Bringing Communities to the Semantic Web and the Semantic Web to Communities

K. Faith Lawrence
University of Southampton
Highfield
Southampton, United Kingdom
kf03r@ecs.soton.ac.uk

m. c. schraefel
University of Southampton
Highfield
Southampton, United Kingdom
mc@ecs.soton.ac.uk

ABSTRACT

In this paper we consider the types of community networks that are most often codified within the Semantic Web. We propose the recognition of a new structure which fulfils the definition of community used outside the Semantic Web. We argue that the properties inherent in a community allow additional processing to be done with the described relationships existing between entities within the community network. Taking an existing online community as a case study we describe the ontologies and applications that we developed to support this community in the Semantic Web environment and discuss what lessons can be learnt from this exercise and applied in more general settings.

Categories and Subject Descriptors

H.5.3 [**Information Systems**]: Information Interfaces and Presentation

General Terms

Human Factors, Design

Keywords

E-Applications, Semantic Web, Communities, Case Study

1. INTRODUCTION

In this paper we will consider applications created to support virtual communities that exist within the Semantic Web. Research into Semantic Web communities has focused on two types: Communities of Practice (COPs) [20, 17] and social networks, especially Web Based Social Networks (WBSN) [12] such as Friend of a Friend [4, 8]. Comparing these types of networks with the definitions of community prevalent is other areas such as virtual communities [18] and offline in the social sciences [21] we have to question whether COPs and WBSNs qualify despite the frequent application of the term.

We propose the concept of a Internet Based Community Network (IBCN) which contains the properties of both a COP and a social network in such a way that it fulfils the definition of a community. Using the case study of the Amateur Fiction Online community we argue that the properties inherent in a known community structure allows for a

new paradigm through which interaction occurs. Leveraging these properties we can provide greater incentive for user interaction and lay the foundation for new types of semantic services.

Below we describe the ontologies and applications that we have designed in collaboration with the above virtual community to support it becoming a IBCN. We consider what lessons we can draw from the community reaction and interaction and how these can be applied to other areas and groups on the Semantic Web.

2. COMMUNITIES ON THE SEMANTIC WEB

2.1 COPs, WBSN and ..?

The idea of an online or virtual community was pushed by early researchers in the field such as Howard Rheingold. He defined such communities as "social aggregations that emerge from the Net when enough people carry on those public discussions long enough, with sufficient human feeling, to form webs of personal relationships in cyberspace."[19]

In 'Widening the Net', Whittaker et al [21] detail the definition for community that came out of a workshop on that subject at the Computer Supported Cooperative Work conference in 1996. Rather than coming up with a specific definition they "settled on an approach of defining the concept by 'prototypical attributes', so that communities with more of these attributes were clearer examples of communities than those that had fewer."[21]. The core attributes were as follows:

- Members have some shared goal, interest, need, or activity that provides the primary reason for belonging to the community

- Members engage in repeated active participation and there are often intense interactions, strong emotional ties and shared activities occurring between participants

- Members have access to shared resources and there are policies for determining access to those resources

- Reciprocity of information, support and services between members

- Shared context (social conventions, language, protocols).

As can be seen this definition is context neutral in that it can be applied to both virtual and offline communities. In comparison Preece [18] defines an online community as containing four components:

- Social interaction

- A shared purpose

- A common set of expected behaviours

- Some form of computer system which both mediates and facilitates communication.

If we compare the two definitions we can see that the first item on the list corresponds to Preece's requirement of shared purpose while the second is roughly parallel to Preece's social interaction. What is interesting here is that the Whittaker list does not use the term "social" instead focusing on active participation leading to shared activities or emotional bonds. In many ways this is a much broader definition than Preece's which is to be expected given their more inclusive agenda. The final obvious match is between the final item on Whittaker's list of a shared context including social conventions which fits with Preece's commonly expected behaviours. The final two items on Whittaker's list are notable partly for their separation. There is an implicit assumption in the third statement that there are shared resources. In the case of a virtual community these resources may well be the information, support and services that are mentioned in the fourth statement. Furthermore in online communities the resources mentioned in the third statement and those mentioned in the fourth are tied into the computer-mediated aspect of that community in that the form of computer-mediation determines what resources, including services, are available for sharing and the level of reciprocity is controlled by a combination of this and the social interactions also mentioned by Preece.

A Community of Practice (COP) gains it community application through the inferred links that develop among a group that have a common task: "What is shared by a community of practice – what makes it a community – is its practice. The concept of practice connotes doing, but not just doing in and of itself. It is doing in a historical and social context that gives structure and meaning to what we do. When I talk about practice, I am talking about social practice."[20] More recently the definition has lost even that much social notation to become "a relatively loose, distributed group of people connected by a shared task, problem, job or practice"[17].

Looking at a semantically based social network such as the WBSNs proposed by Golbeck [12, P.13] we are provided with the following requirements:

- Accessible via web browser

- Relationships/Links between users must be stated. Must be direct links not inferred

- System must support the creation of these relationships/links

- Relationships must be visible and browsable

Comparing these two definitions with those of communities, above, we can see (Table 1) that while a COP or a social network *may* describe a community it is not a guaranteed assumption which can be made about the system described by either type of network. For this reason we propose the identification of a third type of group, the Internet Based Community Network, which might be both a COP and a social network but also fulfils all the definitions for a virtual community. Services can then be designed and run with the assumption that the network it is being run on acts in ways specific to a community. By identifying those attributes and behaviours that are unique to a IBCN we can return to the non-community based networks and consider how other observations seen in IBCNs and seen as beneficial can be applied.

3. CREATING AN INTERNET BASED COMMUNITY NETWORK

3.1 Amateur Fiction Online Community Case Study

The Amateur Writing Online Community is an example of a community which has flourished on the Internet. The community is made up of two mostly separate group-media inspired or fan authors and 'original' authors. Of these two the former are more vocal online because, through necessity, they have eschewed traditional publishing in its official form and thus rely on community published works, or zines, and individual dissemination. While paper based zines still continue to be used as a means of distribution, the ease of electronic publication to share works has led to a massive migration to the Internet. Websites, online archives, mailing lists, IRC, electronic journals, even wikis... the community has embraced the many available methods of online communication.

Studies done with this community have traditionally focused on the social and media [14, 1, 6, 2] aspects rather than the technology which the community has adopted, and in some cases subverted, to support themselves. While analyzing the social interaction we have tried to understand how this relates to the technical issues that face the community and how new technologies could improve on the current systems. Direct observation, questionnaire and interviews were used to gain understanding of the specific difficulties and issues that community members face. Following initial discussions with community members as part of a preliminary assessment, a questionnaire was undertaken to gain greater insight into the issues that could be addressed through technological support. This method was chosen because it allowed us to involve a larger section of the community than would have been possible through alternative methods.

Taken altogether amateur fiction probably represents one of the largest online electronic libraries currently in existence (see Table 2). Unfortunately it is one which doesn't come with a catalogue and is constantly in flux. While the larger archives are fairly fixed in their position, the smaller archives and personal pages are frequently changing address, going down temporarily for maintenance or due to bandwidth limitations or just vanishing. "Can anyone tell me where to find..?" is a frequent question on many lists as is the popular "Can anyone recommend..?" or "I am trying to find a story that contains...".

Table 1: Comparison of Links between Communities of Practice and Social Networks

Community Type	Links	Focus	Preece "community"
Community of Practice	Inferred	Practice	Social Interaction not required
Social Network	Explicit	People	Shared purpose and behaviours not required
IBCN	Explicit and Implicit	People as community members	

Table 2: Small Selection of Multi-Domain Amateur Writing Archives[1]

Archive	Authors	Stories (Poetry)
FanFiction.net	–	1, 087, 412
FictionPress	125, 206	214, 536 (489, 620)
AdultFanfiction.net	35, 871	41, 994
Freedom Of Speech	2, 337	4, 900
Slashfanfiction.net	1, 504	2, 652
WWOMB	1, 431	6, 717

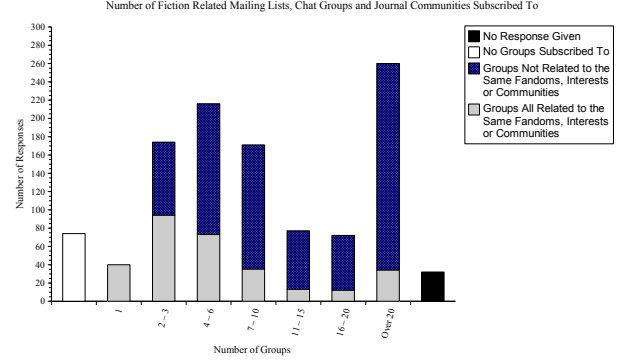**Figure 1: Community Involvement of Respondents**

A large part of this ever changing nature is attributable to the subject matter and the very amateur nature of the enterprise. Complaints, witchhunts and legal threats can force relocation while lack of resources or loss of interest can remove sites entirely. All of which add additional complications to an already diverse system of differing standards and expectations and vocabularies. As a group they represent a community for whom the Semantic Web has a lot to offer, and from whom we can draw many lessons in bringing technology to traditionally overshadowed user groups.

3.2 The Semantic Community

Dr. Golbeck defines a Web-based social network (WBSN) as a social network accessible through a web browser which explicitly states and supports the relationships between people and allows these relationships to be navigated [12, P.13]. How does this differ from an online community? If we consider the definitions put forward by Whittaker et al. [21] and Preece [18] for virtual communities and compare these with WBSNs it can be seen that both have a social aspect and are computer-mediated. What is missing is from the requirements for a WBSN is the shared set of behaviours and the shared purpose. It could be argued that wanting to be part of a WBSN is a shared purpose but that seems an unjustifiably vague bond.

From the other perspective what is preventing a community being a WBSN is the explicitly stated links and the means to make use of them. Electronic journals such as LiveJournal and JournalFen, an online journaling site for adult media fans, provide some measure of WBSN but for the most part the relationships between fans or writers is not explicitly stated. The question is then: why not? Community fellowship, especially mailing-list or group memberships, are frequently used as a networking tool within the community. Friendships are often carried over between the smaller community groups and these links between groups, authors and resources are correlated in general terms via URLs (see Fig. 1). A WBSN may not be an online community and vice versa. What is proposed is the creation of a Internet Based Community Network (IBCN) which combines the attributes of both a WBSN and a virtual community.

Despite the adoption of LiveJournal and JournalFen by

amateur writers, this usage hasn't extended to use of FOAF. This is partly due to ignorance and partly due to lack of incentive. The prospective system doesn't offer anything to a member of the amateur writing community that they are not already getting from the currently used systems. The Semantic Web offers many opportunities that are not otherwise available because, while the community already generates a large amount of human-readable information, it does not also associate machine-readable metadata. This would allow better integration of the distributed systems, improved searching and filtering and more personalised services. These could benefit the experienced user by expending their options and creating new ways with which they could interact with the community as well as aiding the newcomer by easing their introduction into a community which has its own expectations, unwritten rules and obscure terminology.

The Fan Online Persona (FOP) ontology was designed as an extension to FOAF. The extension was created for two main reasons: to clarify the options available and applicable to the needs of amateur writers and their related readers and to de-emphasise those which were not required. The reason for doing this is to create a way of making a persona and community declaration which can be used in a IBCN. This is intended to be in line with Friedman's stated ideal "to maintain a persistent identifier within each social arena without relying on the verification and revelation of true identities"[11, P.177].

While the persistence of a FOP identity cannot be guaranteed the advantages related to reputation in the amateur writing community ensure that is a greater continuity of identities than might be otherwise expected in a non-community based system. "While anonymity prima facie looks like a threat to the building of reputation, it may indeed enhance its relevance. For the sake of reputation, and because of its presence, changing identities has a cost"[5, P.179]. The cost of creating a new identity – whether it is

---

[1]Figures obtained from respective archives on 13/06/05.

in time, money, reputation or some other commodity – is known as identity cost. "Identity cost can take any positive finite value, and users decide whether to white-wash depending on how the identity cost compares to the penalty imposed on free-riders and newcomers"[10, P. 232].

Within the amateur writing community the majority of the resources are going to be freely accessible to anyone, as can be seen by their current availability. This means that even 'untrusted' newcomers will have opportunity to be fully involved in the community process and thus gain trust. It is only if an author is concerned about making her, or his, work freely available, for example, due to the sensitive or adult nature of the subject matter dealt with, that the identity of the person trying to access the work becomes an issue.

Since we are dealing with a IBCN we can assume that there is a shared set of behaviour values to which users need to conform or risk losing reputation. However reputation is gained simply by taking part in the social interaction which defines the community structure. Additional reputation may be earned through creation of works of media which results in their being known as a writer worth reading. Friedman sees "the distrust of newcomers is an inherent social cost of easy identity changes"[11, P. 176]. This fits in remarkably well with the online community system because it is expected for members to know the rules of the community and newcomers need some time to learn what the expected behaviours are.

By studying common practice it was possible to divine what was seen as important by the community. The FOP ontology also covers explicit descriptions related to the types of activities that interaction with the amateur writing community flagged as being important. Through the extension of `foaf:document` and `foaf:groups`, FOP allows users to specify details of their own creations and make recommendations about others' work. The creation, exchange and review of works and ideas are the *raisons d'etre* of the online amateur writing community. Where FOAF files consider work and school information, FOP files dismiss as potentially dangerous and irrelevant when compared to specifying which archives store a persona's work.

To keep the FOP ontology simple, information was restricted to the bibliographic level and leave details related to the content to additional resources. This decision was mainly due to the contentious nature of issues such as ratings and warnings. Since a more complex system was needed to meet the needs in this area it made sense to leave it to a more complex system. The OntoMedia ontology describes the content of media, see below, and is intended to complement the FOP ontology. While not totally integrated the two ontologies were envisioned to work together with the OntoMedia detailing media items that the simple FOP structure lacked and biographical information did not cover.

The extended options also include the ability to indicate the type, theme and restrictions for discussion forums to which the persona belongs or administrates. The options presented in the FOP file were determined through a long term study of the metadata that is attached to works posted to mailing lists or distributed on websites. This additional data provided within the FOP file will open the way for many Semantic Web services such as recommender or notification systems.

It will also allow the social networks of amateur writers to be studied both at the friends- and the collaboration-level.

The tools used to analyse communities of practice, especially those designed with academic publication as a primary case study, will be equally applicable to communities or non-academic authors allowing the creation of virtual reading groups. The development of such groups would provide a possible alternative to the system of having groups dedicated to a single author or theme.

An application to allow users to create FOP files is currently under development.

4. DESCRIBING MEDIA CONTENT

4.1 Why Describe Content?

The Online Amateur Writing community produces many media objects. While the majority of these are textual, images and multimedia items are also created and shared alongside the texts. As mentioned previously, the output of the community is both large and diverse. Infinite Diversity in Infinite Combinations (IDIC) had been one of the guiding principles from before the Internet was invented. While promoting creativity it can lead to difficulties for readers trying to find or avoid a specific type of stories. To aid readers it is customary to post additional information about the story. This goes beyond basic bibliographic details to include hints as to the content of the story such as level of adult detail, the presence of relationships and the presence of any controversial subject matter.

The question of what metadata should be available to readers resulted in a very mixed response from the community (see Fig. 2). Some people wanted as much detail as possible before making a decision on whether or not to read a particular story. Others only wanted to know the basic bibliographic details feeling anything else spoiled the plot. While there was a match between the metadata desired and the metadata given on the basic details the gap between the two grew on more contentious issues (see Fig. 3). Beyond this the language used in the header information was frequently drawn the specific vocabularies used within the community for example: "slash", "PWP", "saffic", "dldr", "MPreg", "gen" and "Pumpkin Pie".

To an outsider or new member of the community these terms can be very opaque in their meaning. Worse, the implied meaning of many of the terms is only vaguely defined and can change depending on the context of their use. Since the terms are evolving and even the community cannot decide on a definitive definition, it became clear that any ontology used would need to be independent of these terms. This has the advantage of not tying the ontology in to one usage and opens the way for personalized vocabulary definitions, something welcomed by the members of the community interviewed since it did not favour one world view over another.

The amateur writing community already spends a significant amount of time associating metadata with the media items that they create (see Fig. 4). This is perhaps aided by the strong overlap between the writers who add the metadata and the readers who make use of it, thus reducing the split between those who do the work and those who make use of it. It is encouraging that the majority of those community members asked responded that they would be willing to consider spending slightly longer adding metadata if it would raise the visibility of their work among its target audience with a strong interest also being shown in metadata
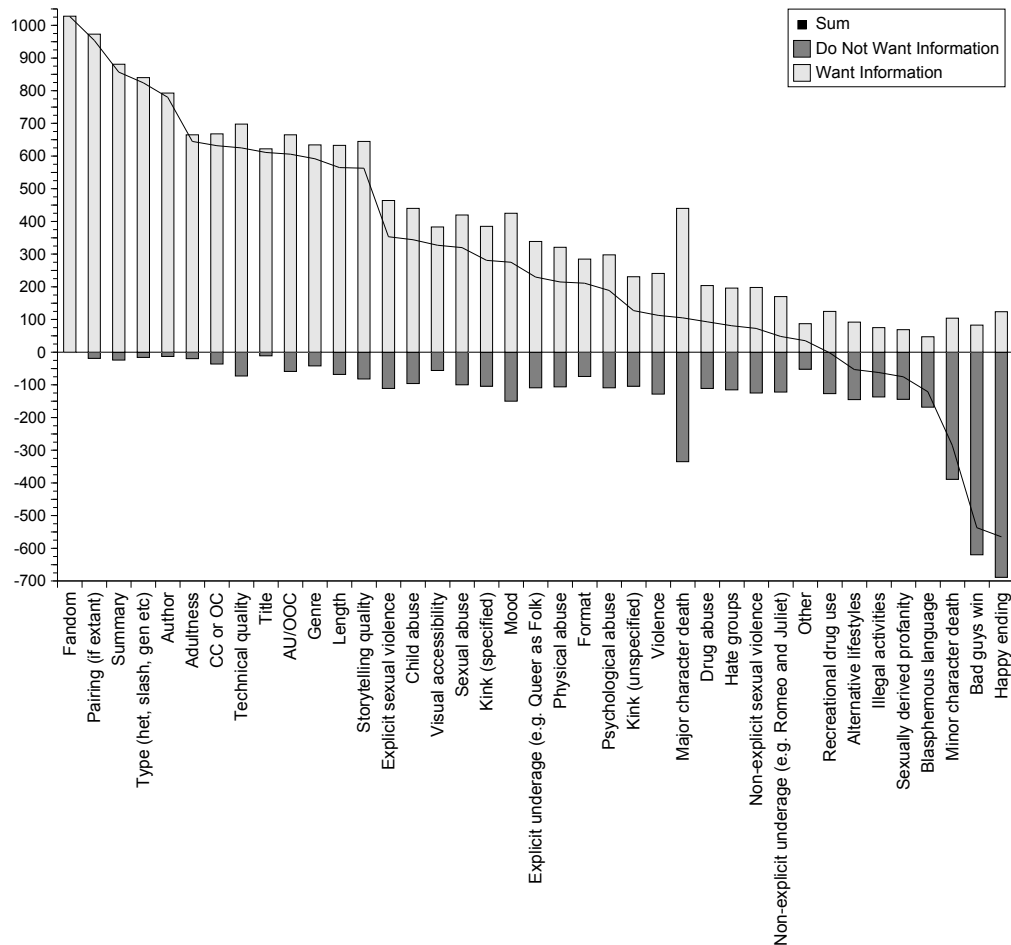
Figure 2: Metadata Preferences

re-usability (see Fig. 5). Having confirmed that the community would be interested in such an ontology, the question arose - which ontology should they use?

4.2 Existing Ontologies

With the aim of modelling the contents of fiction, whatever media format it is presented in, we faced the potentiality of having to deal with anything that the human mind can come up with. The only way to deal with this lack of limitations is to plan for it. By making the ontology modular and expandable we leave the option open for those situations that arise which we had not envisioned. Beyond acknowledging that we are not going to be able to cover every situation, supporting extensibility allows us to reuse existing ontologies where they already exist. The concept of a top level conceptual model to act as a bridge for media ontologies and standards is not new.

The ABC ontology designed by Lagoze and Hunter was intended to "integrate information from multiple genres of multimedia content within digital libraries and archives"[15]. This ontology focused on factual information both within the document and at the level of document creation, provenance and rights management. The strength and weakness of this approach was that it tried to describe everything using the same basic structures and had its roots in factual events. With the OntoMedia model we were concerned not only with the facts but also the possibilities and concepts that existed behind those facts. The ABC model includes the 'abstraction category' as a way of expressing ideas which do not exist in the context of a situation but sees this primarily as a way of binding together various manifestations of the same intellectual expression. Conversely the OntoMedia structure deals primarily with abstractions that exist behind the content. While our structure, like the ABC model, was based on the concepts of entities and events, we specifically focus on the needs of best describing content from both sin-

Normalised Comparison of Calculated Information Desired and Information Provided
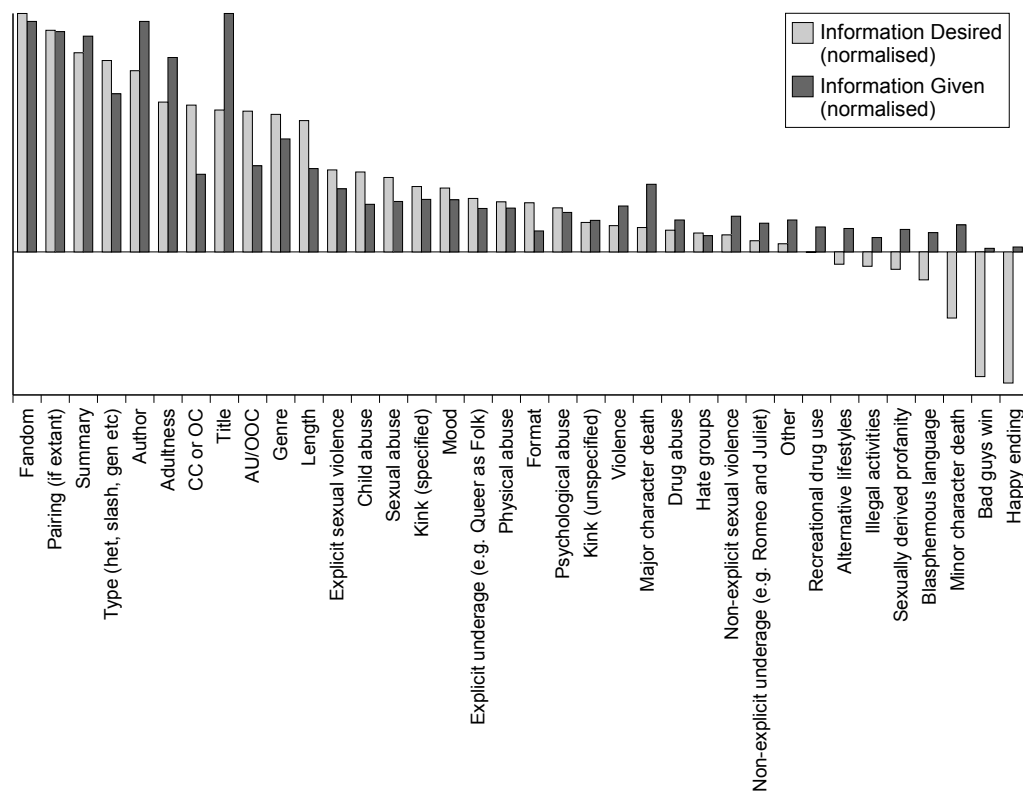

Figure 3: Currently available Metadata

gle and multimedia. In this way the OntoMedia ontology can be integrated with the ABC model if required but otherwise provides augmentation and an alternative method of interfacing with the expression of that content.

The CIDOC Conceptual Reference Model (CRM) was created as a semantic approach to integrated access of cultural heritage data. There are many similarities between the underlying concepts of cultural heritage and fiction – both are concerned with people and events. The main difference between the two being the type of evidence that exists of those manifestations and where they were believed to have taken place. The fictional aspect of the historical narrative has been commented on by figures as diverse as Plato and Churchill. The historical aspect of fictional narrative might refer to an imaginary history, but a history of events none the less.

Despite their initial similarities, the CIDOC Conceptual Reference Model (CRM) is still concerned with equivalent of bibliographic data for physical objects. It can describe in detail the condition, provenance, and attributes of an artefact, such as a red-figure vase, and part of this description includes information on the decorative scenes but without placing the depictions with a narrative context. The summarized scope of the CIDOC CRM is "the curated knowledge of museums"[7]. This intended scope is expanded to being able to describe "all information required for the exchange and integration of heterogeneous scientific documentation of museum collections"[7]. While having a similar model, the description of internal content narrative is beyond the stated scope of the project. Like the ABC ontology the CIDOC CRM works best in parallel with the a content description ontology and a collaboration between the two opens the way for the integration of abstract myths, traditions, and concepts (both written and oral) with the material evidence, allowing for any additional relationships to be explored (see Fig. 6).

4.3 The OntoMedia Model

With The OntoMedia architecture is based on an Event and Entity model (see Fig. 7). We define an Entity as an object or concept. An Event describes an interaction between one or more `ontomedia:Entity`s during which zero or more attributes of those `ontomedia:Entity`s are modified or a new `ontomedia:Entity` is created. An `ontomedia:Entity` may have an attribute set to show that it no longer exists but the `ontomedia:Entity` itself is not destroyed. A number of Events subclasses are defined. These initial divisions were based on the most common types of events found in literature. The subclasses of the Entity construct fall into three different types. Those related to objects both physical and abstract (`ontomedia:Being`, physical, and `ontomedia:Item`, both physical and abstract), those related to spacial models (`ontomedia:Space`) and those relating to time (`ontomedia:-Timeline` and `ontomedia:Occurrence`).
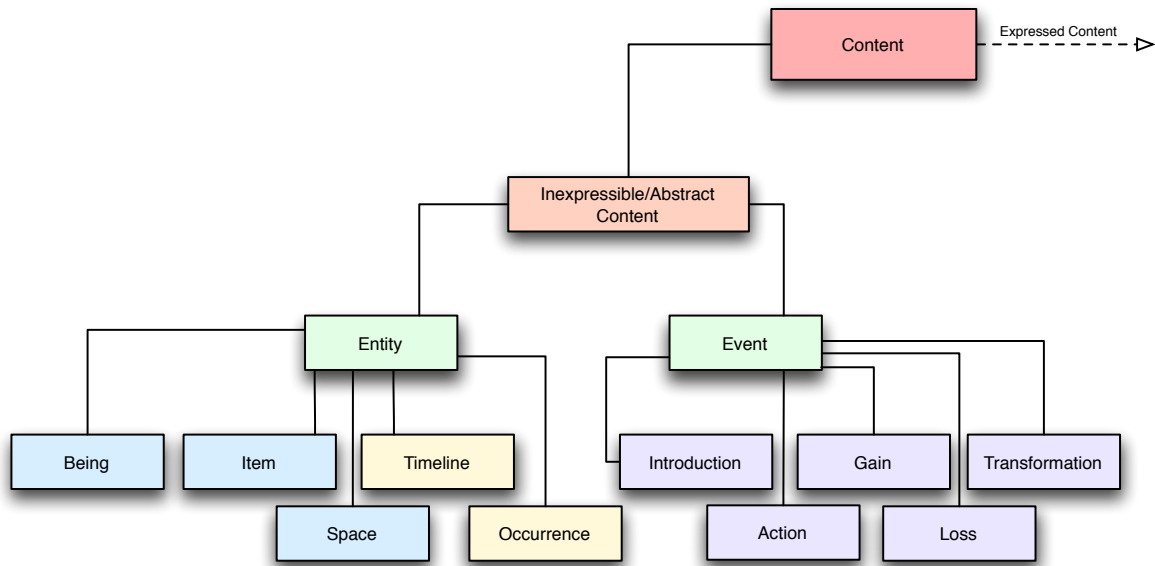
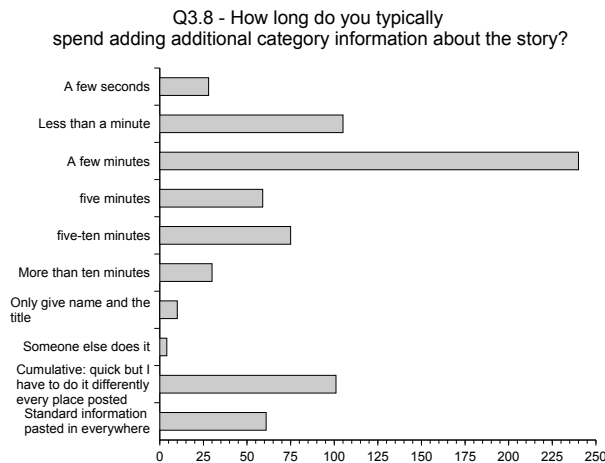**Figure 7: OntoMedia Top Level Structure**

Figure 4: Time Taken to Add Metadata

Figure 5: Additional Time For Semantic Metadata

The class `ontomedia:Item` is subclassed into `ontomedia:‑Physical_Item` and `ontomedia:Abstract_Item`. `ontomedia:‑Abstract_Item`s describe concepts, ideas and other entities which do not have a physical presence. This includes `onto‑media:Trait` and `ontomedia:Being_Trait` which describe traits of the entity such as physical appearance, age, gender, species, material et cetera. An `ontomedia:Physical_‑Item` can be any type of corporeal object.

For use in fiction `ontomedia:Being` class was extended to `ontomedia:Character`. Entities of the `ontomedia:Character` type are defined as fictional entities which have a personality. For example a toaster would be a type of `ontomedia:‑Physical_Item`, a subclass of `ontomedia:Item`, whereas t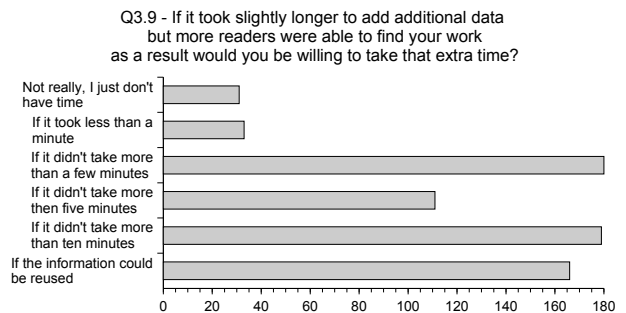he talkie toaster from the television show Red Dwarf and the related books would be classed as an `ontomedia:Character` despite having the physical appearance of a toaster.

An `ontomedia:Occurrence` is a specific instance of an event which occurs within a single `ontomedia:Timeline`. The reason for this is that any `ontomedia:Inexpressible_‑Content` can have a `ontomedia:Timeline` associated with it. It is therefore likely that more than one `ontomedia:‑Timeline` instance will need to be defined for any given work. One event will therefore occur on multiple `onto‑media:Timelines` but the relationship between the events on any given `ontomedia:Timeline` will not necessarily be the same as on any other `ontomedia:Timeline`. For example, if we define one `ontomedia:Timeline` to describe the events which occur in the narrative and another to describe the events chronologically these two will differ whenever we encounter a 'flashback' or one of the characters mentions either a historical event or something from their past. This flexibility regarding time is especially important when describing fictional contexts and events since stories may in-
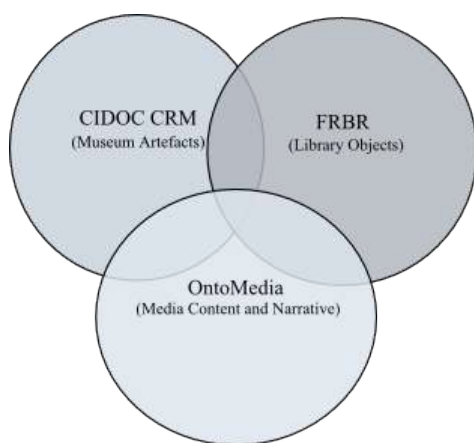
Figure 6: Scope of CIDOC CRM, Bibliographical data and OntoMedia

volve time travel and related paradoxes. While a 1:1 mapping exists between an instance of `ontomedia:Timeline` and an instance of `ontomedia:Occurrence`, a 1:Many relationships exist between any instance of an `ontomedia:Event` and the instances of ontomedia:Occurrence that contextualise it. This can even allow more than one occurrence of the same event to exist on the same timeline, for example if a character meets their future self on their personal timeline this event will occur twice, once when they are the younger version of themselves and once when they are the older version.

5. CREATING AN ONTOMEDIA APPLICATION

5.1 Meditate

Meditate was envisioned to act as a interface between the user and the ontology (see Fig. 8). To require expert knowledge or even familiarity with the underlying ontology immediately limits the number of users for the system. This limitation is especially relevant when the intended user base is made up of people for whom a high technical literacy cannot be assumed.

The basic premise of the program was to automate as much of the metadata creation process as possible and simplify what remained. One of the concepts that had been associated with the OntoMedia from the beginning was that of the creation of Entity and Event stores. This idea presupposed that certain elements would be reused on a frequent basis and by making these elements shareable would ultimately speed up and improve the marking-up experience for the user. This is one area where the nature of amateur fiction, especially the fan fiction sub-genre, lends itself well to the theory. It is frequently said that all the stories have been written and there no original ideas left. Whether this is true or not there are demonstratively well known elements that frequently re-occur: the daring rescue, the destruction of Tokyo or the Chrysler Building, the relationship that develops between the two main protagonists, et cetera. Meditate is intended not only to allow the user

to create new entities but also to interact with entities that have already been created and have been made available. The user will not actually be able to edit the shared entities, instead the entity will act as a template of values for a new instance. This new instance will belong to the user and exist in their `ontomedia:context` but will identify itself as having a `shadow-of` relationship to the instance used as a template.

An early version of the software was shown to a group of potential users at a small British fan convention at the beginning of October (attendance approximately sixty people). The program was demonstrated to a self-selected initial group over the course of the weekend. By the Sunday evening a number of the users felt confident enough to use the software unaided and indeed instruct other fans in its use. Contributions were added by nine people covering twenty TV shows, movies, books and original universes and nearly two hundred characters.

While a few areas were highlighted as needing more simplification or explanation, the overall response was exceedingly positive. So much so that more than one person wanted to know whether they could have a copy of the program. Failing that, whether it was possible for them to collect more data together in a spreadsheet or some similar file which could then be submitted later so that they could continue to add information after they had gone home. When it was explained that the system was not currently capable of interaction with that format the users wanted to know whether it was possible for them to be given an RDF file to use as a template for filling in extra character data. This is a notable contrast to the typical response to data entry.

Beyond this an interesting phenomenon was observed with regard to which characters were entered. The first few users to test the system just added a few random, favourite or funny characters. A few users happened to have lists of characters and their personal information with them - in these cases the characters were added as they appeared in the list. However more people became involved as they were asked for specific information about which they were perceived as knowledgeable or 'expert'. Once involved the experts frequently caused the group to deviate from the lists they were using to take advantage of the additional knowledge base. This soon evolved into something of a game in which the users tried to come up with as many characters as possible within a given domain. The result was that one hundred and twenty-eight of the nearly two hundred characters came from the same series of books with most of those entries coming from a group effort by four individuals. Two conclusions can be drawn from this - firstly that the display and exchange of knowledge can act as a way of gaining reputation within the community as well as bonding members of the community. Secondly, that because of this, community interest can drive people to spending more time than when they are acting as individuals.

5.2 Applying the Semantic Web to the Amateur Community – Harnessing Obsession

One of the current issues facing Semantic Web development is the production of data and metadata. Cory Doctorow [9] highlighted seven problems that stand in the way of reliable metadata annotation. Of those seven, three relate directly to the users of the system - that they lie, are lazy and are stupid. While rarely put in such terms this
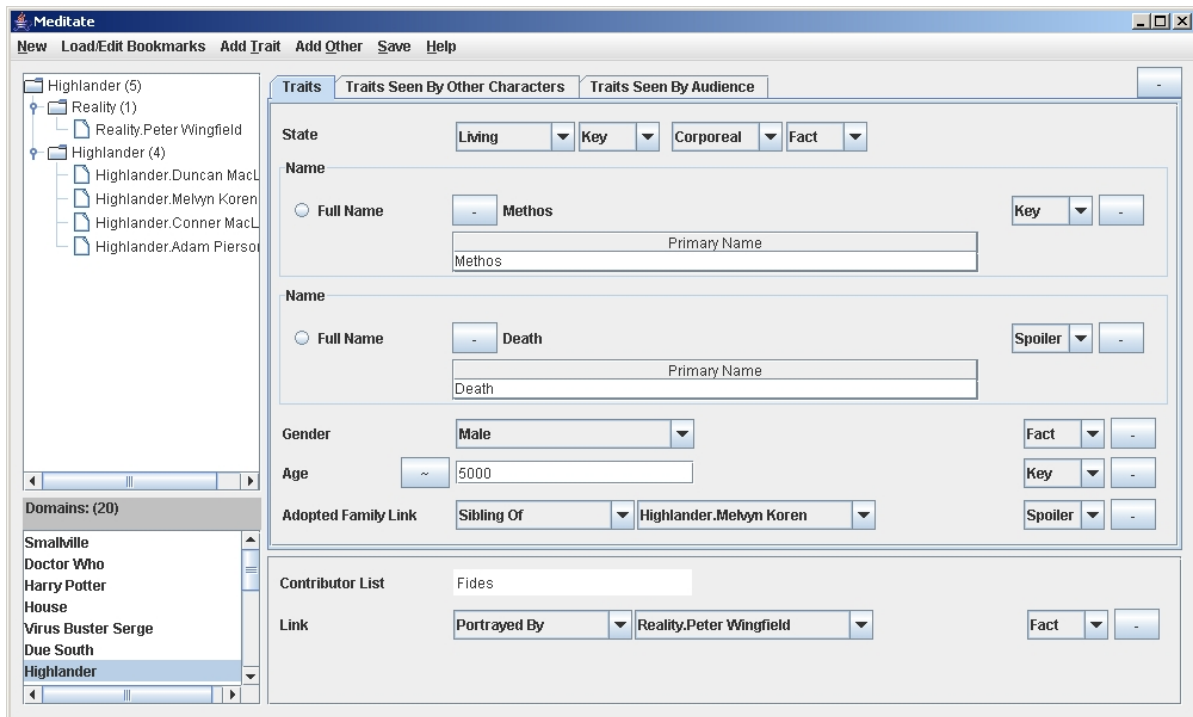
Figure 8: Meditate Screenshot

argument illustrates two recognised dilemmas - how to get people to add data and how to ensure the accuracy of that data.

Work is being done on automatic metadata creation but this currently requires a corpus of material which can be used as a training system and is constrained by advances in natural language processing and artificial intelligence. The more unique the object or the more subtleties of language it contains, the more difficult it will be for an automatic system to correctly identify how the object should be categorised.

The alternative is the manual addition of metadata. People will put time into something for two main reasons - because it is something they care about or because they stand to gain from it. Gains, perceived or real, can take many forms but need to be proportional to the amount of work required. Data entry is universally recognized as being both time consuming and boring and yet given the right incentive people will do it voluntarily. Examples of this can be seen throughout the Internet - the World Wide Web took off when the average user was able to easily create websites about the things that interested them. Amateur authors, and fan authors in particular, are hobbyist orientated, writing for the love of it and their love of the characters. The online communities that support them are based on concept of sharing that love. This translates to the sharing of electronic documents, with accompanying metadata, and the sharing of information. Given the opportunity these types of communities want to, and are, creating and accuracy checking the type of data that will power semantic applications.

6. DISCUSSION

In this paper we have described our work with a commu-

nity whose *raison d'etre* is to share data. This is not a unique attribute. The word 'amateur' is derived from the French verb meaning "to love". There is an increased recognition of the power of the amateur [3], that they can (and often will) spend more time and effort on their interests because of the love that drew them to the interest than they would from a purely professional interest. Humans were motivated by love long before they were motivated by money.

Amateur communities exist through the sharing of the interest that brought them together and, therefore, data related to it. One of the problems that is often noted in Semantic Web research is the difficulty associated with persuading people to create and distribute data. Research into automatic annotation specifically tries to address this problem but has its own issues and limitations. In amateur communities we see a structure where the creation and sharing of data is, if not the goal, then embedded in the very foundations of the community infrastructure.

By working with such communities we have the opportunity to work in an enthusiastic, data rich environment. Further by studying the community interaction we can investigate how the practice of data sharing becomes not just good practice but expected behaviour. The question is then whether factors governing this peer-driven community sharing can be translated to non-community networks.

The community peer pressure towards involvement gains an added dimension when we consider the composition of those involved. Traditionally early adopters of new technologies are seen as men, and often young men. Incidently it is this demographic that is most often targeted with regards to design and publicity. However this is not always the case, for example, Herringetal.[13] argues that "a historical

account of weblogs that accorded a central place to personal journals - as their prevalence merits - would thus identify females as the creators, early adopters, and most characteristic current users of weblogs". While the age range of the community went from approximately eleven to over sixty [16], the community also has a traditionally female bias [14, 1] with the percentage of women rising to the nineties on some areas[2] While we do not suggest this is typical for an amateur community it raises the question of the constitution of the communities that we work with and the potential effect that this might have on adoption.

7. FUTURE WORK

Work is continuing on applications to allow the community to create data objects, nicknamed data crystals. These discreet entities can then be shared and reused throughout the community. We intend to create a browsable, collaborative store of these data crystals with which the community can interact. We can then investigate the use and reuse of these RDF templates as a way of allowing users to annotate their media items with minimal effort and no interaction with the raw data. We maintain that the advantage of this approach will prove to be that, by supporting the creation, reuse and personalisation of commonly desired data crystals, it will speed up the annotation process in the medium to long term while any additional time spent on the short term will be offset by the more user-friendly approach. We believe that this system has potential beyond our initial context and hope that the results will lead into further work in more general areas.

8. CONCLUSION

In this paper we have laid out the requirements for an Internet Based Community Network. Beyond this we have described how, using user-centred design, we have worked with one such potential community to discover how they could be supported through the addition of semantic data to their current work and associated interaction processes. This has resulted in the creation of two ontologies tailored to their needs and the development of software to allow them to interact indirectly with the ontology constrained metadata. Through this work we can find common patterns of usage and interaction requirements that are applicable beyond out IBCN focus.

9. REFERENCES

[1] C. Bacon-Smith. *Enterprising Women: Television Fandom and the Creation of Popular Myth.* University of Pennsylvania Press, Philadelphia, 1992.

[2] N. K. Baym. *Tune In, Log on: Soaps, Fandom and On-line Community.* Sage Publications Ltd, 2000.

[3] BBC Staff Writer. 'amateur culture' set to explode. *BBC News Online*, 18 July 2005.

[4] D. Brinkley and L. Miller. Foaf vocabulary specification. Namespace specification document, FOAF, September 2 2004.

[5] R. Conte and M. Paolucci. *Reputation in Artificial Societies. Social Beliefs for Social Order.* Kluwer Academic Publishers, 2002.

[6] V. J. Costello. *Interactivity and the 'Cyber-Fan': An Exploration of Audience Involvement within the Electronic Fan Culture of the Internet.* PhD thesis, The University of Tennessee, August 1999.

[7] N. Crofts, M. Doerr, T. Gill, S. Stead, and M. S. (eds). Definition of the CIDOC CRM conceptual reference model. Reference document, International Council of Museums, March 2005.

[8] L. Ding, L. Zhou, T. Finin, and A. Joshi. How the semantic web is being used: An analysis of foaf. In *Proceedings of the 38th International Conference on System Sciences*, January 2005.

[9] C. Doctorow. Metacrap: Putting the torch to seven straw-men of the meta-utopia. Website, August 2001.

[10] M. Feldman, C. Papadimitriou, J. Chuang, and I. Stoica. Free-riding and whitewashing in peer-to-peer systems. In *Proceedings of the ACM SIGCOMM Workshop on Practice and Theory of Incentives in Networked Systems*, pages 228 – 236. ACM Press, 2004.

[11] E. J. Friedman and P. Resnick. The social cost of cheap pseudonyms. *Journal of Economics and Management Strategy*, 10(2):P. 173 – 199, 1999.

[12] J. Golbeck. *Computing and Applying Trust in Web-based Social Networks.* PhD thesis, University of Maryland, April 2005.

[13] S. C. Herring, I. Kouper, L. A. Scheidt, and E. L. Wright. Women and children last: The discursive construction of weblogs. *Into the Blogosphere: Rhetoric, Community, and Culture of Weblogs*, 2004.

[14] H. Jenkins. *Textual Poachers: Television Fans and Participatory Culture.* Routledge, New York and London, 1992.

[15] C. Lagoze and J. Hunter. The abc ontology and model. *Journal of Digital Information*, 2(2), November 2001.

[16] K. F. Lawrence, M. O. Jewell, M. M. Tuffield, A. Prugel-Bennett, D. E. Millard, M. S. Nixon, m. c. schraefel, and N. R. Shadbolt. Ontomedia - creating an ontology for marking up the contents of heterogeneous media. In *Ontology Patterns for the Semantic Web ISWC-05 Workshop*, November 2005.

[17] K. O'Hara, H. Alani, and N. Shadbolt. Identifying communities of practice: Analysing ontologies as networks to support community recognition. In *IFIP World Computer Congress. Information Systems: The E-Business Challenge*, 2002.

[18] J. Preece. *Online Communities - Designing Usability, Supporting Sociability.* John Wiley and Sons, Ltd., 2000.

[19] H. Rheingold. *The Virtual Community: Homesteading on the Electronic Frontier.* Secker and Warburg, 1993.

[20] E. Wenger. *Communities of Practice: Learning, Meaning and Identity.* Cambridge University Press, 1997.

[21] S. Whittaker, E. Isaacs, and V. O'Day. Widening the net: Workshop report on the theory and practice of physical and network communities. In *SIGCHI Bulletin*, volume 3, July, 1997 p29.

---

[2]The previously mentioned convention at which Meditate was demonstrated was attended by approximately sixty people of which two were men and a small number identified as 'other'.