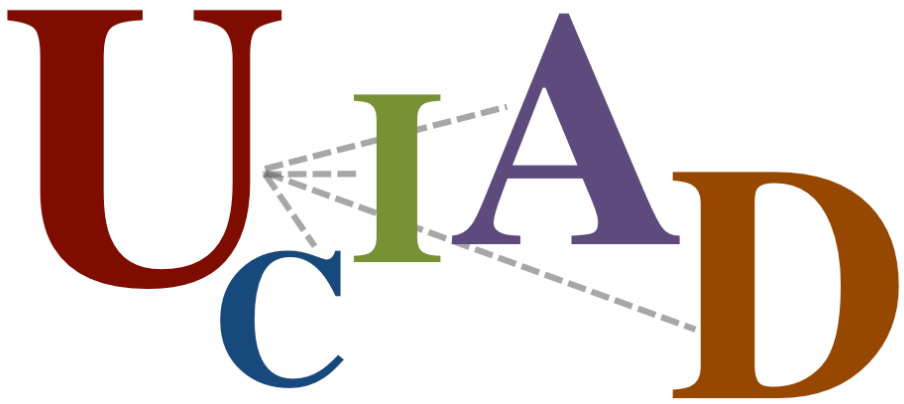
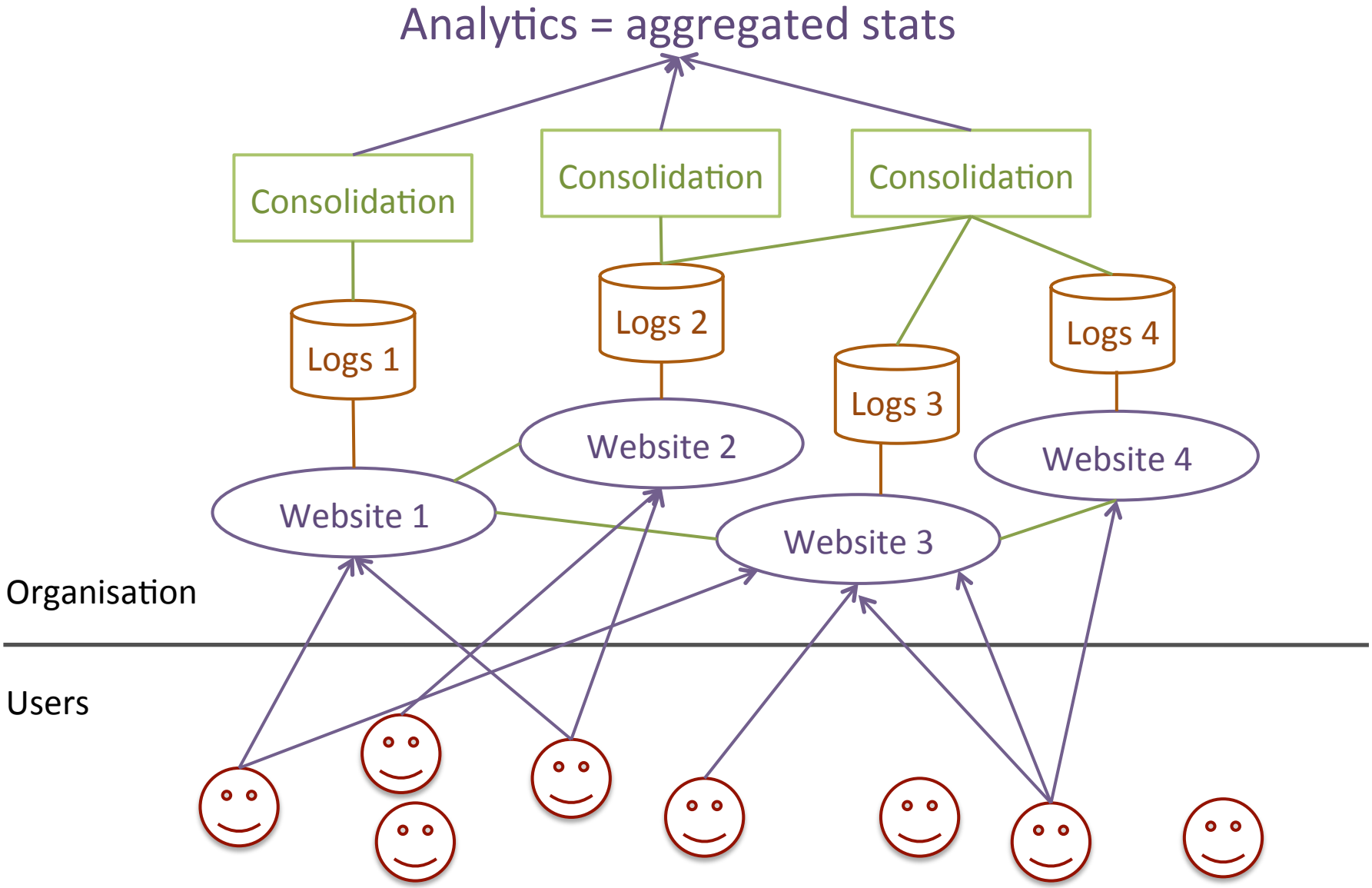
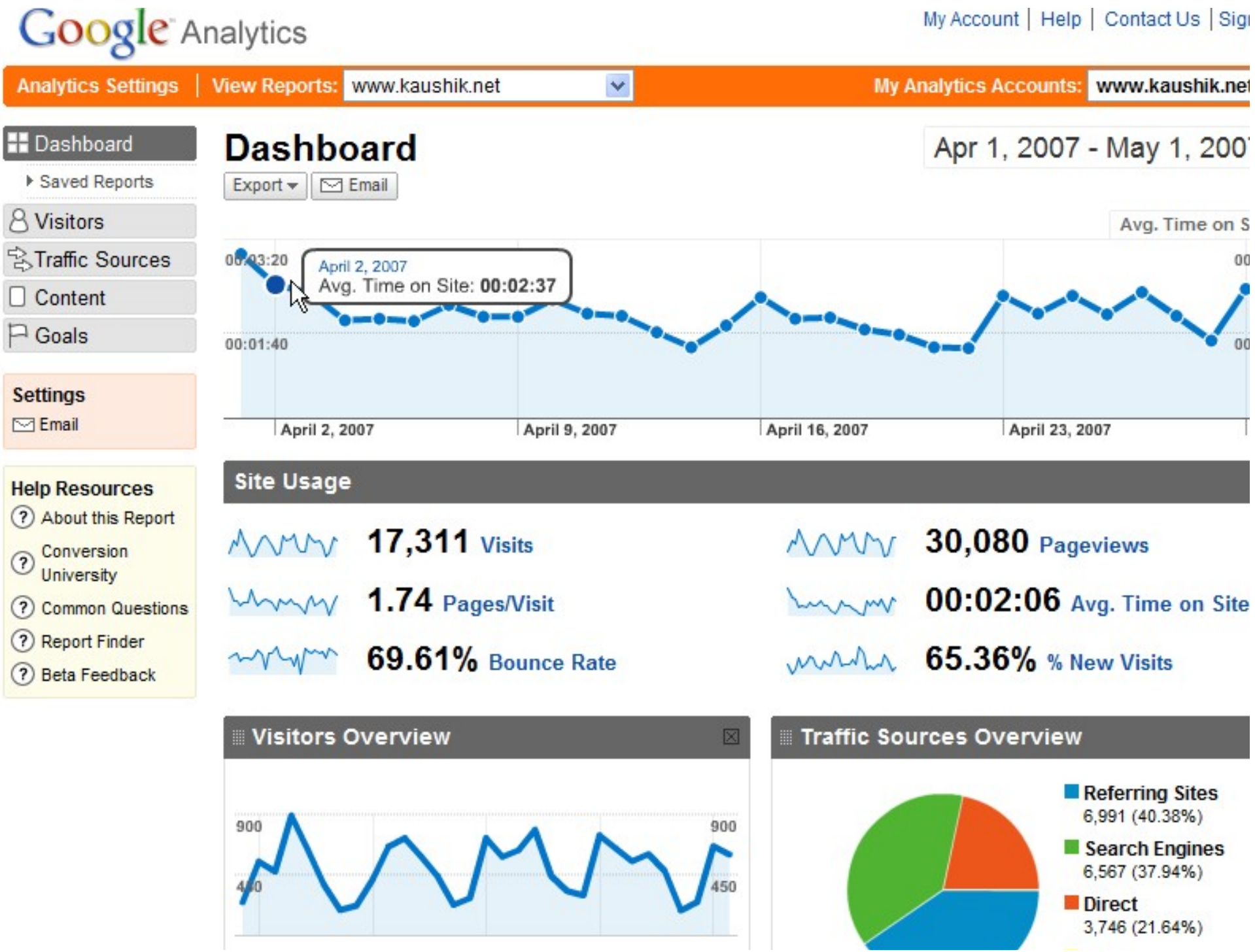



# User Centric Integration of Activity Data

Mathieu d'Aquin
Knowledge Media Institute
The Open University

# Organisation Centric Activity Data




Google Analytics
My Account | Help | Contact Us | Sig
Analytics Settings | View Reports: www.kaushik.net | My Analytics Accounts: www.kaushik.net
Dashboard
Saved Reports
Visitors
Traffic Sources
Content
Goals
Settings
Email
Help Resources
About this Report
Conversion University
Common Questions
Report Finder
Beta Feedback
Dashboard
Export
Email
Apr 1, 2007 - May 1, 200
Avg. Time on S
April 2, 2007
Avg. Time on Site: 00:02:37
00:01:40
April 2, 2007
April 9, 2007
April 16, 2007
April 23, 2007
Site Usage
17,311 Visits
30,080 Pageviews
1.74 Pages/Visit
00:02:06 Avg. Time on Site
69.61% Bounce Rate
65.36% % New Visits
Visitors Overview
900
450
Traffic Sources Overview
Referring Sites
6,991 (40.38%)
Search Engines
6,567 (37.94%)
Direct
3,746 (21.64%)

# Consumer/user centric data

ECONOMIC VIEW

## Show Us the Data. (It's Ours, After All.)

By RICHARD H. THALER
Published: April 23, 2011

"NO one knows what I like better than I do."

This statement may seem sel
but the revolution in informa
technology has created a gro
of exceptions. Your grocery s
knows
probab

PROJECT DANUBE
OPEN-SOURCE SOFWARE FOR IDENTITY & PERSONAL DATA SERVICES

data liberation

NYT: Companies should give usage data to customers

By Phil Wolff, on April 25th, 2011%

mydata project "recognises personal data as a personal asset"

publication date: Apr 18, 2011

*Give Me My Data helps you reclaim and reuse your Facebook data.*

Give Me My Data

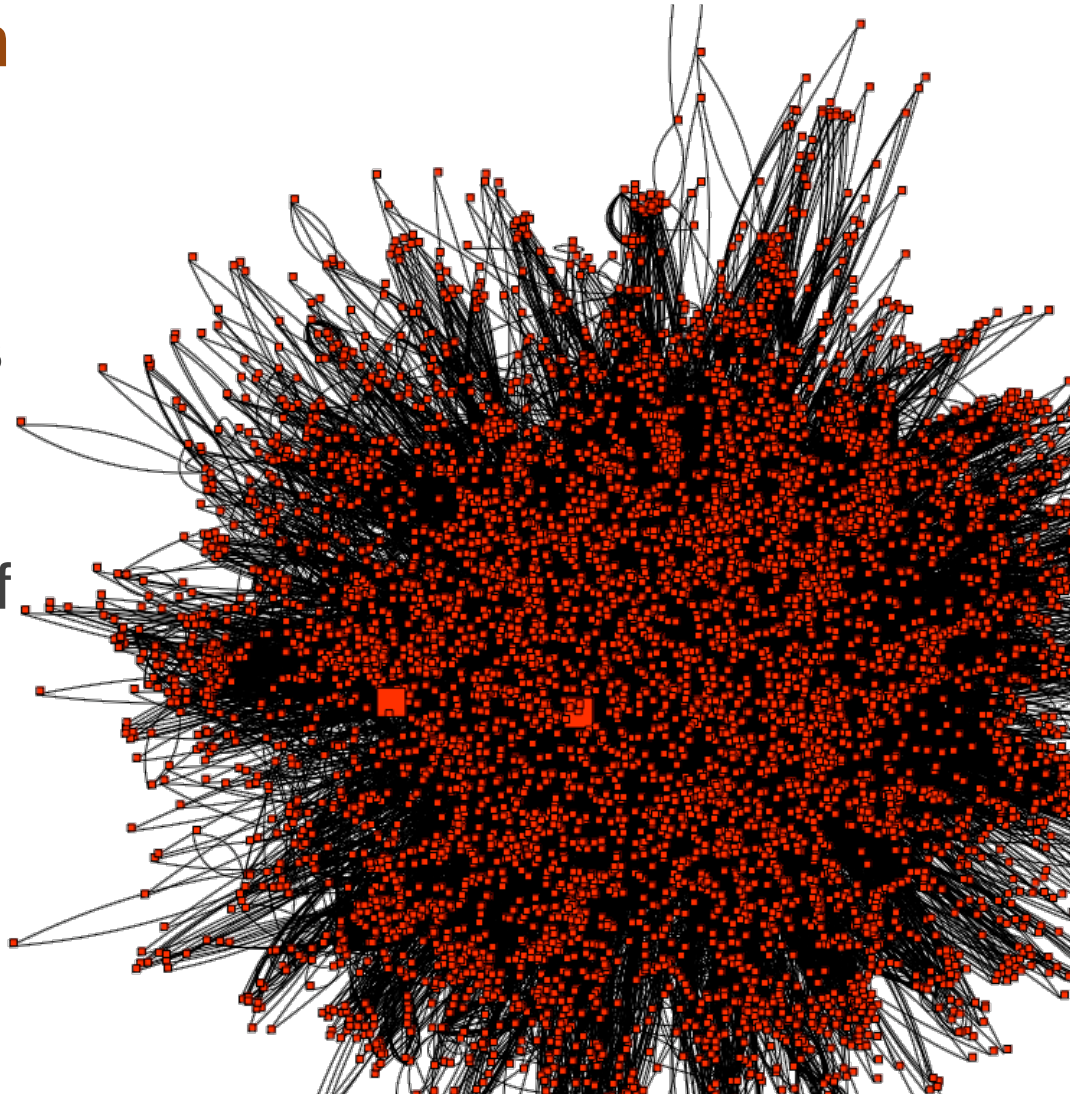
Google takeout

**Welcome to Google Takeout.**

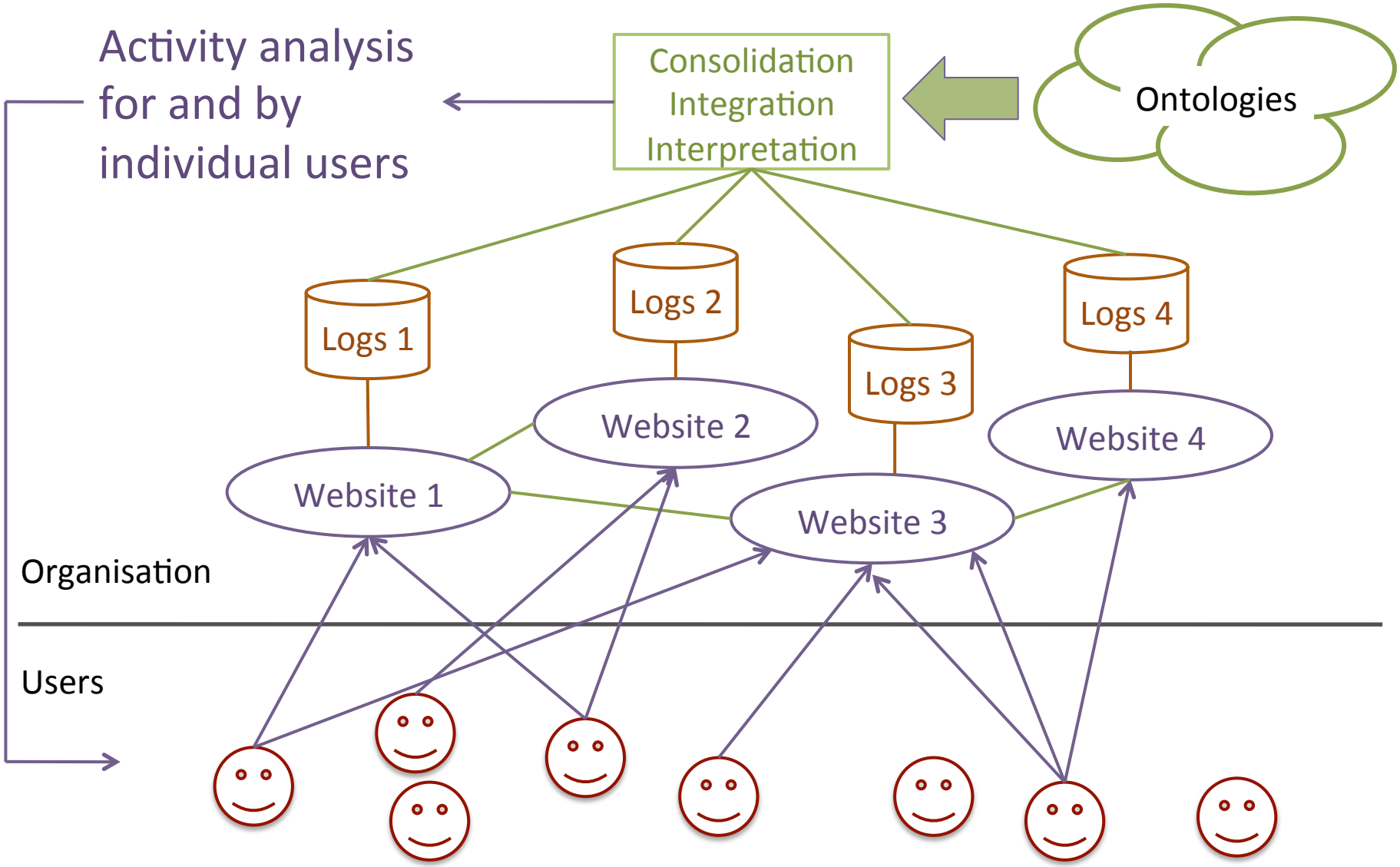
Google Takeout allows you to download a copy of your data stored within Google products.

# Challenges in user centric **activity** data

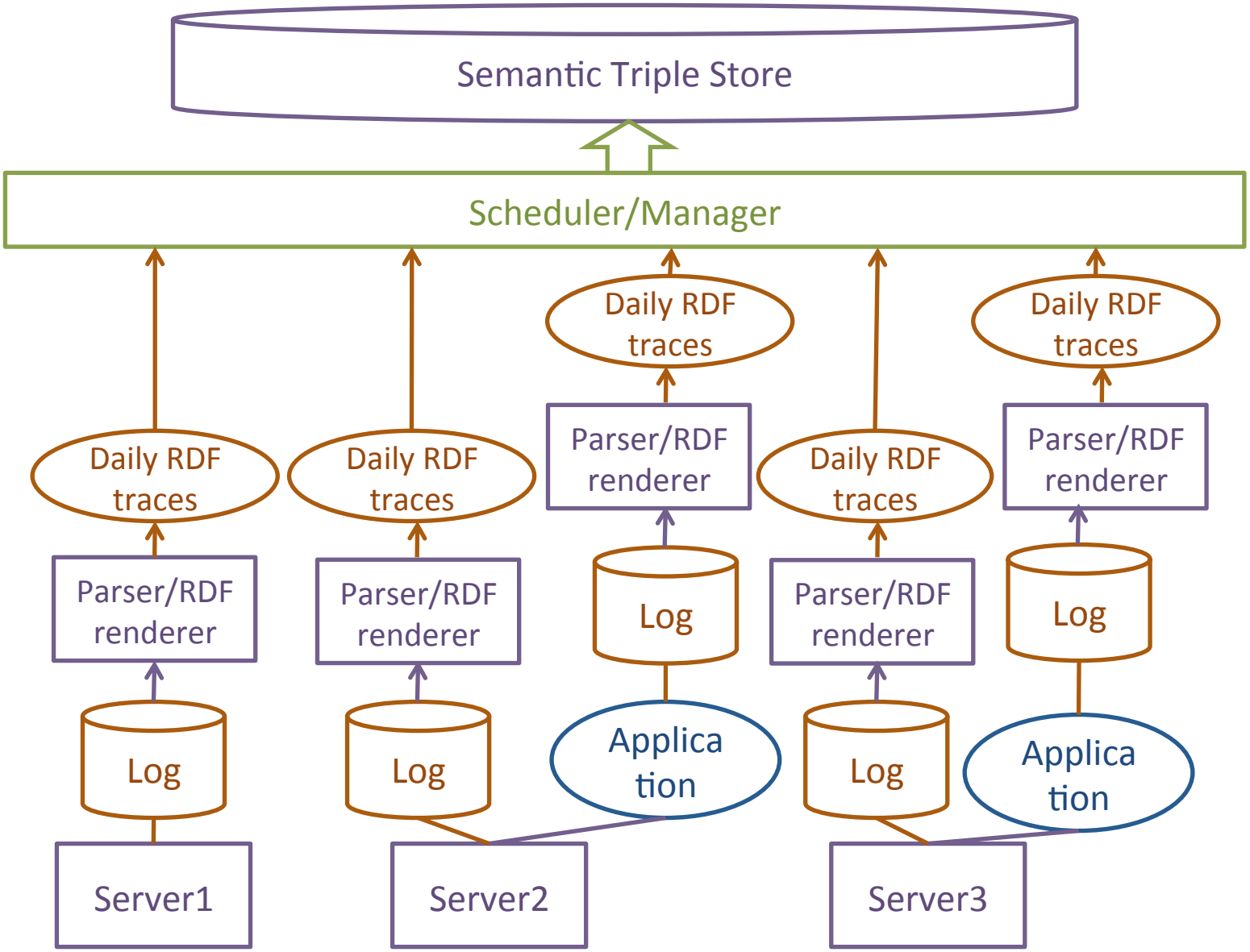
- Activity data that sit in logs are
  - Heterogeneous – different models for different sites/systems
  - Raw – uninterpreted
  - Horribly big – thousands of pieces of information generated every minute
  - Hard to exploit, understand, analyze



# User Centric Activity Data

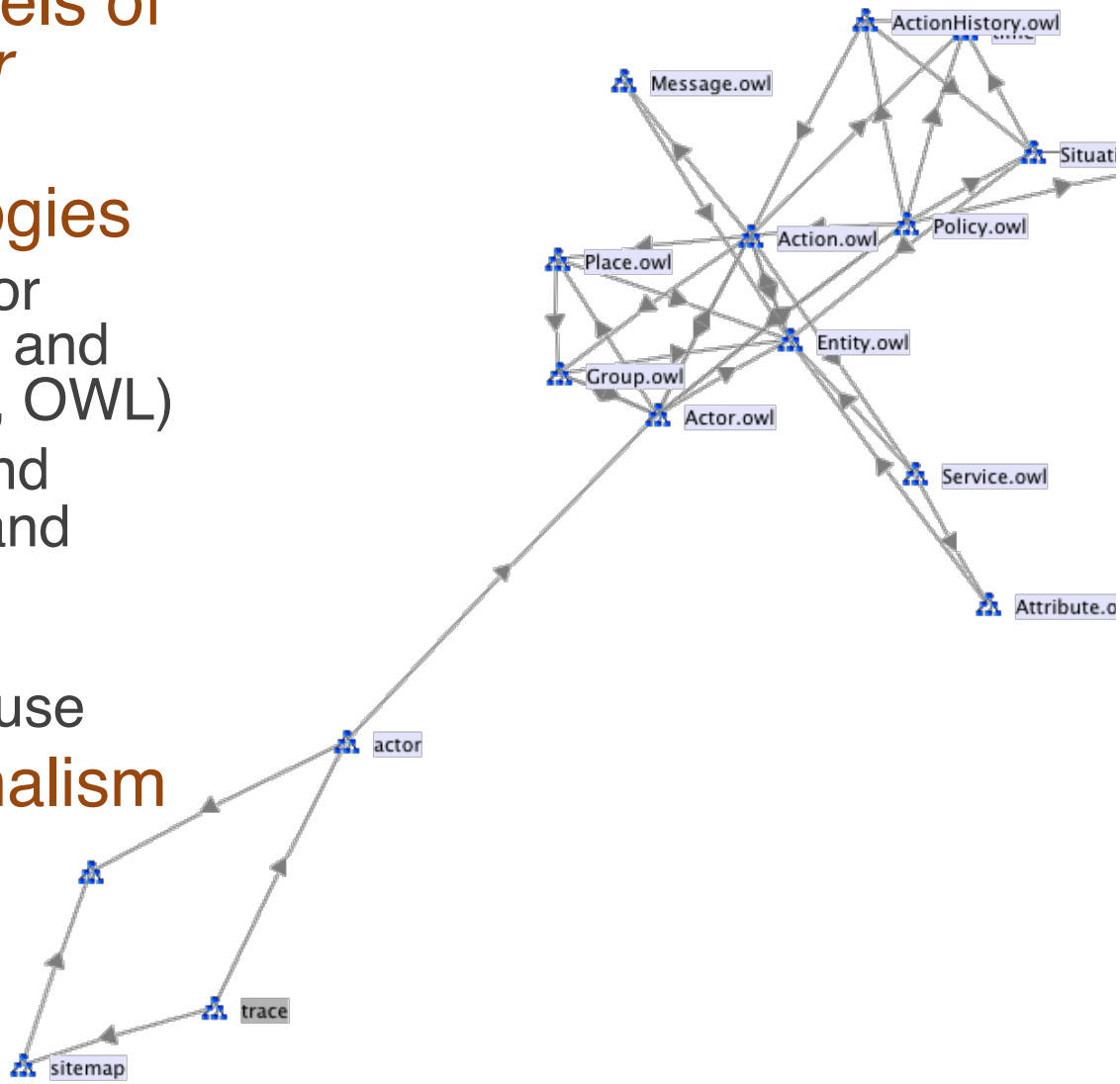


# Technical infrastructure



# Ontologies

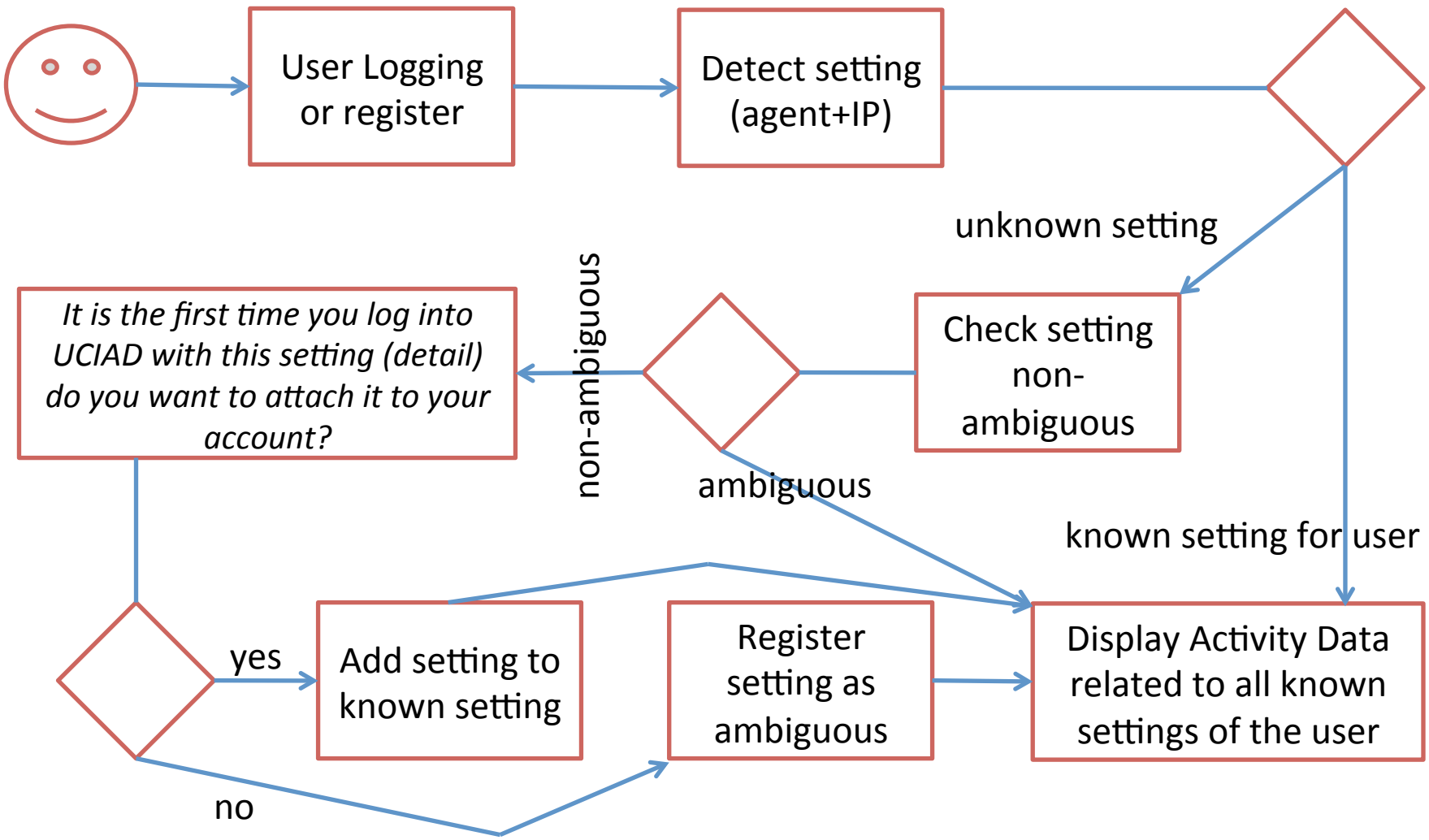
- Formal conceptual models of a domain: *online user activity*
- Semantic Web technologies
  - Standard languages for expressing ontologies and ontological data (RDF, OWL)
  - Tools to manipulate and work with ontologies and semantic data (NeOn Toolkit, OWLIM)
  - Many ontologies to reuse
- Adhere to a logical formalism → inferences

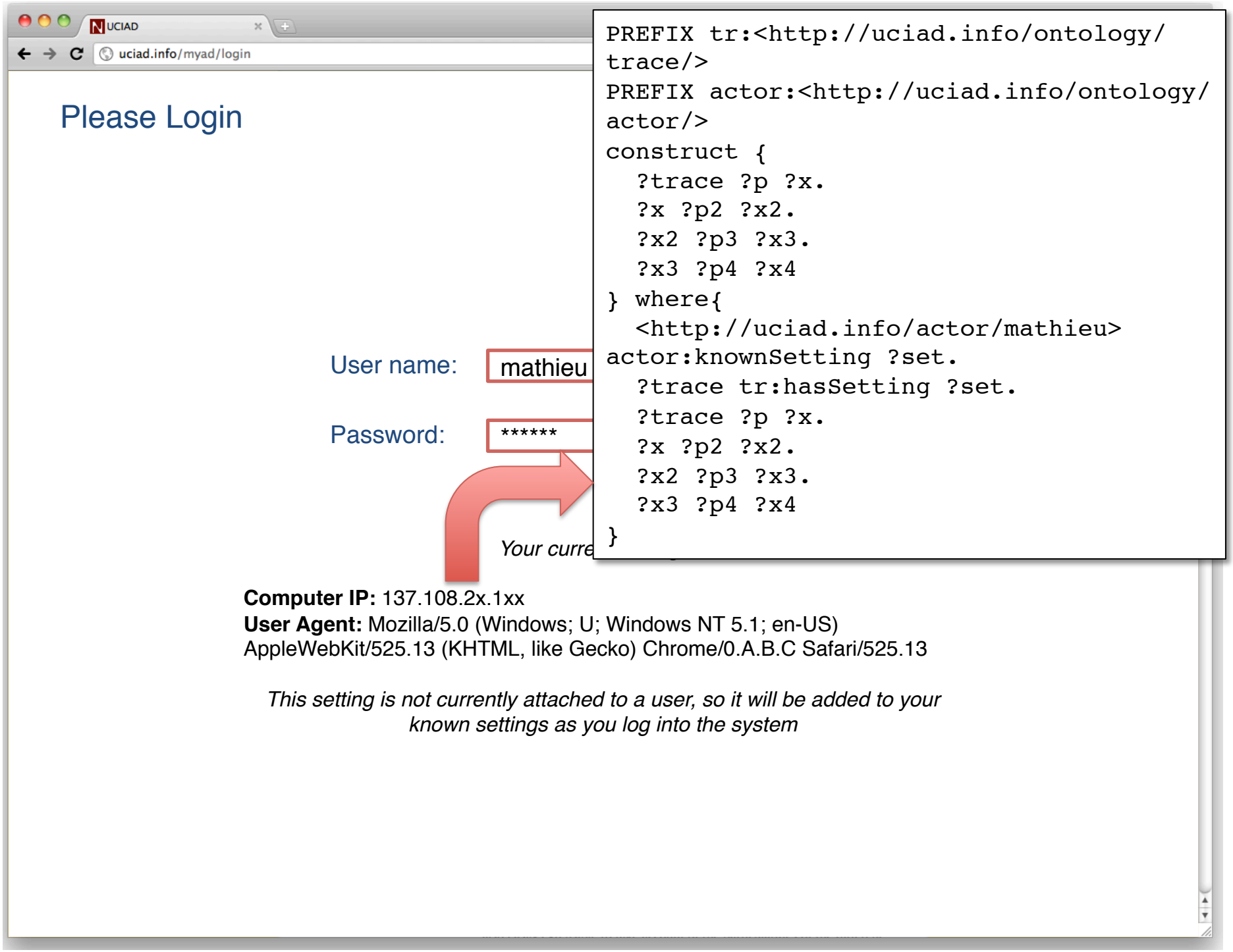


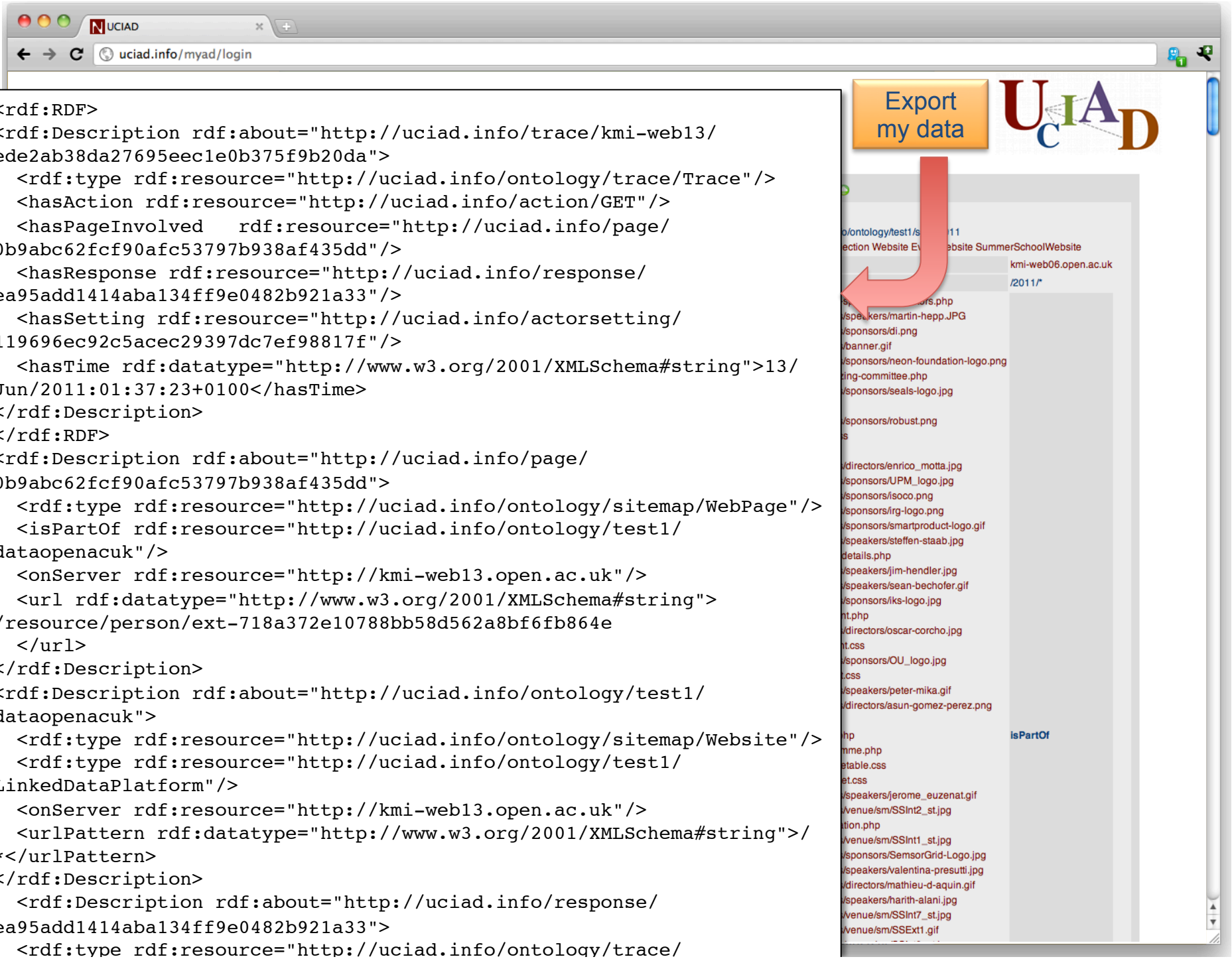
# Ontologies

Actor
TraceAction
hasActor
hasAction
TraceResponse
hasResponse
ParameterValue
Trace
hasPageInvolved
WebPage
hasSetting
ActorSetting
hasAgent
fromComputer
ActorAgent
Computer
Attribute
Thing
Document
OnlineAccount
Thing
Thing
made
Actor
holdsAccount
Place
mbox
topic_interest
hasRole
PersonRole
hasTeamPartner
isEncompassedBy
openid
Actor
account
hasPlace
Document
Place
OnlineAccount
capableOfPerforming
controlledBy
weblog
Action
isPartOf
hasPart
interest
Policy
Document
Entity
Entity
ActorSetting
fromComputer
Document
Computer
hasAgent
ActorAgent
Thing
topic
Thing
primaryTopic
WebPage
isPartOf
WebPageCollection

# User support




UCIAD
uciad.info/myad/login
Please Login
User name:
mathieu
Password:
******
Your curre
Computer IP: 137.108.2x.1xx
User Agent: Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US) AppleWebKit/525.13 (KHTML, like Gecko) Chrome/0.A.B.C Safari/525.13
This setting is not currently attached to a user, so it will be added to your known settings as you log into the system
PREFIX tr:<http://uciad.info/ontology/trace/>
PREFIX actor:<http://uciad.info/ontology/actor/>
construct {
?trace ?p ?x.
?x ?p2 ?x2.
?x2 ?p3 ?x3.
?x3 ?p4 ?x4
} where{
<http://uciad.info/actor/mathieu>
actor:knownSetting ?set.
?trace tr:hasSetting ?set.
?trace ?p ?x.
?x ?p2 ?x2.
?x2 ?p3 ?x3.
?x3 ?p4 ?x4
}

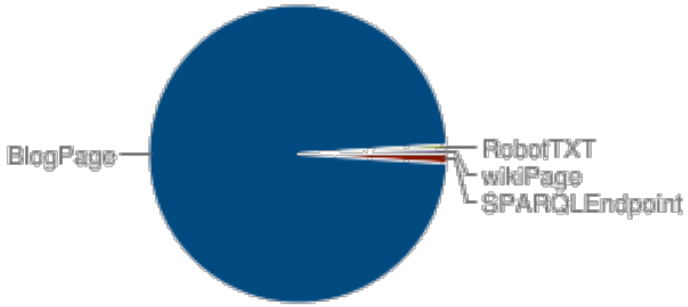



```
<rdf:RDF>
<rdf:Description rdf:about="http://uciad.info/trace/kmi-web13/
ede2ab38da27695eec1e0b375f9b20da">
  <rdf:type rdf:resource="http://uciad.info/ontology/trace/Trace"/>
  <hasAction rdf:resource="http://uciad.info/action/GET"/>
  <hasPageInvolved   rdf:resource="http://uciad.info/page/
0b9abc62fcf90afc53797b938af435dd"/>
  <hasResponse rdf:resource="http://uciad.info/response/
ea95add1414aba134ff9e0482b921a33"/>
  <hasSetting rdf:resource="http://uciad.info/actorsetting/
119696ec92c5acec29397dc7ef98817f"/>
  <hasTime rdf:datatype="http://www.w3.org/2001/XMLSchema#string">13/
Jun/2011:01:37:23+0100</hasTime>
</rdf:Description>
</rdf:RDF>
<rdf:Description rdf:about="http://uciad.info/page/
0b9abc62fcf90afc53797b938af435dd">
  <rdf:type rdf:resource="http://uciad.info/ontology/sitemap/WebPage"/>
  <isPartOf rdf:resource="http://uciad.info/ontology/test1/
dataopenacuk"/>
  <onServer rdf:resource="http://kmi-web13.open.ac.uk"/>
  <url rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
/resource/person/ext-718a372e10788bb58d562a8bf6fb864e
  </url>
</rdf:Description>
<rdf:Description rdf:about="http://uciad.info/ontology/test1/
dataopenacuk">
  <rdf:type rdf:resource="http://uciad.info/ontology/sitemap/Website"/>
  <rdf:type rdf:resource="http://uciad.info/ontology/test1/
LinkedDataPlatform"/>
  <onServer rdf:resource="http://kmi-web13.open.ac.uk"/>
  <urlPattern rdf:datatype="http://www.w3.org/2001/XMLSchema#string">/
*</urlPattern>
</rdf:Description>
  <rdf:Description rdf:about="http://uciad.info/response/
ea95add1414aba134ff9e0482b921a33">
  <rdf:type rdf:resource="http://uciad.info/ontology/trace/
```

# Example

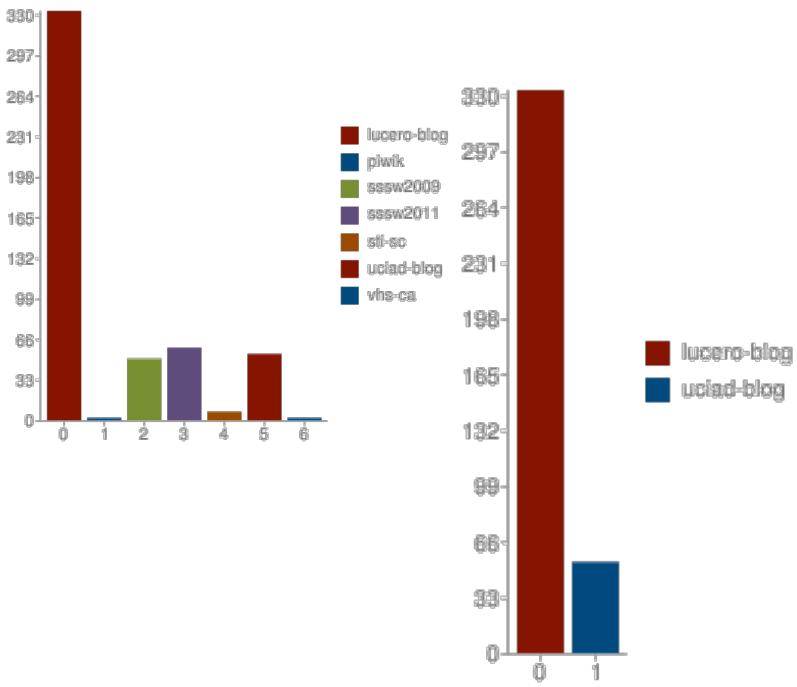
## In the ontology:

- UCIAD-Blog and LUCERO-Blog are Blogs (Website)
- A BlogPage is a page which is part of a Blog
- An activity onBlog is an activity happening on a Blog Page



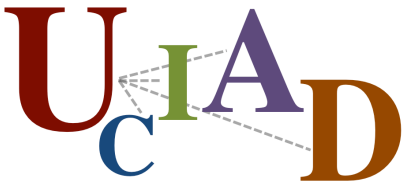

## Result:

- Can look specifically at activities happening on a Blog and specialize them (same applies to Wikis, and other types of websites)



# Issues left to resolve

- Scalability
  - OWLIM triple store can handle billions of triples
  - But struggle with millions when inference is “on”
  - → 1 repository without inference with all historical data, 1 with inference with 1 week of data only, and 1 with inference for registered users
- User management and privacy
  - Ensuring that the user who logs in from a particular setting is the one having the activity is difficult (e.g., in the case of shared computers)
  - Is this really a problem?
  - Check ambiguity – ask verification questions – moderate?
- Licensing
  - Overall data: privacy issues (is k-anonymity actually applicable? Would it work?)
  - Overall data: institutional issues (can we show the traffic on our websites to everybody)
  - User data export: what license?




# More info

UCIAD Blog: http://uciad.info

Code base: http://github.com/uciad

Twitter: #uciad

@mdaquin

# Team

- **Dr Mathieu d'Aquin** – Research fellow, KMi – project director
- **Stuart Brown** – Web developments and online communities, communication services – member of the steering group, liaison with online services
- **Salman Elahi** – Resarch assistant and PhD student, KMi – developer/researcher
- **Prof Enrico Motta** – Professor of knowledge technologies, KMi – Chair of the steering group