Microsoft

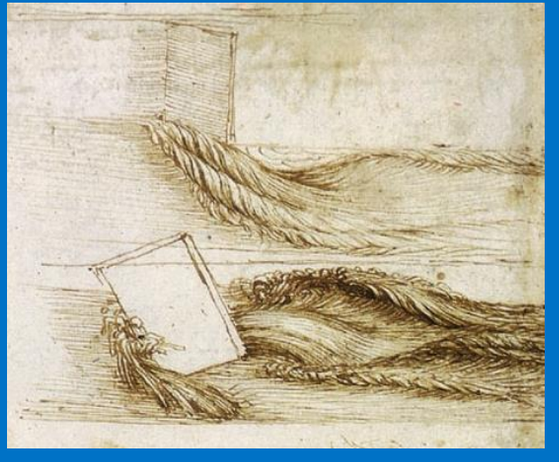
eResearch in Action - Accelerating Knowledge, Enabling Decisions

Dr Kenji Takeda (kenji.takeda@microsoft.com)
Microsoft Research

@azure4research

The Nature of Scientific Discovery

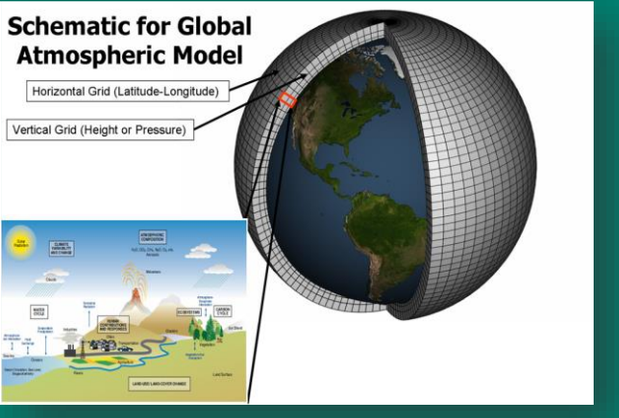
Experiment



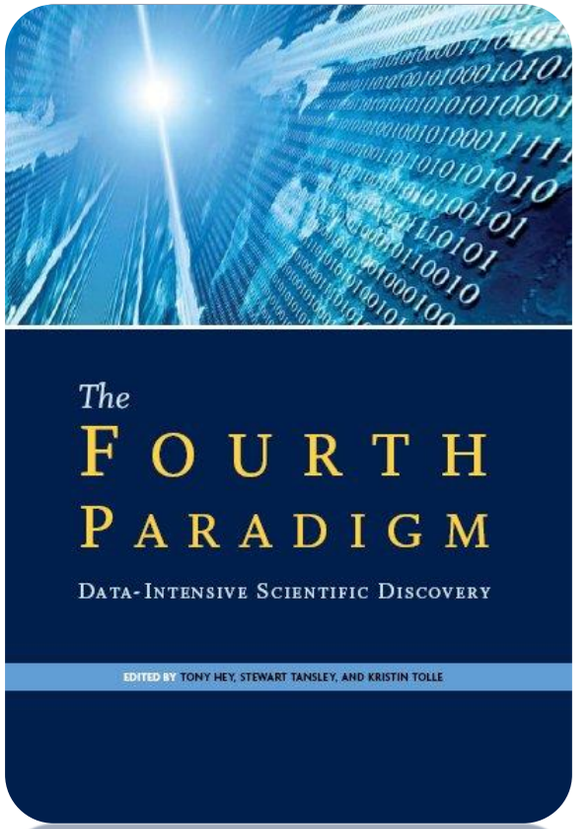
Theory

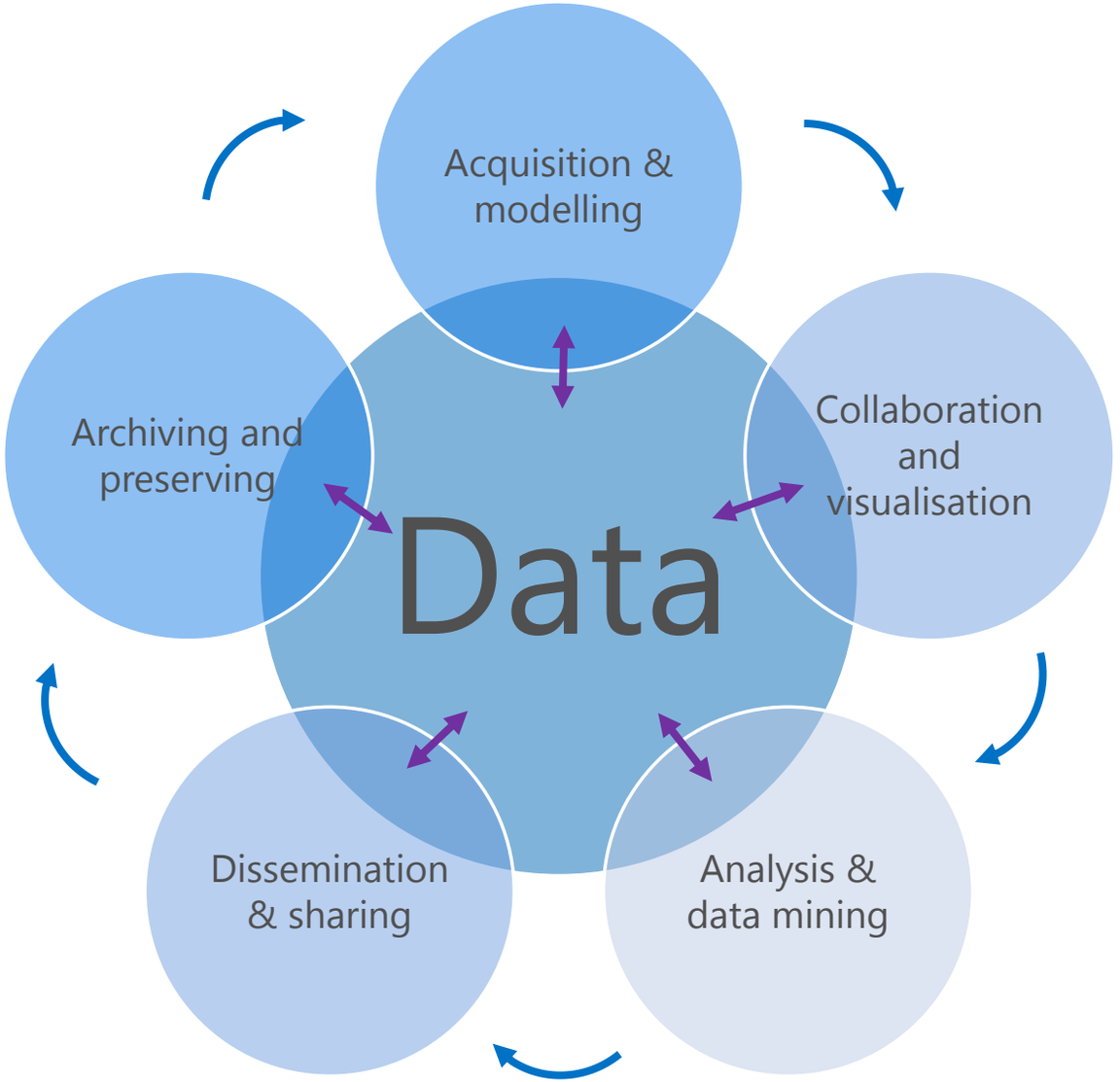
$$\rho \frac{Dv}{Dt} = -\nabla p + \nabla \cdot \boldsymbol{T} + \boldsymbol{f}$$

Computation



Data-intensive Research

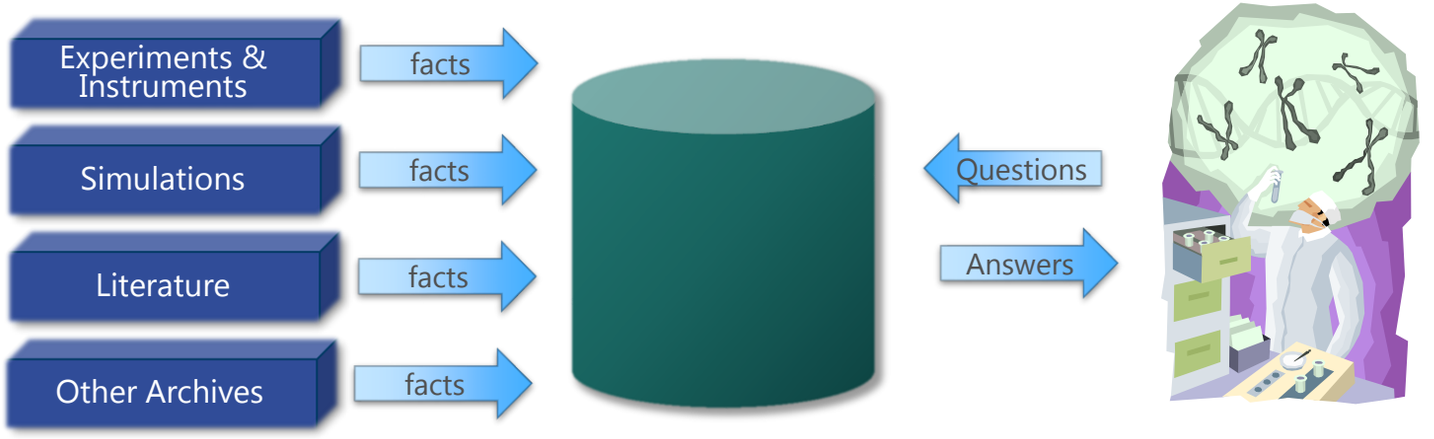




fourthparadigm.org

X-Info

- The evolution of X-Info and Comp-X for each discipline X
- How to codify and represent our knowledge

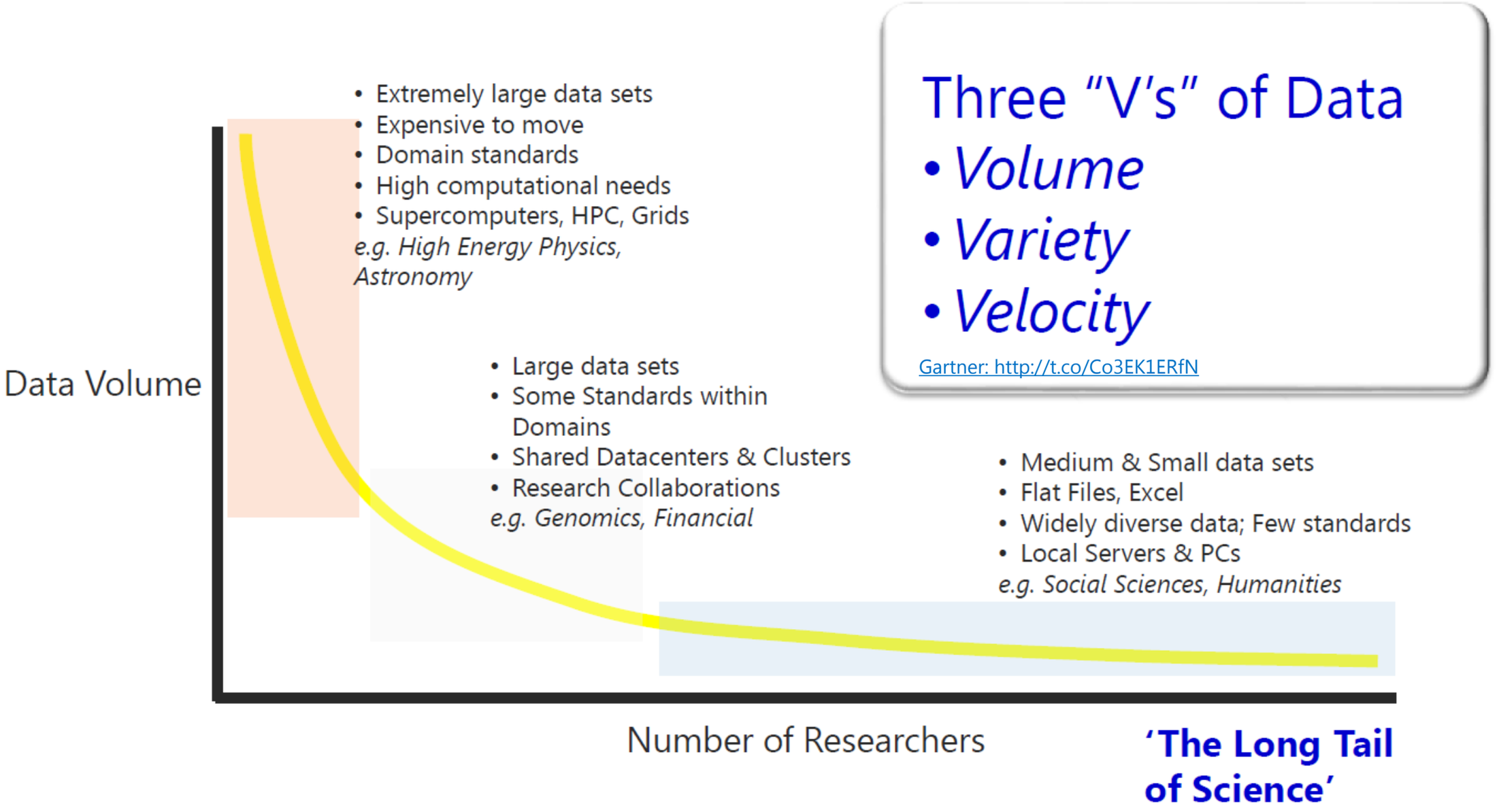


The Generic Problems

- Data ingest
- Managing a ~~petabyte~~ Exabyte
- Common schema
- How to organize it
- How to *re*organize it
- How to share with others
- Query and Vis tools
- Building and executing models
- Integrating data and Literature
- Documenting experiments
- Curation and long-term preservation

Thanks to Jim Gray

Data-Intensive Research & Big Data



Improving information value

Data >>> Information >>> Insight

Monitoring

Collation

Quality assurance

Aggregation

Distribution

Integration

Analysis

Reporting

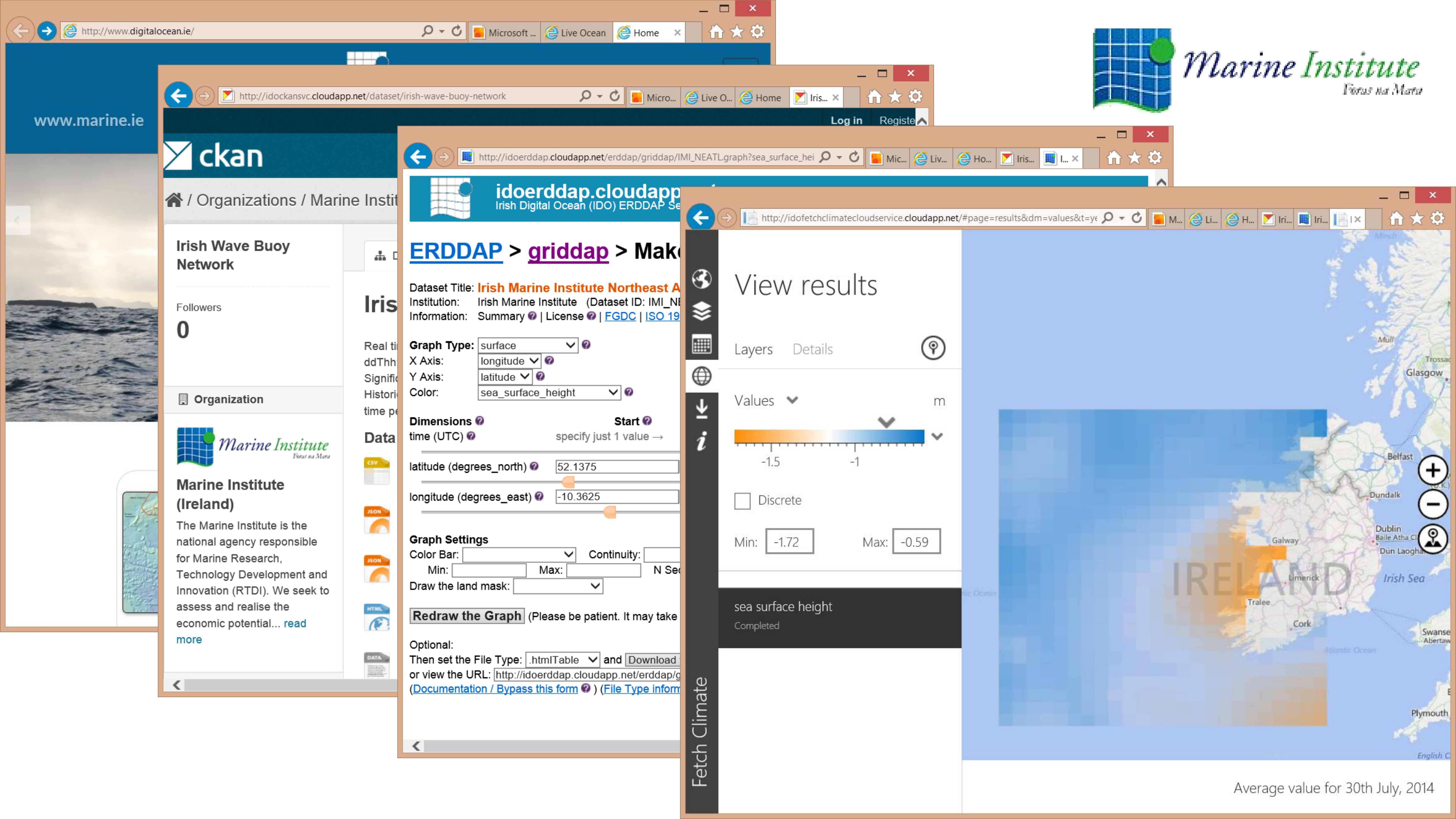
Forecasting

Sometimes done well, generally discoverable and available, but could be improved

Done poorly to moderately, not easy to find

Done poorly, but a few notable counter-examples

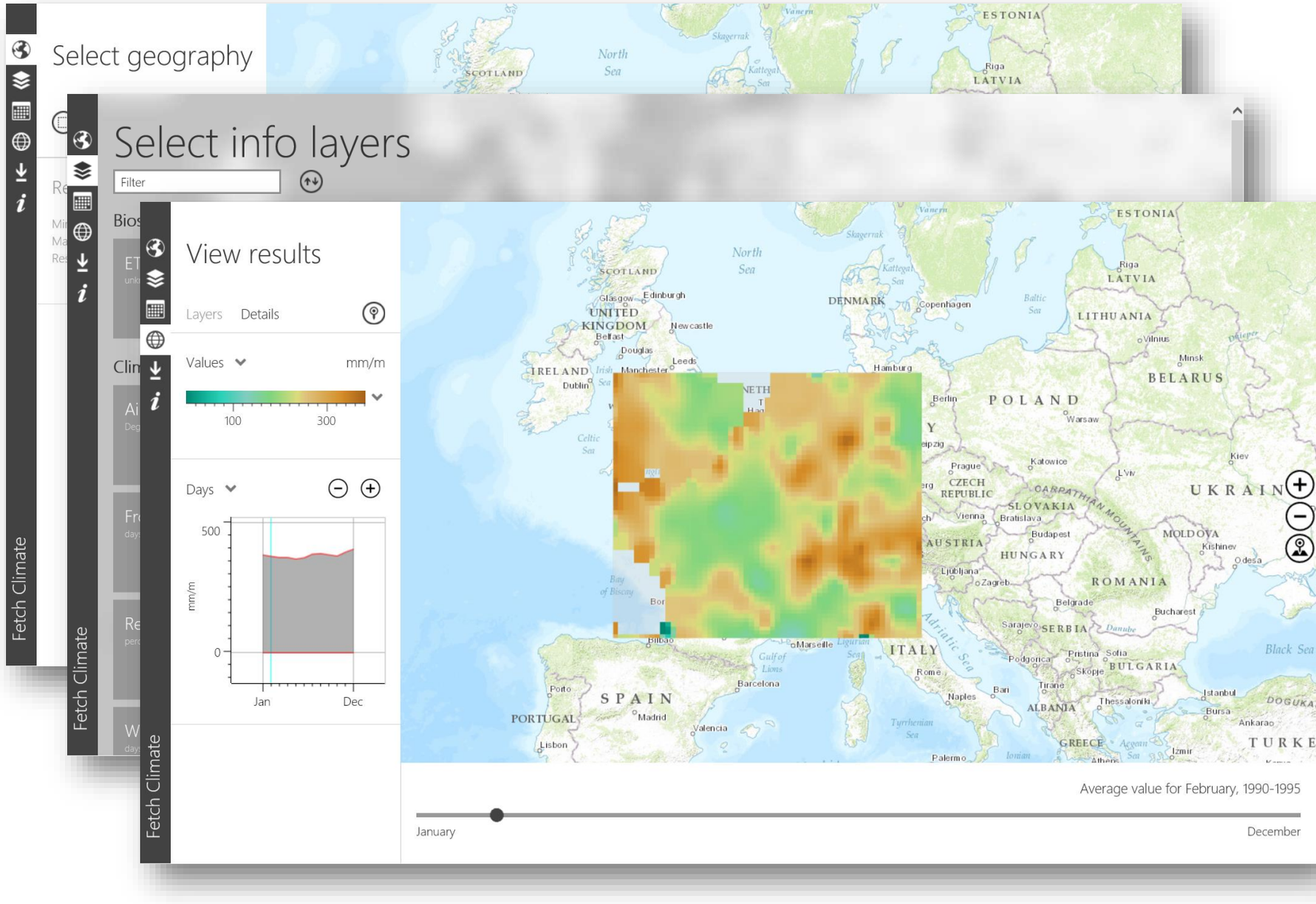
(I. Zaslavsky & CSIRO, BOM, WMO)


Marine Institute
Foras na Mara
http://www.digitalocean.ie/
www.marine.ie
http://idockansvc.cloudapp.net/dataset/irish-wave-buoy-network
ckan
Organizations / Marine Instit
Irish Wave Buoy Network
Followers
0
Organization
Marine Institute (Ireland)
The Marine Institute is the national agency responsible for Marine Research, Technology Development and Innovation (RTDI). We seek to assess and realise the economic potential... read more
http://idoerddap.cloudapp.net/erddap/griddap/IMI_NEATL.graph?sea_surface_hei
Irish Digital Ocean (IDO) ERDDAP Se
ERDDAP > griddap > Make
Dataset Title: Irish Marine Institute Northeast A
Institution: Irish Marine Institute (Dataset ID: IMI_NE
Information: Summary | License | FGDC | ISO 19
Graph Type: surface
X Axis: longitude
Y Axis: latitude
Color: sea_surface_height
Dimensions
Start
time (UTC)
specify just 1 value →
latitude (degrees_north) 52.1375
longitude (degrees_east) -10.3625
Graph Settings
Color Bar:
Continuity:
Min:
Max:
Draw the land mask:
Redraw the Graph (Please be patient. It may take
Optional:
Then set the File Type: .htmlTable and Download
or view the URL: http://idoerddap.cloudapp.net/erddap/g
(Documentation / Bypass this form) (File Type inform
http://idofetchclimatecloudservice.cloudapp.net/#page=results&dm=values&t=ye
Fetch Climate
View results
Layers
Details
Values
m
-1.5
-1
Discrete
Min: -1.72
Max: -0.59
sea surface height
Completed
IRELAND
Glasgow
Belfast
Dundalk
Dublin
Galway
Limerick
Tralee
Cork
Irish Sea
Plymouth
Average value for 30th July, 2014

FetchClimate

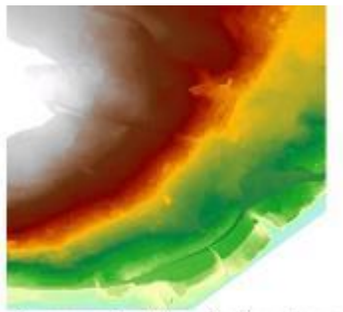
- Intelligent environmental information service
- Automatically:
  - Selects best data source to answer the query
  - Regrids results
  - Calculates uncertainty

www.fetchclimate.org

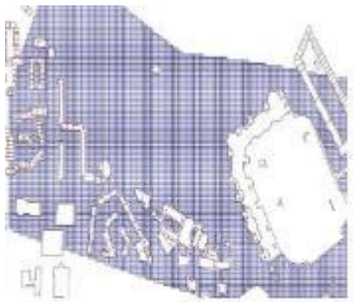


Urban flood risk management

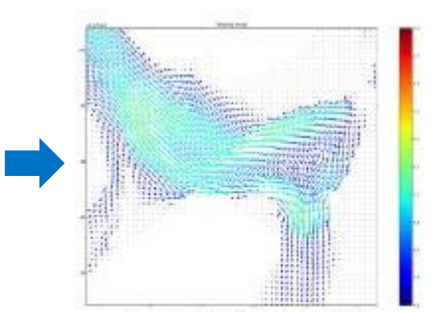
- Improving **decision-support** via Flood Risk Assessment tool
- On-demand CityCAT model running in Microsoft Azure
- Drive FRA studies for mitigation



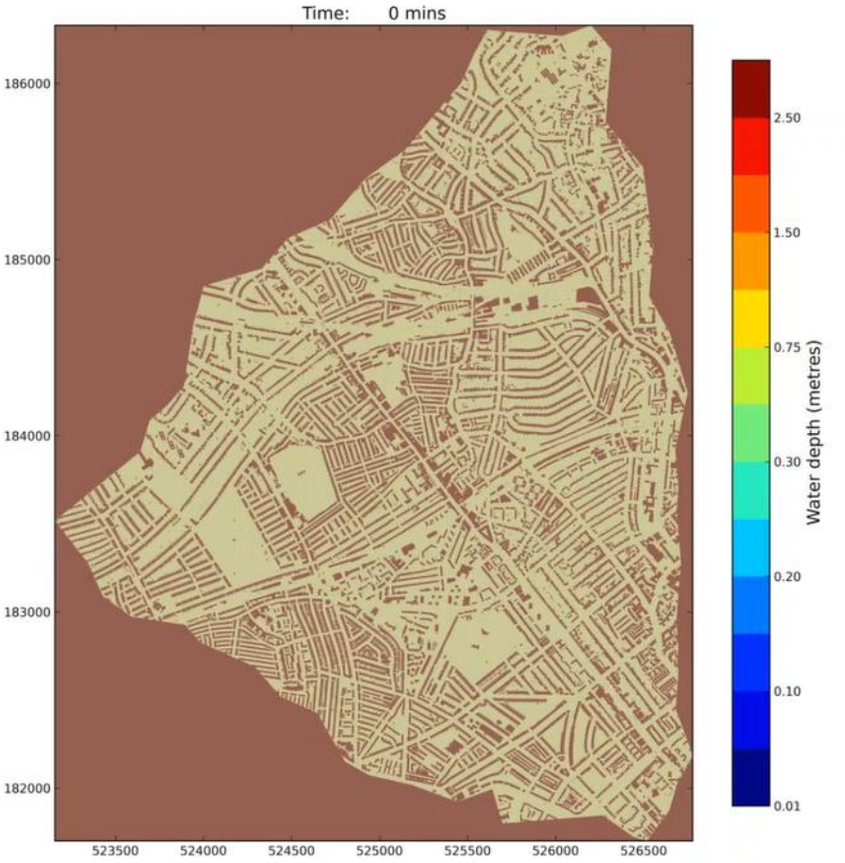
An example of terrain given by a Digital Terrain Model (DTM)



An example of the CityCAT grid without buildings



An output map showing flow velocity field



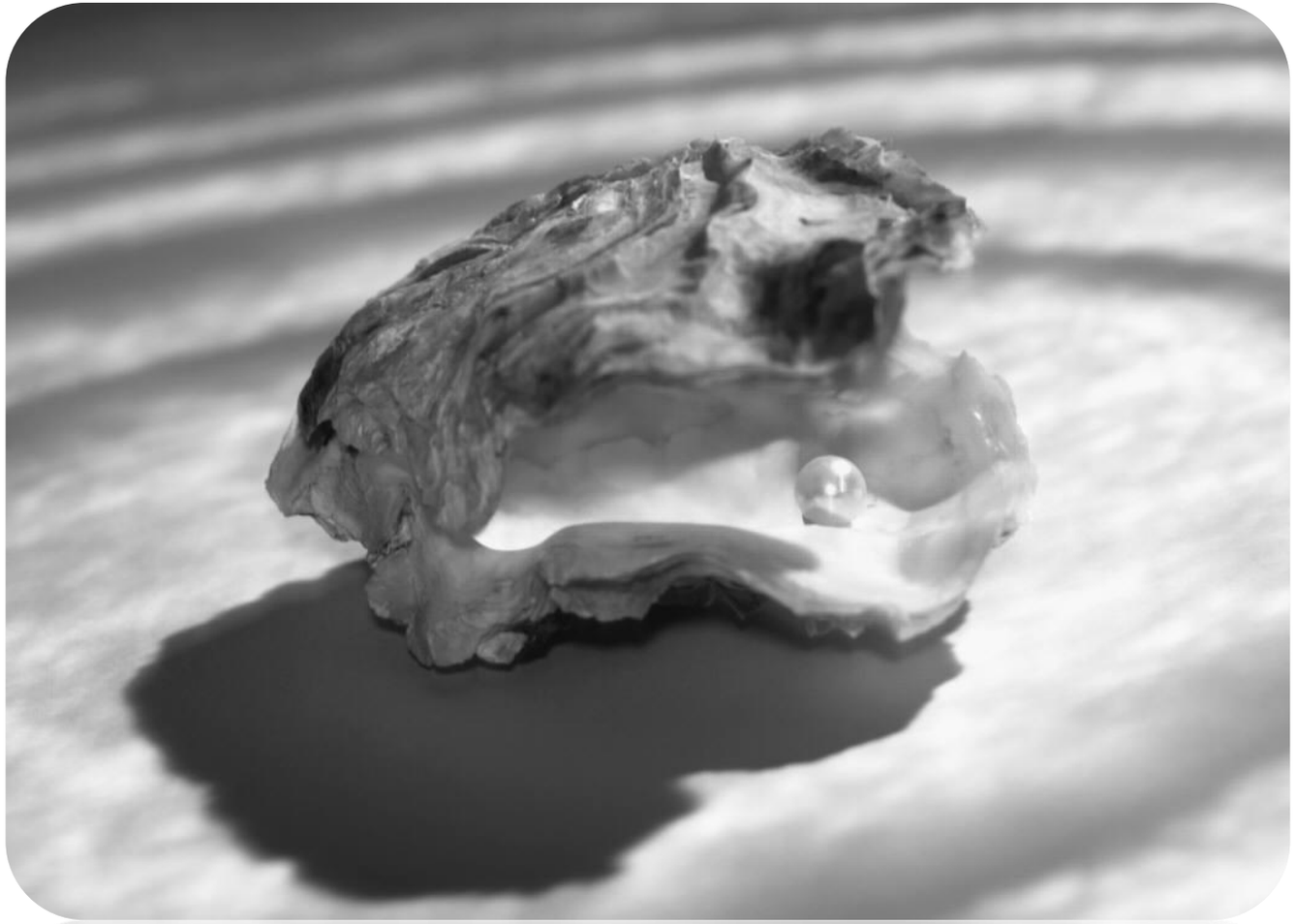
Water depth map of London. Storm event of 60 minutes and 100 years return period

Newcastle University

EPSRC

Microsoft Azure

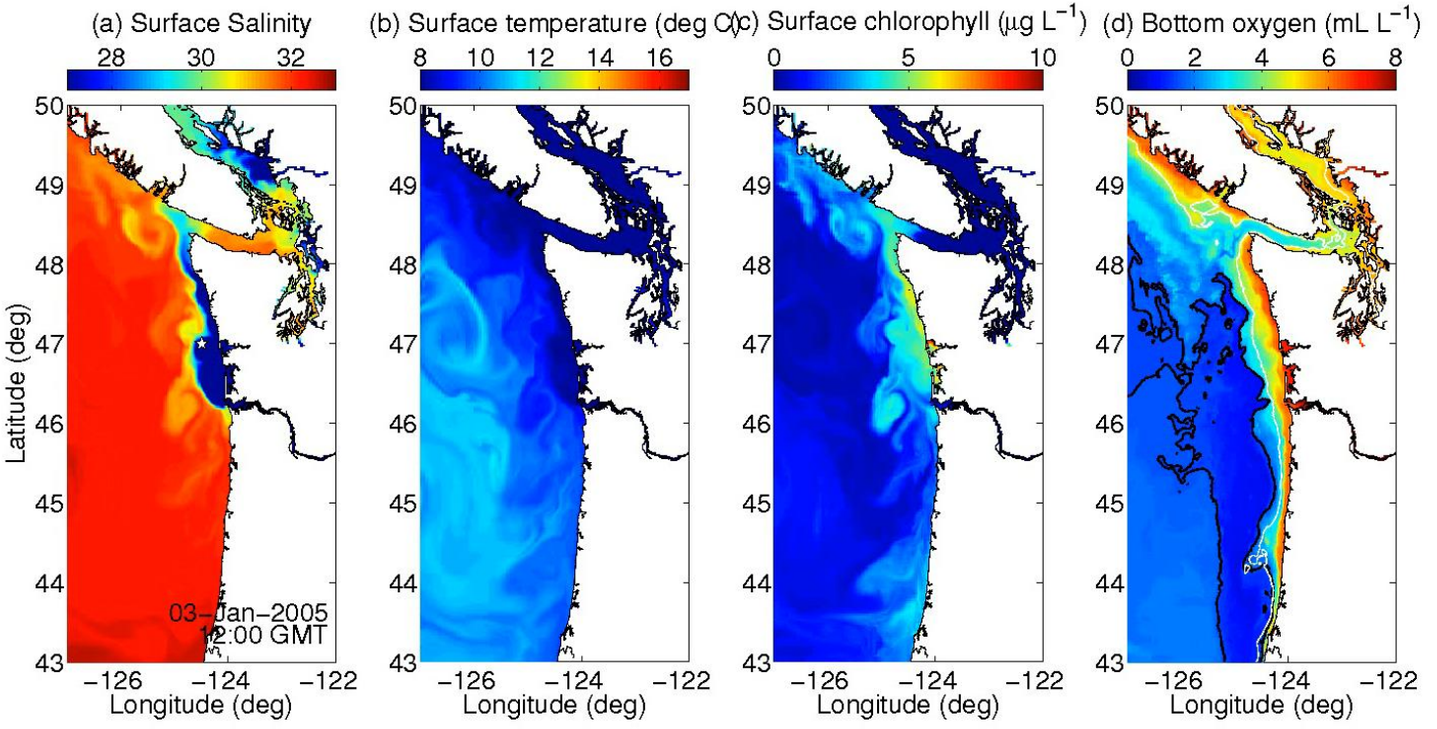
LiveOcean



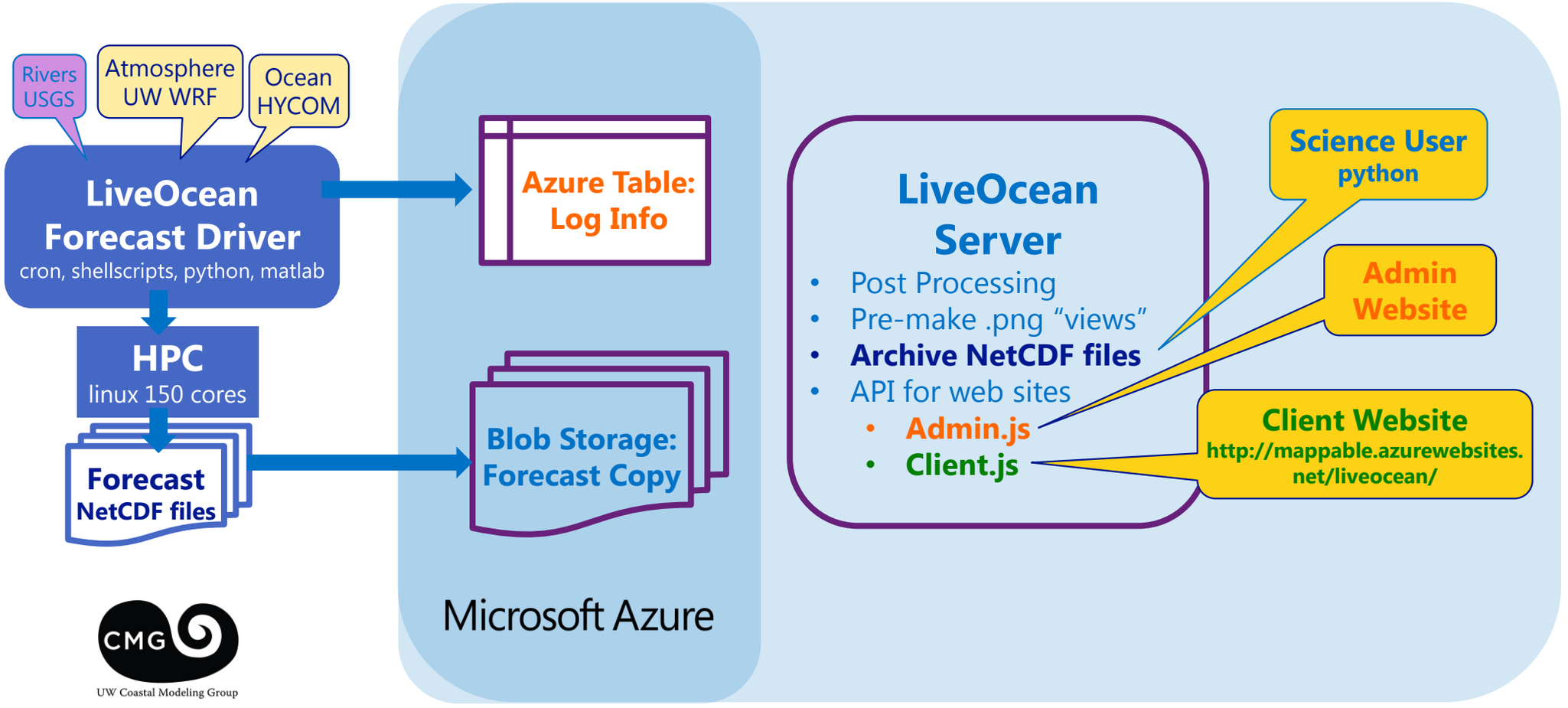
UW Coastal Modeling Group

Parker MacCready: *Univ. of Washington*
Rob Fatland:, Wenming Ye, Nels Oscar, *Microsoft Research*

LiveOcean


(a) Surface Salinity
28 30 32
(b) Surface temperature (deg C)
8 10 12 14 16
(c) Surface chlorophyll (μg L−1)
0 5 10
(d) Bottom oxygen (mL L−1)
0 2 4 6 8
Latitude (deg)
50 49 48 47 46 45 44 43
03–Jan–2005
12:00 GMT
−126 −124 −122
Longitude (deg)
(f) τmaj (N m−2)
W (10−2 m2s−1)
downwelling
upwelling
along–shore winds
0.5 0 −0.5
0 50 100 150 200 250 300 350
year day 2005

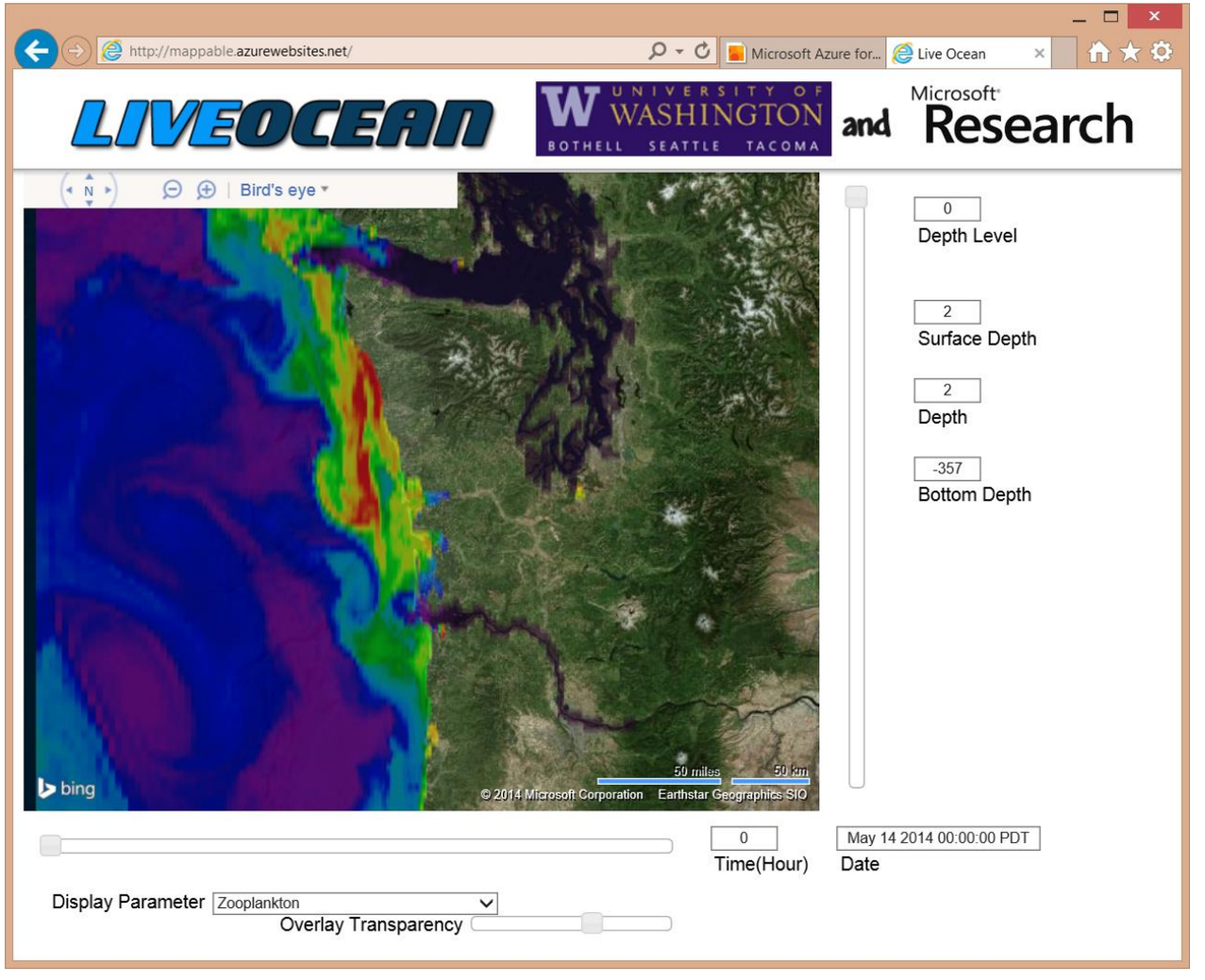
LiveOcean: Hybrid Architecture



LiveOcean on Azure

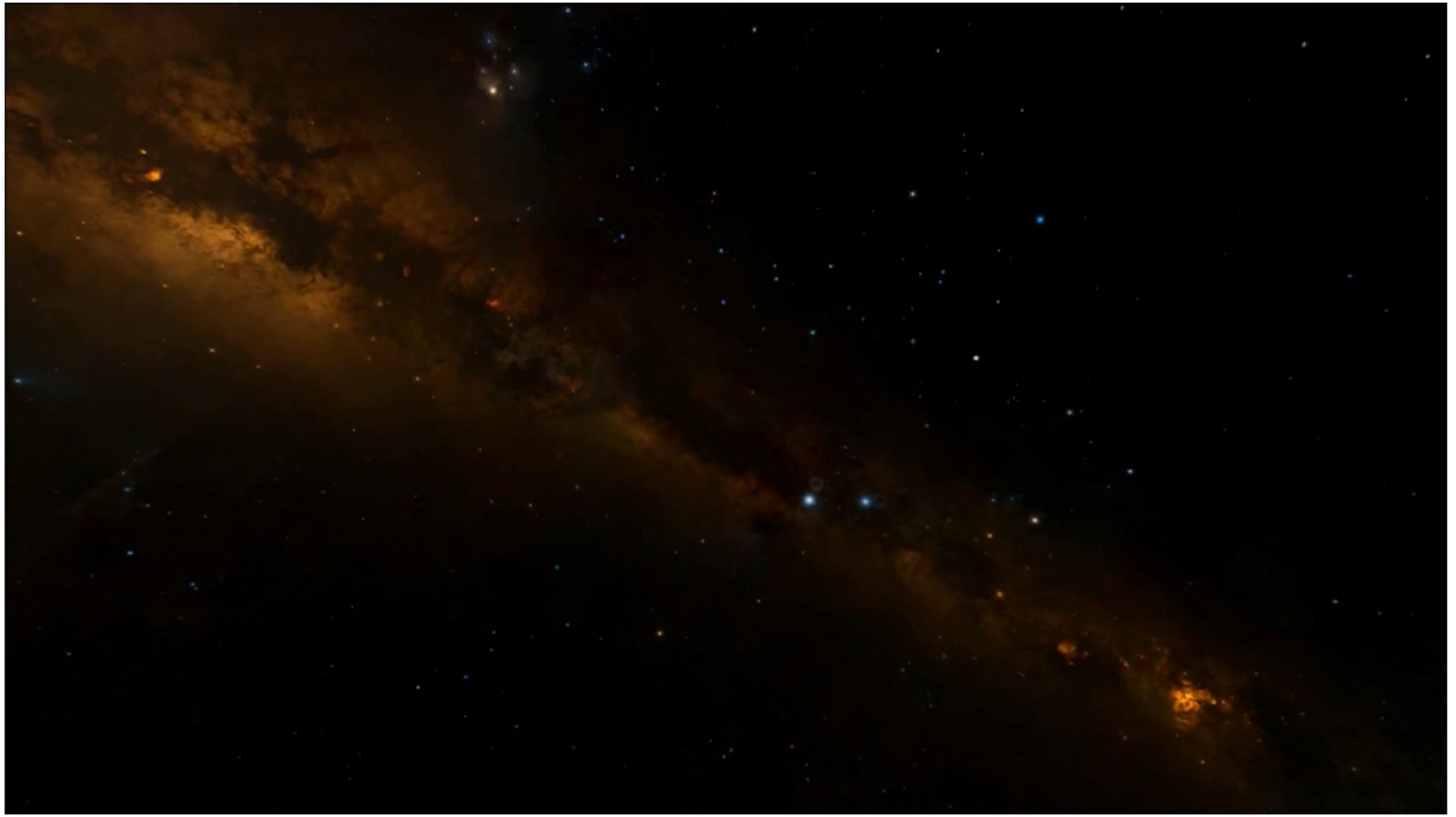
What are advantages of using the cloud?

- Reliable and scalable
- Simplifies: Collaboration with others who work on content delivery (e.g. Client Website)
- Enables: Nesting On-Demand
- Enables: Particle Tracking On-Demand
- Enables: Comparison with Real-Time Observational Data (OOI, IOOS)
- Enables: Comparison with Data from Other Models



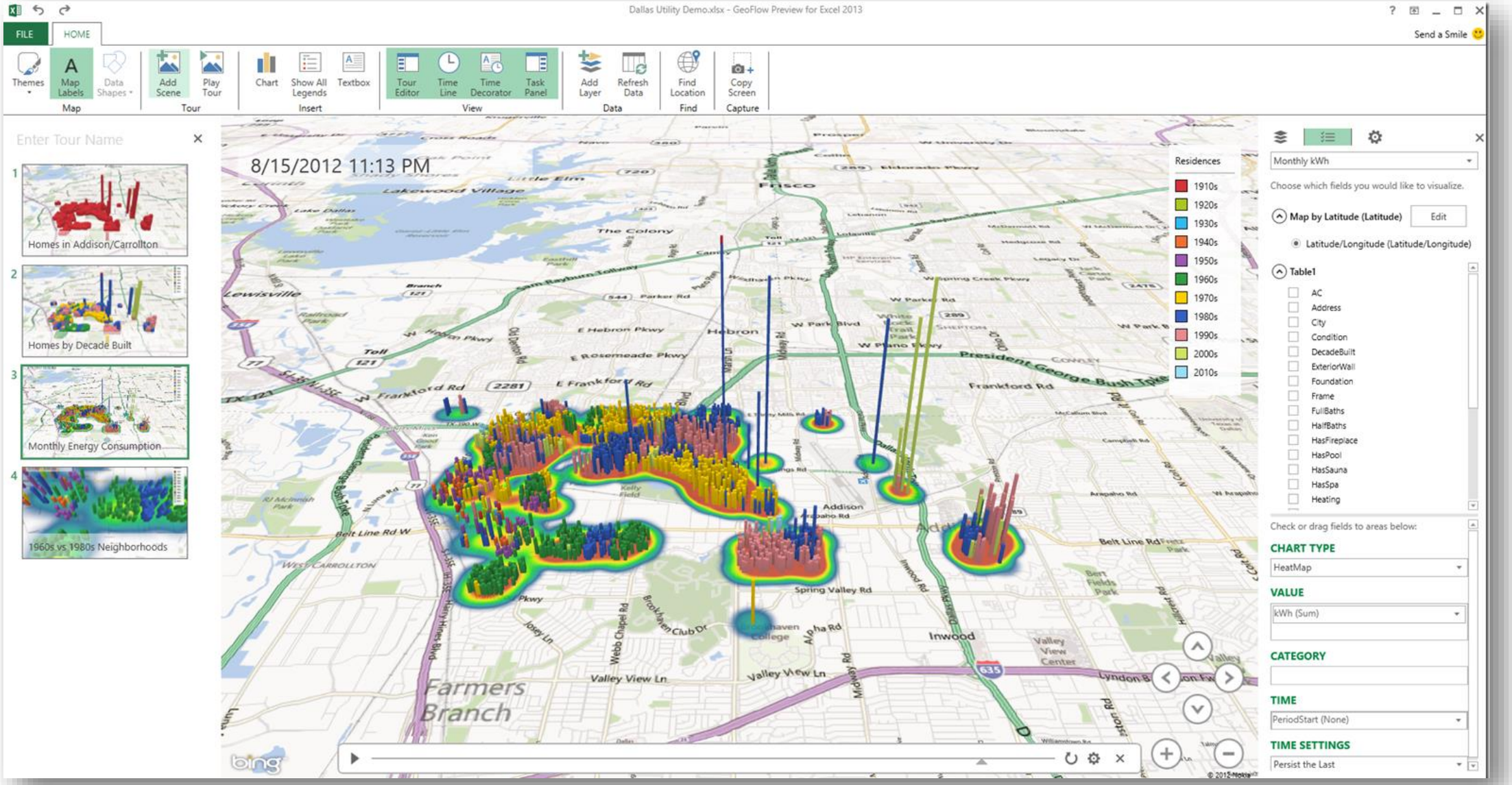
World Wide Telescope

Big data requires new types of data visualization

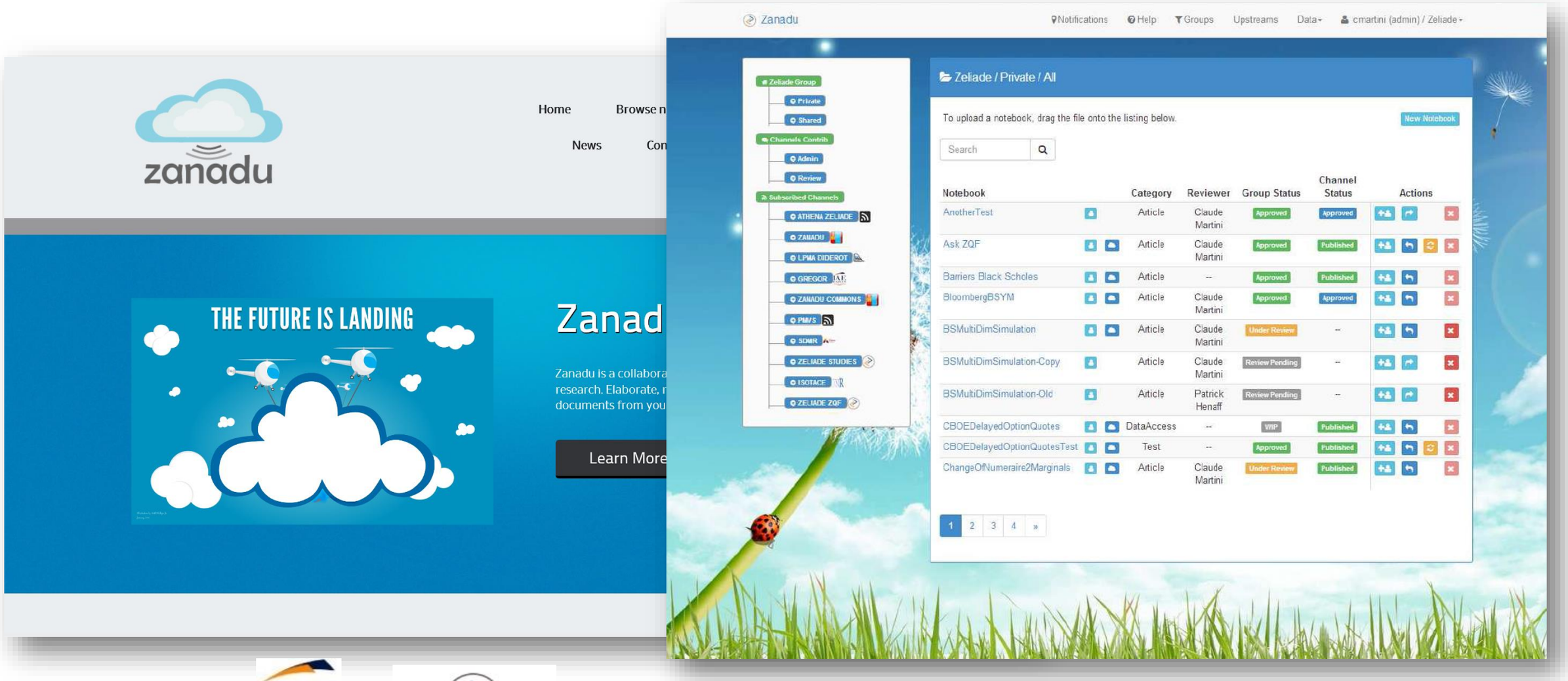


www.worldwidetelescope.org

Excel PowerBI



Reproducible Science

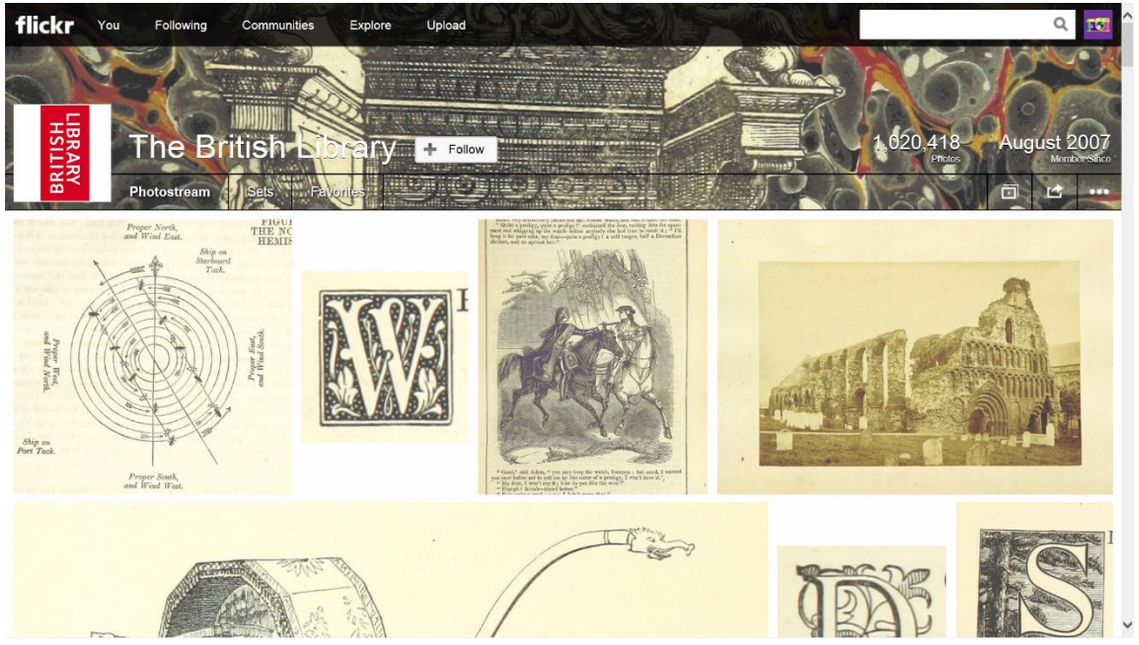


IAE PARIS
PANTHÉON-SORBONNE

Zeliade Systems

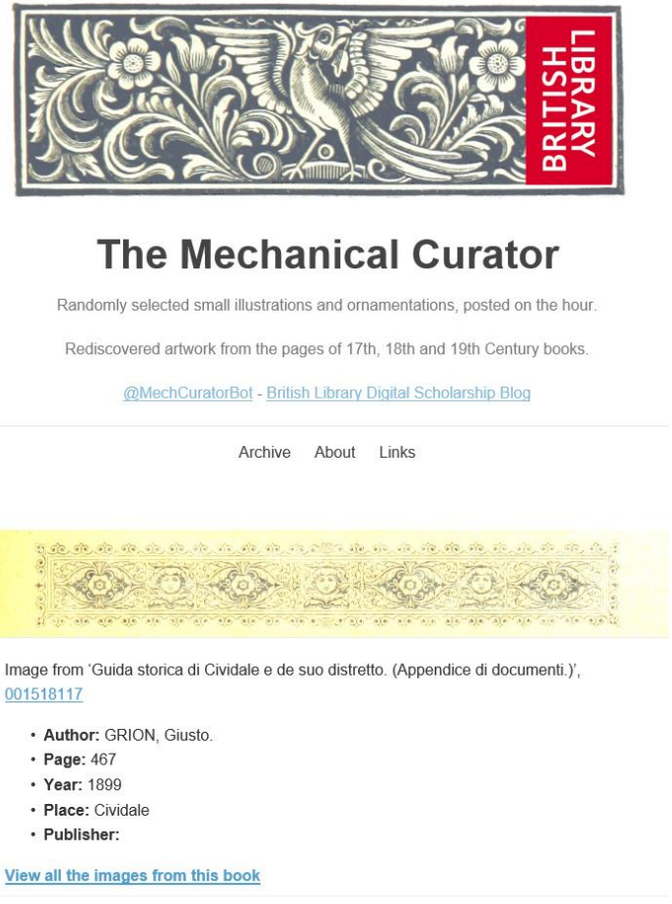
http://zanadu.io

Unlocking Humanities Research



British Library Labs cloud analysis of digital catalogues, including 19th Century books scanned by Microsoft.

@MechCuratorBot



mechanicalcurator.tumblr.com

Microsoft Azure

Opportunity

The Cloud democratizes access to scale & economies of scale


Economies
of Scale

5.8+ billion
worldwide queries each month
bing

250+ million
active users
SkyDrive

400+ million
Active accounts
Outlook.com

2.4+ million
emails per day
Microsoft Exchange Hosted Services

8.6+ trillion
objects in Microsoft Azure storage
Microsoft Azure

48+ million
users in 41 markets
XBOX LIVE

50+ million
active users
Microsoft Office Web Apps

1 in 4
Enterprise customers
Office 365

50+ billion
Minutes of connections handled each day
skype

200+ Cloud Services

1+ billion customers · 20+ million businesses · 90+ markets worldwide

app services

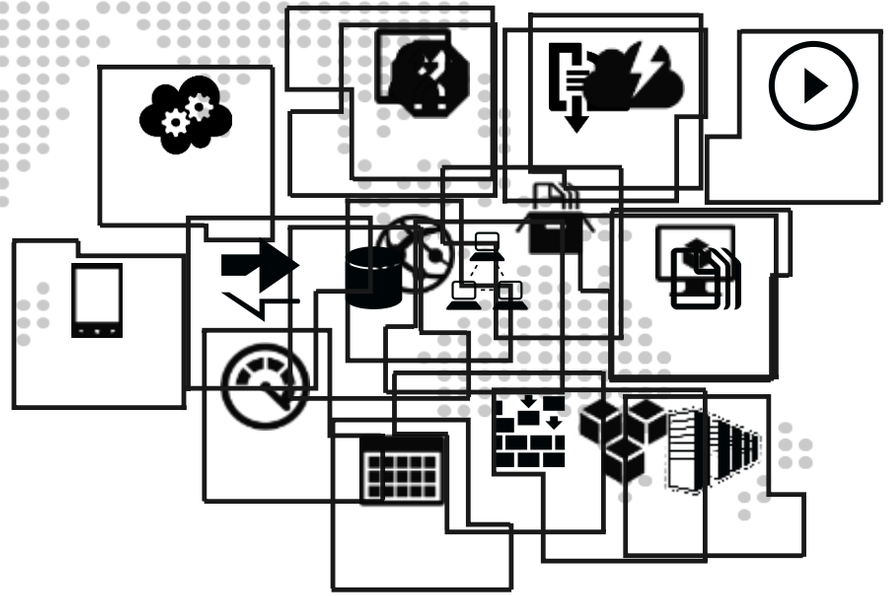
cloud services
caching
identity
service bus
media
mobile services
web sites
integration
hpc
analytics

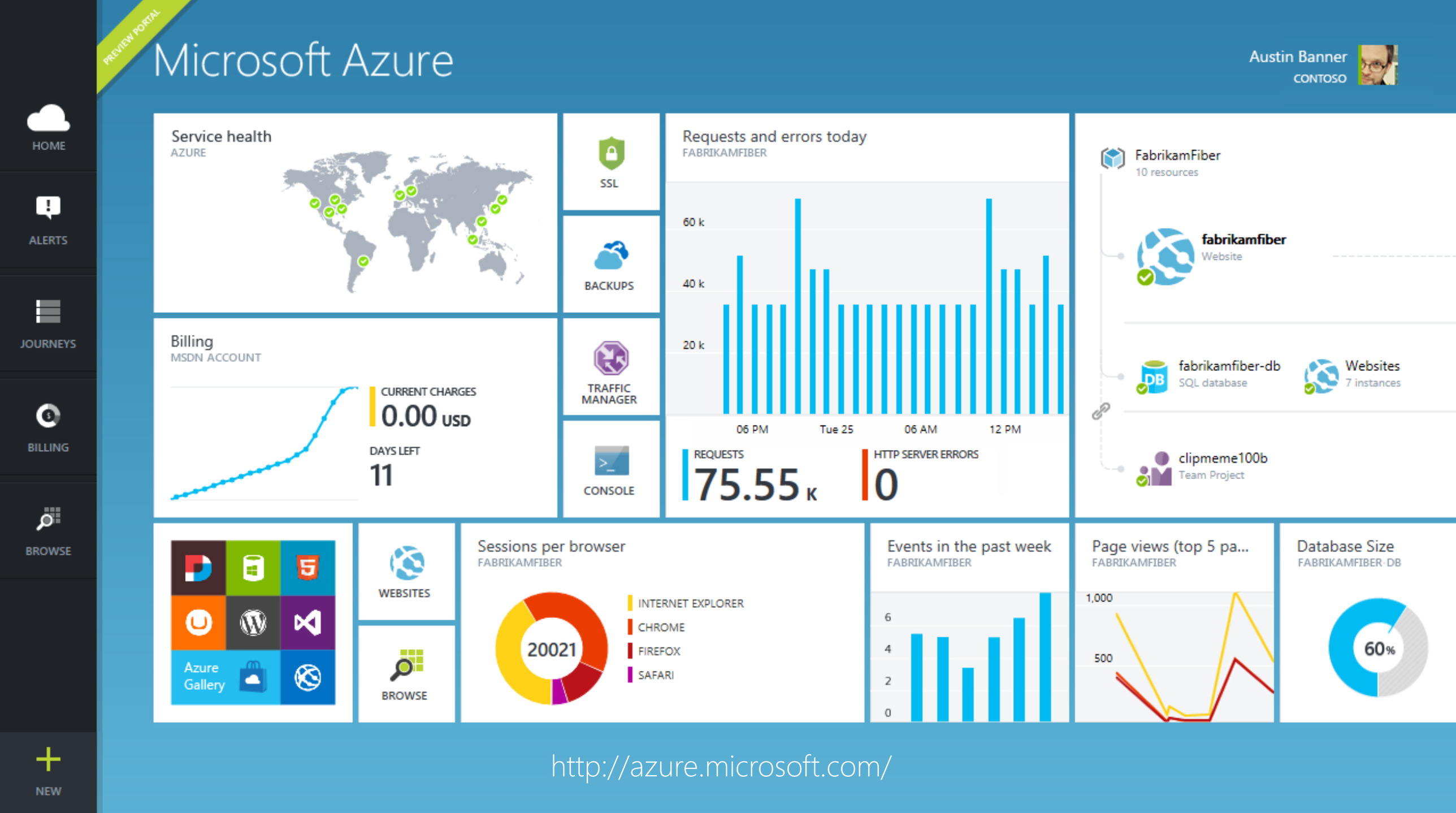
data services

SQL database
HDInsight
table
blob storage

infrastructure services

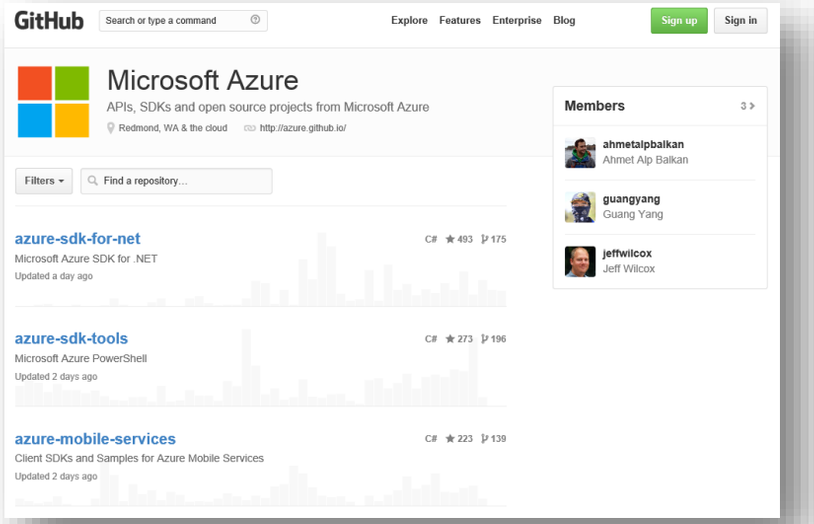
virtual machines
virtual network
vpn
traffic manager
cdn

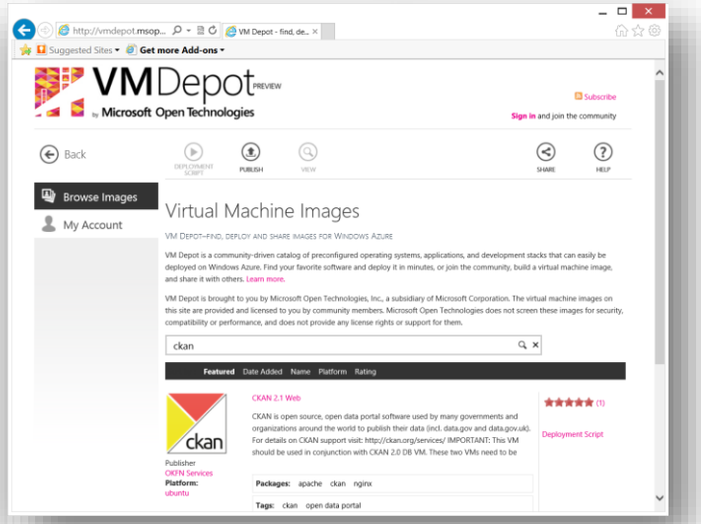




http://azure.microsoft.com/

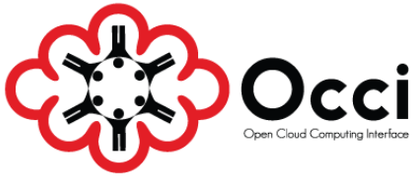
Open and Flexible





European Grid Infrastructure

OpenNebula.org
Flexible Enterprise Cloud Made Simple

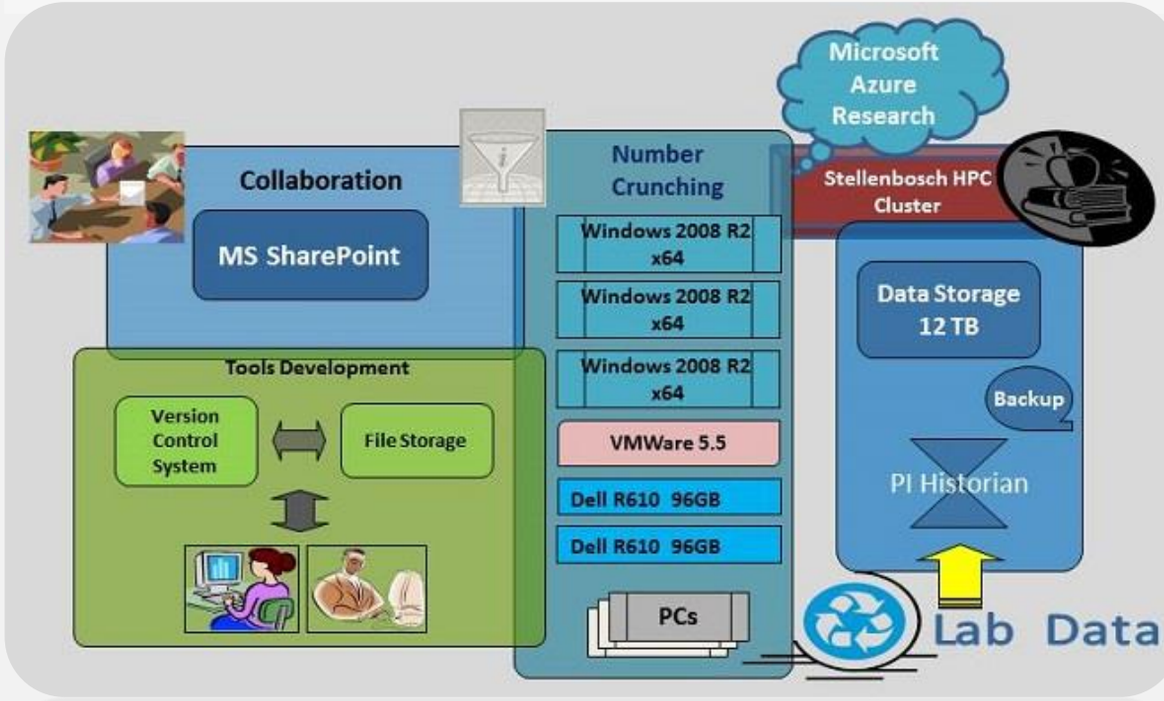


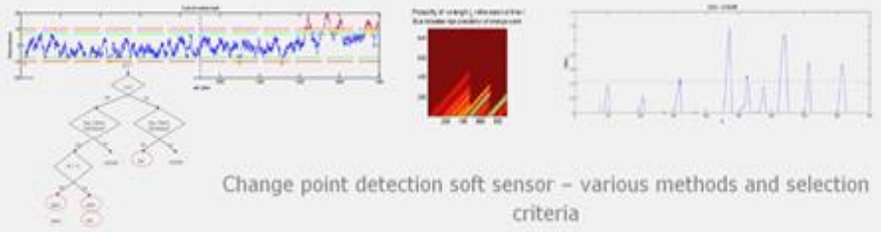
Department of Process Engineering

Home to Chemical Engineering and Mineral Processing at Stellenbosch University

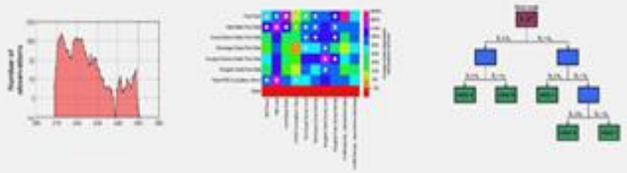
English

Process Monitoring and Systems

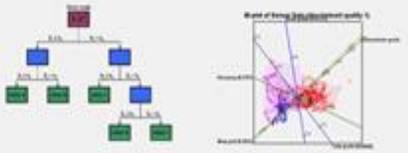




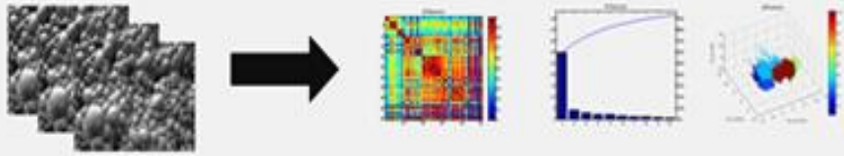
Change point detection soft sensor – various methods and selection criteria



Milling circuit dependencies

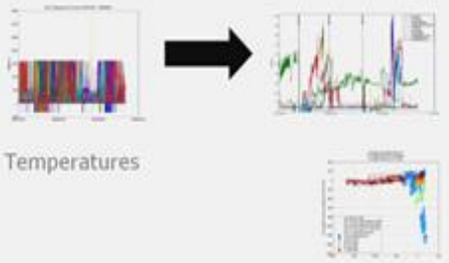


Variable importance soft sensor – various methods



Froth images

Operational states



Temperatures

Thermal Condition Soft Sensor

Scalability on-demand

A-series

- 1-16 cores
- 0.75-112GB RAM
- 20-605 GB HDD
- Up to InfiniBand 40Gbit/s RDMA network (MPI)

D-series

- 1-16 cores
- 3.5-112 GB RAM
- Up to 800GB SSD

G-series

- 32 cores
- 468 GB RAM
- 6.5 TB SSD

hadoop

Spark

APACHE STORM™
Distributed • Resilient • Real-time

python™

Genomics at scale

David Heckerman,
Microsoft Research

What next?

Research-as-a-Service

RaaS

SaaS

PaaS

IaaS

Cloud Services

Research collaboration and data lifecycle services

Data management, application services, collaboration tools.

Programming abstractions, database support, runtime systems

Virtual machines, reliable storage, provisioning tools, network bandwidth

Research Marketplace

Analytics services and expert consulting

Domain specific applications and data access

Advanced development tools and libraries to SaaS developers

Specially configured virtual machine templates

Democratising Research

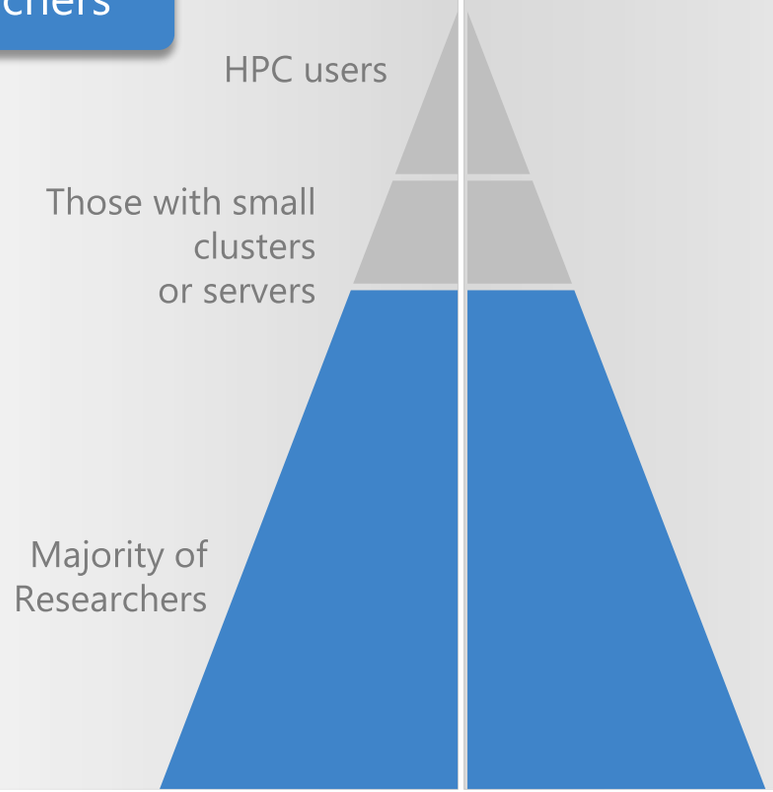
Today

Majority of Researchers

Use laptops & desktop computers

Overwhelmed by data

Finding analysis ever more difficult; sharing even harder



www.azure4research.com

Democratising Research

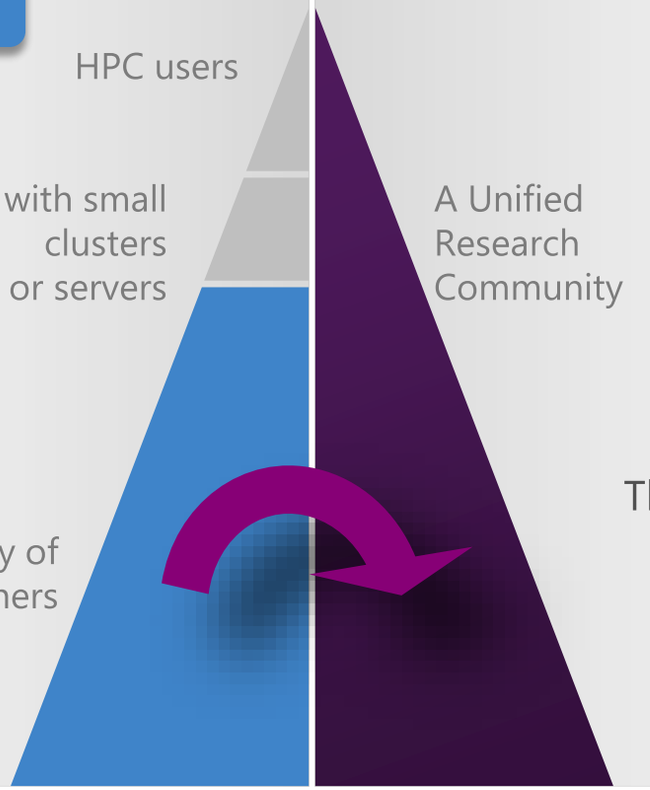
Today

Majority of Researchers

Use laptops & desktop computers

Overwhelmed by data

Finding analysis ever more difficult; sharing even harder



Tomorrow?

Paradigm Shift

Powerful tools

Data and analysis tools in the cloud
Cycles, storage, support

Building communities around research results

The ability to marshal needed resources on demand
Without caring or knowing how it gets done…

Accelerating discovery

www.azure4research.com

Azure for Research Awards

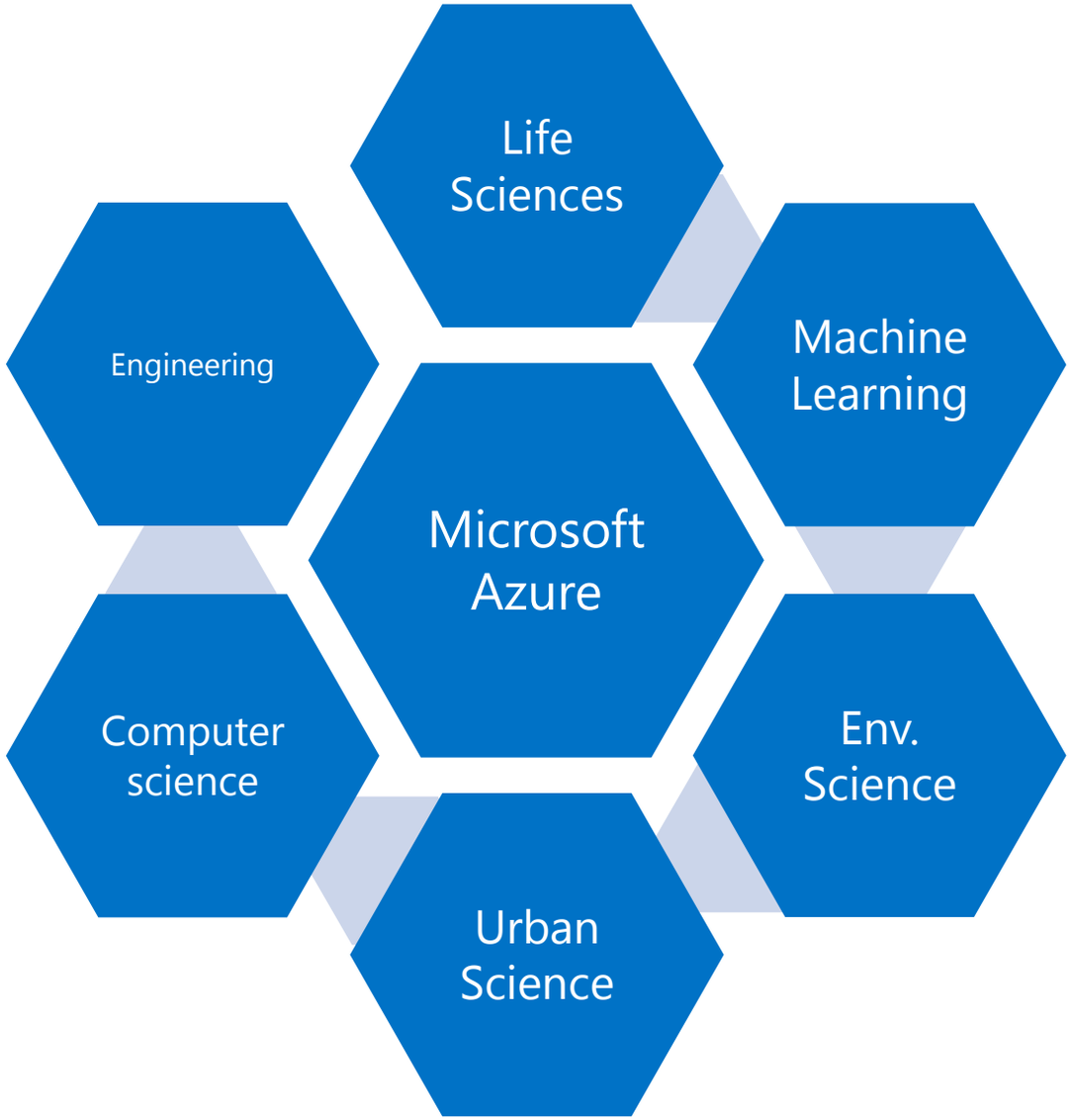
>350
Worldwide

15 Dec 2014
every 2 months

200k
core hrs

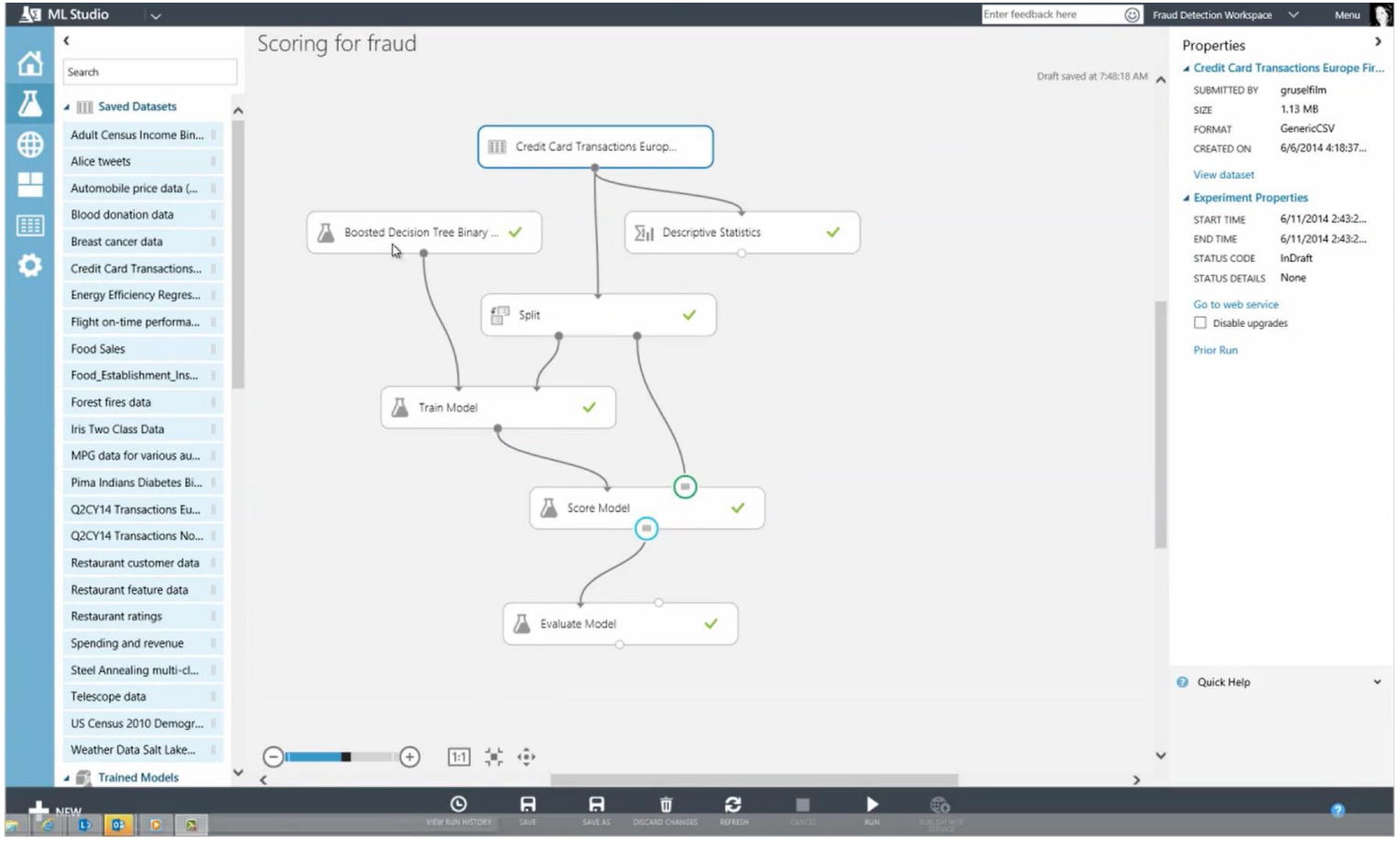
20TB
storage

Azure4Research.com



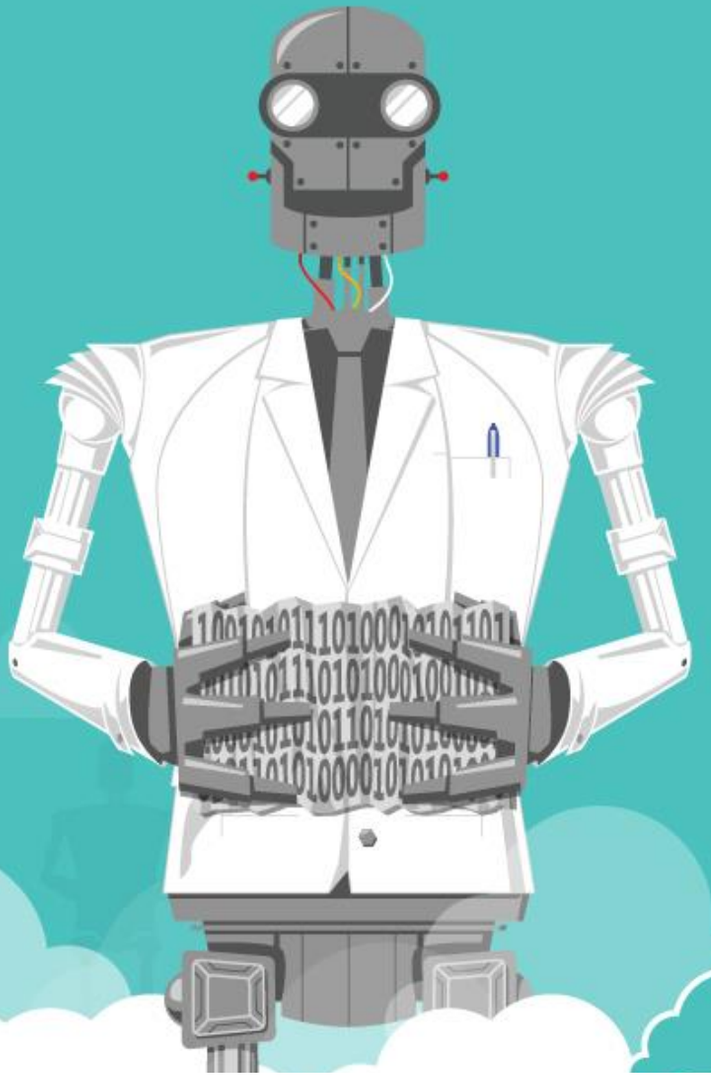
Azure Machine Learning

Tutorial, Friday 28 Nov, http://aka.ms/azure4researchml-za



Microsoft Azure for Research

- Azure Research Awards
- Azure for Research Training
  - Azure (VMs, Storage, Big Data) - Thu 27 Nov - http://aka.ms/azure4researchza
  - Machine Learning – Fri 28 Nov - http://aka.ms/azure4researchml-za
- Technical resources & curriculum



Microsoft Azure for Research Group

@azure4research & *#azureresearch*

www.azure4research.com

Thank you

kenji.takeda@microsoft.com

http://research.microsoft.com
http://azure4research.com