



# Ontology building as a social-technical process: a case study

[Link to publication record in Manchester Research Explorer](#)

## Citation for published version (APA):

Lin, Y-W., Poschen, M., Procter, R., Kola, J., Job, D., Harris, J., Randall, D., Sharrock, W., Ure, J., Lawrie, S., Rector, A., & Goble, C. (2008). Ontology building as a social-technical process: a case study. In *Oxford e-Research Conference*

## Published in:

Oxford e-Research Conference

## Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

## General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

## Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact [uml.scholarlycommunications@manchester.ac.uk](mailto:uml.scholarlycommunications@manchester.ac.uk) providing relevant details, so we can investigate your claim.



This paper was presented at the conference:

## **Oxford eResearch 2008**

11-13 September 2008

University of Oxford

Conference website:

<http://www.oii.ox.ac.uk/microsites/eresearch08/index.cfm>

Conference papers collection:

<http://ora.ouls.ox.ac.uk/objects/uuid%3A64aa6f39-7e81-4d42-a008-ee2d7524bd67>

Conference organisers:

Oxford Internet Institute

University of Oxford

1 St Giles, Oxford OX1 3JS

<http://www.oii.ox.ac.uk/>

email: [enquiries@oii.ox.ac.uk](mailto:enquiries@oii.ox.ac.uk)

Oxford e-Research Centre

University of Oxford

7 Keble Road, Oxford OX1 3QG

<http://www.oerc.ox.ac.uk/>



# Ontology building as a social-technical process: a case study

Yuwei Lin<sup>1</sup>, Meik Poschen<sup>1</sup>, Rob Procter<sup>1</sup>, Jay (Subbarao) Kola<sup>2</sup>, Dominic Job<sup>5</sup>, Jonathan Harris<sup>5</sup>, Dave Randall<sup>4</sup>, Wes Sharrock<sup>3</sup>, Jenny Ure<sup>6</sup>, Stephen Lawrie<sup>5</sup>, Alan Rector<sup>2</sup>, Carol Goble<sup>2</sup>

<sup>1</sup>National Centre for e-Social Science, University of Manchester

<sup>2</sup>Department of Computer Science, University of Manchester

<sup>3</sup>Department of Sociology, University of Manchester

<sup>4</sup>Department of Sociology, Manchester Metropolitan University

<sup>5</sup>Division of Psychiatry, University of Edinburgh

<sup>6</sup>School of Informatics, University of Edinburgh

## Abstract

The NeuroPsyGrid Project aims to build an ontology of psychosis that will be usable by practitioners and scientists across a range of different disciplines as a common and formalised knowledge framework. This paper provides a socio-technical perspective on the collaborative work involved in the development of such an ontology of psychosis. After introducing the context and nature of the project, we will elaborate on the challenges of such an undertaking, focussing on the process of ontology building and the socio-technical aspects of collaboration, including the meaningfulness of documents as boundary objects.

## 1. Introduction

Increasingly, scientists are finding that they need to be able to discover, access, manipulate, link, analyze and share vast amounts of data if they are to tackle new research challenges successfully (Hey and Trefethen, 2003). Satisfying this need has driven the development of a new generation of research infrastructure ('e-Infrastructure') composed of networked and interoperable digital research resources and computational services.

Ontologies, which provide a means to represent formally, and in a machine-readable form, metadata ('data about data'), i.e., definitions of the entities and their relationships within a domain of knowledge, have emerged as a key technology for building e-Infrastructures.

In the systematics communities examined by Hine (2008), ontologies have already been successfully used for integrating and standardizing experimental data from sources in the genetic, cognitive and neuroanatomical knowledge domains, the most prominent example being the Gene Ontology<sup>1</sup>; an extensive database of open biomedical ontologies and related information, including mappings between ontologies can be found at The Open Biomedical Ontologies<sup>2</sup> website. While it has been recognised that ontologies are increasingly needed to capture, integrate, and manage metadata, the challenges increase when attempting to build such a common and formalised knowledge framework that can be used across different disciplines and sectors.

---

<sup>1</sup> See <http://www.geneontology.org/>

<sup>2</sup> See <http://www.obofoundry.org/>

This paper, based on our experience of the NeuroPsyGrid project, aims to explore from a socio-technical perspective some of the challenges of building such an ontology. Drawing on one focus group, ethnographic observation and interviews with project members, in this paper we will investigate the socio-technical issues relevant to the challenges of ontology development by examining the practices of those involved and the practical nature of ‘sorting things out’ (Bowker and Star, 1999; Randall et al., 2007). We focus on a range of different kinds of ‘documents’ co-written, co-edited and shared between domain experts and ontologists to exemplify how this kind of cross-disciplinary and collaborative boundary work can be facilitated.

## **2. The NeuroPsyGrid Project**

In mental health research there is continually emerging evidence linking clinical features of psychosis and psychotic disorders with genetic, neuroanatomical and neuropsychological findings. The meaningful integration of these diverse findings is essential to obtain a holistic understanding of psychiatric disorders. An important obstacle to such integration is the use of different assessments, scales and structured interviews for measuring and recording symptoms and assisting in diagnosis. A reliable method of combining research data produced by these diverse measures would increase the statistical power and validity of findings thereby enhancing understanding of psychiatric disorders and the prediction of their occurrence.

The NeuroPsyGrid project, which is a collaboration between mental health researchers and ontologists at Manchester and Edinburgh universities, aims to facilitate this by building an ontology of psychosis that will be usable by practitioners and scientists across a range of different disciplines (e.g., psychology, psychiatry, neurology).

The NeuroPsyGrid project proposes that building an Ontology of Psychosis can:

1. harmonise and accelerate work towards a psychosis database, common metadata and a simple ontology for relevant clinical and biological data, with a particular focus on changes in these measures around the time of onset; and
2. ensure that these data acquisition, storage and access arrangements are compatible with other major national developments such as the Mental Health Research Network (MHRN) and the Mental Health Minimum Data-set in the UK, and international initiatives such as the Brain Informatics Research Network (BIRN) in the USA; and thus
3. put UK clinical neuroscience researchers generally in a better position to be able to mount multi-centre clinical studies and trials with neuro-imaging biomarkers on this informatics platform.

The ontology will be used to combine existing datasets from people at high genetic risk of schizophrenia and people diagnosed as in their first episode of psychosis.

## **3. Challenges of developing an ontology of psychosis**

The two datasets to be combined are the Edinburgh High Risk Study (EHRS) of schizophrenia, and the Outlook and Eden projects. They contain detailed clinical, behavioural, cognitive (and imaging in EHRS) data, usually as serial assessments over time intervals of years, months, weeks respectively. That information can be used to:

- establish that people do or do not have a psychosis (Schizophrenia, Bipolar Disorder, Psychotic Depression) at entry into the various studies
- predict their likelihood of developing schizophrenia (in the EHRS)
- predict their chance of getting better (Outlook)
- predict whether or not they will respond to treatment (Eden)

However, the data in these two datasets were collected through different means by different medical workers based on different assessments and scales (e.g., PSE<sup>3</sup>, PANSS<sup>4</sup>, SCID<sup>5</sup>), and bore different data provenance issues. This denotes the first challenge of building a common ontology.

Psychiatric symptoms and descriptive terms form the base of clinical diagnoses and categorization of research groups. An ontology of psychosis terms adequately representing diverse scales and measurements was therefore considered fundamental to obtaining interoperability between different measurements. Moreover, in clinical research, there is continual discovery and understanding of the domain is frequently being updated. That said, the heterogeneity of different research measurements, the semantic heterogeneity of symptoms and terms, the evolving domain knowledge all contribute to the second challenge of building a common ontology.

To ensure that such an ontology is acceptable and understandable by diverse groups of people, making the ontology ‘comparable’ (Bowker and Star, 1999) is key. Comparability refers to a ‘regularity in semantics and objects’ (1999: 231) and thus pertains almost by definition to ontologies. What is important in this context, however, is the degree to which this stability is, in practice, obtainable. Most ontology design hitherto has been aimed at relatively homogeneous communities, where underlying concepts (if not terminology) stand a good chance of being commonly held. Ontologies that have to serve more heterogeneous situations and purposes may turn out to serve one user group more successfully than another – a problem that has been well-attested to within the field of medical informatics.

The NeuroPsyGrid project team consists of a distributed team of about ten researchers (including PIs) from different disciplines (psychology, neurology and computer science). Most of the ontology development is being done by a domain expert based in Edinburgh and an ontologist based at Manchester. The domain expert is in charge of liaising and communicating with a number of expert practitioners in the domain (we refer this person as the domain expert because he is the one directly working with the ontologist, bringing in the domain knowledge). The ontologist is in charge of turning the domain knowledge into a machine-readable ontology. The distributed and

---

<sup>3</sup> Present State Examination (Wing et al., 1974)

<sup>4</sup> Positive and Negative Syndrome Scale (Kay et al., 1987) is a medical scale used for measuring symptom severity of patients with schizophrenia.

<sup>5</sup> Structured Clinical Interview for DSM-IV, is a semi-structured interview for making DSM-IV psychiatric diagnoses.

cross-disciplinary nature of the team signals that the development of a common ontology is deemed to be contingent, and denotes the third challenge in this task.

Our observations of, and discussions with, the NeuroPsyGrid team members have drawn attention to the role played by a variety of documents (in different forms and formats, either electronic or hardcopy), written, co-written, exchanged, co-edited, reviewed, discussed, modified, versioned, shared, published and archived throughout the process, in facilitating collaboration between domain experts and ontologists. These documents include the original grant proposal, documentation of existing terms, a list of controlled vocabularies, the actual ontology itself (written in the Web Ontology Language (OWL)<sup>6</sup>, mainly using the Protégé ontology editor<sup>7</sup>), data, high level use cases representing data collection and use, meeting minutes, action plans, abstracts, emails, academic papers, presentation slides, and even ethnographers' fieldnotes. We may view these documents as serving as 'boundary objects' (Star and Griesemer, 1989; Brown and Duguid, 2001) within the NeuroPsyGrid team as they work to develop a common understanding of their goal and the artefact (the ontology) which will eventually embody it. Boundary objects "are both plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites" (Star, 1989). On the one hand, these documents provide team members with a sufficient degree of 'interpretive flexibility' to accommodate their different backgrounds; on the other hand, these documents provide sufficient structure to support the establishing of a common objective and a shared understanding how to proceed toward it.

#### **4. Documents and collaboration**

Given the important roles that various documents play in the ontology development process, we find it useful to follow the 'social life' of these documents (Brown and Duguid, 1996; 2000), to see how they evolve through the interactions between the domain expert and the ontologist as the project unfolds.

##### *4.1 Action plan: a structuring and managerial device*

As described earlier, the NeuroPsyGrid development team is distributed with domain experts mainly based at Edinburgh and ontologists based at Manchester. To manage the project, the team has regular teleconferences and some periodic face-to-face meetings. The domain researchers and the ontology developer maintain periodic face-to-face meetings (usually every 2 months) and regular email and telephone communications. Noting down, discussing and agreeing on next steps and future plans with a defined common target is an important and self-evident strategy in this team.

One example of action plans that are shared amongst the team members is the one being circulated after the kick-off meeting in February 2008. The PI summarised the discussion at the meeting and circulated the notes which included some action plans setting out the next step of the project.

The action plan document states:

---

<sup>6</sup> W3C OWL Web Ontology Language Overview: <http://www.w3.org/TR/owl-features/>

<sup>7</sup> Stanford Center for Biomedical Informatics Research: <http://protege.stanford.edu/>

The plan was to focus on clinical elicitors (questions for deciding whether or not someone has a particular symptom), severity indicators and their attached descriptors/definitions in these data sets. It was proposed at the kick-off project meeting to start with a range of assessments that were used in all of the above projects: including clinical interview schedules (such as the Present State Examination, Structured Clinical Interview for DSM-IV) and casenote methods (OPCRIT) for establishing the presence or absence of particular symptoms and diagnoses, as well as symptom severity rating scales (such as the Positive And Negative Symptom Scale – PANSS) and the operationalised criteria and diagnostic algorithms for specific psychosis sub-categories in the Diagnostic and Statistical Manual – fourth edition (DSM-IV) and the International Classification of Diseases – 10<sup>th</sup> edition (ICD-10 [both research and clinical versions]).

As decided at the project kick-off meeting, the team would start from ‘mapping’ these different types of data in three levels, symptoms, severity, diagnosis; map the relationships within and between classes of data; and construct a user-friendly interface, so that clinical predictions can be made from that data. It was thought that the primary test of the success of the resulting ontology would be whether clinical queries could be made and returned (answered) with this system.

In a similar vein, in the end of each teleconference, the PI (the chair of the meeting) would draw up some notes including action plans. These minutes and action plans serve as an ordering device for ensuring the orderly character of work (Button and Sharrock, 1996). The action plans informed the team members of the progress of the project, and allowed them to achieve the formatted arrangements of the project.

Interestingly, there is another kind of action plan written and co-written by people on a less managerial level than the PI, the chair of the meeting. At an interview, the ontologist reported that he kept a plan himself, and shared that with the domain expert whenever it got updated. In this document, he documented his design decision, the rationale, and reminded himself of the to-do items. Describing this document, he said:

“If I make a design decision, I document the rationale for the decision. How it would look like when we finish. It should help someone understand the development of the ontology if they were to read through the document and then look at the ontology. That’s the kind of things we’ve got in this document. This is more a reference document for me than for anyone else. For example, if there is an interesting line of thought, which I can’t follow up right away (subject to the factors e.g. I’ve got deadline to meet and I’ll follow up this later), I’ll write it down in here, to maybe use it later.”

The ontologist’s action plan is less formal and general than the PI’s action plan (inscribed in the meeting minutes) and addresses the concrete issues of his work practice. It drives the project forward by prioritising development tasks and by providing a record of modifications and improvements as well as a guideline for aims and collaboration. This (living and evolving) document was private on the one hand, written in the ontologist’s own language, but on the other hand it was shared with the domain expert via email every time it was updated to keep him informed of the progress, the new ideas and the next move of the project. This action plan, kept by the

ontologist, symbolises a boundary object shared between the ontologist and the domain expert. It showed the demarcation of two different kinds of work, and it also enabled the two to synchronise.

#### 4.2 *PSE and PANSS, SCID scales*

Problematic issues are highlighted by the attempt to map terms between two psychiatric symptom measures commonly used in research and diagnosis: the PANSS and the PSE. Structural differences between these measures include the fact that the PANSS scores symptoms on a scale of severity from 1 (absent) to 7 (extreme); whilst the PSE scores symptoms from 0 to 2 where 0 = absent and 1 and 2 are generally but not always distinguished by severity of symptom. For example, for non-affective verbal hallucinations (PSE item 63) symptoms are rated according to quality – 1 for voices with a pleasant or neutral tone and 2 for voices with a hostile or negative tone. As a consequence, the ratings for many symptom terms cannot be equated between these two measures. Here, the mapping can only be between presence of hallucinations in PANSS and presence of hallucinations in PSE regardless of severity or quality.

Semantic differences between the measures manifest in symptoms which may overlap in definition but have differences which question the validity of mapping these items in an ontology. For example, ‘inefficient thinking’ (in the PSE) and ‘disturbance of volition’ (in PANSS) both concern indecision about simple matters and impediments to thought and behaviour consequential to this. However, ‘inefficient thinking’ does not refer to observed motor activity disruption whereas moderate severe to extreme ratings of ‘disturbance of volition’ do. Here, mapping is only possible between the PSE item and minimal to moderate scores on the PANSS item. For other items, type of symptom description can affect semantic content.

During the initial stages of the project, in order to find out what items on the PSE, PANSS or SCID scale can be considered as entities in an ontology, which then would be mapped into a common ontological knowledge framework, the ontologist and the domain expert sat down next to each other with two hardcopies of identical documents (PSE, PANSS or SCID interview assessment scale sheets used by psychiatrists to question the patients) at hand. Both of them highlighted those items they thought were qualified to be an entity in the ontology to represent the explicit knowledge inscribed in these different assessment scales of psychosis.

Although the assessment scale sheets explicitly presented some knowledge in the domain of psychosis (e.g., terminology, symptoms), what was invisible is the evaluator’s (usually a psychiatrist’s) subjective judgment of how seriously the patient’s symptoms was and selected a number to score. In this case, what was invisible was the ontologist’s and the domain expert’s tacit knowledge of making sense of and judging the importance of each word. Their different disciplinary backgrounds and the understandings of what count as an ontological entity entail that disagreements were unavoidable in this exercise. As the ontologist admitted: “And, of course, there were differences. We wanted to find out on how many of these entities we agree.”

Among the many disagreements, a classic one concerned ‘Catatonic behaviour’ in the SCID scale. The ontologist wanted to represent the domain knowledge as complete as



possible, while the domain expert bore the psychiatrist practice in mind (thinking how the question would be scored and which bits would be of importance in this context). For Catatonic behaviour there are items about Motoric immobility (in this case, catalepsy or stupor) and four scoring options: 0 inadequate information; 1 absent or false; 2 sub-threshold; 3 threshold or true. If someone marks up Motoric immobility, only the score is ticked, but we do not know which one of the symptoms we are looking at (is it catalepsy or stupor that makes up the motoric immobility?). This may be irrelevant to the scoring and to the use of the assessment scale in practice, but not including the items of catalepsy or stupor would mean losing some data in terms of representing the domain knowledge – and maybe this information could be needed at a later point in time, using new or additional assessment criteria. And there were many more data entries like this in the datasets.

While the ontologist was thinking of categorising these items, the domain expert reacted: “you can have them as entities in the ontology, but I can’t actually do anything with them because they are not directly scored on an assessment sheet (SCID in this example).” In the end those items have been included as metadata annotations, which are not directly bound to the other entities in either ontology of PSE or SCID, because they are not currently directly useful for scoring or the mapping job to be done in this project.

This is emphasised by the ontologist in the following:

“If some categories/descriptions are not scored in SCID it doesn’t make sense of recording them. But me being an ontologist, I don’t think in that way. If you want to develop an ontology you have to capture entities at an atomic level. I think I said ‘hang on, I know there’s something called ‘impoverishment of speech’ and I know from my medical training what a patient has in terms of impoverishment.’ I thought that was conveying meanings. I can also see this is material you need if you want a complete domain knowledge. But the domain expert didn’t think it would be a useful entity used in the domain (that’s what ontology is about – the entities in the domain). In the end I added these items in, not as PSE or SCID, but as a metadata item.”

The disagreements and agreements were found through looking at the paper-based psychosis scales together and show the non-triviality of the process of collaborative ontology building, including the challenge of the mapping of different assessment scales. Again, these paper-form assessment sheets serve as boundary objects enabling the domain expert and the ontologist to exchange and negotiate their different understandings of the domain knowledge. At the same time the findings of this process are entered into the action plan, keeping track of what has been discussed and implemented, and (re-)defining the next steps. Both documents complement the ontology itself in order to progress the project in a structured and successful way.

#### *4.3 Spreadsheet data*

The following example demonstrates that some tools or techniques used in ontology development have been selected or invented because of situated problems (i.e., domain workers’ habits of using specific types of data formats and tool dependency).

The domain expert circulated the data mainly in Excel spreadsheets to the ontologist, who, from his previous work experience, knew of a Protégé plug-in<sup>8</sup> for automatically turning the spreadsheet data into classes in an ontology. In using both the Protégé and this plug-in, which he refined further for his purposes, the ontologist claimed that he used “a hybrid approach for ontology building”:

“When I was working on my previous project, I created a plug in for Protégé. People don’t really use Protégé, but they want to use ontologies. And they have a lot of data as spreadsheets. Because this is cannot straightforwardly be imported into Protégé, I created a plug-in to import spreadsheet data into an ontology. Before you could not easily transform the data just like that. It’s a hybrid approach I would say [i.e. using different tools and formats]. For the NeuroPsyGrid project, I extended this plug-in to import data from textual files into an ontology using a bit of Java code. Each time the domain expert sent me an update, what I did is to replace the old file with the new file and updated the ontology automatically. It can be any text file, e.g., an Excel spreadsheet or a simple comma separated file.”

Documentation in spreadsheet form which was getting updated frequently inspired the ontologist to adopt this ‘hybrid approach’ to developing ontologies to better cope with the requirements of working collaboratively, i.e., having to rely on different formats and tools used in the project. It started from the resistance to doing a laboriously coded or annotated manually job, but ended up in a Protégé plug-in.

This plug-in was built because the domain expert was used to documenting the work in a spreadsheet form. In order to automate the process of transforming the data in a spreadsheet into an ontology, this plug-in was created. There are other examples of this kind of plug-in for Protégé (e.g., DataMaster<sup>9</sup>) being created and their existence is very much dependent on the data formats used in the project.

#### 4.4 *Scholarly communications*

The ontologist reported that writing an academic paper (roughly 2 months after the project started) really helped in providing a deeper understanding of the project and of the different views and approaches involved:

“When I started to write it down and formalise the paper – why is this domain difficult? – that helped formalise my understanding of the domain. I got comments from the domain expert. That was really helpful. He would ask ‘what would this mean to a psychiatrist and I tried to answer that question’. That was incredibly useful. During this process we realised that we needed to tackle data sharing first rather than other tasks (define domain entities – which is about how entities are defined formally – no, that’s not what we’d like to do at the moment). It was quite interesting.”

---

<sup>8</sup> Available here: <http://www.co-ode.org/downloads/protege-x/plugins/#excel>

<sup>9</sup> For more information see <http://protegewiki.stanford.edu/index.php/DataMaster>

## 5. Summary and future work

Our study of the NeuroPsyGrid project team's work to develop a common ontology of psychosis has documented a number of challenges faced by the team:

1. obscure definitions of terms in the domain
2. data provenance, different means of collection and data formats, disciplinary divergence, and the use of different assessments and scales
3. changes of practices and scales over time through new findings in the field
4. managing a distributed project

We have seen how project team members have oriented to the challenge of working across disciplinary boundaries. We have also seen how the NeuroPsyGrid project team have improvised solutions to support their collaboration, including making use of a range of documents in different forms and formats to help them exchange information, negotiate meanings and develop a common understanding.

Table 1 summarises the lessons learned so far by the NeuroPsyGrid project team.

What problems have been resolved?	<ol style="list-style-type: none"> <li>1. Difficulties in mapping concepts at different levels of granularity and scale.</li> <li>2. Challenges in ontological representation of clinical findings in psychosis (though this can never be completed).</li> <li>3. Representation and acquisition methods for quantitative (statistical) and qualitative (domain experts confidence) measures of validity of ontological items.</li> </ol>
What problems / challenges remain?	<ol style="list-style-type: none"> <li>1. Ontological modelling of disorders/diseases in the domain of psychosis. This will make use of the ontological models of clinical findings referred to above. However, this reveals the additional challenge of representing anatomical, physiological and functional concepts in neurosciences.</li> <li>2. Validating the mappings generated in the previous step against real data and also by consulting domain experts.</li> <li>3. Temporal modelling of psychiatric disorders that often have an on and off clinical course.</li> <li>4. Integration with other ontologies – Gene ontology, Brain anatomy ontology, brain function, neuropsych ontology.</li> <li>5. Testing of the homo/heterogeneity of the ontology.</li> <li>6. Adding extensive annotations/descriptions to the classes, as there are no electronic versions of the PSE, SCID, PANSS, and their respective manuals available.</li> </ol>
What problems were not anticipated?	<ol style="list-style-type: none"> <li>1. The amount of work needed to map concepts from different assessment scales, which required cross-referencing different multiple individual items per category, cross-referencing books and data, and discussions with psychiatrists.</li> <li>2. Ontological modelling of the mappings with respect to OWL axioms that could be used and the ability of DL reasoners to cope with modelling strategies.</li> </ol>

Table 1. Summary of lessons learned from the NeuroPsyGrid project.

We are continuing our study of the NeuroPsyGrid project as it proceeds towards its objective of developing a common ontology for psychosis. Our next step will be to commence interviews with domain users. As the ontologist observed:

“You can build a ‘perfect ontology’ but without any users, it will not be very useful. A good ontology is the one that gets adopted and used.

## References

- Button, G. and Sharrock, W. (1996). Project Work: The Organisation of Collaborative Design and Development in Software Engineering. Computer Supported Cooperative Work: *The Journal of Collaborative Computing* 5: 369-386
- Bowker, J. and Star, S. (1999). Sorting things Out: Classification and Its Consequences. Cambridge MA, MIT Press.
- Brown, J.S. and Duguid, P. (1996). The social life of documents. First Monday, May 1996. URL (retrieved on 12 August 2008): <http://www.firstmonday.dk/issues/issue1/documents/>
- Brown, J.S. and Duguid, P. (2000). The social life of information. Harvard Business School Press.
- Brown, J.S. and Duguid, P. (2001). Knowledge and Organization: A Social-Practice Perspective. *Organization Science* 12(2): 198-213.
- Hey, A. and Trefethen, A. (2003). ‘The Data Deluge: An e-Science Perspective’, in Berman, F., Fox, G. C. and Hey, A. (Eds.) *Grid Computing - Making the Global Infrastructure a Reality*, chapter 36, pages pp. 809-824. Wiley and Sons.
- Hine, C. (2008). *Systematics as Cyberscience: Computers, Change and Continuity in Science*. Cambridge, MA: MIT Press.
- Randall, D., Sharrock, W., Rooksby, J., Lin, Y., Procter, R. (2007). Ontology Building as Practical Work: Lessons from CSCW. 3<sup>rd</sup> e-Social Science conference, Ann Arbor, October.
- Star, S.L. (1989). The structure of ill-structured solutions: Boundary objects and heterogeneous distributed problem solving. In *Distributed Artificial Intelligence*, Vol. 2. Pitman, London.
- Star, S.L. and Griesemer, J.R. (1989). Institutional Ecology, ‘Translations’ and Boundary Objects: Amateurs and Professionals in Berkeley’s Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, 19, 387-420.
- Ure, J., Procter, R., Lin, Y., Hartswood, M. and Ho, K. (2007). Aligning Technical and Human Infrastructures in the Semantic Web: a socio-technical perspective. 3<sup>rd</sup> e-Social Science conference, Ann Arbor, October.