Standing on the Shoulders of the Trusted Web: Trust, Scholarship and Linked Data.

Matthew Gamble
School of Computer Science
University of Manchester
m.gamble@cs.man.ac.uk

Carole Goble
School of Computer Science
University of Manchester
carole.goble@manchester.ac.uk

ABSTRACT

The web of linked data is incompatible with the modern "selfish scientist". What is missing is a mechanism that supports both *what* scientists share, and *how* they share. Solutions must be informed by social, technical and cultural issues surrounding the sharing of scientific data in the web of linked data. We propose the adoption of social trust techniques to share a new emerging class of scientific digital object - *Research Objects*. We suggest a mechanism for introducing social trust metrics into the distributed social web to facilitate access control to aggregations of linked data resources. Through the application and analysis of two established trust metrics, we then present the grounding of the Colleague of a Colleague (Cocoa) trust metric suited to the sharing of scientific knowledge delivered as Research Objects.

Keywords

Trust, Scholarship, Linked Data, Sharing, Science

1. INTRODUCTION

Whilst the web was initially designed to assist the dissemination of scientific knowledge and research, some argue that this is the one area in which the web has so far been least disruptive[10]. The semantic web, and in particular the principles of linked data[6], provide a platform through which the exchange and discovery of scientific knowledge can thrive and support the new wave of data intensive scientific discovery[22]. The realisation of this potential is however limited to the level of adoption by the scientific community, and scientists share their data rarely and selectively[16], with levels of sharing varying dependent upon discipline[11]. Scientists are reluctant to reveal their data chiefly fearing that they will not receive the appropriate credit[35] due to the lack of a standard attribution and citation model for scientific data and resources.

Sharing of data can be beneficial and lead to unforeseen collaboration and discovery (cf. Penzias and Wilson's Nobel Prize winning discovery of cosmic background radiation[31]). For the academic concerned with credit, trust plays a crucial role in deciding who to share data with, under what circumstances and during different stages of the scholarly life cycle[16]. In the case of Penzias and Wilson, it was through a trusted intermediary that sharing was instigated. Social networks for scientists such as myexperiment.org[1] are emerging in an attempt to replicate this social scientific community online and encourage collaboration and sharing. As these sites begin to expose their resources into the web of linked data we must meet challenges such as trust, authentication, and authorization and provide the same scalability, efficiency, and utility that has made the Web a success.

In this paper we aim to highlight key technical, social, and cultural issues regarding the sharing of scientific data openly in the web of linked data. These issues suggest that current trust models employed in the social scientific web are either too open and permissive, or too pessimistic and restrictive for the community to benefit from linked data.

The community would be better served by a balanced optimistic trust model, informed by two primary concerns: how scientists share and what they are sharing. We explore social trust metrics typically applied to benefit the consumer of resources and data in making trust assessments[27, 26, 19]. We explore their application from the perspective of the producer to allow a "just in time" analysis of who to share scientific data with. We also adopt the research objects model (see section 2.2) as the mechanism for the encoding and delivery of scientific knowledge and data. The advantages of research objects are twofold, providing assets of collaborative, compound scientific data as well as a means of exposing this data into the web of linked data.

2. SHARING IN THE SOCIAL SCIENTIFIC WEB

Across scientific disciplines, trends in co-authorship demonstrate that academic practice is becoming increasingly collaborative [21]. Harnessing and facilitating this global collaborative trend is the recent application of social networking techniques to the development of Virtual Research Environments (VREs) such as myexperiment.org [14], OpenWetWare[2], HubZero[29] and the Scientific Collaboration Framework (SCF)[12]. Their success has shown that scientists are increasingly prepared to share their experimental data and resources and in turn discover and reuse resources that have been shared by other scientists.

These systems have to be tuned to the motivations for scientists to share their experimental work and data, both pre- and post-publication on the web [30].

Web Science Conf. 2010, April 26-27, 2010, Raleigh, NC, USA.
.

[1] http://www.myexperiment.org
[2] http://www.openwetware.org

2.1 How to Share: Open, Pessimistic or Optimistic?

An entirely open and altruistic approach to data sharing is impractical and undesirable for many. The reward systems for academics raise concerns that by openly sharing their resources and data they risk being scooped and losing credit[35], or their data being misinterpreted[16]. Proponents of open sharing of data instead argue that the potential for wider collaboration and the ability to instantly push ideas out into the community, outweighs the risks.

The Datacite initiative[3] aims to improve the ability to cite scientific research sets and increase their acceptance as legitimate scientific contributions in their own right. Yet until reward systems are updated to reflect these contributions, scientists will remain reluctant to reveal their data.

Investigating the theory of embeddedness in relation to scientists' decisions to share [16] suggests scientists, motivated by these parallel and competing factors, are more willing to share personal data with colleagues within their own *trusted social network*, defined as relations with others that are direct or, via a form of *transitive trust* (see section 3.1), through a number of intermediate nodes. Trust and self-efficacy have been identified as crucial considerations with regards to knowledge sharing in virtual communities[24] and with these closer colleagues, scientists feel they have greater control over the shared data and its interpretation. Crucially they also trust that they will receive the appropriate credit upon any subsequent publication derived from the shared data. The decision to share a resource is also dependent upon its stage in the scholarly life-cycle, with scientists much less likely to share data that is incomplete or supporting an upcoming publication[16]. Furthermore the social network of a member of the academic community is not a static construct. Instead it is dynamic and context sensitive, where individuals may be members of a number of communities (e.g. research groups, projects), and sharing decisions may be dependent upon this context.

Current trust and sharing models employed in the social scientific web broadly fall into two classes which we define as *open* trust models or *pessimistic* trust models.

Open Trust. Inspired by Open Science ideals promoted by groups such as Science Commons[2] and the recently proposed Panton Principles for open data sharing[1], sites like OpenWetWare.org adopt an entirely open and permissive approach to data sharing. By default all pages created by users on OpenWetWare.org are visible to everyone. Although this entirely permissive approach to sharing is admirable, and its success encouraging, the majority are still concerned with openly sharing their data, methods and ideas.

Pessimistic Trust. Alternatively a collaborative VRE such as SysMO SEEK[4] has taken into consideration the complexities of dynamic group and project memberships and controls access to resources through carefully and manually constructed access control lists, using white lists and black lists for individuals and groups. This requires the owner of a resource to explicitly state that they allow another user access. We view this as a pessimistic and distrustful approach to data sharing.

The manual construction of access control lists may scale for small numbers of users or shared resources, and the effort may be acceptable for the most sensitive of resources. However as data sharing shifts to the scale of linked data we require a more scalable solution; one where the user maintains a sense of control and acts as a middle ground between open and pessimistic models.

Optimistic Trust. We define an optimistic trust model as one that attempts to provide a balance between entirely open and pessimistic models of trust. The user maintains a level of control over individual resources however the requirement to explicitly state trust in an entity (individual or group) is removed and instead an attempt to infer members of their trusted social network is made.

2.2 What to Share: Research Objects

To successfully facilitate the sharing of scientific data and knowledge, we must choose an appropriate and scalable representation that reflects the collaborative and compound nature of scientific investigations on the web. Whilst the digital exchange of data is now straightforward, the digital exchange and transfer of scientific knowledge in collaborative environments has proven to be a non-trivial task[7], requiring tacit, and rapidly changing expert knowledge – much of which is lost in traditional methods of publication and information exchange.

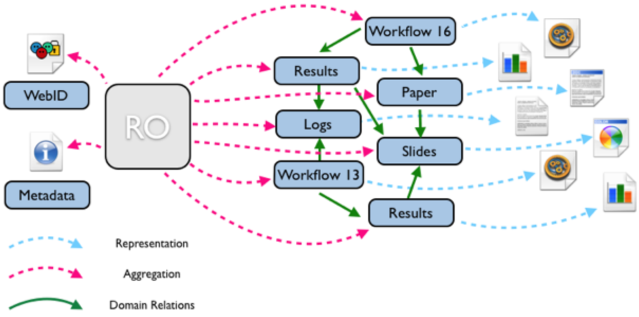

Figure 1: Research Object aggregation structure.

An emerging paradigm for distributing and sharing scientific knowledge on the web, *research objects*[3] are semantically rich aggregations of web and linked data resources (see Figure 1) that are being constructed and applied in e-laboratories projects [9, 13]. Moving away from the flat and inflexible "paper metaphor" for scientific publication, research objects aim to provide rich artefacts that encapsulate the components of an investigation and provide assets of reproducible research that can be shared within and across e-laboratory and VRE solutions. Recognising that during an investigation scientists will work with multiple content and data types in disparate locations across the web, research objects allow the aggregation of resources into one logical entity bringing together the data, methods, and crucially for sharing, people involved. The contents of this aggregation can then be enriched by describing relationships between resources and describing the aggregation as a whole. Current emerging implementations (such as myExperiment's 'Packs') have adopted the Open Archives Initiative Object Reuse and Exchange Specification (OAI-ORE)[32] and can be serialised in RDF to create rich linked data resources, and exposed into the web of linked data.

[3] http://www.datacite.org/

[4] http://seek.sysmo-db.org

Research objects provide us with a suitable mechanism for sharing scientific knowledge as linked data resources. However the nature and composition of research objects – compound content, mixed authorship, mixed stewardship, dynamic resources – introduces some interesting challenges to the realisation of an optimistic trust model.

2.3 Where to Share: Socially Aware Cloud Storage

As we use research objects to expose content through APIs and endpoints such as rdf.myexperiment.org beyond the walls of collaborative environments and into the web of data, we encounter issues of trust, identity, attribution, authentication and authorization. The socially aware cloud storage approach[4] provides a linked data solution to integrating and sharing data across heterogeneous "social networking silos" and into the distributed social semantic web. Realised in the social semantic web access control system [23] it provides a mechanism for authenticating and authorizing users requesting access to linked data resources. The FOAF+SSL protocol [33] is adopted for decentralised authentication. FOAF+SSL provides a simple, distributed, web scale authentication mechanism by linking a user's FOAF[8] file with Public Key Infrastructure to created a trusted unique WebID (the HTTP URI of the FOAF file). It also simultaneously provides the user's social metadata via their FOAF file, which the creators identify could be harnessed to create a rule-based authentication method. Authorization is then managed through the introduction of an RDF access control list (ACL) metadata file linked to each web document, that describes levels of access (read, write and control) for FOAF agents. This RDF based access control method currently suffers from the same pessimistic approach to trust - requiring manual construction of access control lists and an explicit statement of trust in an entity to access a resource.

To introduce an optimistic trust metric into social semantic web access control system, we propose that each web document (e.g. research object) have an access control rule (ACR) also associated with it as a linked data resource. This ACR can then be applied to generate an ACL for the resource as shown in Figure 2.

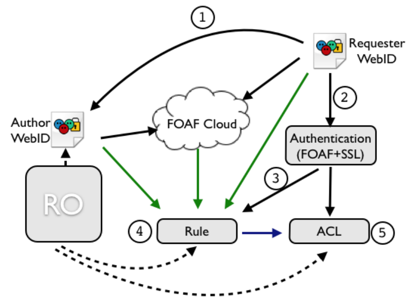

Figure 2: Introducing social trust to access control - extension of the architecture discussed in [5].

The authorization and authentication process is then performed as follows:

1. An individual complete with their FOAF+SSL WebID requests access to the document.
2. Authentication is performed through the same initial FOAF+SSL challenge.
3. If authenticated the ACR is resolved for the document.
4. The WebID of the research object author and any additional metadata required from the document is applied and the ACR computes over the author's social metadata to generate an ACL for this resource.
5. The original authorization step occurs and the requester's access rights are looked up in the ACL.

The use of WebIDs provides a potential mechanism to incorporate multiple contexts, using a different WebID for each. The management of a proliferation of WebIDs is however a concern and would affect the success of such an approach.

3. COCOA: AN OPTIMISTIC TRUST METRIC

Trust is a complex phenomenon and the meaning of the concept itself is the subject of significant study[28]. It is therefore traditional for trust systems and those working with trust to adopt an appropriate definition. Our choice of definition taken from [20] is motivated by the previously discussed concerns with regards to sharing scientific data:

> "[Trust is] the belief in the competence of an entity to act dependably, securely and reliably within a specified context."

Where the context is data sharing, the entity is the consumer of the data, and to act dependably, securely and reliably is to do so in possession of the shared data.

3.1 Social Trust

As the social web and social networking have proliferated, social trust metrics have emerged[18] to exploit the rich metadata available in social network graphs and to help answer the question of who to trust online. Both the insight and the challenge of social trust metrics such as TidalTrust[19] is their foundation on the notion of transitive trust (see Figure 3) where trust of an unknown individual is inferred through a series of direct relationships.

Faced by the volume of users in the social web[5], social trust metrics attempt to use the graph structure of social relationships to estimate how much one user should trust another.

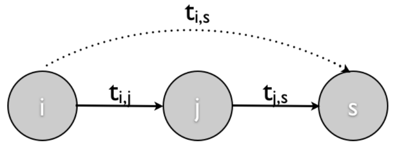

Figure 3: An example of inferring transitive trust $t_{i,s}$ from two direct trust assertions $t_{i,j}$ and $t_{j,s}$ [19].

Social trust metrics are typically applied to aid the consumers of resources in making trust decisions. We instead explore their application to the producers of resources, specifically linked data aggregations delivered as research objects in the distributed social semantic web.

[5] http://www.facebook.com/ currently states over 400 million active users

3.2 Developing the trust model

For the generation of an ACL we wish to identify a trusted community of colleagues using the available social trust metadata. Ford-Fulkerson maximum network flow and Spreading Activation Energy (SAE) model techniques have been applied to the identification of web communities[15] and have subsequently been adapted into social trust metrics to perform the task of identifying trusted communities - Advogato[27] using network flow, and Appleseed[37] using SAE. These trust models are defined generally by [25] as *flow models* and are categorised by [37] as local group trust metrics.

By applying features of both Appleseed and Advogato flow models we aim to develop a suitable trust metric. Our metric and the features we adopt from the two trust metrics must be informed by both the technical and socio-cultural constraints concerning sharing of scientific data in the distributed social semantic web.

Technical constraints

- *Distributed.* Situated in the distributed social semantic web, social metadata is hosted in distributed FOAF files. The metric must be capable of computing over distributed information.
- *Large-scale.* The size of the social graph is variable and potentially large. The metric must therefore scale well.

Socio-cultural constraints

- *Transitive trust.* To reflect an individual's trusted social network we must follow the principles of transitive trust and favour nodes closer to the source.
- *Self efficacy and control.* The user must maintain the feelings of control over the shared resources and self-efficacy, with outcomes of interactions intuitive and predictable.
- *Variable sensitivity of data.* We must support resource dependent levels of sharing.

Though also aware of additional constraints; the *multi-authored* nature of research objects, the *dynamic* and *context sensitive* communities in which trust decisions must be made, we choose to focus on the above constraints for the initial grounding of our trust metric.

3.3 General Energy Flow Algorithm

Our proposed trust metric is a hybrid of both the Appleseed and Advogato trust metrics. The algorithm more closely resembles Appleseed's SAE model and adopts a general energy flow algorithm that recursively computes $e_{x \to y}$, the energy (or trust) flowing between two nodes x and y in our social graph where x has asserted that they trust y. At each recursive stage node x retains a proportion of the energy, increasing its trust rank, and then propagates the remaining energy to subsequent nodes in the graph. This is repeated for each sibling node where energy flows into the sibling. Termination occurs when convergence factors are satisfied. The output is a set of discovered nodes each with a trust rank.

3.4 Cocoa outline

We now present the outline of the Cocoa trust metric:

We have a set of individuals $F = \{a_1, a_2 ... a_n\}$ where each individual $a \in F$ is represented by their unique FOAF+SSL WebID.

Each individual a is associated with a set of trusted individuals $T_a = \{a_1, a_2 ... a_n\}$ produced from the FOAF file metadata where each individual $a_j \in T_{a_i}$ is present in a `foaf:knows` assertion in a_i's FOAF file.

The Appleseed metric applies a single dampening factor $d \in \mathbb{R}^+ : d < 1$ at each recursive step of the metric to achieve trust decay[37]. Where $in(x)$ represents the trust flowing into node x it retains $(1 - d) \cdot in(x)$ to contribute to its trust value and propagates $d \cdot in(x)$ to subsequent nodes. We adopt the dampening factor d from Appleseed but modify the semantics of the dampening factor to represent the *sensitivity* of the resources being shared, such that each resource has an associated sensitivity d. It therefore follows that the higher the sensitivity of the resource the lower one should set the value of d and the less that trust will propagate through the system.

The Appleseed metric allows individuals to explicitly assign weighted trust statements indicating a variable level of trust. Advogato instead performs a pre-processing step over the social graph that computes a capacity in a node, based upon its proximity from the source. As a modification to the pre-processing, we incorporate a function $f(c, l) \to (0, 1)$ that computes a similar capacity where l is the distance from the source (see Figure 4) and c is a user defined capacity factor $c \in \mathbb{R}^+ : c < 1$ where the user can define their confidence in transitive trust statements. A capacity function should behave similarly to the dampening factor such that for lower values of c the less the trust can propagate through the system, favouring nodes closer to the source. A suitable example for a capacity function is $f(c, l) \to c^l$. The capacity function essentially replaces the weighted trust statements used by Appleseed and provides inferred trust statements where our confidence in the ability of nodes to assert trust reduces as the proximity from the source increases.

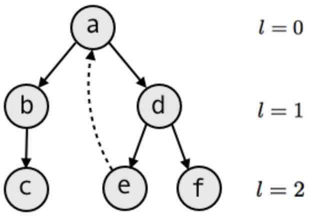

Figure 4: Backwards trust propagation from node e to the source node a.

Backward trust propagation is employed as in Appleseed where each node is given an implicit trust relationship to the source (see Figure 4). The practical advantage of backwards propagation is the removal of potential dead ends where trust flow may be trapped. Rather than weight the backward propagation as a full trust assertion, we instead weight it according to our capacity function. As a result all capacity values for sibling nodes are the same.

Finally as in Appleseed the metric is initiated with an initial energy in^0 into the seed node for whom the trusted network is being calculated. The recursive calculation is then as follows:

$$e_{a_i \to a_j} = d \cdot in(a_i) \cdot \frac{f(c, l_{a_i})}{|T_{a_i}|}$$

Termination occurs as in Appleseed when two convergence factors are met:

- By monitoring the set of discovered nodes at each iteration if no new nodes are discovered from iteration i to iteration $i+1$.

and

- When the trust values for each node in the monitored set converge such that the change in trust value from iteration i to iteration $i{+}1$ is less than a given accuracy threshold.

Appleseed also suggests a number of additional factors to aid termination by setting a maximum path length or maximum number of nodes to unfold. There is also scope for a minimum trust threshold for trust propagating between two nodes.

The output from the metric is a set of discovered nodes along with a trust value for each, where higher trust scores indicate a stronger membership in the trusted community.

3.5 Feature choice

Feature choice from both Appleseed and Advogato was informed by the technical and socio-cultural constraints identified above. Table 1 outlines the key features present in both trust metrics and where they were adopted in the Cocoa trust metric.

Table 1: An overview of features present in Advogato and Appleseed and their adoption in Cocoa.

Feature	Advogato	Appleseed	Cocoa
Weighted Trust	n	y	n
Inferred capacity	y	n	y
Normalization	n	y	y
Deterministic	n	y	y
Dampening Factor	n	y	y
Partial Graph	y	y	y
Backward Propagation	n	y	y

We relate each feature adopted back to the constraints as follows:

- **Weighted trust statements.** Appleseed along with other trust metrics [19] employ weighted trust statements that allow an individual to express a degree of trust (or distrust) in another individual in the system. In a distributed system such as the semantic web where trust is computed by a third party we must assume that all trust information is publicly available. For weighted trust values this introduces privacy concerns as any individual may see the trust values that have been assigned to them. We therefore choose not to adopt weighted trust statements.
- **Deterministic computation** Advogato in its application of Ford-Fulkerson techniques suffers from non-deterministic outcomes of trust attribution. The decision to trust a node in a graph is dependent upon the order in which trust flow is distributed[37]. This can be seen to be unintuitive and in conflict with our desire for self-efficacy and control. Appleseed provides a deterministic computation due to its use of a modified recursive SAE model.
- **Inferred Capacity.** In Advogato a capacity is assigned to each node in a pre-processing step. Capacity is calculated based upon the distance from the seed node and weighted by the average out-degree at each level. The weighting by average out-degree gives the power to individual nodes to affect capacity and therefore trust assignment globally at their level. With the modified inferred capacity value computation $f(c, l)$ we reduce the ability of a node to influence the spread of trust through the system that should not be under their direct influence. We also provide an additional level of user control to manipulate the trusted community independently.
- **Dampening Factor.** The purpose of the dampening factor is modified to represent the sensitivity of the resource being shared. The higher the sensitivity the lower the propagation of trust through the network, resulting in a narrower trusted network. This provides resource level control of sharing.
- **Backwards Trust Propagation.** Trust flow from backwards propagation can be seen to satisfy our requirement of trust transitivity by propagating it back up to the source, and subsequent distribution favouring nodes closer to the source.
- **Trust Normalization.** We adopt trust normalization by weighting the capacity according to the total number of trust assertions made by an individual. Normalization is regularly employed in energy flow based approaches to reduce the influence of eager trust dispensers[37].
- **Partial Graph.** Both Appleseed and Advogato as local group metrics are able to compute their trust rankings based upon a local subsection of the social graph. This is a desirable feature adopted to enable the trust metric to scale well in relation to the growth of the global social graph.

With these features we believe the proposed Cocoa trust metric is well suited to the calculation of a resource dependent trusted social network for scientists, which can then be used to populate an ACL.

4. RELATED WORK

Our aim is to apply social trust metrics that incorporate transitive trust to the task of bootstrapping access control to online resources. There are a number of examples in the literature that attempt to address access control by exploiting social relationships. [36] presents a system with similar objectives, providing a linked data approach to controlling access to images on the popular photo sharing site Flickr through the use of semantic web rules written in the AIR policy language. Whilst the rules reason over a user's FOAF file, the authentication is performed using OpenID and the system does not employ any social trust metrics. However the use of access control rules provides the scope for optimistic sharing of resources. Interestingly the system uses a tag ontology to achieve context sensitive access control, where a user may produce a general rule for sharing of any resources that are tagged with a particular context. Whilst we could potentially achieve multiple contexts through additional WebIDs describing different sets of relationships, the use of a controlled vocabulary is an interesting and potentially useful approach to adopt. The Lockr system[34] also attempts to decentralise and decouple social networking and access control information from current online social networks. In doing so it proposes signed 'social attestations' as a mechanism for representing and asserting a *type* of relationship between two agents, where agents are uniquely identified by their public keys. Access control is mediated through social access control lists which dictate for each resource the public keys of those who may access a resource, along with the particular *type* of relationship they must have with the owner. This can be seen to be a particularly pessimistic model where a resource owner must explicitly create a social attestation and state in a social access control list a particular type of relationship.

5. DISCUSSION AND FUTURE WORK

Here we have introduced the theoretical grounding of Colleague of a Colleague (Cocoa), a simple and scalable trust metric informed by the unique challenges of sharing scientific data openly in the web of linked data and a means of introducing this metric into existing socially aware access control mechanisms.

Our next steps are to implement Cocoa and compare its performance against other trust metrics in real sharing scenarios in the social scientific web, using myExperiment and SysMO as sources of relevant data.

We have focused initially on a single seed or author in our trust network. When considering the multi-authored nature of research objects we must develop procedures to combine trusted networks addressing questions of precedence and conflicting trust rankings. We have also been concerned with a generally static community. A challenge faced by the social scientific web (e.g. SysMO) is that projects, groups, affiliations and therefore trust relationships modify over time. We must therefore investigate the effects of a changing community on our trust metric.

Many other trust-based applications are able to claim that performance is at least as good as a trust-less counterpart[17]. In their application to data sharing it is not necessarily the case that a system employing a social trust metric will perform as well as a pessimistic approach. With limited ability to redress occurrences of accidental sharing, we place optimistic trust models in a spectrum of approaches to sharing scientific data. With pessimistic models appropriate for data of the utmost sensitivity and open models for those wishing to achieve maximum exposure, optimistic models are positioned for the many data sharing scenarios informed by competing concerns. By using an optimistic model to create an ACL we also envision degrees of optimism. The process may be left to be entirely automated or it can instead be an assistive, one-off step at the stage of exposing data, generating a static ACL that can be consulted and repaired before use.

It is by establishing this simple, distributed, and optimistic trust model for sharing rich aggregations of scientific linked data, that we plan to explore methods of data sharing in the social scientific web, and develop a platform to investigate further challenges of trust, attribution, identity, licensing and provenance to the utilisation of linked data by the scientific research community.

6. REFERENCES

[1] Panton principles. http://pantonprinciples.org.

[2] Science commons. http://sciencecommons.org.

[3] S. Bechhofer, D. De Roure, M. Gamble, C. Goble, and I. Buchan. Research Objects: Towards Exchange and Reuse of Digital Knowledge. In *The Future of the Web for Collaborative Science (FWCS)(WWW'10)*, Raleigh, NC, 2010.

[4] T. Berners-Lee. Socially aware cloud storage. http://www.w3.org/2010/Talks/0303-socialcloud-tbl/, 2009.

[5] T. Berners-Lee. Minutes of the w3c social web xg group 03-03-2010. http://www.w3.org/2010/03/03-swxg-minutes.html, 2010.

[6] C. Bizer, T. Heath, and T. Berners-Lee. Linked data the story so far. *International Journal On Semantic Web and Information Systems, Vol. 5, Issue 3*, 2009.

[7] N. Bos, A. Zimmerman, J. Olson, J. Yew, J. Yerkie, and E. Dahl. From shared databases to communities of practice: A taxonomy of collaboratories. *J Computer-Mediated Communication*, 12(2):article16, 2007.

[8] D. Brickley and L. Miller. Foaf vocabulary specification 0.97. http://xmlns.com/foaf/spec/, 2010.

[9] I. Buchan, J. Ainsworth, S. Bechhofer, S. O'Brien, C. Goble, and A. Rector. Public Health e-Labs: An Ethical Model and Architecture for Distributed Epidemiology Using Healthcare Records. In *7th Annual Public Health Information Network (PHIN) Conference*, Atlanta, GA, 2009.

[10] M. Clarke. Why Hasnt Scientific Publishing Been Disrupted Already?,Jan 4th 2010. http://scholarlykitchen.sspnet.org/2010/01/04/why-hasnt-scientific-publishing-been-disru249.

[11] M. H. Cragin, C. L. Palmer, J. Carlson, and M. Witt. Data Sharing, Small Science, and Institutional Repositories. In *UK e-Science All Hands*, Oxford, 2009.

[12] S. Das, T. Green, L. Weitzman, A. Lewis-Bowen, and T. Clark. Linked Data in a Scientific Collaboration Framework. In *The 17th International World Wide Web Conference (WWW2008)*, number April, Beijing, China, 2008.

[13] D. De Roure and C. Goble. Lessons from myExperiment: Research Objects for Data Intensive Research. In *Microsoft eScience Workshop*, Pittsburgh, US., 2009.

[14] D. De Roure, C. Goble, and R. Stevens. The design and realisation of the myExperiment Virtual Research Environment for social. *Future Generation Computer Systems*, (May 2008), 2009.

[15] G. Flake, S. Lawrence, C. Giles, and F. Coetzee. Self-organization and identification of Web communities. *Computer*, 35(3):66–70, March 2002.

[16] R. Giordano. The Scientist: Secretive, Selfish or Reticent? A Social Network Analysis. In *E-Social Science conference*, Ann Arbor, MI, 2007.

[17] J. Golbeck. Weaving a Web of trust. *Science (New York, N.Y.)*, 321(5896):1640–1, September 2008.

[18] J. Golbeck. *Computing with Social Trust and Reputation.* Springer, London, 2009.

[19] J. Golbeck and A. Mannes. Using trust and provenance for content filtering on the semantic web. In *Proceedings of the WWW'06 Workshop on Models of Trust for the Web (MTW'06)*, Edinburgh, Scotland, UK, 2006.

[20] T. Grandison and M. Sloman. A Survey of Trust in Internet Applications. *IEEE Communications Surveys & Tutorials*, 2001.

[21] E. Hand. 'Big science' spurs collaborative trend. *Nature*, 463(7279):282, January 2010.

[22] T. Hey, S. Tansley, and K. Tolle, editors. *The Fourth Paradigm: Data-Intensive Scientific Discovery.* Microsoft Research, Redmond, Washington, 2009.

[23] J. Hollenbach, J. Presbrey, and T. Berners-Lee. Using RDF Metadata To Enable Access Control on the Social Semantic Web. In *Workshop on Collaborative Construction, Management and Linking of Structured Knowledge (CK 2009) (ISWC 2009)*, Washington, DC., 2009.

[24] M. Hsu, T. Ju, C. Yen, and C. Chang. Knowledge sharing behavior in virtual communities: The relationship between trust, self-efficacy, and outcome expectations. *International Journal of Human-Computer Studies*, 65(2):153–169, 2007.

[25] a. Josang, R. Ismail, and C. Boyd. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618–644, 2007.

[26] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina. The Eigentrust algorithm for reputation management in P2P networks. *Proceedings of the twelfth international conference on World Wide Web - WWW '03*, page 640, 2003.

[27] R. Levien. *Computing with Social Trust*, chapter Attack-Resistent Trust Metrics, pages 121–132. Human-Computer Interaction Series . Springer London, London, 2009.

[28] H. Mcknight and N. L. Chervany. *Trust in Cyber-societies*, volume 2246 of *Lecture Notes in Computer Science*, chapter Trust and, pages 27–54. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.

[29] M. McLennan and R. Kennell. Hubzero: A platform for dissemination and collaboration in computational science and engineering. *Computing in Science and Engineering*, 99(PrePrints), 2009.

[30] B. Nelson. Data sharing: Empty archives. *Nature*, 461(7261):160–3, 2009.

[31] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *The Astrophysical Journal*, 142:419, 1965.

[32] A. Pepe, M. Mayernik, C. L. Borgman, and H. Van De Sompel. Technology to Represent Scientific Practice: Data, Life Cycles, and Value Chains. *World Wide Web Internet And Web Information Systems*, pages 1–22, 2009.

[33] H. Story, B. Harbulot, I. Jacobi, and M. Jones. FOAF+SSL: RESTful Authentication for the Social Web. In *SPOT2009 European Semantic Web Conference*, Heraklion, 2009.

[34] A. Tootoonchian, S. Saroiu, Y. Ganjali, and A. Wolman. Lockr : Better Privacy for Social Networks. In *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, page 169180. ACM, 2009.

[35] M. M. Waldrop. Science 2.0. *Scientific American*, 298(5):68–73, 2008.

[36] A. Yeung, C. Man, L. Kagal, N. Gibbins, and N. Shadbolt. Providing Access Control to Online Photo Albums Based on Tags and Linked Data. In *AAAI Spring Symposium on Social Semantic Web: Where Web 2.0 Meets Web 3.0*, Stanford, CA., 2009.

[37] C. Ziegler and G. Lausen. Spreading Activation Models for Trust Propagation. In *Proceedings of the IEEE International Conference on e-Technology, e-Commerce, and e-Service*, page 8397. Citeseer, 2004.