineering

**Jim Gray’s legacy**
- The Fourth Paradigm

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

**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

Jim Gray's legacy
• The Fourth Paradigm

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

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

--
Jim Gray's legacy
• The Fourth Paradigm

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

- Instrumentation
  - Resolution increasing
  - Speed increasing
  - Cost / energy dropping

- Automation
  - laboratories, observatories, businesses, households, ...

- Simulations & Analyses

- Policy & standards
Data Growth

- Storage capacity
  - Kryder’s law
- Instrumentation
  - Resolution increasing
  - Speed increasing
  - Cost / energy dropping
- Automation
  - Laboratories, observatories, businesses, ...
- Simulations & Analyses
- Policy & standards

Faster than Moore

> Moore

> Moore

> Moore

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
EUROPEAN INTEGRATED DATA ARCHIVES (EIDA)

Japan Meteorological Agency

US Incorporated Research Institutions in Seismology

VERCE

Data Intensive Research

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

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

UvA: How do we help people exploit their data? 11 June 2015

2015 “VERCE delivers a productive e-Science environment for seismology research”
VERCE Platform, Components Interaction

Data Management

Data Archives

Science Gateway
Community Applications

Alessandro Spinuso, KNMI
VERCE Platform, Components Interaction

Data Management

Data Management

Data Archives

Science Gateway
Community Applications

Alessandro Spinuso, KNMI
VERCE Platform, Components Interaction

1 - Raw data acquisition
3 - MISFIT

Data Management

Small Clusters

Prov

Science Gateway
Community Applications

Alessandro Spinuso, KNMI
VERCE Platform, Components Interaction

1 - Raw data acquisition
2 - HPC Simulation (model stage-in)
3 - MISFIT

Data Management

Small Clusters

HPC

StgIn

Job Manager

 Prov

Science Gateway
Community Applications

Alessandro Spinuso, KNMI
VERCE Platform, Components Interaction

1 - Raw data acquisition
2 - HPC Simulation (model stage-in)
3 - MISFIT

Data Management

Small Clusters

HPC

StgIn

Job Manager

StgOut

P

Results and provenance management

Metadata and Provenance Archive and Services

Web API

Runtime Provenance messaging

StgIn

data

metadata

StgOut

StgOut

StgOut

Science Gateway Community Applications

Prov

Prov

Alessandro Spinuso, KNMI
VERCE Platform, Components Interaction

1 - Raw data acquisition
2 - HPC Simulation (model stage-in)
3 - MISFIT

Data Management

Data Archives

Small Clusters

HPC

StgIn

Job Manager

StgOut

P

Web API

Results and provenance management

Metadata and Provenance Archive and Services

Interactive Validation and Visualisation throughout the process

Science Gateway Community Applications

prov

Alessandro Spinuso, KNMI
VERCE Platform, Components Interaction

1 - Raw data acquisition
2 - HPC Simulation (model stage-in)
3 - MISFIT

Data Management

HPC

StgIn
Job Manager
StgOut

Prepared for

Small Clusters

Data Archives

Data Management

Web API

Results and provenance management

Metadata and Provenance Archive and Services

Interactive Validation and Visualisation throughout the process

Science Gateway Community Applications

GridFTP
Globus
UNICORE

Prov

OGC
FDSN

Alessandro Spinuso, KNMI
Virtual Earthquake and Seismology Research Community in Europe

Virtual Environment for 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
Virtual Earthquake and Seismology Research Community in Europe

Virtual Environment for 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

Alessandro Spinuso, KNMI
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)

Alessandro Spinuso, KNMI
Human issues!


Nature 14 Sept. 2011

Scientists on trial: At fault?

In 2009, an earthquake devastated the Italian city of L’Aquila and killed more than 300 people. Now, scientists are on trial for manslaughter.

Stephen S. Hall

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; Giuseppina 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
In 2020, volunteers collected more than 9,156,606 hours collecting bird observations.

Prof. Steve Kelling, Cornell Lab of Ornithology
Optos retinopathy diagnosis
Workflows as a DI strategy
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
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
Why use a workflow?

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

Empower the *domain* experts
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

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

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

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

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

### 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