

Abstract Terrains: Position Paper

Mike Robinson
Sageforce Ltd., England
E-mail: robinson@gmd.de

Workshop *Collaboration in Virtual Environments - Methods and Experiences*, Fourth European Conference on Computer Supported Cooperative Work, Stockholm, September 11-14

Abstract: Multiple technical, social and philosophical contradictions are at the core of VR systems. These are not avoidable, and need to be confronted by ensembles of computer and social scientists in a strategic development program. This position is briefly illustrated by reference to large, distributed, politically important, and recurrently controversial datasets.

One of the most popular coffee table books of the '20's was *Tertium Organum* by P.D. Ouspenski. Like Benedikt (1992), it took off from the idea of dimensionality and re-analysed life, the universe, and everything, surfing from the Book of Revelations to Kant and Riemann, looking anew at the phenomenal world, at the nature of living beings, the spectres of different realities, and intimations of cosmic consciousness, higher moralities, and the eruption of mysticism into everyday life. There is even a handy pull-out table of past and future evolution..... Discussion of Virtual Reality dives into the same muddy waters where the pre-suppositions of thought are challenged. Once this happens, contradictions multiply. And, of course, from a contradiction, *anything follows*.

The position taken here is that contradictions in conceptualising VR are not accidental but central. As for instance with “abstract terrains”, the title of this paper. It would be useful to have a text listing the contradictions, apparent contradictions, and conflicting positions in VR practice and theory. An example would be the conflicting positions taken by Benedikt (1992) and England (1995a,b) over several issues, for instance whether it is possible/desirable for an object to be in more than one place at the same time, or for two objects to occupy the same place. Another would be the conflicting positions taken on whether a virtual world should be “objective” — the same for all its inhabitants — or whether “subjectivity” is desirable/necessary

(Robinson, 1994; England, 1995b; Snowden, 1995). But this is not the place for a comprehensive survey of contradictions and conflicts, only to start on one specific issue: abstract terrains.

Abstract means you can't go there. Self evident really. Concretizing the abstract is usually a recipe for fallacy. As in the Platonic concretization of principles as forms (a triangle is a triangle, and recognisable as such, by virtue of reflecting the form of a triangle), which moves from being a philosophical mistake to a ludicrous and dangerous political position¹. Much sorrow followed. Tinkering with dimensions and fundamental categories *is* dangerous as well as intellectually stimulating. So abstract means you can't go there, by definition.

Terrain means you can go there, and so can other people. A terrain is precisely where people go, meet, talk, do things.

VR “landscapes” are abstract terrains. A fundamental distinction has been abolished. Many purely practical questions are raised, and are a very long way from being answered. The canons, conventions, and etiquettes of encounters and meetings in cyberspace are almost unexplored, and the little exploration that has been done raised many problematic features (Benford et al., 1994, 1995). Even the fundamentals of moving about and taking simple actions are hard and controversial issues (England, 1995b). It is already clear that a major design effort will be needed to provide mappings of abstract spaces, databases, and the like into VR. Usable and pleasant environments will take serious and diverse architectural and aesthetic talent — Le Corbusier meets Dali! — technical knowhow and vision — Isambard Brunel meets William Blake!

But what *are* the abstract spaces that will become “Populated Information Terrains” (Benford and Mariani, 1994)? My own work is with large, international, politically important databases, such as the World Health Organisation's Mortality Database, also known as the ICD or International Classification of Diseases, and with the European Union's Employment Statistics. (*Some salient feature of the ICD are briefly described in the Appendix*) These datasets are politically important because they inform decision making at Governmental levels. What policies to adopt? Health, employment, penal, economic..... What resources and funding is made available? Such datasets, formal and largely static in themselves, can be highly controversial in almost every way. The data itself — what gets in — is a highly complex and contested social matter, involving multiple conflicts of interest, ongoing compromises, local contingencies, and recurrent practical difficulties (Star and Griesemer, 1989; Bowker and Star, 1994; Star and Ruhleder, 1994). How the datasets are read and interpreted, selected, filtered, presented is another little-understood story of which the sociological surface has hardly been scratched. We know that the interpretation of both the ICD and employment

¹The belief that it is possible to visit/know these forms led to the belief that those that can are better, and those that can't are worse, and therefore should rule over the others whether the others like it or not, and thus Athenian Democracy should be abolished (Stone, 1988)

statistics frequently calls for knowledges — sometimes practical, sometimes local, sometimes expert — that is not *in* the database, that is not *referenced* in the database, that is not *accessible* via the database, and that is often *illegitimatised*, cannot be made public. It is only available within a community of practice that is self-organising, self-policing, and largely “invisible” and intangible from the “outside” (Star, 1992). And once the data is collected and interpreted, the policies that emerge can be controversial and challenged. An army of politicians, researchers, press and media can push into, re-invent and re-read all the problematics of input and interpretation. The political geography of such post-modern informational socio-political processes and emergent power relations (Bowers, 1992) are as dangerous as they are fascinating.

Inventing and improving, designing and hacking the new worlds of VR are activities with almost everything to learn. There is already no shortage of moralists pronouncing on the basis of the underlying contradictions (which of course makes them all correct!). The COMIC Project has crystallized an emerging trend to mesh ethnographers with computer scientists in the evolution and growth of CSCW systems. The invaluable lesson is how to hold together different disciplines without compromising them, and without requiring everyone to become a jack-of-all-trades. Projecting datasets such as the ICD and EU employment statistics into VR, into public cyberspace will need to expand this collaboration to demographers and political scientists. As the process expands it will probably need to be bootstrapped and realised through the same medium it is helping to create. The strategic program for VR development will demand a unique and unprecedented combination of computer and social scientists.

- Benedikt, Michael. 1992. *Cyberspace, First Steps*. Cambridge, MA: MIT Press.
- Benford, Steve, John Bowers, Lennart Fahlen E., John Mariani, and Tom Rodden. 1994. Supporting Cooperative Work in Virtual Environments. *The Computer Journal* 37 (8): 653-668.
- Benford, Steve, Adrian Bullock, and Steve Proberts. 1995. *Minutes of the first virtual COMIC meeting*. Draft University of Nottingham, 13/4//95 1995. COMIC-NOTT-4-21.
- Benford, Steve and John Mariani. 1994. Populated Information Terrains: Virtual Environments for Sharing Data. COMIC NOTT-4-12 (Draft), 1994.
- Bowers, John. 1992. The Politics of Formalism. In *Contexts of Computer Mediated Communications*, ed. M.Lea, Hassocks: Harvester/Wheatsheaf.
- Bowker, Geoffrey and Susan Leigh Star. 1994. Knowledge and infrastructure in international information management: problems of classification and coding. In *Information Acumen: The Understanding and Use of Knowledge in Modern Business*, ed. Lisa Budd-Frierman, 186-213. London: Routledge.
- England, David. 1995a. Abstract Data Shadows. *COMIC ftp site: DRAFT* .
- England, David. 1995b. Critique of Approaches to Interaction and Visualisation. *COMIC ftp site: DRAFT* .
- Robinson, Mike. 1994. *Conceiving Odysseus: Social Dimensions & Multiple Populated Worlds*. COMIC Final Report 1994, University of Lancaster, 1994.
- Snowdon, Dave. 1995. What You See Is Not What I See: Subjectivity in Virtual Environments. *COMIC ftp site: DRAFT* .

- Star, Susan, Leigh and Karen Ruhleder. 1994. Steps towards an Ecology of Infrastructure. In *CSCW '94, Chapel Hill, N. Carolina, USA*. ACM.
- Star, Susan Leigh. 1992. The Trojan Door: Organisations, Work, and the 'Open Black Box'. *Systems Practice - forthcoming* .
- Star, Susan Leigh and James R. Griesemer. 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.
- Stone, I.F. 1988. *The Trial of Socrates*. Boston: Little, Brown.

Appendix

The International Classification of Diseases (ICD)

Some Background.

The ICD is a worldwide list of causes of death and disease collected, administered, and published by the World Health Organisation (WHO). It is about a hundred years old, has been revised ten times, and data has been held in electronic form since 1950. The ICD is essential information to government health departments, public health offices, statistical bureaus, and hospitals all over the world. It informs decisions on national and international health policies, on the allocation of medical resources, whether and how to control epidemics and endemic illnesses, and on appropriate research directions. (Bowker and Star, 1991)

The ICD exists electronically as a set of SQL databases based in 6 regional offices, each dealing with data from all WHO member states, but from a different perspective. For example, data on malaria is held at Geneva, on AIDS at another office, on general mortality at another, etc. So far there are no central guidelines for setting up and maintaining these databases, but these are being created at the moment. The current version of the ICD is the 10th. revision, intended to be more flexible than the 9th., which had been found to be inflexible, for example, in dealing with HIV data which had not been anticipated in its categories.

There are plans to make the ICD Mortality Database available on the WWW through the NASA CIEZIN server (it is currently distributed on magnetic tape), but despite the fact that the ICD is regarded by the WHO as public domain, a publishing agreement has not yet been reached with CIEZIN.

Technologically, the ICD started as a mainframe database, and is now moving to a Client Server Architecture. When it becomes available on the Internet (it is now past the stage of β -testing) NASA will provide the generalised end user interfaces, and the WHO will provide database specific front ends.

In house (WHO) users vary from those who need policy level summary reports (a high level of sophistication, including natural language presentations is envisaged) to those who need full technical detail. Outside users include scientific workers (e.g. those investigating cancer of the ovaries), NGO Research Trusts, Me more interested in aggregate data for policy purposes), and the press/media (who sometimes have little awareness of problems inherent in statistical data).

The ICD is a static database, meaning that it is centrally maintained and not incremented or changeable by the data contributors. This avoids many of the problems of distributed databases, and there are no plans to change this.

There are national variations in the way that data is collected, and there are variations in the way that national data is held and provided. For instance, some countries use modifications of the ICD lists which may only partly overlap with the main lists. This is discussed in (Bowker and Star, 1991; Bowker and Star, 1994). Also some countries withhold detailed data for various reasons (e.g. data protection laws) and only provide summary statistics.

Data, in varying degrees of completeness, is held from 1950 onwards, and the ICD currently comprises some 50 million atomic data items organised on 5 axes: country; year; cause of death; age; sex. These axis categories are not always self evident. County, for example, is not necessarily a political entity. The UK is a construction from aggregating data on England, Wales, Scotland, & Northern Ireland. Similarly, age (groups) have been subject to conversions since some countries hold data in 10, and others in 5 year periods.

Issues of ICD content

The International Classification of Diseases is a *list* of the worldwide causes of death and disease. It is distributed in book form to statistical bureaus, public health offices, etc., and is used in planning and implementing epidemic control, reduction of infant mortality, etc. As noted earlier, it is currently available on magnetic tape, and will soon be available on the WWW. It is probably significant for our purposes — historically, and vis a vis use — that it is classified as a list rather than a database ((Bowker and Star, 1991; Bowker and Star, 1994)p.74)

Tensions and conflicts inherent in the ICD will be discussed in more detail later, and are central for our ICD/VR usefulness explorations. These tensions and conflicts are exemplified with reference to the power and resources that flow from the adoption of one classification rather than another. Typical interest groups are *medical specialists* claiming the value of new sorts of treatment, *public health officials* claiming the value of sanitation in cities, and *economists* claiming the beneficial effects of a rise in the standard of living. An apparently simple cause of death in terms of symptoms (infant diarrhoea) may be accounted for in term of an implicated bacteria, a sanitation category, or a nutrition category. Whether such categories are available, whether they are used, and how they are aggregated have consequences for future allocation distribution.

Other conflicts are present in the way information is collected and coded, and in variations of the speed with which it is returned. Different countries typically have different reporting methods and practices at individual and national levels. Two of the many national examples given are: a period when the USSR did not record causes of death in areas with less than 10,000 inhabitants; “the curious case of Japan’s low rate of fatal heart attacks”. For the latter, many dietary and life style explanations were sought before it was realised that the aggregate of “heart attacks” *and* “strokes” was not very different to other countries. There was a cultural bias in Japantowards “strokes” as

a sign of an overworked brain, and *against* “heart attacks” as a sign of demeaning physical labour.

Different individual (reflecting national) practices are also illustrated by very different reporting practices on abortion, suicide, and stillbirth/miscarriage. In this case, the different practices are clearly related to very different national laws, insurance conditions, and ethical traditions. The tensions produced by different values, needs, and practices of doctors, epidemiologists, and statisticians have been mentioned. They are compounded by the conflicting purposes of insurance companies, industrial firms, and pharmaceutical companies, etc. who also have roles in the ICD. Other conflicts arise from national desires for status or political control over the ICD, and between highly detailed information suitable for developed and computerised countries and basic trend information more suited to the problems of developing countries.

The whole thing, as Bowker & Star aptly remark,

“is not so much a list of causes of death as a series of dynamic compromises between a wide range of players in a number of dimensions”

((Bowker and Star, 1991)P.76)

The ICD, whether as book, list, or database is essential for both policy makers and researchers. It is an authoritative version. And like all authoritative versions, it hides different interpretations, underlying compromises, political struggles. Further, it is difficult to dig under the authoritative version for finer grained, original, specific data, or for specific contributors/reporters. Many of those using the data are aware of its limitations, of local qualifications, and or biases. But this is local, expert knowledge, and is not available in any formal or systematic way. The question is what difference it would make, if any, to project it into Cyberspace using VR, 3-D representation, and related techniques.

(from Robinson 1995, “The Glass Mountain”, COMIC DRAFT)