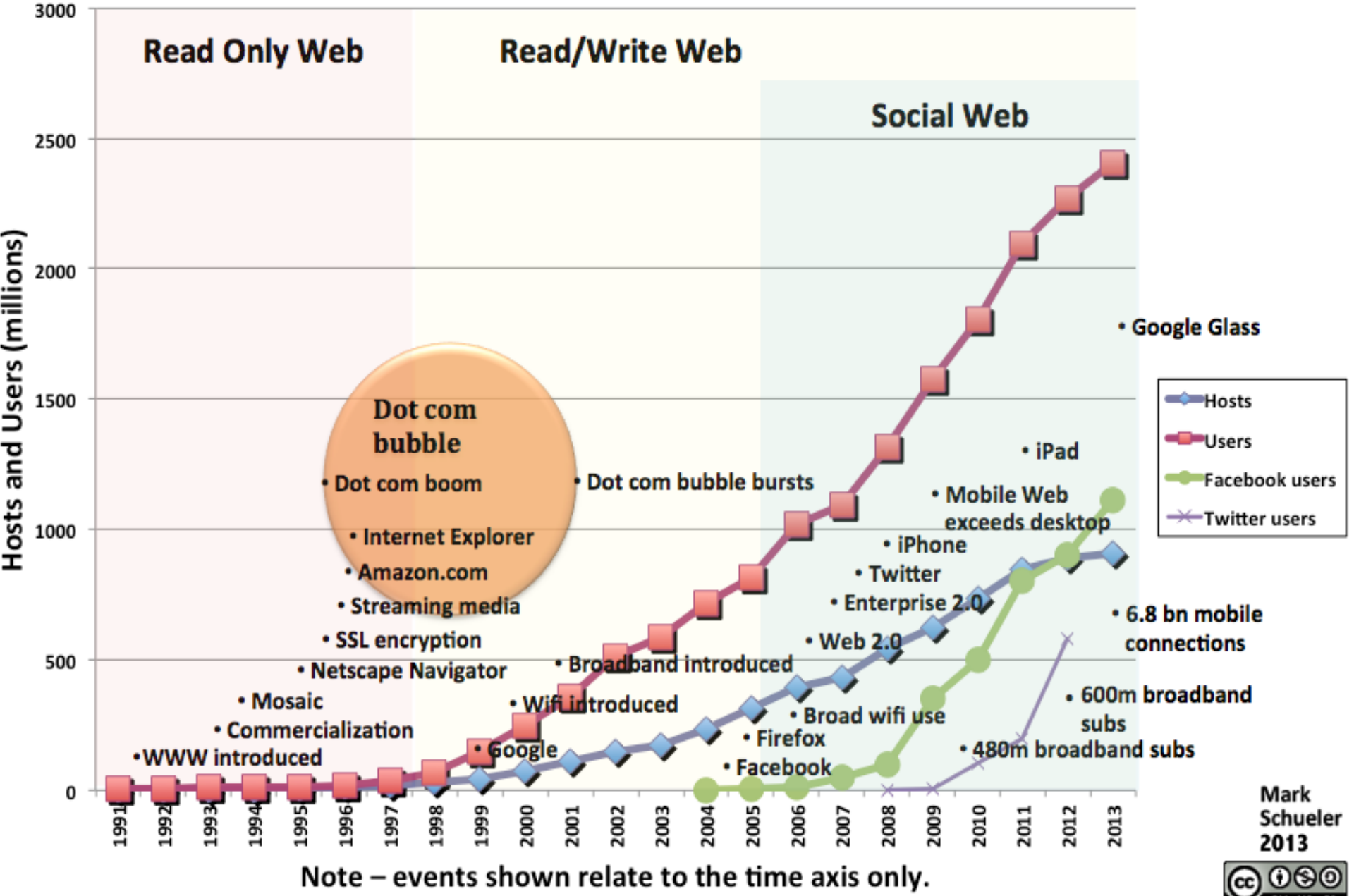
UNIVERSITY OF Southampton

Observing the Web

Professor Dame Wendy Hall
University of Southampton
15th September 2014

@DameWendyDBE

Internet Growth - Usage Phases - Tech Events



Note – events shown relate to the time axis only.

Mark Schueler 2013

UNIVERSITY OF Southampton

The Web as a Social Machine

Tim Berners-Lee didn't create the Web

It is socio-technical

It is co-constitued by technologists, machines and society

Web Science is the theory and practice of social machines?

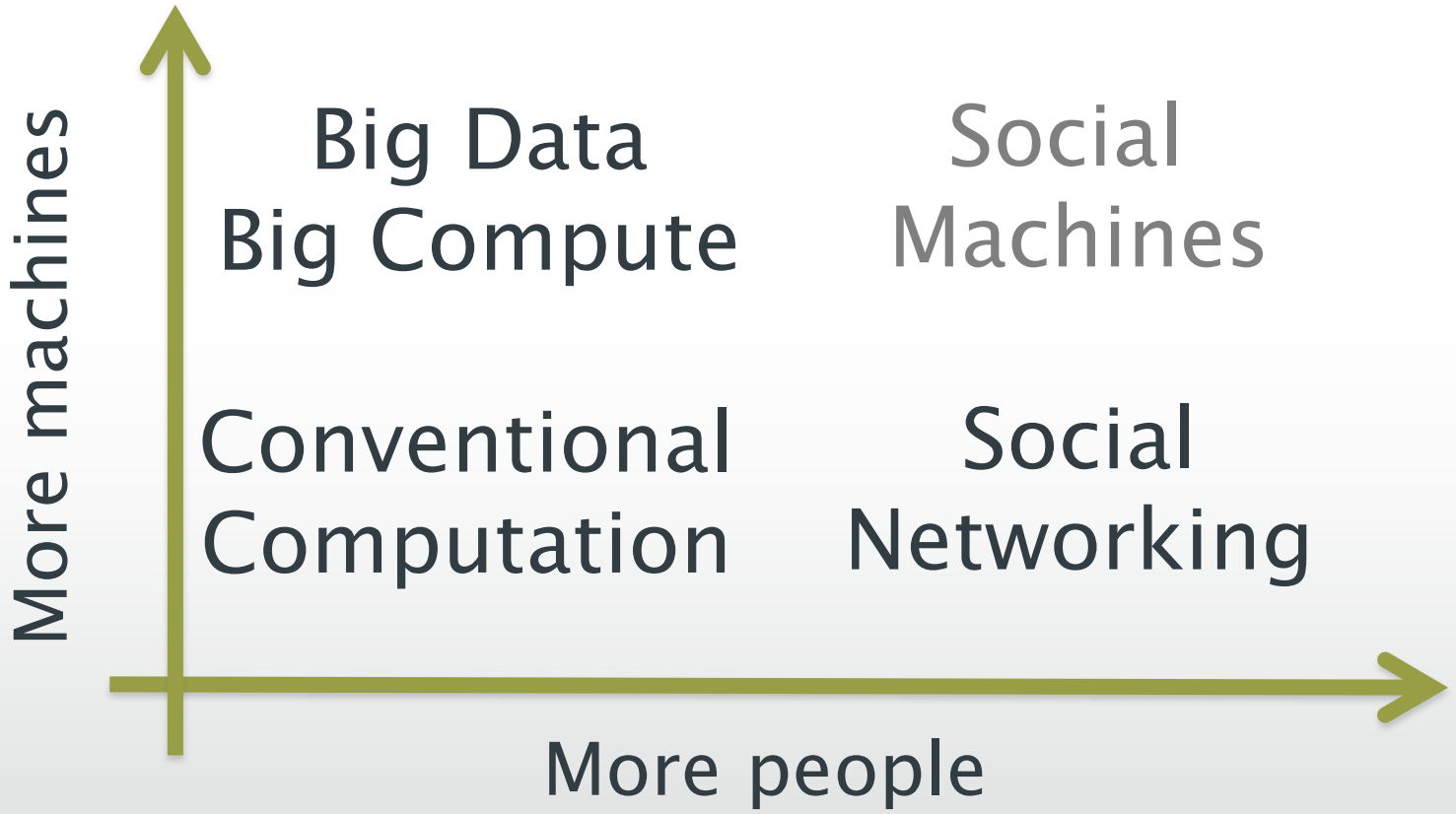
“Real life must be full of all kinds of social constraints – the very processes from which society arises. Computers help if we can use them to create abstract social machines on the Web, processes in which people do the creativity and the machine does the administration. The stage is set for an evolutionary growth of new social engines. The ability to create new forms of social process would be given to the world at large and development would be rapid”

Tim Berners-Lee and Mark Fischetti, Weaving the Web, 1999

Examples of social machines

- The Web, Google, Facebook, Twitter, Wikipedia ……
- Captcha (Luis von Ahn)
- Trip Advisor
- Galaxy Zoo
- Ushahdi – open source project which allows users to crowd source crisis information to be sent via mobile
- The OpenStreet Map of Haiti created after the earthquake
- The list goes on … Amazon, e-Bay, YouTube ….design your own

Social Machines in Context



SOCIAM

The Theory and Practice of Social Machines

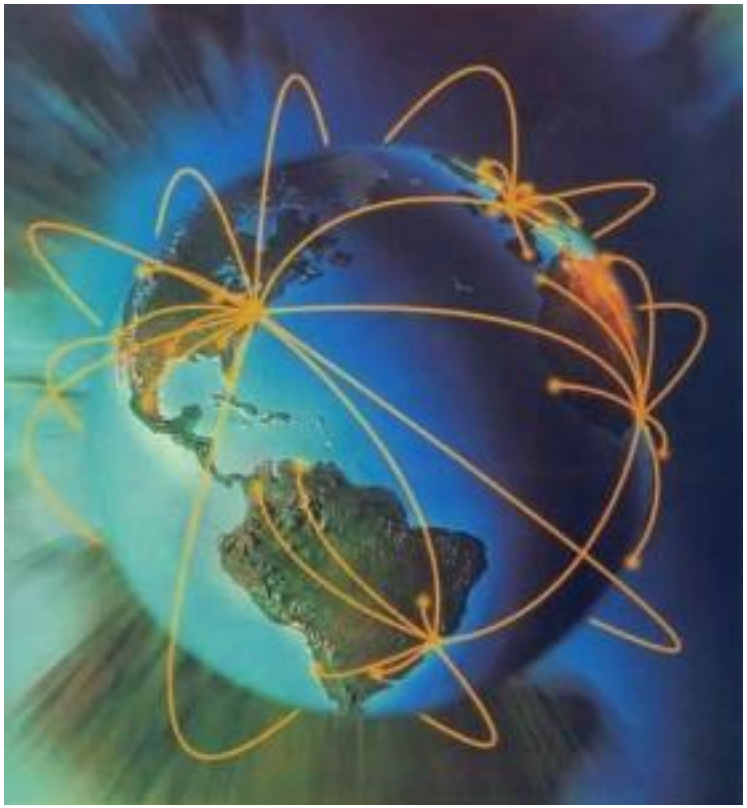
Nigel Shadbolt

Social Machines are NOT Turing Machines

- they do contain conventional algorithmic components but much else is different
- a social machine will start with an incomplete specification that grows and evolves to cover more of the problem via interaction
- a social machine achieves participation through local incentives which become reinforced as the…
- incentive for an individual to supply data to the algorithm increases as more individuals participate
- a social machine has a notion of completeness that is a social rather than mathematical issue
- a social machine will not usually have a notion of the correct output or termination… rather it runs continuously

What will SOCIAM do
Theme 6 Web Observatory

- Understand Social Machines through an **observatory** that observes, monitors and classifies social machines - both those of the project and more widely on the Web - as they evolve;
- it will also act as an early warning facility for new disruptive social machines elsewhere on the Web;
- to understand how Social Machines reach tipping points, longitudinal observational data will reveal how they grow once launched;
- whether they coalesce into larger machines or fragment into micro machines that still have utility;
- what signals need to be observed, what is a fair and faithful sample of Web behaviour;
- this is likely to call attention to appropriate governance, ethical and legal issues.



Physical and
Applied Sciences

UNIVERSITY OF
Southampton

The Web Observatory

Tiropanis, Hall, Shadbolt, DeRoure, Contractor & Hendler

"The Web Science Observatory"
IEEE Intelligent Systems, May 2013

Web Science across continents

- Astronomers obtain a very high resolution picture of the sky from small telescopes a long distance apart.
- Many labs, contributing across the globe, help build an accurate picture of human activity at planetary scale.
  - *transcending parochial social, political, economic, legal interpretations*

Web Observatory: Global partnerships

- Partners contribute their insight and experience, and benefit from the network and business intelligence insights
- Observatory events are hosted bringing together thought leaders to learn from each other – next one in LA in March – planning a summit later in 2014
- Data sets, **open or closed**, can be shared under t's and c's
- Analytics and tool sets can be contributed
- Joint research and projects can be agreed
- This enables longitudinal research

understanding web evolution:

- observation
- experimentation

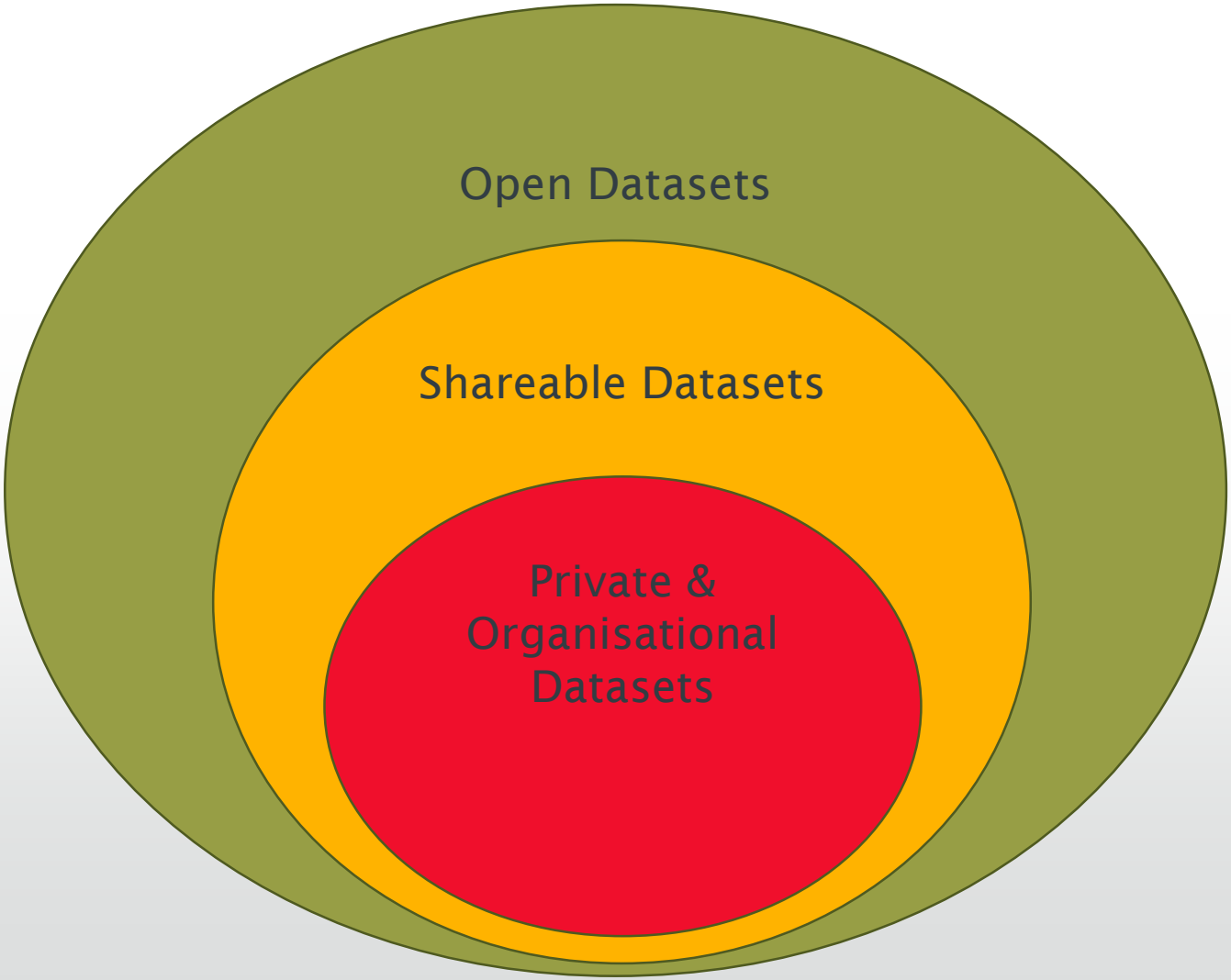
Sharing on the Web Observatory

- Sharing Datasets
- Sharing Ways of Interacting with Datasets
- Sharing Analytics

EXPOSING BOTH DATASETS AND ANALYTICS

LINKING EXPLICITLY ANALYTIC TOOLS TO DATASETS USED

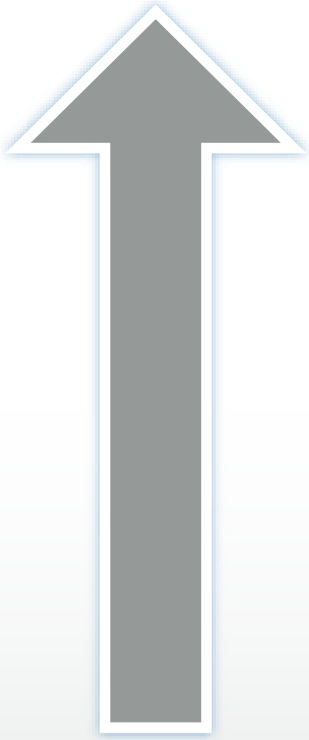
Levels of Sharing



Web Observatory Objectives

- Listing and describing datasets and archives of activity on the Web
- Analytics and visualisations based on identified and shared datasets
- Harmonisation of infrastructures and standardisation – W3C Community group
- Support for analytics on a global, distributed scale
- Provision for safe harbours for data analysis of public and private data

Web of Observatories



Harmonisation with Other Observatories	Social Innovation – Computational Social Science	Interdisciplinary e-Science
Catalogues of Analytic and Visualisation Tools	Open Standards	Novel Analysis and Visualisation Methods
Semantic Catalogues	Harvesting and Archival	Harmonised Access

Open Data Cloud

e-Science Infrastructures

Internet Archive

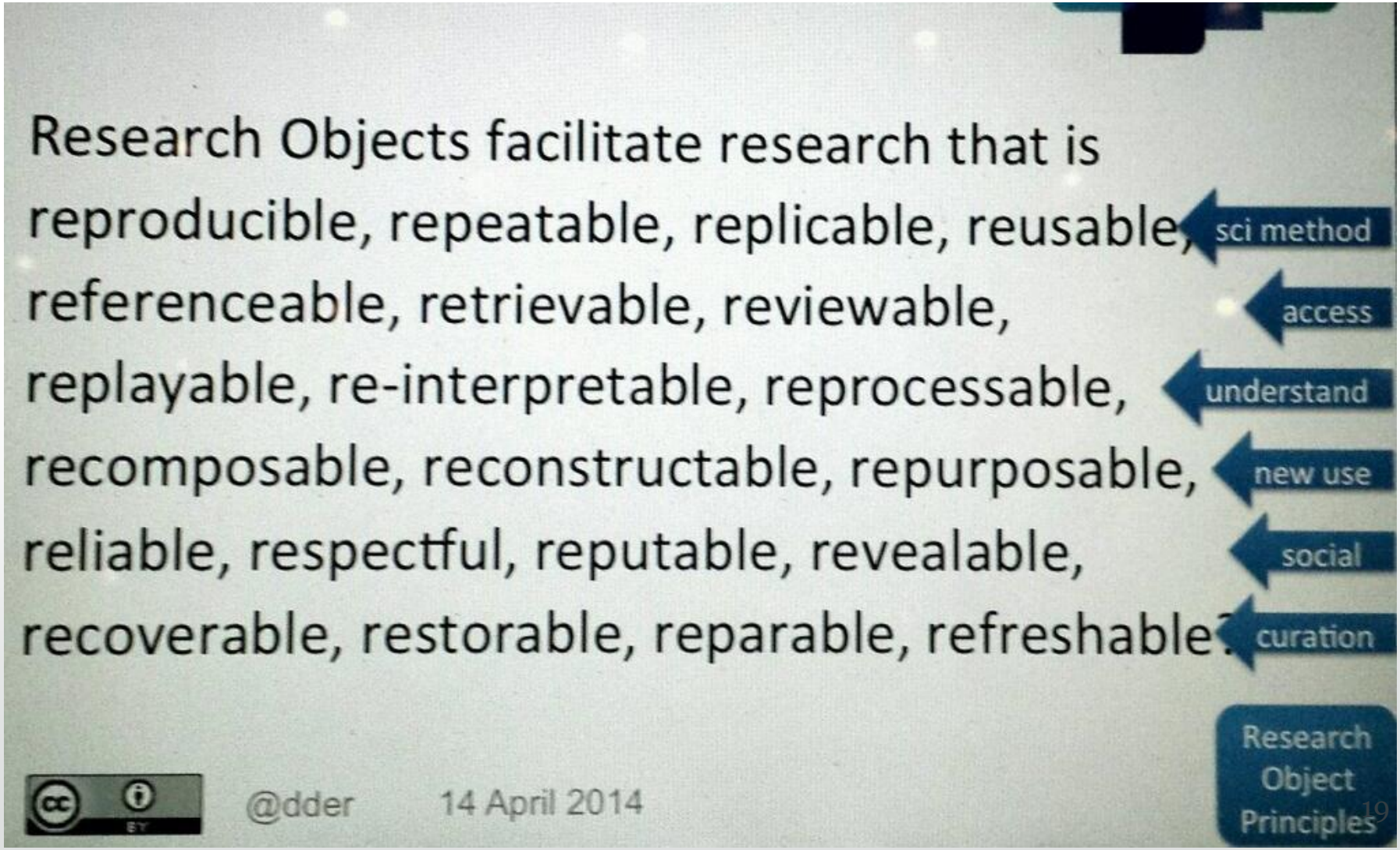
Open Web Analytics

Social Media

Repositories with Authorised Access

Tiropanis, T., Hall, W., Shadbolt, N., de Roure, D., Contractor, N., & Hendler, J. (2013). The Web Science Observatory. *Intelligent Systems, IEEE*, *28*(2), 100–104. doi:10.1109/MIS.2013.50

The case for research on the WO - R dimensions



UNIVERSITY OF Southampton

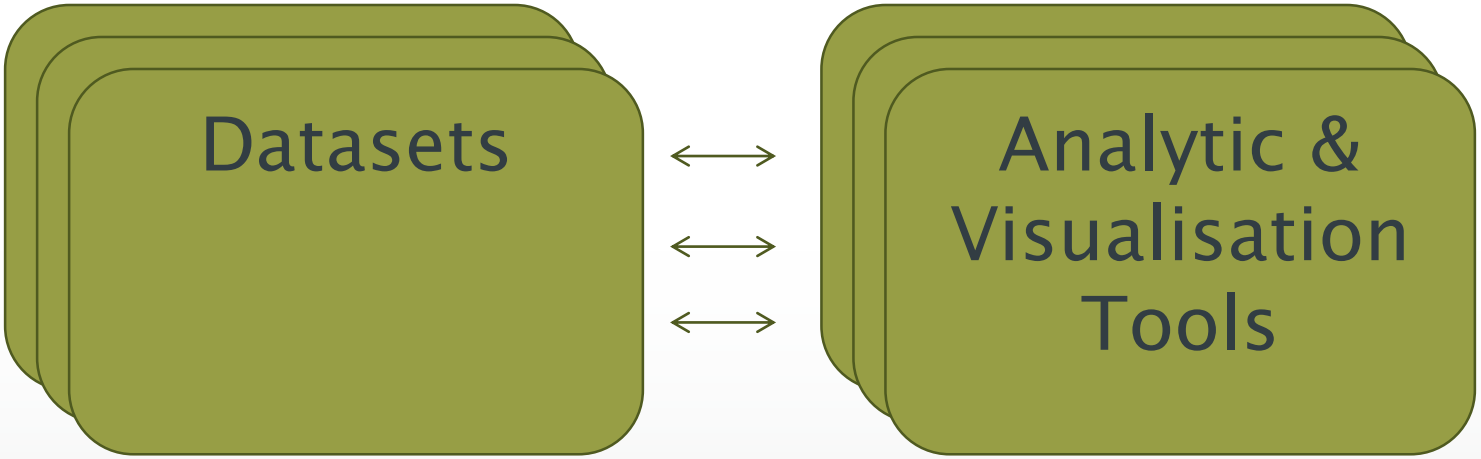
We are building a social machine to observe social machines

Physical and
Applied Sciences

UNIVERSITY OF
Southampton

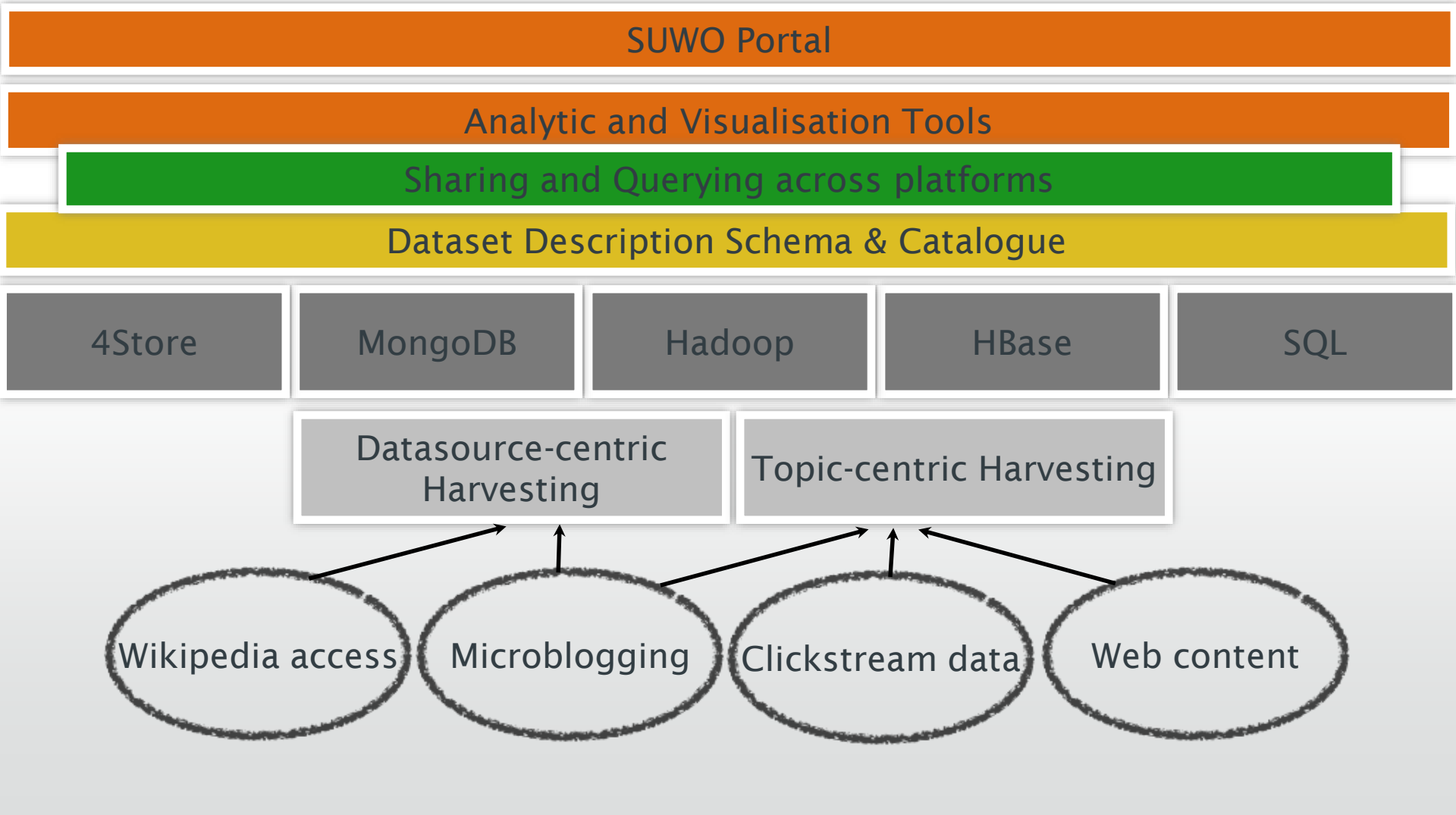
The Southampton Web Observatory

Southampton Web Observatory

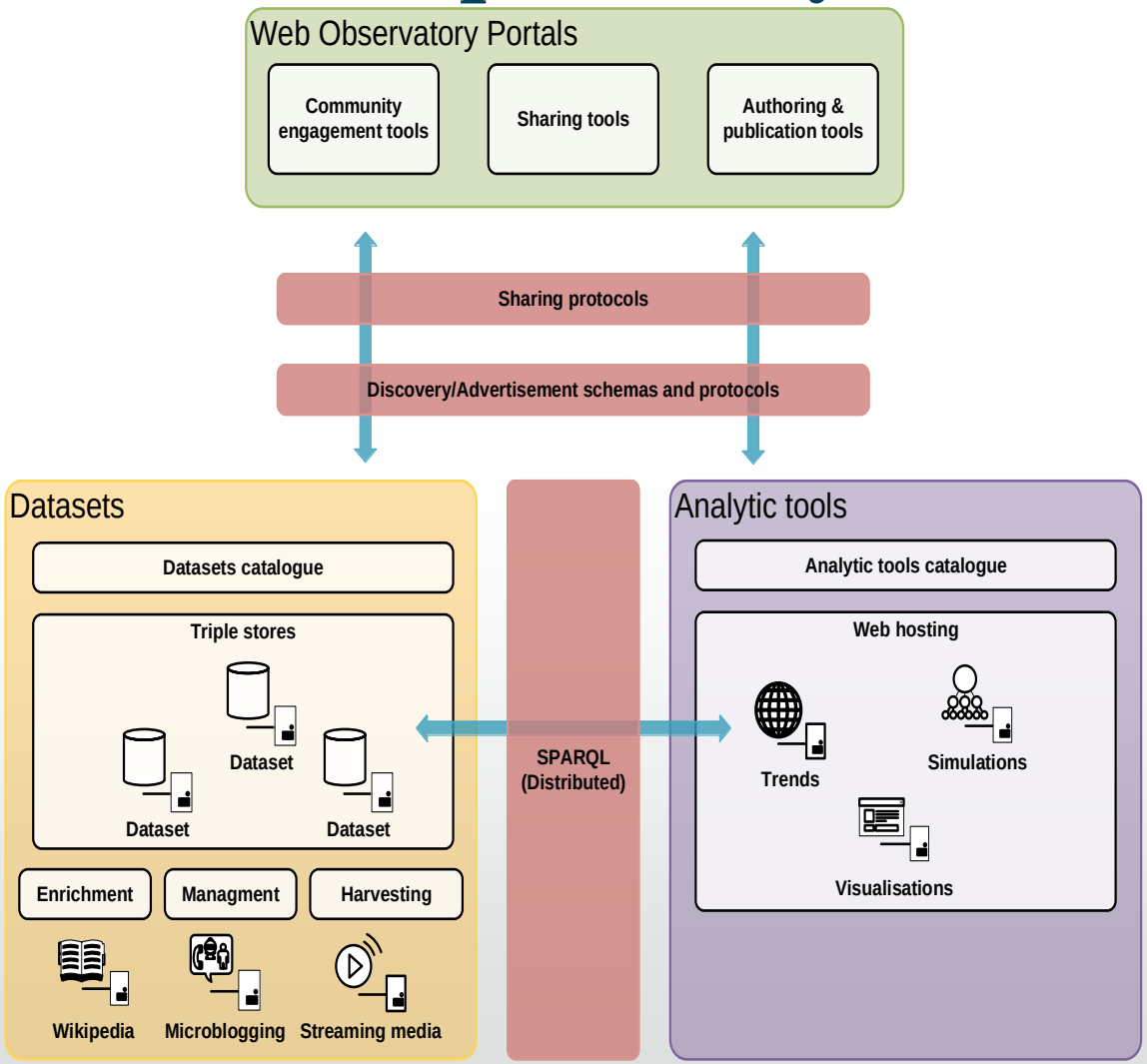


Sharing
Distribution
Optimisation
Interoperability

SUWO Infrastructure

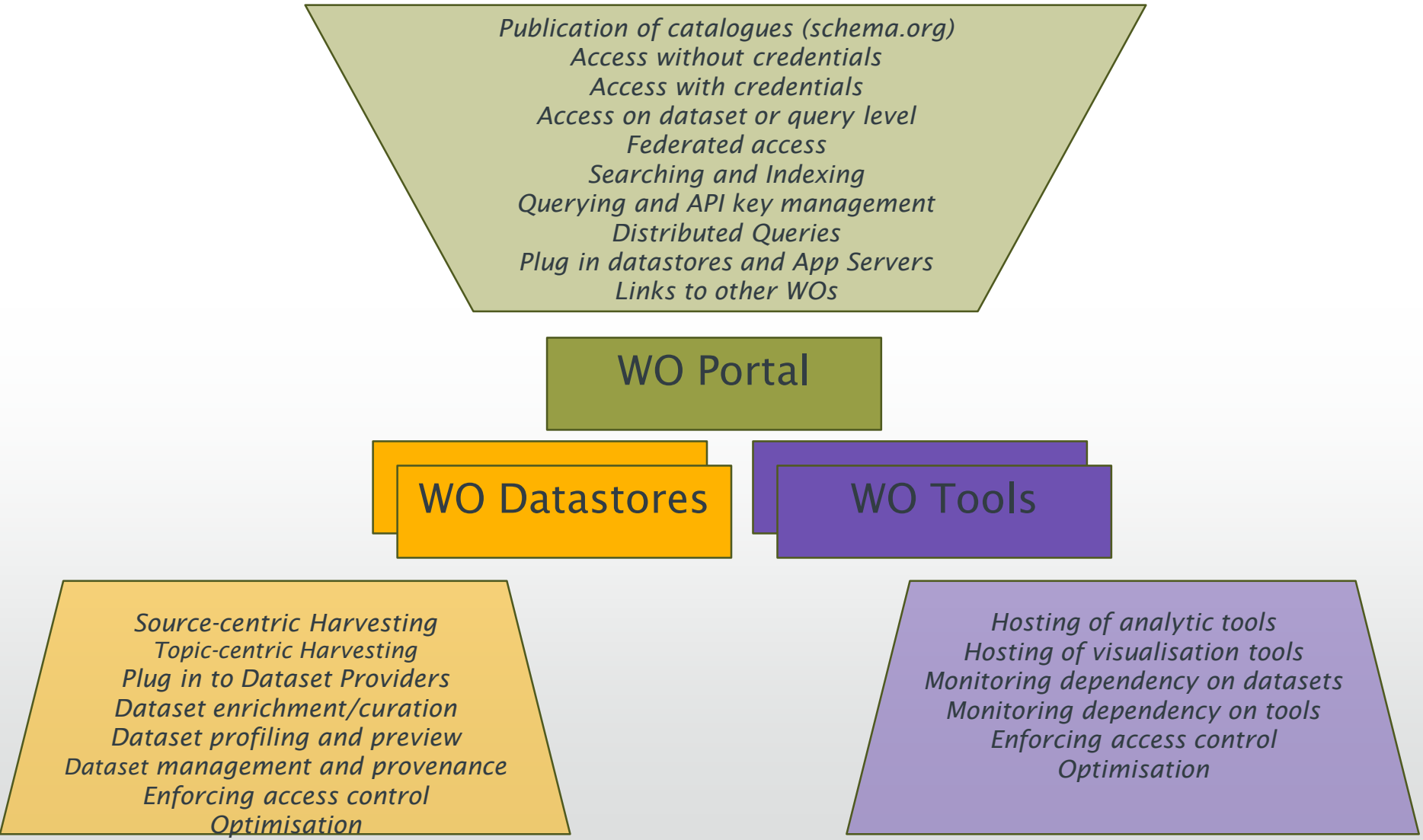


The case for interoperability (Linked Data)

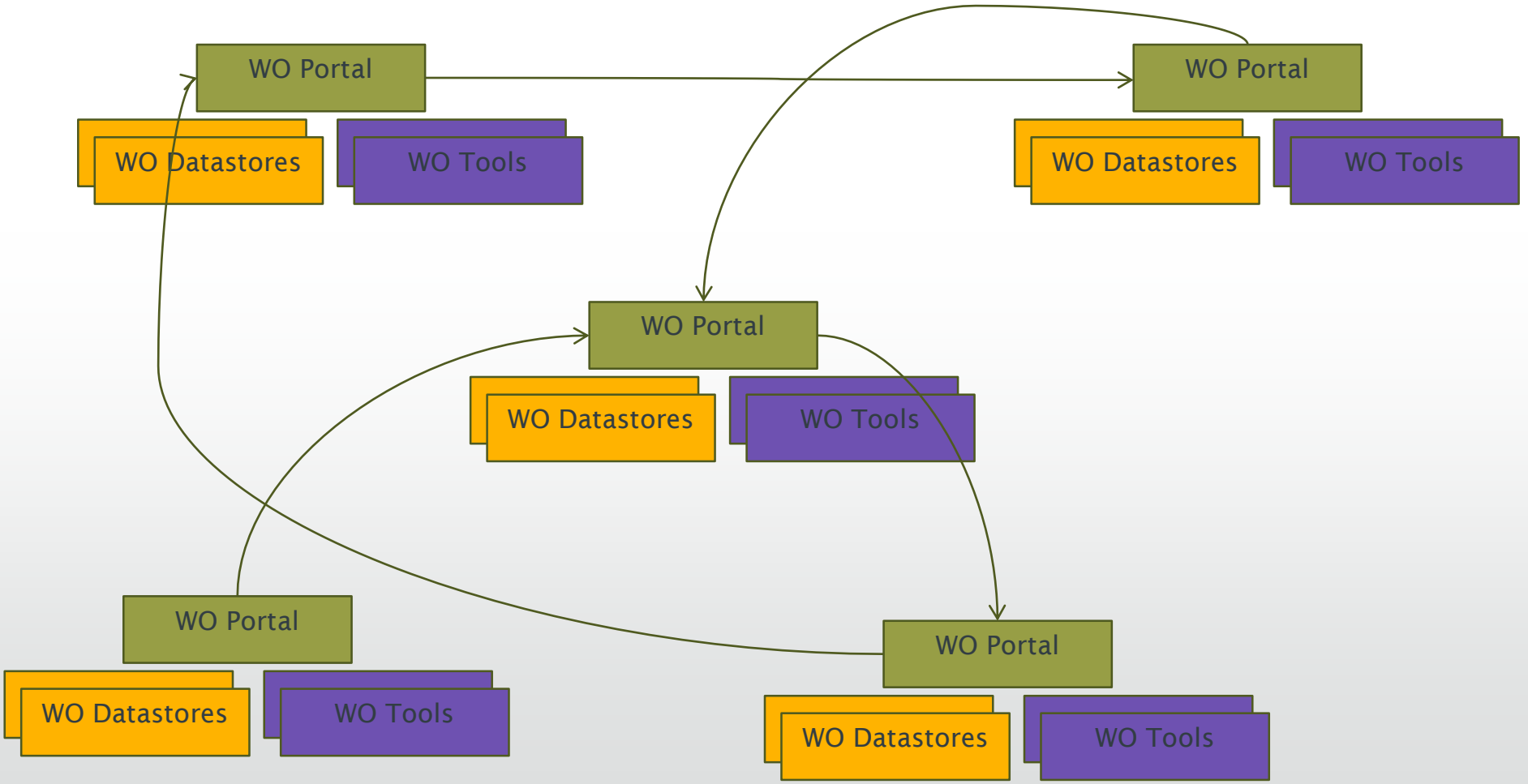


Hall, W., Tiropanis, T., Tinati, R., Wang, X., Luczak-Rösch, M., & Simperl, E. (2014). The Web Science Observatory - The Challenges of Analytics over Distributed Linked Data Infrastructures. *Computer Networks*, *96*, 29–30.

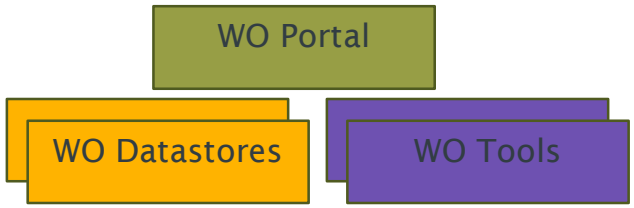
The three components of SUWO



SUWO Architecture & the Web of Observatories



The Web of Observatories - Stakeholders



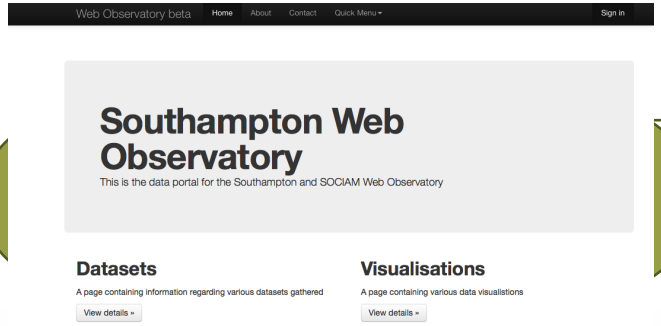
WO provider

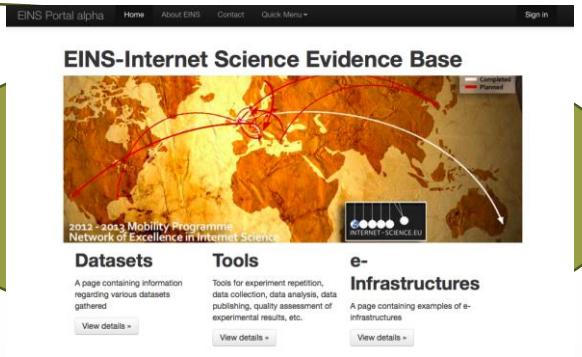
- Portal provision and maintenance
- Datastore and Application server plugin maintenance
- Maintenance of links to other observatories

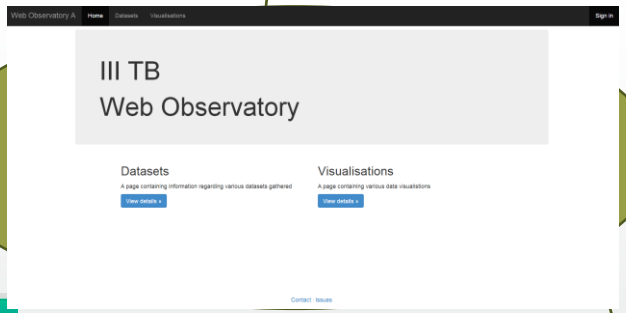
WO curator

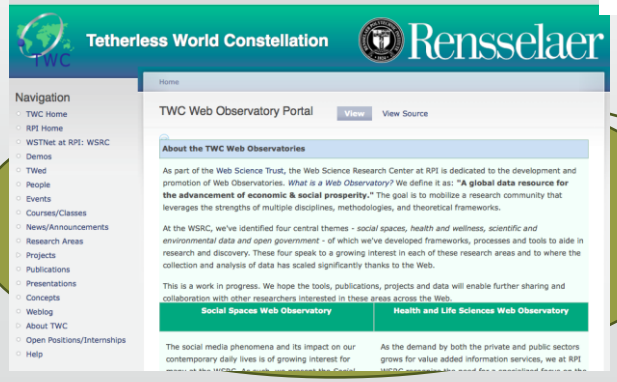
- They have users/customers
- Secure cloud for organisational datasets – safe harbours
- Liability Indemnification
- Curation of datasets
- Access control management
- Provision of analytic tools
- High performance computing

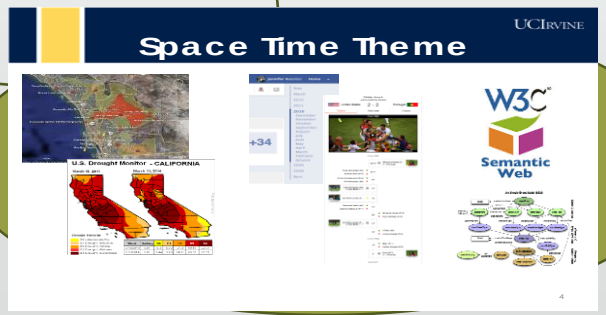
SUWO Architecture & the Web of Observatories



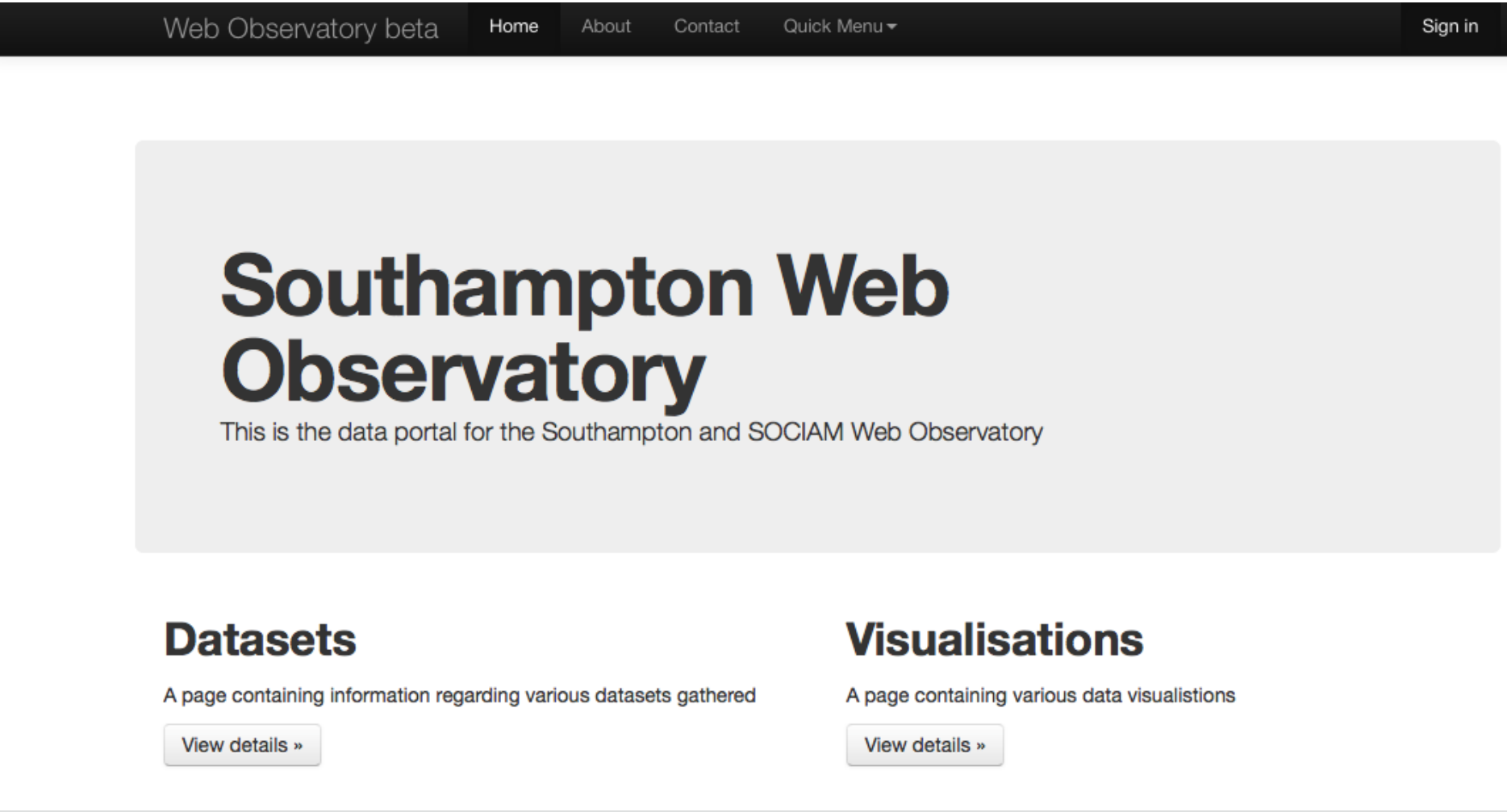








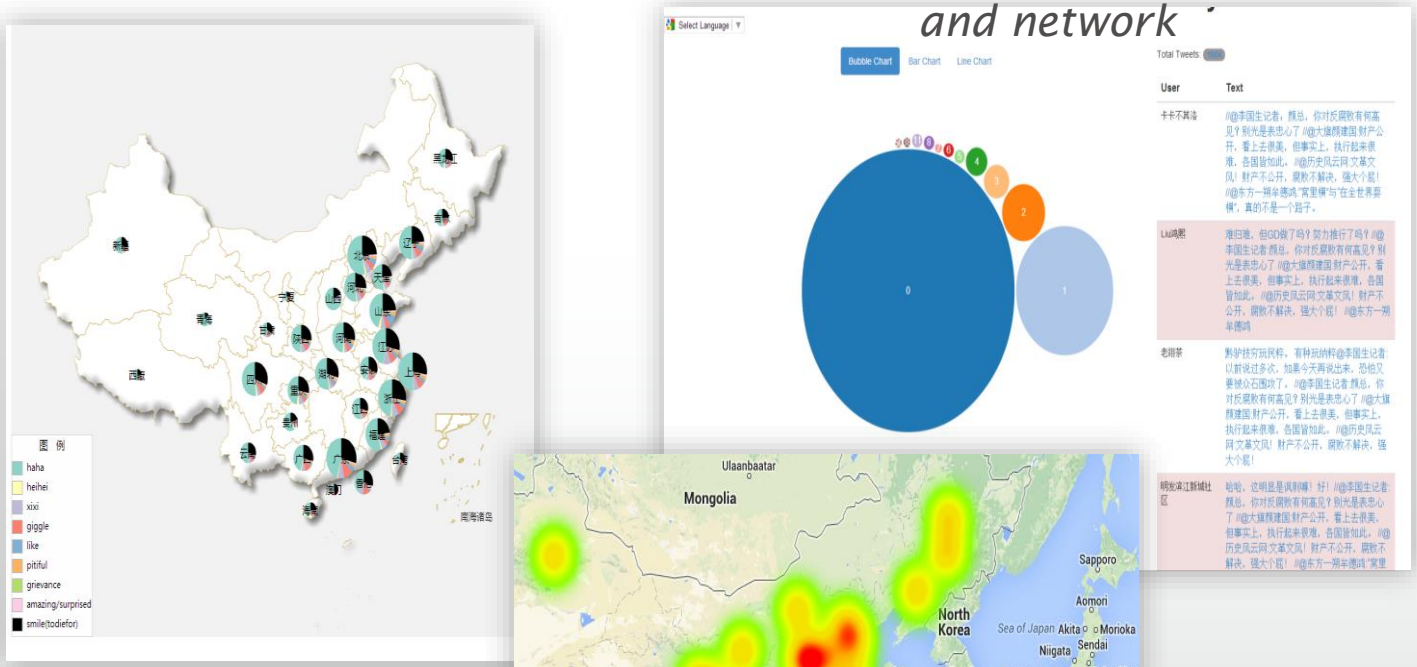
http://web-001.ecs.soton.ac.uk



Observatory in action

- Developing a live environment to observe and analyze in real-time

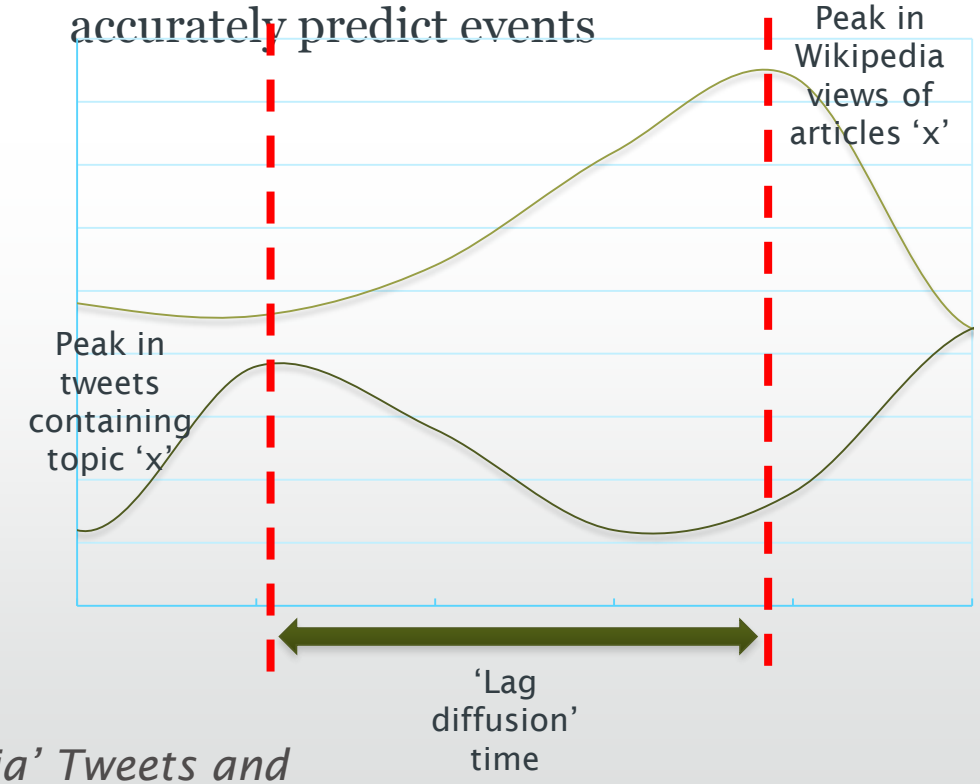
Weibo: Anti-Corruption messages and network



Chinese 'Salt Crisis': Visuals of Humour

Observations: Twitter and Wikipedia

- Observing the effects of real-world events across multiple sources
  - Combining multiple streams of real-time information to better understand and more accurately predict events

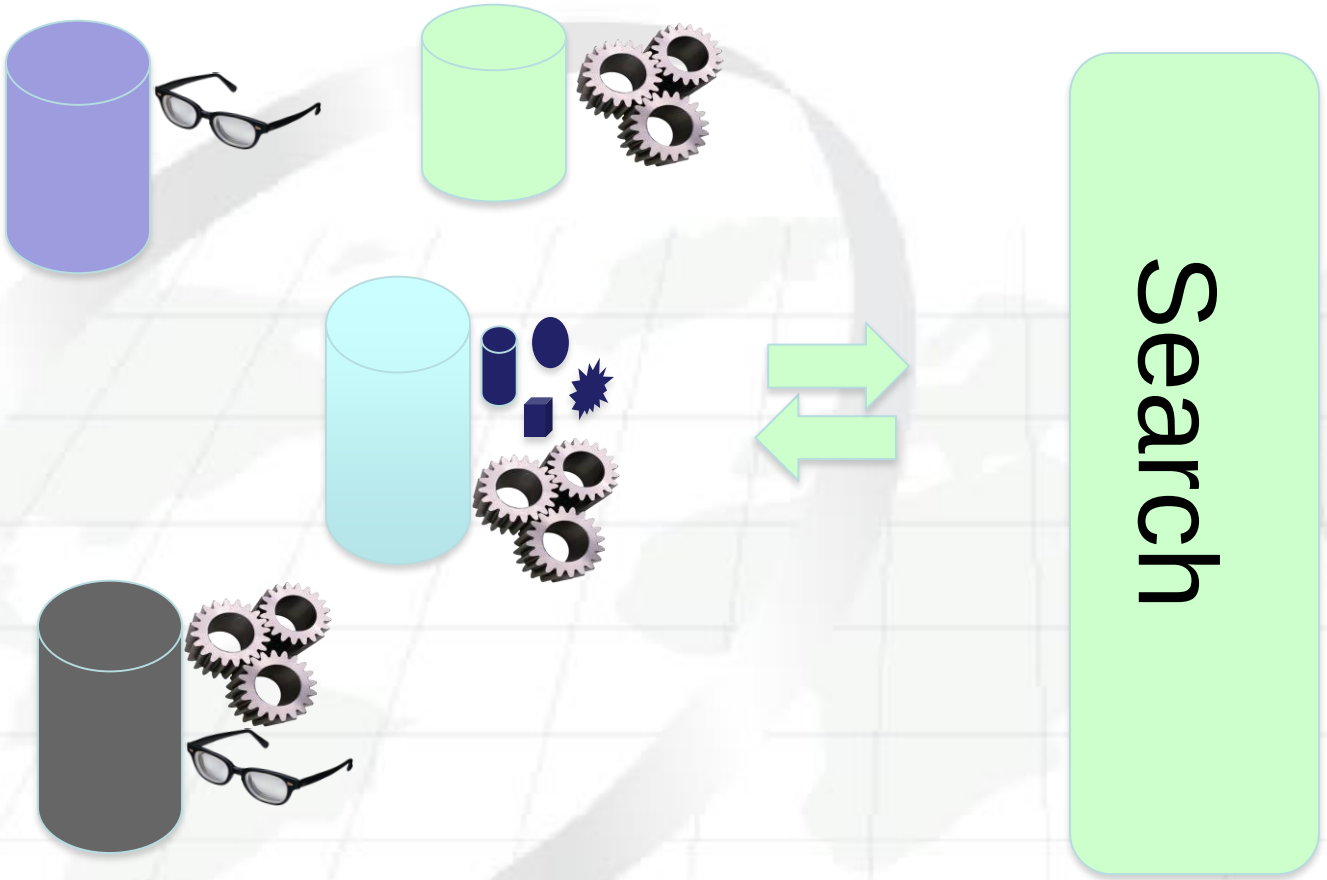


Twitter 'Wikipedia' Tweets and Wikipedia Page Views

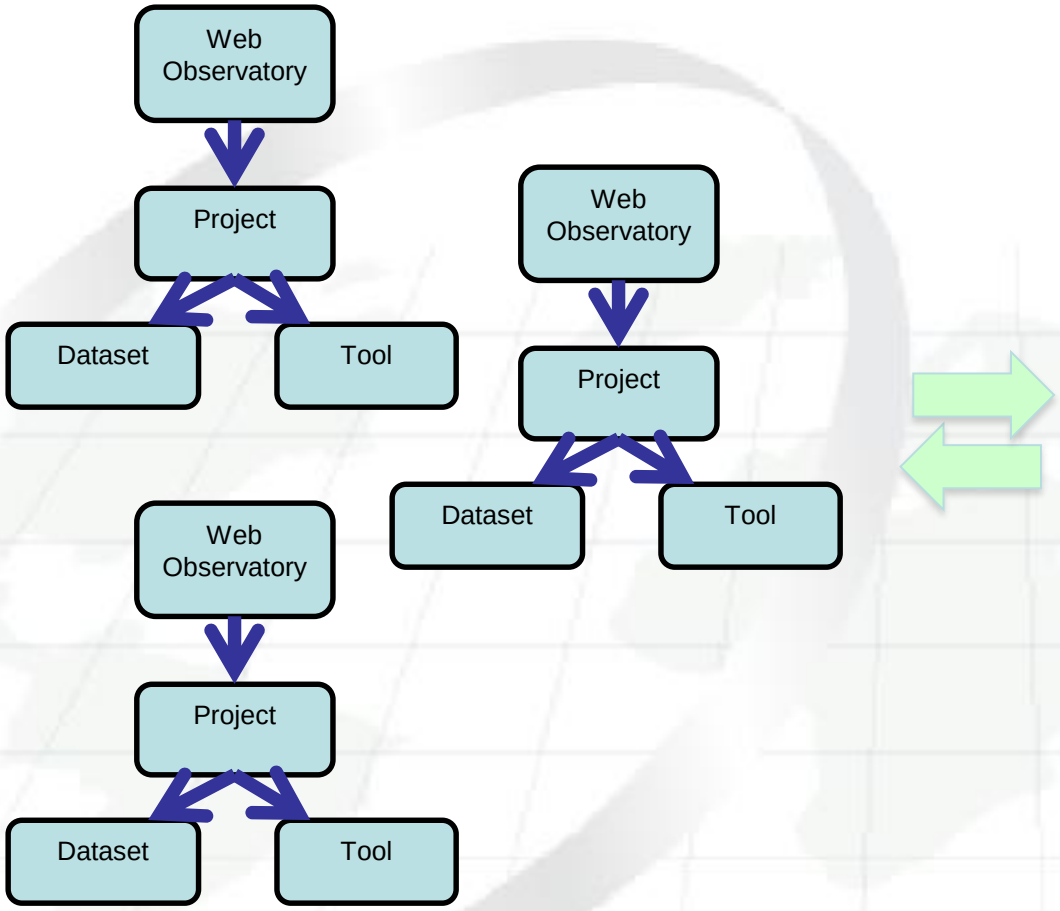
Schema.org

- An initiative launched by the leading search engine providers to create and support a common set of schemas for structured data markup on Web pages.
- These vocabularies enable the metadata to be more machine readable, allowing for better search, discover and display this information

Without Schema.org:



With Schema.org:



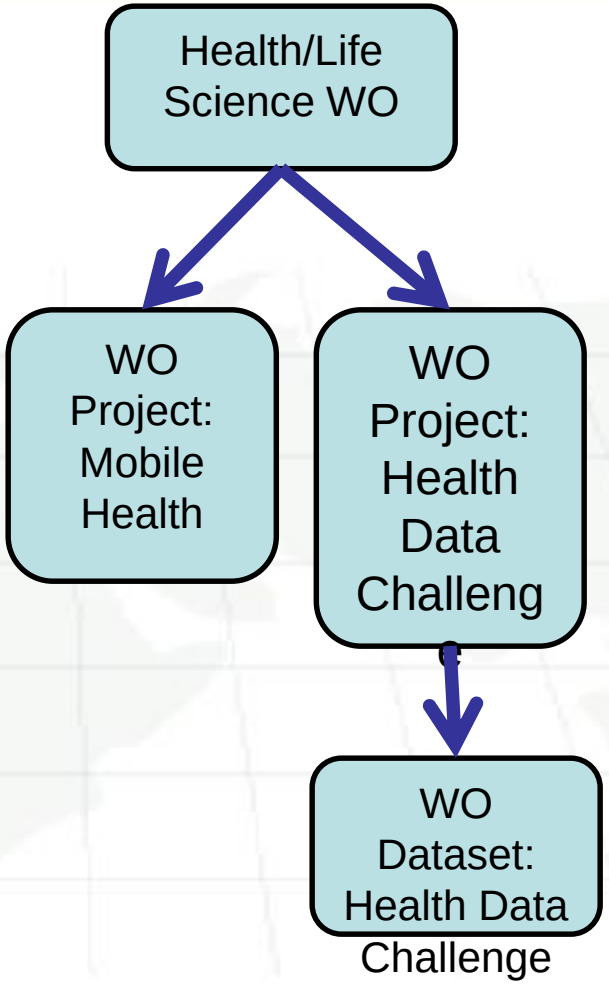
Search

Schema.org vocabulary demo

Health and Life Sciences Web Observatory

As the demand by both the private and public sectors grows for value added information services, we at RPI WSRC recognize the need for a specialized focus on the health and life sciences. The following are projects and tools we've used to tackle some of these issues.

1. **Projects**
   - Mobile Health
   - Repurposing Drugs with Semantics
   - Health Data Challenge
   - Health Web Science
2. **Datasets and Demos**
   - Health Data Challenge Sparql Endpoint
3. **People**
   - Joanne Luciano: Health Web Science
   - Deborah McGuinness: Health Web Science, Mobile Health and Repurposing Drugs with Semantics
   - Jim McCusker: Repurposing Drugs with Semantics & Health Data Challenge
   - Evan Patton: Mobile Health
   - Tim Lebo: Health Data Challenge



https://www.w3.org/wiki/WebSchemas/SchemaDotOrgProposals

https://www.w3.org/wiki/WebSchemas/WebObsSchema

UNIVERSITY OF Southampton

What next?

The Age of Data

- A Web of linked data was always part of Tim's original vision
- Machines can process and interpret linked data to make inferences about that data leading to a more intelligent (semantic) Web
- Open data leads to greater transparency, efficiency, and economic and social value, as demonstrated by the UK's Open Data Institute. Linked open data is even more powerful
- Big data – the Web has enabled the generation of lots of data that we are hungry to analyze and share

UNIVERSITY OF Southampton

The Future as the Web turns 25

- Amazing technical developments ahead but also major challenges – net neutrality, internet governance, cybersecruity, privacy, trust, ….
- Who has the right to do what with our data – fallout from the Snowden affair and development of personal data store
- Global Commission on Internet Governance. A Magna Carta for the Internet?
- Observing the Web to protect and develop the Web We Want
- Can we do this in real-time?

UNIVERSITY OF Southampton

Internet Archives and the Web Observatory

- Two sides of the same coin
- Every internet archive can contribute data to the web observatory project and the web observatory project can support research into internet archives
- Use the schema so we can include a listing of your datasets in the catalogue
- Enables analytics and visualisations which can also be shared
- Collect, curate, collaborate

Observing the Web

The ambition is to map the digital universe

Web Science Trust Network of Laboratories

The **Web Science Network of Laboratories (WSTNet)** combines some of the world's leading academic researchers in Web Science, with new academic programmes that will enhance the already growing influence of Web Science. The member Labs will provide valuable support for the ongoing development of Web Science. There are now 14 WSTNet labs:

Southampton, UK
MIT, USA
North-Western, USA
Tsinghua Graduate School, Shenzhen, China
DERI, Galway, Ireland
KAIST, Korea
Hannover, Germany
Oxford, UK
RPI, USA
Anaheim School of Communication, USC, USA
VU, Amsterdam, The Netherlands
Koblenz, Germany
Rio, Brazil
Indianna, USA