NCeSS
National Centre for e-Social Science

# The ethical implications of data as representation

Annamaria Carusi
University of Oxford Computing Laboratory


E-Cap Conference, 2006,
Trondheim, Norway

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

# eScience and the Grid

- e-Science: using technologies to support research;
- Grid (1960s- 1990s): computational grids, that is, distributed sets of computers co-operating on a calculation.

# Grid 2006

- Grid (now and what we're working towards working towards): reconfiguring access to
  - data/information;
  - other researchers;
  - computational resources.
- That is, Virtual Organisations

# Examples:

- Medical images and other information;
- Social science images (and sometimes, other information);
- Statistics;
- Anthropological artefacts and images.

# e-Diamond

‘eDiaMoND is an ambitious 2 year flagship pilot e-Science project of £4.1 m funded through EPSRC/DTI and IBM SUR4 grants. The project team is made up of academic and industrial collaborators over 12 sites. The aim of eDiaMoND is to demonstrate the value of grid technology to the NHS. **To this end the project is building a grid-enabled, federated database of annotated, digitised mammograms and patient information, which will be piloted at four participating UK screening centres** (Power et al., 2004). A number of prototype grid-enabled applications have also been developed to demonstrate how the database architecture can support screening work, radiologist training (Soutter et al., 2003) and epidemiological studies. The database will also be used for computer science research, including the development and evaluation of image analysis algorithms, search mechanisms for data mining, and image standardisation techniques. The eDiaMoND database embodies the e-Science vision of encouraging innovative approaches to research by enabling data generated at screening to be shared between a variety of disciplines (radiology practice and training, epidemiology and breast imaging) and made available to the whole breast care community, regardless of where or how it was generated.’ (Jirotka et al, 2005)

# Social science images

- Research in private spaces (such as homes, schools) as well as public spaces, such as museums, streets, etc.
- For education, very large data bases are extremely useful, as they allow research to be conducted which connects local, particular environments, to distant environments. This includes very large data bases of images. Images of children / adults: different formal ethical processes.
- Ethnomethodologists standardly use video and images to conduct their research into situated practices. For these researchers, a grid makes it possible to collaborate in the interpretation of these images with non-colocated colleagues.
- **Multi-modal analysis of natural language / gesture (transcripts + video): computations that 'model the shape, appearance and motion of the face and head' (Carter et al, forthcoming)**

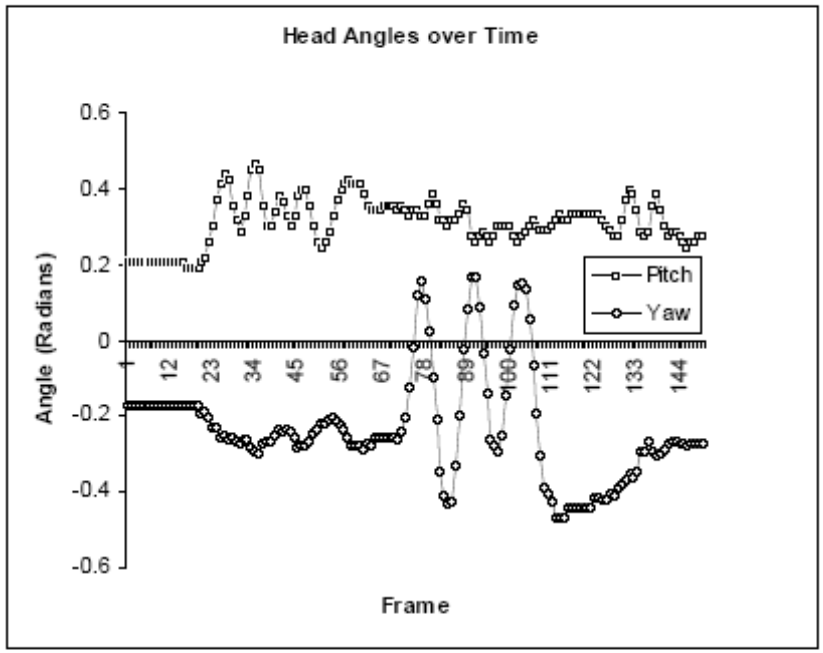



**Figure 3.** Head angles over time, with a nod, followed by a large head shake, then a smaller nod

Carter et al, forthcoming, p.18

# British Household Panel Survey

The British Household Panel Survey began in 1991 and is a multi-purpose study whose unique value resides in the fact that:

it follows the same representative sample of individuals - the panel - over a period of years;

it is household-based, interviewing every adult member of sampled households;

it contains sufficient cases for meaningful analysis of certain groups such as the elderly or lone parent families;

The wave 1 panel consists of some 5,500 households and 10,300 individuals drawn from 250 areas of Great Britain. Additional samples of 1,500 households in each of Scotland and Wales were added to the main sample in 1999, and in 2001 a sample of 2,000 households was added in Northern Ireland, making the panel suitable for UK-wide research.

http://www.iser.essex.ac.uk/ulsc/bhps/

# BHPS

- Same representative sample of people: not necessarily same individuals but same type.
  - Stability; responsibility; trust between the project members and the group.
- Interviewing methods: qualitative data -> quantitative
- Large data sets, anonymised (database of respondents is separated from the one on the results)
- **Intended for third party analysis from inception**
- **Standardisation of practices**

What does it take thus to arrive at data usable by BHPS users? To name/code variables and decide of coding frames for verbatim responses. From the point of collection, answers formulated by respondents in words will be visualized and circulated in the form of numbers. To transfer in batches the 2000 flat files collected in a survey database mirroring the structure of questionnaires and to collapse them into thirty variables. The renaming of the files is part of this first major conversion as well as the passage from CAPI software to SPSS to an in-house one. Files are loaded into a SIR database where the cleaning of data is operated. Flats files collapsed into 30 variables then get re-expanded again when loaded into a third users database: 'Having all the data, we transform them again because people, most analysts, want a much flatter structure' commented the computer manager. 'Derived variables, weighting, imputations are put on' (ibid.). What was first in CAPI and passed in SPSS and SIR is made then accessible again through SPSS, SAS and STATA software. Variables names are kept consistent across years as well as the terms used for indexing. It is all this process of conversion of words into numbers; of successive flattening and restructuring; of renaming and renumbering that transform materials collected into visible, manageable, communicatable and intelligible data for its community of users. (p.8)

# Cumaa / Anthropological museums

- Anthropological artefacts
- Images, video, audio, text.
- Objects
- **Acquisition**

*'Andrew [Moutu] has put a picture on the cover of a publication. He will be fined for that, because the artefact shows a ritual/secret process. This is despite the fact that it was a three-dimensional artefact sold to Andrew'* (for another description of such divergence in conceptions of exchange and ownership see Strathern 2005). 'How do you know this is sensitive knowledge? – the problemis that you might not know until it gets public […] People think that access is to take everything you have and put it online. If we did that, we would alienate many of the communities we deal with […] For instance, we have a photo in the Williamson Collection showing relationships among people that these same people deny having had. They see in these pictures the political rather than the daily Tibetan practices […] Also, the museum is thought to be a neutral place for many communities because it's so far from them. They sometimes prefer their objects to be there rather than on the island of their neighbours. Then to put everything online would be the best way to ensure not working with them anymore' (Anita Herle, Cumaa.See also Herle 1994). No conflict ever arose from Cumaa having publicized specific objects but Cumaa's 'ethical obligation to respect creators and communities of origin' nevertheless constitutes a barrier to the circulation of content over e-media (Cumaa AHRB Grant Proposal: 7). (Chimera report, p.13)

# Data consisting of collections of the following:

- Questionnaires, polls, surveys
- Documentation (letters, reports, emails, etc.)
- Interviews and focus groups (verbal)
- Observations (including images, video, audio)
- Field notes

  etc …

- Interpretive methods:
  - coding
  - annotations
  - comments, remarks, application of interpretive principles of the theory
  - algorithms  for statistics or modelling
  - ontologies for semantic web aspects

  etc.

- Ethical challenges (confidentiality, anonymity, profiling, informed consent of research participants)
- Grey areas created by e-social science
  - Put a stop to it?
  - Ethical, social, institutional implications of that

# My focus:

- Data that represents human subjects
- The nature of the representational relation, that in virtue of which data represents (anything).
- Question: is there a connection between that representational relation and ethical concerns:
  - Towards what or whom ought ethical concern to be directed
  - What form should that ethical concern take?

# Research ethics

- Human subjects model of medical and social sciences
  - Subjects as agents
    - the represented
  - Privacy, confidentiality, anonymity
- Humanities model
  - Subjects as authors
    - representers
  - Copyright issues
- ‘Representational model’
  - Researchers’ & subjects’ relation to representational content

# Identity

- ‘Thin’ identity: a particular individual
- ‘Thick’ identity: the person (persona; being-for-self/others; the subject of an autobiography; person-qua-representative of a medical condition; of a group of people)
- ‘Ensemble’ identities: Collective identities

# Representation as …

- From philosophy of art
- Object of ethical concern for data subjects is representation as …
- But there are different ‘representation as …’ statements to be derived from representations

# Representational content : an ecological theory

- Causality
- Intentionality
  - In the broad sense: connections of meaning, aboutness; figure/ground; gestalt;
  - In the narrow sense: intendings;
  - In the full panoply of intendings: conscious / non-conscious; more-or-less determinate; higher-order /lower-order etc.
- Situatedness
  - Interactions
  - Contexts
  - Lifeword

# Media of representation

- Physical media;
  - Printed, electronic, audience possibilities (one to one, one to many, many to many, etc)
- Abstract media:
  - Language and other symbolic systems, rhetoric, figures and tropes, codes, conventions and constructions.
  - Genres; discourse; history.

# Digital media / electronic data bases

- manipulation
- standardisation and reification (Kluge 2001; Strathern 1999),
- the way in which data enters into a digital context (was ‘born digital’ or does it become digital),
- whether the data was intended for the digital world, and if so, with what understandings of representation, and of private and public spaces (Nissenbaum 2004),
- the decontextualising and recontextualising possibilities of the digital world.

# Digital media / electronic data bases

- standardisation and reification (Kluge 2001; Strathern 1999),
- the way in which data enters into a digital context (was ‘born digital’ or does it become digital),
- whether the data was intended for the digital world, and if so, with what understandings of representation, and of private and public spaces (Nissenbaum 2004),
- the decontextualising and recontextualising possibilities of the digital world.

‘By reification I simply intend to point to the manner in which entities are made into objects when they are seen to assume a particular form (‘gift’, ‘exchange’). This form in turn indicates the properties by which they are known and, in being rendered knowable or graspable through such properties, entities appear as “things”’ (Strathern, 1999: 13)

Once it has been generated, the existence and functioning of this patientanalogue is independent of the patient and functions independently as basis for interventions and decisions-making. Therefore, to all intents and purposes, it has acquired what amounts to a functionally independent status. In this sense, we are beginning to see the ontological reification of the patient record. (Kluge 2001:33)

- Interpretational possibilities elided or occluded
- ‘Othering’

But also:

- Breaking existing norms of standardisation;
- Demythologisation
- Emergence of different aspects, narratives
- Re-appropriation

Access allowing ….

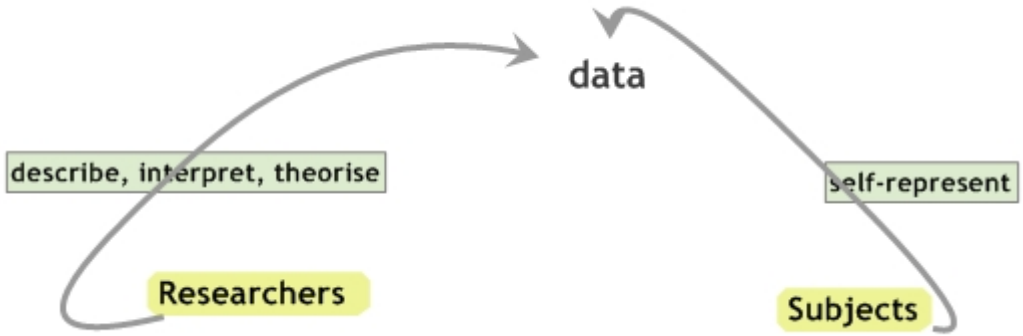
# Digital media / electronic data bases

- standardisation and reification (Kluge 2001; Strathern 1999),
- the way in which data enters into a digital context (was ‘born digital’ or does it become digital),
- whether the data was intended for the digital world, and if so, with what understandings of representation, and of private and public spaces (Nissenbaum 2004),
- the decontextualising and recontextualising possibilities of the digital world.

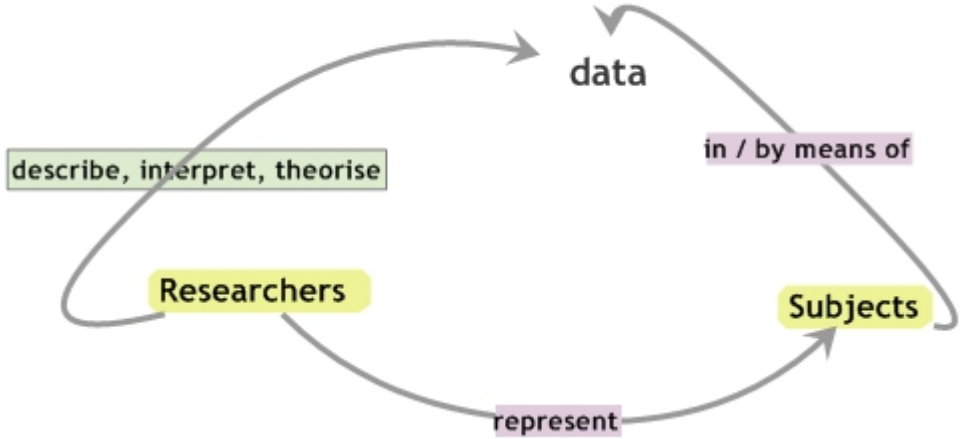
# Relations of representation

- Who represents whom;
- Who interprets whom

# Research on subjects on the Internet



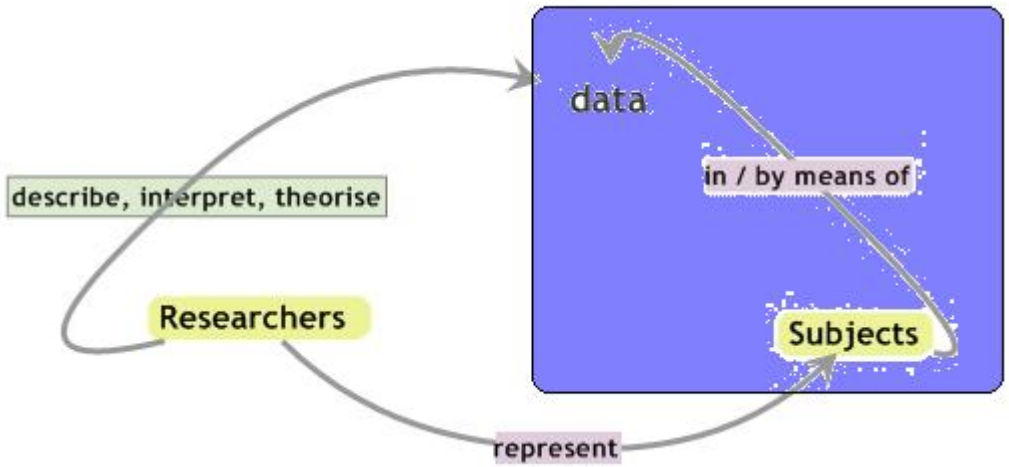
# Research in e-science



# Relation between representational content and identity (in either sense)

- **Naturalism**
- Isomorphism
- Figural
- Statistical
- Construction
- Interactional

# Naturalism: ethical focus



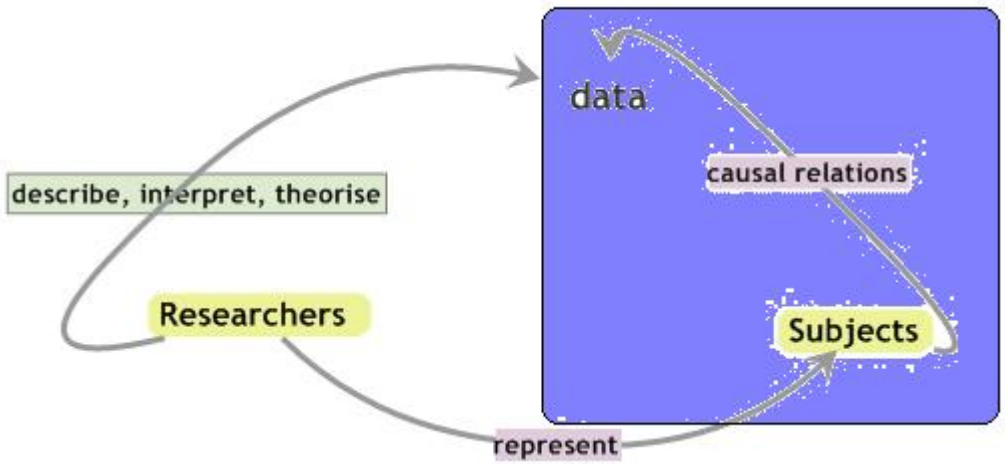
# Relation between representational content and identity (in either sense)

- Naturalism
- **Isomorphism**
- Figural
- Statistical
- Construction
- Interactional

# Isomorphism

- They bring together items that in paper-based records would have remained separate;
- They allow for a ‘joined-up’ (in space and time) ‘view’ of the person qua patient;
- They can by ‘combined and manipulated in a holistic fashion’;
- They are not simply copies of features of the patient’s health status, but actually share some features (the combinatorial features of representational elements) with the elements represented. [‘picture theory of meaning’]
- They is therefore a relation of isomorphism between integrated electronic patient records and the patient.
- Electronic patient records are the analogues of patients (Kluge 2001)
- Not all of the patient (‘most notably […] , not the patient as a person’ 2001:30) but only those aspects which have generated data in the health care professional / patient encounter (2001:30), or that are a **causal antecedent** of the data (2001:31 *my emphasis*)

# Isomorphism: ethical focus



# Problems

Isomorphic with what?

- Not only the patient but also *encounters* with health-care professionals
- Written records (eg those kept by GPs) are written for other health care practitioners (Heath, et al). Their representational content is determined as much by the addressee relationship as it is by the relationship with the patient.

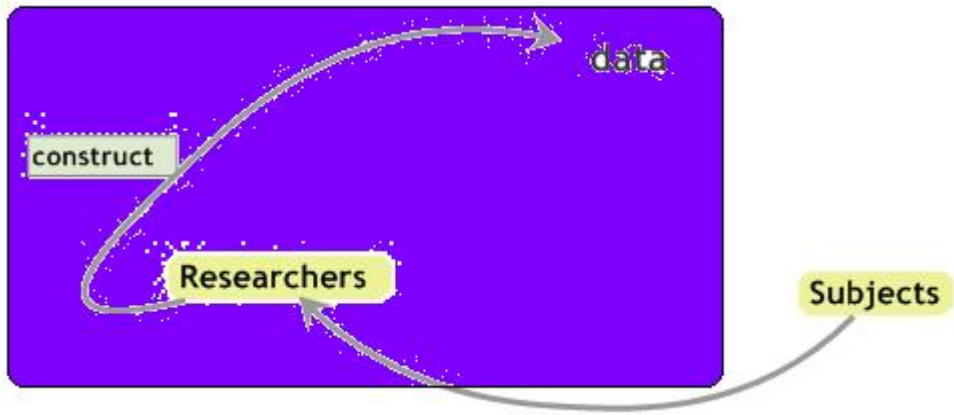
# Relation between representational content and identity (in either sense)

- Naturalism
- Isomorphism
- Figural
- Statistical
- **Construction**
- Interactional

# Relation between representational content and identity (in either sense)

- Naturalism
- Isomorphism
- **Figural**
- Construction
- Social / interactional
- Ecologies of meaning

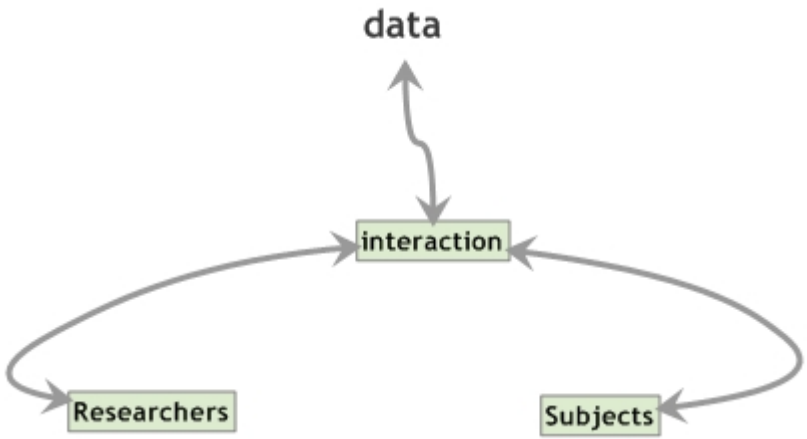
# Constructions: ethical focus



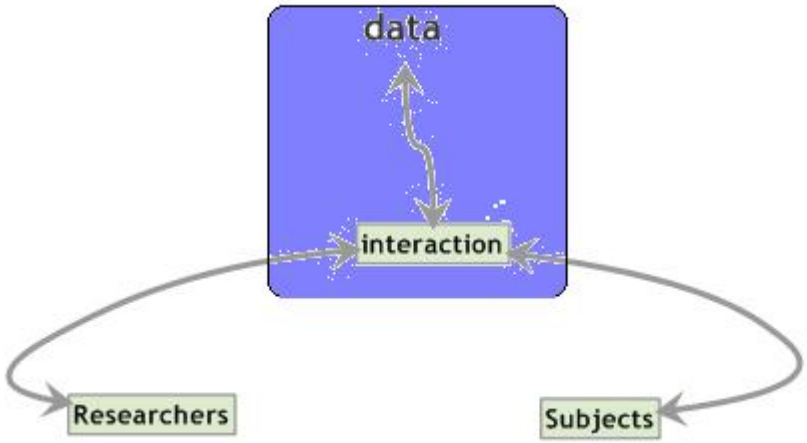
# Relation between representational content and identity (in either sense)

- Naturalism
- Isomorphism
- Figural
- Statistical
- Construction
- **Interactional**

# Interactionism



# Interactionism



# Relation between representational content and identity (in either sense)

- Naturalism / Isomorphism
  - Identity at stake: thin/thick identity [of data subject]
  - No sharing, no re-use without informed consent
- Figural
  - Identity at stake: 2$^{o}$ Thin identity [of data subject]; 1$^{o}$ thick identity intrinsic
  - Loose constraints on sharing (due to interpretive underdetermination)
- Statistical
  - Identity at stake: 1$^{o}$Thin identity [of data subject], 2$^{o}$ thick identity
  - Loose constraints on sharing
- Construction
  - Identity at stake: 1$^{o}$ researcher identity (*qua* researcher), 2$^{o}$ Thin/thick identity (of data subject)
  - Loose contstraints on sharing
- Interactional
  - Identity at stake: 1$^{o}$ thick identity [of data subject & researcher], 2$^{o}$ thin identity of researcher
  - Joint constraints on sharing

# References


- Bella Dicks, B. S. a. A. C. (2006). "Multimodal ethnography." Qualitative Research **6**(1): 77-96.
- Bruckman, A. (2002). "Studying the amateur artist: A perspective on disguising data collected in human subjects research on the Internet." Ethics and Information Technology **4**: 217-231.
- Carlson, S. (2006). Entangled data: Interim Report. Chimera Working Paper Number 2006-01, Chimera, University of Essex.
- D'Agostino, G., Hinds, C., Jirotka, M., Meyer, C., Piper, T. & Vaver, D. (2006). On the Importance of Intellectual Property Rights for eScience and Integrated Health Record. Integrated Healthcare Workshop Edinburgh, UK.
- Goguen, J. (1997). Toward a social, ethical theory of information. Social Science, Technical Systems and Co-operative Work: Beyond the Great Divide. S. L. S. Geoffrey C. Bowker, William Turner, Les Gasser. Mahweh, N.J., Lawrence Erlbaum**:** 27-56.
- Hinds, C., Jirotka, M., Rahman, M., D'Agostino, G., Meyer, C., Piper, T. & Vaver, D. (2005). Ownership of Intellectual Property Rights in Medical Data in Collaborative Computing Environments. First International Conference on eSocial Science, Manchester, UK.
- Jirotka, M. P., Rob; Hartswood, Mark; Slack, Roger; Simpson, Andrew; Coopmans, Catelijne; Hinds, Chris; Voss, Alex (2005). "Collaboration and Trust in Healthcare Innovation: The eDiaMoND Case Study." Computer Supported Cooperative Work **14**: 369-398.
- Kindon, S. (2003). "Participatory video in geographic research: a feminist practice of looking?" Area **35**(2).
- Pink, S. (2001). "More visualising, more methodologies: on video, reflexivity and qualitative research
  - " The sociological review **49**(4): 586-599.
- Strathern, M. (1999). Chapter 2: Pre-figured Features. Property, substance and effect: Anthropological Essays on Persons and Things. London & New Brunswick, Athlone Press.
- Larry Gross, J. S. K., Jay Ruby, Ed. (2003). Image ethics in the digital age. Minneapolis, London, University of Minnesota Press.
- Pink, S. (2001). Doing Visual Ethnography: Images, Media and Representation in Research
  - London, Sage
- Tomaselli, K. G. (1996). Appropriating images: The semiotics of visual representation. Højbjerg, Intervention Press.

- Bruckman, A. (2002). "Studying the amateur artist: A perspective on disguising data collected in hu..." Ethics and Information Technology; **4**(3): 217-231.
- Ronald Carter, S. A., Dawn Knight, Tony Pridmore, Steven Mills, and M. H. Sahar Bayoumi, and Andy Crabtree (forthcoming). "Developing Multi-modal Corpora of Natural Language " Computer Supported Co-operative Work.
- Kluge, Eike-Henner W. The Ethics of Electronic Patient Records. New York: Peter Lang, 2001. Questia. 19 June 2006 <http://www.questia.com/PM.qst?a=o&d=107264728>.
- Ess, C. (2002). "Introduction
  - " Ethics and Information Technology; **4**(3): 177-188.
- Heath, D. (1998). "Locating Genetic Knowledge:Picturing Marfan Syndrome and its traveling constituencies." Science, technology and human values **23**(1).
- Heath, Christian, and Paul Luff. Technology in Action. Cambridge, England: Cambridge University Press, 2000. Questia. 19 June 2006 <http://www.questia.com/PM.qst?a=o&d=105492445>.
- Kindon, S. (2003). "Participatory video in geographic research: a feminist practice of looking? ." Area **35**(2): 142-153.
- Nissenbaum, H. (1998). "PROTECTING PRIVACY IN AN INFORMATION AGE: THE
  - PROBLEM OF PRIVACY IN PUBLIC." Law and Philosophy **17**: 559-596.
- Nissenbaum, H. (2004). "Privacy as Contextual Integrity." Washington Law Review **79**: 119-157.
- O'Riordan, E. H. B. K. (2002). "Ethics of Internet research: Contesting the human subjects research model
  - " Ethics and Information Technology **4**(3): 233-247.
- Strathern, M. (2004). "The whole person and its artefacts." Annual Revew of Anthroplogy
  - **33**: 1-19.
- White, M. (2002). "Representations or people?" Ethics and Information Technology **4**(3): 249.
- Pink, S. (2001). "More visualising, more methodologies: on video, reflexivity and qualitative research." Sociological Review.