

Digital Archiving as Information Production: Using Experts and Learners in the Design of Subject Access

Journal:	<i>Journal of Documentation</i>
Manuscript ID:	JD-Apr-2012-0047.R1
Manuscript Type:	Article
Keywords:	Classification schemes, Archives, Electronic media

SCHOLARONE™
Manuscripts

Review

1.0 Introduction

Archives exist as one of the major channels through which culture is transmitted from generation to generation. Through the act of archiving materials that are perceived to be of cultural value are preserved for future display, appreciation, and study. Traditionally performed by information professionals and subject experts the work of archiving is today being greatly influenced by the near-pervasive presence of digital, networked, information environments. These networked information environments are impacting on the work of archiving in at least two main ways: a) the materials being archived are as likely to be in digital form as they are to be in tangible form, and b) the labour involved in producing a digital archive can be divided not only between a single or small group of information professionals, and subject experts, but can also involve the contributions of a larger group of archival users. In summary, while digital archives present information professionals and subject experts with new archival challenges of selection, organization, and display, a networked information environment can also constitute a democratic opportunity to involve users in the production of the digital archive and its value-added distribution.

Informed by these new digital challenges, the purpose of this article is to develop a framework that considers digital archiving as a form of networked information production (Benkler, 2006), in which the different stages of producing a digital archive are modularized and distributed across different actors. The framework is applied and developed within the context of designing a digital archive for the electronic artwork *Rider Spoke*. More specifically the framework is applied and developed within the context of designing a subject scheme that provides its users with consistent yet relevant access to the content of the archive. The article is structured as follows. A description of the materials being archived is presented first, along with a discussion of the specific challenges which electronic artworks present for bibliographic control. This is followed by a review of current approaches to designing subject access to archives of electronic artworks. The approach and methods used to design a subject scheme for the *Riders Have Spoken* archive is then explained and discussed. The article concludes with implications of the information production framework for the design and development of other digital archives.

2.0 *Rider Spoke*

Rider Spoke (http://www.blasttheory.co.uk/bt/work_rider_spoke.html) is a pervasive location-based game that lasts for about an hour and can be played by up to one hundred cyclists at any one time; the game was created by British artists *Blast Theory* (<http://www.blasttheory.co.uk/bt/index.php>). After a brief introduction at which game players pick up a bicycle, a tablet computer that is mounted on the bicycle's handlebars, and a pair of headphones, players will enter a 'mixed reality' gaming experience. As part of the game, cyclists will journey around a physical urban space. As they cycle around they will at the same time be prompted by the artist, at particular points during the journey (via a program stored on the console) to find a hiding-place and to either record messages or to listen to messages left by other cyclists or Riders (see Figure 2). While a rider will always be asked to record a short description about themselves, any succeeding questions will be randomly selected by the program. For example, one question asks: "I want you to look for a flat or a house and find a

Understanding Subject Access as Information Production

1
2
3 window that you would want to go through. I want you to stare into that window and
4 tell me what you see and tell me why you want to go through that window”. Other
5 questions might relate to acts of spectatorship and the elicitation of personal
6 narratives. A final question invites the rider to make a promise and to shout this into
7 the air. The game ends with the cyclist riding back to the starting-point and returning
8 the bicycle and the technology to the organizers. Through the virtual exchange of
9 such messages the *Rider Spoke* game therefore acts as a meeting-point for people who
10 do not encounter each other in real life. *Rider Spoke* has been running since 2007 and
11 has been staged among other venues at the following locations: Brighton, London,
12 Athens, Budapest, Sydney, Adelaide, Liverpool and Linz. The record of a *Rider*
13 *Spoke* performance consists of audio-recordings of cyclists’ utterances, wi-fi data, and
14 GPS traces. These materials have been supplemented with video recordings, artistic
15 commentaries and interviews.
16

17 18 2.1 What is *Rider Spoke* about? 19

20 The first task in designing a digital archive is, once the materials have been selected,
21 to provide facilities that enable users to browse or search its content. Designing these
22 facilities involves exerting some form of bibliographic control over the materials,
23 which in turn implies understanding the subject content of the materials. What a work
24 or document is ‘about’ has been a topic of considerable interest to information science
25 (e.g. Maron, 1977; Svenonius, 1994, 2000):
26

27
28 *“it would seem that the most colossal labor of all involved in*
29 *organizing information is that of having to construct*
30 *an unambiguous language of description – a language that imposes*
31 *system and method on natural language and at the same time*
32 *allows users to find what they want by the names they know”*
33 (Svenonius, 2000: 14).
34
35

36 Since understanding what a work is about involves some degree of normalization, it is
37 an issue that is particularly pertinent to humanistic works. Here a premium is placed
38 on interpretation rather than positivist notions of the objective content of a work, and
39 a consensus on terminology is harder to establish than in the sciences. This can be
40 illustrated with reference to the different perspectives one might use as starting-points
41 for the construction of an ‘unambiguous language of description’ for the *Rider Spoke*
42 work. For example one might take Blast Theory, the creators of the work, as an
43 authority on the meaning of *Rider Spoke* and its archival materials. Much of what is
44 meaningful about the work resides however in the players’ responses to the artists’
45 questions. Hence, given the interactive and participative nature of the game, the
46 creators of the work are only able to provide a partial perspective on its subject matter.
47 Alternatively one might consider a first-person perspective, and use the utterances of
48 a sample of riders as a starting-point for the development of terminology. While a
49 promising starting-point it once again does not seem the securest of foundations for
50 the development of ‘unambiguous language of description’. One might also draw on a
51 third person perspective; and develop one’s terms from the vocabularies of schema
52 that currently exist within the electronic arts field. Finally one can consider taking
53 advantage of a networked information environment and its possibilities for supporting
54 more self-organizing approaches to bibliographic control:
55
56
57
58
59
60

Understanding Subject Access as Information Production

1
2
3
4
5
6
7
8
9
10
11
12
13
“The rise of the Internet is affecting the actual work of organizing information by shifting it from a relatively few professional indexers and catalogers to the populace at large. In other words, the work is becoming deprofessionalized [...] to the extent that the bibliographic universe can be organized by keyword access and beyond that by the voluntary efforts of individuals who mount information on the Web it is self-organizing. While not consciously teleological, a self-organizing bibliographical universe nevertheless succeeds in meeting the bibliographic objectives in part, occasionally, and somewhat randomly. And for many documents and many users that is all that is needed”

14 (Svenonius, 2000: 26). And since these comments were made tagging and
15 folksonomies have become a common and cost-effective addition to the organizing of
16 information. A networked information environment can also support the
17 modularization of the tasks involved in organizing information, and thereby enable
18 collaboration between different groups involved in the bibliographic process. Before a
19 methodology and exercise is described that combines the analyses of users and expert
20 is described, a review of existing approaches to subject access in the electronic arts is
21 first presented.
22
23

24 **3.0 Current Approaches to Subject Access in the Electronic Arts**

25
26 Existing bibliographic tools for subject access in the electronic arts range from simple
27 lists of terms to more complex thesauri that utilize the functions of a fully developed
28 subject language. While lists of terms and vocabularies can enable the identification
29 and grouping of like works, fully developed subject languages fulfill these and more
30 sophisticated browsing and search objectives by adding a semantics that relates
31 different subject categories and/or terms, and syntax for combining different terms.
32
33

34 The vast majority of subject schemas in the electronic arts have been developed by
35 organizations for the purposes of exercising bibliographic control over large
36 collections of artworks. The main objective has therefore been the development and
37 manipulation of a general vocabulary that enables the identification, location, and
38 collocation of works that are relevant to a subject request. The standard subject
39 schema, and the most frequent starting-point for the development of others, is the
40 Getty Museum’s Art & Architecture Thesaurus (AAT) ([http://www.getty.edu/](http://www.getty.edu/research/tools/vocabularies/aat/index.html)
41 [research/tools/vocabularies/aat/index.html](http://www.getty.edu/research/tools/vocabularies/aat/index.html)). The AAT is a general classification
42 scheme with terminology drawn from existing authority lists, and the literature of art
43 and architectural history. The AAT is developed in conjunction with a scholarly
44 advisory team. The methods of construction used to design the AAT draw on existing
45 MESH standards for the arrangement of headings and hierarchies and the flagging of
46 preferred terms. The AAT has a category semantics that organizes and differentiates
47 the subject of art and architecture into seven facets (associated concepts; physical
48 attributes; styles and periods; agents; activities; materials; and objects) and further
49 sub-hierarchies. The ‘objects’ facet is the most populous in terms of hierarchies. A
50 relational semantics links the terms of one hierarchy with another. Therefore enabling
51 the display of poly-hierarchical arrangements, in which the terms of one hierarchy
52 exists in multiple contexts as well as relations of equivalence with other terms. Each
53 of the AAT hierarchies is browsable, while the thesaurus as a whole can be searched
54 via standard Boolean operators. In summary the terminology of the AAT is
55 underpinned by literary and scholarly warrant, while the functions of a fully
56
57
58
59
60

Understanding Subject Access as Information Production

1
2
3 developed subject language are fulfilled by a category and relational semantics, and a
4 syntax for retrieval purposes. While the AAT, and its construction techniques, is
5 without doubt the standard reference point for the indexing of humanistic works, the
6 generic nature of its terminology places limitations on its utility for electronic
7 artworks. Indeed it is the issue of vocabulary, i.e. what terms to use and how a
8 consistent vocabulary can be developed and assignment, that first faces the designer
9 of a subject language in the area of electronic art:
10

11
12 *“In the young field of electronic art, there is still a lack of established, agreed-*
13 *upon terminology for describing the content of works, projects, activities and*
14 *documents [...] The field of electronic art is unique in its extreme*
15 *interdisciplinarity. Unlike strictly discipline-based, often scientific thesauri, a*
16 *thesaurus for electronic art will typically cover an extremely broad range of*
17 *terms, related to many aspects of contemporary society, from architecture and*
18 *transportation to Internet technology, from politics and activism to medical*
19 *science. For most of these fields and disciplines, an electronic art thesaurus*
20 *will not go into small detail but rather include general terminology” (V2_*
21 *2004).*
22

23
24 As the author points out the terminology developed for gaining access to electronic
25 artworks is for interdisciplinary reasons often generic, rather than specific to a
26 particular work. As part of the *Capturing Unstable Media Project* from which this
27 comment is taken, a thesaurus of keywords was developed which acted as a source for
28 describing the works contained in the V2_ organization’s archive of electronic art.
29 The V2_ thesaurus was built through a combination of “top-down” and “bottom-up”
30 methods:
31

32
33 *“as top-down methodology, existing thesauri were used as inspiration for the*
34 *basic structure and for many relationships with V2_’s thesaurus. Besides this,*
35 *most of the terms were developed from actual (bottom-up) analysis of V2_’s*
36 *archive data” (V2_ 2004).*
37

38 Unlike the AAT terminology, the warrant for the terminology of the V2_ thesaurus is
39 the work(s) themselves rather than an existing scholarly vocabulary. In construction
40 however the V2 thesaurus closely follows the techniques of the AAT thesaurus. The
41 V2_ thesaurus has a category semantics that arranges the vocabulary into eight facets,
42 with six of the AAT’s seven facets retained and two further facets, ‘geographical
43 locations’, and ‘themes’, added. In contrast to the AAT the V2_ hierarchies are less
44 numerous however because of the V2_ organization’s wish for a flatter type of
45 knowledge structure; an approach reflected in a preferred emphasis in their thesaurus
46 design on a relational semantics and the development of a semantic network between
47 the terms of different hierarchies. Finally although the V2_ thesaurus and its
48 hierarchies can be browsed they cannot be searched. In summary, while retaining
49 some of the AAT’s construction techniques, the work-based warrant underpinning its
50 vocabulary and the flatter representation employed in the organization of its subject
51 knowledge make it quite a different schema from that of the AAT.
52
53

54
55 A further recent initiative in the field of electronic arts is that of the Documentation
56 and Conservation of Media Arts Heritage (DOCAM) research alliance (2005-2010).
57 Funded by the Daniel Langlois Foundation for Art, Science, and Technology, this
58 initiative identified and implemented five research axes relevant to the preservation of
59
60

Understanding Subject Access as Information Production

1
2
3 media arts, one axis of which was ‘terminology’. One of the outcomes of the
4 DOCAM initiative has been the production of a ‘Glossaurus’ (<http://www.docam.ca/glossaurus>) consisting a list of terms and an additional semantics, hence the name
5 ‘glossaurus’. The source and warrant for the vocabulary were case studies of works
6 of electronic art; hence warrant is work-based rather than literary-based. Its semantics
7 consists of a set of hierarchically organized concepts and terms used by the DOCAM
8 project and a set of relational terms. In pragmatic terms, the glossaurus is intended as
9 a cataloguing rather than a retrieval tool, and there is therefore no syntax. Finally,
10 although neither currently has sophisticated subject facilities but because of their
11 contribution to the field of electronic art, it is worth mentioning the Rhizome
12 Organization and its *ArtBase* (<http://rhizome.org/>) and the *Ars Electronica Archive*
13 (<http://www.aec.at/about/en/archiv/>). The *ArtBase* can be browsed and searched via
14 artist defined keywords, while the *Ars Electronica Archive* maintains an online
15 catalogue and picture archive of works that have appeared at the Ars Electronica
16 festival. A keyword search facility of full-text descriptions of the works contained
17 within the archive is available.
18
19
20

21
22 In summary, there are a number of ways in which subject access for works of
23 electronic art differs from that developed within the arts more generally a) One of the
24 major issues is terminology. The nascent and interdisciplinary nature of the field
25 means that any terminology is underpinned by work rather than literary warrant; with
26 terms and a vocabulary developed almost exclusively from the ground up b) In
27 keeping with the arts more generally, the emphasis in subject language design is
28 normally on the development of a semantics for browsing purposes rather than a
29 syntax for retrieval purposes.
30

31 4.0 Information Production

32
33 Digital networks afford the capacity to upload and distribute material outside of
34 centralized control or market mechanisms. This has led to new forms and modes of
35 ‘information production and exchange’ (see fig. 1) in which individuals can share,
36 distribute and consume information in a manner that is collaborative and non-
37 proprietary (Benkler, 2006). Benkler (2006) identifies three main functions or
38 components of information production:
39
40

41 *“First, there is an initial utterance of a humanly meaningful statement.*
42 *Writing an article or drawing a picture, whether done by a professional or an*
43 *amateur, whether high quality or low, is such an action”* (Benkler, 2006: 68).
44

45
46 The initial utterance can consist either of an individual utterance in any medium (e.g.
47 text, podcast, video clip) or a collection of utterances (e.g. Wikipedia).
48

49 *“Second, there is a separate function of mapping the initial utterances on a*
50 *knowledge map. In particular, an utterance must be understood as “relevant”*
51 *in some sense, and “credible”* (Benkler, 2006: 68).
52

53
54 The second component of information production concerns the sorting and evaluation
55 of the available utterances in accordance with criteria of relevance and credibility. For
56 Benkler, relevance is the process of mapping the utterances on to a subjective
57 knowledge map, the outcome of which may be a tag or a review for example; while
58 credibility is the process of mapping the utterances on to an objective knowledge map.
59
60

Understanding Subject Access as Information Production

Prior judgments of subjective relevance can be aggregated and displayed as part of an objective knowledge map. This might take the form of a numerical rating system, or further organized moderation of the initial judgments e.g. Wikipedia. The final component of networked information production consists of value-added distribution. As Benkler points out the question of distribution is something of a non-issue in a networked environment. He nevertheless points to examples of information products, where the primary means of adding value is the ability of networks to distribute material to which the further contributions of others can be made. In this regard he cites the example of Project Gutenberg, where distributed proofreaders add value to the initial digitization of texts.

While the ‘information production and exchange’ framework can be described as an integrated whole, it is important to point out that because of the digital format in which the material appears in a networked information environment each of the individual functions can easily be performed by different individuals or groups of people:

“In the mass-media world, these functions were often, though by no means always, integrated. NBC news produced the utterances, gave them credibility by clearing them on the evening news, and distributed them simultaneously. What the Internet is permitting is much greater disaggregation of these functions” (Benkler, 2006: 69).

In the next section we describe an exercise that focuses on the relevance/accreditation function in the production of a digital archive, while at the same time situating this function within a broader information production and exchange chain.

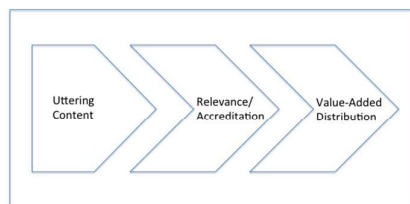
4.1 Information Production and Subject Access

This section describes an approach to designing subject access that retains a distinction between subjective and objective notions of aboutness, and Benkler’s distinction between subject relevance and objective credibility. In keeping with the possibilities of the information and exchange chain the relevance and accreditation ‘sub-functions’ are modularized and performed by different people. A description of the methods used to perform the relevance sub-function is described first and this is

followed by a description of the methods used to perform the accreditation sub-function. *Relevance*. In order to generate a vocabulary relevant to users, an exercise was devised in which a selection of material from the archive would be tagged and/or annotated by a group of students. The materials consisted of four videos, each of between 40 mins and 1 hr in length. The broad subject of each video relates to different perspectives or ‘trajectories’ through the *Riders Have*

Spoken archive (Benford et al., 2009) e.g. a technical design perspective on the game, an artistic viewpoint on the game. Each video shadows a player as they participate in the game; and is accompanied by commentary written by the designer of the trajectory. Seventy-four postgraduate students from the Information School at the

Figure 1 Information Production and Exchange Chain



(Benkler, 2006)

Understanding Subject Access as Information Production

1
2
3 University of Sheffield were invited to tag the four videos from the *Riders Have*
4 *Spoken* archive as a voluntary exercise. Students were evenly distributed across the
5 four videos and each participant was invited to generate up to ten tags; with each tag
6 or annotation representing a point of interest in the content of the video for viewer.
7 The time was also noted. Forty-six students completed the exercise and this generated
8 356 user tags. As a collection these tags and annotations represent the terms and
9 vocabulary upon which a subsequent subject analysis was conducted and a subject
10 language developed. *Accreditation*. A key function of any subject language is the
11 ability to categorize i.e. differentiate between and group together works or materials
12 that are either similar in content or have some shared attribute. From a bottom-up
13 perspective tools such as tag clouds and folksonomies can aggregate users'
14 contributions in this regard. Although more participative, and democratic, in intent
15 these tools are limited by definition in their ability to provide consistent subject
16 descriptions; hence Benkler's distinction between relevance and credibility. For a
17 more objective knowledge mapping to occur some more objective feature of the work,
18 or a concept needs to be employed as part of a top-down approach. The concept of
19 trajectory chosen here is that of trajectory. This is a concept drawn from the work of
20 the sociologist Anselm Strauss. A trajectory can be defined as "*any experienced*
21 *phenomenon as it evolves over time [...] and the actions and interactions contributing*
22 *to its evolution*" (Strauss, 1993: 53). Strauss cites 'an engineering project', a 'chronic
23 illness', 'dying' and a 'national revolution' as examples of trajectories. What
24 underpins the concept is a general theory of action that proposes that sociological
25 phenomena unfold over time, and that while the course of such phenomena may be
26 anticipated they are not entirely predictable; with the course of an engineering project
27 being shaped for example by the actions and interactions of the participants. The
28 applicability and use of the concept here relates to a number of viewpoints that can be
29 taken on the *Rider Spoke* work a) the first person viewpoint of the cyclist who by
30 participating in the game will clearly come away with an experience that gradually
31 unfolds as s/he cycles around the city making recordings and listening to those of
32 others b) the second person viewpoint of the artist narrator who addresses the cyclist
33 through the asking of the questions that act as a key structuring device for the
34 unfolding action of the game (a 'trajectory scheme') and c) the third person
35 perspectives of the game designer (a 'projected trajectory' through a city space from
36 which a cyclist may deviate for example), or subsequent listeners and viewers of the
37 materials in the archive, who may wish to comment on the actions and interactions
38 of the cyclists (segmentation of a trajectory into 'phases' for example); along with
39 what these unfolding actions and interactions shape or are shaped by e.g. spatial
40 environment. Strauss further explicates the idea of trajectory with reference to 'orders'
41 that he defines as "analytic abstractions that summarize what, with regard to an
42 evolving phenomenon, the actions and interactions, are directed at shaping". These
43 orders include: spatial, technological, sentimental and aesthetic orders that are shaped
44 by the unfolding actions and interactions of the *Rider Spoke* participants. These orders
45 are defined here and formed the basis for the main facets of the subject schema. These
46 orders or facets acted as the objective knowledge map onto which the additional
47 utterances or relevance judgments of the students would be mapped on to. 'Space'
48 refers to a spatial order that summarizes "how objects are arrayed in given spaces;
49 how actions take place or are supposed to take place in certain spaces"; technology
50 refers to a technological order and refers to "action that requires machinery or
51 equipment or other 'hard' technology; bit technological order is equally characteristic
52 of any kind of action-there are always at least procedures that constitute significant
53
54
55
56
57
58
59
60

Understanding Subject Access as Information Production

1
2
3 “soft” technology; ‘sentiment’ refers to a sentimental order that refers to matters such
4 as “moods, motivations, organizational climates, and interpersonal relationships”; and
5 aesthetics refers to “proper style, or appropriately aesthetic standards as conceived by
6 actors”.

7
8 Four of the orders or top-level “facets” were taken from Strauss’s trajectory
9 framework to form the top-level categories or facets of the subject scheme. This was
10 supplemented by a further facet taken from the AAT thesaurus; that of ‘Agents’.
11 Classical content analysis (Bauer, 2000) was then used to group and organize the
12 students’ tags by like subject matter into facets, hierarchies and sub-hierarchies.
13 Following a preliminary analysis the four initial facets were supplemented with one
14 further facet taken from the AAT, ‘agents’. A brief description of the content of each
15 facet and its hierarchies is presented here while a tabular representation of the current
16 subject schema is presented in *Appendix 1. Agents*. This was the most well-populated
17 facet accounting for half (50.56%) of the tagged content. Tags were grouped into four
18 separate hierarchies representing different kinds of agency relevant to the unfolding
19 action or recording of the game: Blast Theory, Game Participants, Game Non-
20 Participants and Researchers. Of the four hierarchies, the Game Participants hierarchy
21 accounted for the vast majority of the content (78.33%), which incorporated users’
22 observations of the actions of the key participant in the game i.e. the Rider; and the
23 key action performed being recording which accounted for nearly half of the observed
24 actions (47.22%). *Space* was the second most well-populated facet accounting for a
25 further quarter (24.71%) of the tagged content. Tags were grouped into three separate
26 hierarchies: observations of the built environment, the natural environment, and
27 transport. Users’ observations of the built environment within the film was the most
28 numerous (61.36%) and differentiated of the hierarchies containing observations of
29 architecture, shops, street signs etc. With *Agents* and *Space* accounting for three-
30 quarters (75.27%) of the observed content, the remainder of the content was
31 accounted for by aesthetic observations (10.11%) e.g. categorized under an *Aesthetics*
32 *Facet*; observations of the use of technology in the playing and recording of the game
33 (9.83%) categorized under a *Technology Facet*; and finally users’ observations of the
34 moods of riders, bystanders, or comments as to the effect of the viewing experience
35 on their own mood (4.79%).

36 5.0 Discussion

37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
Once the archival materials have been identified or selected, the initial task of a
subject indexer is to source and/or develop terms and a vocabulary for the materials
being described. In this regard the nascent and interdisciplinary field of electronic art
presents the subject indexer with considerable challenges. To date the conventional
solution has been to adopt an approach from the ground up and to use the work itself,
rather than an existing scholarly vocabulary as an initial basis for terminological
development. The approach taken to the construction of the subject language for the
Riders Have Spoken archive is one informed by a networked information production
framework and in particular the relevance/accreditation function of this framework.
The virtue of adopting this framework from a subject indexing perspective has been a)
to provide a systematic underpinning for combining the subjective and objective
aspects of aboutness as part of the same logical framework, and b) to demonstrate the
potential for modularizing and distributing responsibility for the development of these
aspects that a networked information environment affords.

Understanding Subject Access as Information Production

1
2
3 More specifically within the relevance/accreditation function of information
4 production, the approach adopted here in the construction of a subject language for
5 the *Riders Have Spoken* archive has also been one that has, at the vocabulary stage,
6 been developed from the ground up. Rather than using the work however as a basis
7 for the development of vocabulary, the terms have been developed from the grounded
8 up in the sense that initial responsibility for the development of terms has been
9 handed over in a democratic sense to users. This has the virtue of developing terms
10 and a vocabulary that are relevant to users. In contrast to vocabulary development, the
11 category semantics of a subject language for accessing works of electronic art have
12 conventionally been developed in a top-down fashion i.e. the main organizing
13 categories have been identified through the recognition of existing standards, and
14 predominantly those of the AAT. While the approach adopted in developing a
15 category semantics for the *Riders Have Spoken* subject vocabulary can be termed to-
16 down, it can nevertheless be distinguished from previous approaches in the sense that
17 the relevance of using the organizing concept of trajectory, and its categories —along
18 with the consistency of these categories — can be justified from within the particular
19 context of the work itself; rather than having been a set of categories drawn from a
20 necessarily generic standard. The test of any subject language however will remain its
21 usefulness for accessing the artistic work and its materials. And while the emphasis of
22 the study has been on the innovative development of a relevant yet consistent subject
23 schema, a necessary next step will be to conduct an evaluation of the performance of
24 that subject language, and whether the materials retrieved meets the archival users'
25 needs.
26
27
28

6.0 Conclusion and Implications

31
32 The current study has focused on the construction of a relevant yet consistent subject
33 language that aids in providing access to a digital archive of materials for the
34 electronic artwork *Rider Spoke*. The approach that has been taken to the development
35 of this subject language is one that has been informed by a networked information
36 production and exchange framework that considers digital archiving as a form of
37 information production. The implications of this framework relate both to the
38 processes used in the production of the digital archive, as well as the character of the
39 digital archive produced. In terms of processes the framework enables us to perceive
40 information organization not only as a form of relevance/accreditation production but
41 also as one informational activity among several inter-related activities that in a
42 networked information environment can be performed by different people. As a
43 number of authors have remarked (e.g. Svenonius, 2000; Rafferty & Hilderley, 2005,
44 Winget, 2009) a networked information environment affords the opportunity to draw
45 on the contributions of others. The framework also draws attention to the character of
46 a digital archive that, unlike a traditional archive, can be considered —from initial
47 utterance through relevance/accreditation to value-added distribution — less of a final
48 product than an artifact that is extendable through ongoing creative invention.
49
50
51
52
53
54

References

56
57
58 Bauer, M.W. (2000). "Classical content analysis: a review". In: Bauer, M.W. &
59
60

Understanding Subject Access as Information Production

1
2
3 Gaskell, G. eds., *Qualitative Researching with Text, Image, and Sound*, 131-151.
4 London: Sage.

5
6 Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms*
7 *Markets and Freedom*. London: Yale University Press.

8
9 Maron, M.E. (1977). "On indexing, retrieval and the meaning of about", *Journal of*
10 *the American Society for Information Science*, **28**, 38-43.

11
12 Rafferty, P. & Hilderley, R. (2005). *Indexing Multimedia and Creative Works: The*
13 *Problems of Meaning and Interpretation*. Aldershot: Ashgate Publishing.

14
15 Strauss, A.L. (1993). *Continual Permutations of Action*. Chicago: Aldine Press.

16
17 Svenonius, E. (1994). "Access to nonbook materials: the limits of subject indexing for
18 visual and aural languages", *Journal of the American Society for Information Science*,
19 **45** (8), 600-606.

20
21 Svenonius, E. (2000). *The Intellectual Foundation of Information Organization*.
22 London: MIT Press.

23
24 V2_ (2004). "Capturing Unstable Media", available at:
25 <http://capturing.projects.v2.nl/index.html>

26
27 Winget, M. (2009). "Describing art: an alternative approach to subject access and
28 interpretation". *Journal of Documentation*, **65** (6), 958-976.

Understanding Subject Access as Information Production

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

AESTHETICS

Built environment**Natural environment****Music****Colour**

AGENTS

Blast Theory*Narrator**Rider Spoke***Riders***Cycling**Listening**Recording***Non-Participants***Pedestrians**Diners**Other cyclists***Videographer**

SENTIMENT

Bystanders*Happy***Riders***Fun***Viewer***Fear**Gloom**Horror**Interest*

SPACE

Built environment*Architecture**Bridge**Buildings**Clock**Fountain**Graffiti**Shops**Signs**Street**Tower*

Understanding Subject Access as Information Production

*Townsquare***Natural environment***Grass**Mountain**Park**River**Trees**Twilight**Weather***Transport***Cars**Tram**Van*

TECHNOLOGY

Game Technology*Bicycle**Camera**Cellphone**GPS**Tablet**Wifi***Recording Technology***Audio**Video*

Digital Archiving as Information Production:
Using Experts and Learners in the Design of Subject Access

1.0 Introduction

Archives exist as one of the major channels through which culture is transmitted from generation to generation. They preserve materials perceived to be of cultural value for future display, appreciation, and study. Until quite recently the work of archiving has been performed by subject experts and information professionals alone. This is a situation that is rapidly changing as a digitally networked environment provides a platform not only for the production of a digital archive but also for access to a larger group of users who can be enrolled in the work of producing the archive. While digital archives present both subject experts and information professionals with new archival challenges of selection, organization, and display, the digitally networked environment that underpins them is also a democratic opportunity to involve users both in the production of the archive and to contribute to its value-added distribution.

The purpose of this article is to develop a framework that treats digital archiving as a form of networked information production in which the different stages of producing the archive are modularized and distributed across different actors (Benkler, 2006). The framework is applied within the context of the design of a digital archive for a participative and interactive artwork called *Rider Spoke*. More specifically the framework is applied within the context of designing subject access to the materials in the *Riders Have Spoken* digital archive. The article is structured as follows. First, a description is provided of the game and the materials to be archived, along with a discussion of the specific challenges that electronic artworks present for bibliographic control. Second, a review is presented of previous approaches to designing subject access for archives of electronic artworks. Third, the approach and methods used to design subject access to the *Riders Have Spoken* archive are described and discussed. The article concludes with identifying some of the implications of networked information production for the design and development of digital archives.

2.0 *Rider Spoke*

Rider Spoke (http://www.blasttheory.co.uk/bt/work_rider_spoke.html) is a pervasive location-based game for cyclists. Created by British artists Blast Theory (<http://www.blasttheory.co.uk/bt/index.php>) the game lasts for about an hour and involves multiple participants riding around a city space, responding to questions and leaving recorded messages for other cyclists to pick up. After a brief introduction during which the players of the game pick up a bicycle, a tablet computer mounted to the bicycle's handlebars, and a pair of headphones, the players enter a 'mixed reality' gaming experience. As the cyclist sets off on their journey around the city space, they are prompted by the artists, via a tape-recorded question, to find a quiet area or hiding-place and using the tablet to record a short description about themselves. As the cyclist continues to ride around the space, they are prompted by further questions randomly selected from a set of questions devised by the artists. One question asks for example: "I want you to look for a flat or a house and find a window that you would want to go through. I want you to stare into that window and tell me what you see and tell me why you want to go through that window". Other questions pertain to acts of spectatorship, voyeurism, and the elicitation of personal narratives. At each point the rider has the opportunity either to record their response to the question or to listen to

Digital Archiving as Information Production:
Using Experts and Learners in the Design of Subject Access

the messages that other unknown cyclists have left in that geographical position. Through this form of technologically mediated exchange share messages, often intimate in nature, are shared between strangers who will never meet in real life. The final question invites the cyclist to make a promise and to shout this into the air. The game ends with the cyclist riding back to the place from where they set off, and returning the bicycle, tablet, and headphones to the organizers. In summary the *Rider Spoke* game and its mobile networked technologies, i.e. tablet, wifi, GPS, act as a meeting-point and a mechanism for the virtual exchange of messages between people who would not otherwise encounter each other in real life. *Rider Spoke* has been running since 2007 and has been staged among other venues at the following locations: Brighton, London, Athens, Budapest, Sydney, Adelaide, Liverpool and Linz. The record of a *Rider Spoke* performance consists of audio-recordings of cyclists' utterances, wifi data, and GPS traces. These materials have since been supplemented with video recordings, artistic commentaries and interviews. The materials used here relate to video documentaries of the staging of the work in Linz in 2009.

2.1 What is *Rider Spoke* about?

One of the first tasks facing the designer of a digital archive is, once the materials have been identified and selected, to design the facilities that enable users to browse or search the archival content. Designing these facilities involves exerting some form of bibliographic control over the materials; which in turn involves some understanding something of their subject content. Judging what a document or work is *about* has been a topic of considerable interest to information scientists (e.g. Maron, 1977; Svenonius, 2000, 1994):

“it would seem that the most colossal labor of all involved in organizing information is that of having to construct an unambiguous language of description – a language that imposes system and method on natural language and at the same time allows users to find what they want by the names they know”
(Svenonius, 2000: 14).

This labour of establishing an ‘unambiguous language of description’ for a text or work is more difficult to establish in the arts and humanities than it is in the natural sciences. Indeed it can be considered anathema to the former. While the phenomena of natural science, its objects and their attributes, can be readily agreed upon and enumerated; no such stable description is available to the interpreter of humanistic works, except in general terms. At stake is the ‘significance’ of the artwork, and how significance can reside in the reader’s, viewer’s, or participant’s, experience of the work, rather than the creator’s or critic’s interpretation of what the work is ‘about’ (Vickery, 2008). The issue of significance can be illustrated with reference to the possible perspectives that a curator might use as a starting-point for the construction of an ‘unambiguous language of description’ for *Rider Spoke*. One could take the creators of the work, Blast Theory, as an authority on the meaning of *Rider Spoke* and its materials for example; much of what is meaningful about the work resides with the participants though and in the cyclists’ responses to the artists’ questions. Given the participative and interactive character of the work, its creators are only able to provide a partial perspective on its possible subject matter. Alternatively, one could consider a

Digital Archiving as Information Production:
Using Experts and Learners in the Design of Subject Access

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

first-person perspective, and use a sample of the cyclists' utterances as a starting-point for terminology. While having the virtue of identifying words and terms in the art work itself this approach doesn't seem to be a secure foundation for the development of an 'unambiguous language of description'. Additionally the cyclists spoken utterances weren't the primary focus of this version of the archive. One could also draw on a third person information scientist's perspective; and develop one's terms from the vocabularies or schema currently existing within the electronic arts. Finally one could however consider taking advantage of a networked information environment and the capacity of such an environment to support more self-organizing approaches to bibliographic control:

"The rise of the Internet is affecting the actual work of organizing information by shifting it from a relatively few professional indexers and catalogers to the populace at large. In other words, the work is becoming deprofessionalized [...] to the extent that the bibliographic universe can be organized by keyword access and beyond that by the voluntary efforts of individuals who mount information on the Web it is self-organizing. While not consciously teleological, a self-organizing bibliographical universe nevertheless succeeds in meeting the bibliographic objectives in part, occasionally, and somewhat randomly. And for many documents and many users that is all that is needed" (Svenonius, 2000: 26).

Since the time of these comments, tagging and folksonomies have become common and cost-effective solutions to the problem of facilitating access to information. A networked information environment can also support the modularization of the tasks involved in facilitating access, and enable collaboration between different groups involved in the bibliographic process. Before a methodology and exercise is described that combines the contributions of experts and learners in addressing the problem of designing subject access, a review is presented first of existing approaches to designing subject access for electronic artworks.

3.0 Designing Subject Access for the Electronic Arts: Current Approaches

As previously mentioned the task of establishing a consensus on the significance of a text or work is one that is much more difficult to achieve in the arts than in the sciences. Given the heterogeneous construction of humanistic works, the differing experiences they instill in the reader, and the different interpretations that can be placed on them, the development of an 'unambiguous language of description' that reduces their subject matter to the enumeration of its basic attributes of the work can be accomplished only at a general level e.g. artist, time period, genre etc. For some purposes, e.g. cataloguing collections of artworks, a general level of description is sufficient for practical purposes. For the description of particular single works a general level of description exists at too high a level of abstraction. Given the problem of establishing a definitive perspective on the significance of an artwork and establishing therefore a definitive account of its subject matter, a common strategy has been to adopt a faceted approach to their classification. In contrast to an enumerative approach to classification in which all the possible classes of a subject are exhaustively enumerated from the 'top down'; a faceted approach adopts an analytico-synthetic approach in which a subject is analysed into its basic isolate

Digital Archiving as Information Production:
Using Experts and Learners in the Design of Subject Access

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

subjects and strings formed ‘bottom up’ through synthesis. The need for enumeration leads to a lot of pre-coordinated compound classes that are placed in hierarchical superordinate and subordinate semantic relationships. A faceted classification scheme on the other hand dispenses with the enumeration of compound subjects, separates out a subject into its basic facets and sub-facets and places a greater emphasis on post-coordinate syntax for the formation of compound subjects (Broughton, 2004). The faceted approach to classification is mostly commonly associated with the work of Ranganathan (1963), and the work of the UK Classification Research Group (e.g. Langridge, 1976; Vickery, 1960, 1968). Dissatisfied with current classification schemes that were unwieldy in the way they demanded exhaustive enumeration of their subjects and long schedules, Ranganathan famously originated the faceted approach after a visit to Selfridges toy store, where he came upon a toy meccano set; in which complex designs could be assembled out simple isolates (La Barre, 2010). As will be seen there are a number of issues that confront the designer of a classification scheme for a particular single electronic artwork, that make the use of an enumerative approach unrealistic, and the development of a faceted approach appropriate. These concern an interdisciplinary rather than homogeneous subject field and the development of terms and a vocabulary in order to populate these facets that does not rely on literary warrant; the latter being a problem that compounded in the performing arts when the work being classified is either wholly, or partially, visual and non-text-based in its traditional sense.

The vast majority of existing subject schemas in the electronic arts have been developed by galleries, museums and other institutions for the purposes of exercising bibliographic control over collections of artworks rather than a particular single artwork. Therefore the main objectives have been the development of a general level of description and the manipulation of a general vocabulary that enables the institution and the user to identify, locate, and collocate works that are relevant to a subject request. The standard for subject schemas in the arts is the *Getty Museum’s Art & Architecture Thesaurus* (AAT) (<http://www.getty.edu/research/tools/vocabularies/aat/index.html>). In the AAT the subject of Art & Architecture is divided initially into seven facets and then twenty-two hierarchies. The top-level facets are physical attributes; styles and periods; agents; activities; materials; objects; and associated concepts; with an *Objects* facet being the most populous in terms of hierarchies. Its methods of construction draw on existing MESH standards for the arrangement of headings and hierarchies and the flagging of preferred terms; with a relational semantics linking the terms of one hierarchy with another. This enables the display of poly-hierarchical arrangements, in which a term within one hierarchy can exist both in a hierarchical relationship with broader and narrower terms from other hierarchies, as well as related equivalent terms. AAT hierarchies are browsable, and the thesaurus as can be searched as a whole with standard Boolean operators. The facets are hierarchies are populated with terms taken from existing authority lists and the literature of art and architectural history. The AAT has also been developed in conjunction with a scholarly advisory team. The basis for the development of the terminology, and the solution for identifying terms of relevance to works in the Getty Museum collection and other institutions is therefore scholarly literary warrant. While its facets provide a possible starting-point for the development of categories for single artworks, the general level of description ensure that its terms lack viability beyond pointing to the attributes of similar artworks at anything other than a generic level of scale.

Digital Archiving as Information Production:
Using Experts and Learners in the Design of Subject Access

In the field of electronic art, issues of significance, heterogeneous construction, and choice of facets, are pertinent for bibliographic control and subject access, as they are for the arts in general. For reasons of interdisciplinarity and a lack of a canonical scholarly language, the issue of vocabulary and warrant are particularly problematic when designing subject access to such works:

“In the young field of electronic art, there is still a lack of established, agreed-upon terminology for describing the content of works, projects, activities and documents [...] The field of electronic art is unique in its extreme interdisciplinarity. Unlike strictly discipline-based, often scientific thesauri, a thesaurus for electronic art will typically cover an extremely broad range of terms, related to many aspects of contemporary society, from architecture and transportation to Internet technology, from politics and activism to medical science. For most of these fields and disciplines, an electronic art thesaurus will not go into small detail but rather include general terminology” (V2_ 2004).

This quotation is taken from a report published by the V2_ organisation as part of its *Capturing Unstable Media Project*. As part of that project a thesaurus of keywords was developed to act as a resource for the subject description of works contained in its archive of electronic art. Like the AAT, the V2_ is generic in its terms and descriptions, partly due to a focus on collections and partly for reasons of vocabulary that permit communication across disciplinary boundaries; it also adopts a faceted approach to subject access. Unlike the AAT however the V2_ thesaurus takes a different approach to the problem of terminology and populating the facets; with the thesaurus designed through a combination of ‘top-down’ and ‘bottom-up’ methods:

“as top-down methodology, existing thesauri were used as inspiration for the basic structure and for many relationships with V2_’s thesaurus. Besides this, most of the terms were developed from actual (bottom-up) analysis of V2_’s archive data” (V2_ 2004).

As a top-down methodology, in construction the V2 thesaurus closely follows the techniques of the AAT thesaurus. The V2_ thesaurus has a category semantics that arranges the vocabulary into eight facets, with six of the AAT’s seven facets retained and two further facets, ‘geographical locations’, and ‘themes’, added. In contrast to the AAT the V2_ hierarchies are less numerous however because of the V2_ organization’s wish for a flatter type of knowledge structure; an approach reflected in a preferred emphasis in their thesaurus design on a relational semantics and the development of a semantic network between related terms across different hierarchies. Unlike the AAT terminology however, as an exercise in ‘actual (bottom-up) analysis’ the terms used in the V2_ thesaurus are based not on a pre-existing scholarly vocabulary, thus on literary warrant, but are derived from the V2_ curator’s own interpretation and analysis of the art works in the V2_ archive. Constrained by a lack of existing scholarly vocabulary the terms in the V2_ thesaurus are underpinned by ‘empirical warrant’ and observation of the artworks by the archivist (Svenonius, 2000).

The issue of terminology continues to be debated. For example, a recent initiative in the field of electronic arts, the Documentation and Conservation of Media Arts

Digital Archiving as Information Production:
Using Experts and Learners in the Design of Subject Access

Heritage (DOCAM) research alliance (2005-2010), funded by the Daniel Langlois Foundation for Art, Science, and Technology, identified the issue of ‘terminology’ among its five research axes relevant to the preservation of media arts. This axis resulted in the production of a ‘Glossaurus’ (<http://www.docam.ca/glossaurus>) that consists of a list of relevant subject terms organized into a hierarchy, with some relational semantics. The criterion for admittance to the glossaurus is again empirical warrant through inspection and analysis of case studies of electronic artworks, rather than literary warrant and an existing scholarly vocabulary. Intended as a cataloguing rather than a retrieval tool, the glossaurus therefore has no syntax. Finally, because of their contribution to the field of electronic art, rather than the development of sophisticated subject facilities, it is worth mentioning the Rhizome Organization and its *ArtBase* (<http://rhizome.org/>) which can be browsed and searched via artist defined tags; and the *Ars Electronica Archive* (<http://www.aec.at/about/en/archiv/>) which maintains an online catalogue and picture archive of works that have appeared at the *Ars Electronica* festival. A keyword search facility of full-text descriptions of the works contained within the archive is available.

In summary, for reasons of heterogeneity and interpretive flexibility it is difficult, if not anathema, to attempt to definitively establish the significance of humanistic works. As a result an enumerative approach that exhaustively lists and pre-coordinates compound subjects has been proven to be unfeasible. What has proven to be feasible is to develop a general language of description through the use of a faceted approach. While standards currently exist in the arts & humanities e.g. AAT that can be drawn upon, such general schema will nevertheless need to be adapted for the classification of particular single works and at the same time address issues of terminology and warrant. An exercise is now described in which a faceted approach is taken to the design of subject access, with terms and a vocabulary developed in concert with learners; and as such based on user warrant rather than on either literary warrant or the empirical warrant of archival experts.

4.0 Information Production

A digitally networked environment affords the capacity to upload and distribute material outside of centralized control or market mechanisms. This has led to new forms and modes of ‘information production and exchange’ (see fig. 1) in which individuals can share, distribute and consume information in a manner that is collaborative and non-proprietary (Benkler, 2006). Benkler (2006) identifies three main functions or components of information production:

“First, there is an initial utterance of a humanly meaningful statement. Writing an article or drawing a picture, whether done by a professional or an amateur, whether high quality or low, is such an action” (Benkler, 2006: 68).

The initial utterance can consist either of an individual utterance in any medium (e.g. text, podcast, video clip) or a collection of utterances (e.g. Wikipedia).

“Second, there is a separate function of mapping the initial utterances on a knowledge map. In particular, an utterance must be understood as “relevant” in some sense, and “credible” (Benkler, 2006: 68).

The second component of information production concerns the sorting and evaluation

Digital Archiving as Information Production: Using Experts and Learners in the Design of Subject Access

of the available utterances in accordance with criteria of relevance and credibility. For Benkler, relevance is the process of mapping the utterances on to a subjective knowledge map, the outcome of which may be a tag or a review for example; while credibility is the process of mapping the utterances on to an objective knowledge map. Prior judgments of subjective relevance can be aggregated and displayed as part of an objective knowledge map. This might take the form of a numerical rating system, or further organized moderation of the initial judgments e.g. Wikipedia. The final component of networked information production consists of value-added distribution. As Benkler points out the question of distribution is something of a non-issue in a networked environment. He nevertheless points to examples of information products, where the primary means of adding value is the ability of networks to distribute material to which the further contributions of others can be made. In this regard he cites the example of Project Gutenberg, where distributed proofreaders add value to the initial digitization of texts.

While the ‘information production and exchange’ framework can be described as an integrated whole, it is important to point out that because of the digital format in which the material appears in a networked information environment each of the individual functions can easily be performed by different individuals or groups of people:

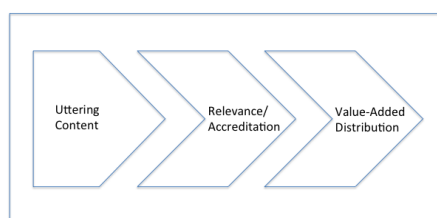
“In the mass-media world, these functions were often, though by no means always, integrated. NBC news produced the utterances, gave them credibility by clearing them on the evening news, and distributed them simultaneously. What the Internet is permitting is much greater disaggregation of these functions” (Benkler, 2006: 69).

In the next section we describe an exercise that focuses on the relevance/accreditation function in the production of a digital archive, while at the same time situating this function within a broader information production and exchange chain.

4.1 Information Production and Subject Access

This section describes an approach to designing subject access that retains a distinction between subjective and objective notions of aboutness in the form of

Figure 1 Information Production and Exchange Chain



(Benkler, 2006)

Benkler's (2006) distinction between subjective relevance and objective credibility. In keeping with the possibilities of the information production and exchange chain, the ‘relevance and accreditation’ sub-function is further modularized further and performed by different people. A description of the methods used to perform the relevance sub-function is described first; this is then followed by a description of the methods used to perform the accreditation sub-

function. The first of these sub-functions addresses the problem of generating terms and a vocabulary by drawing on learners as a resource. The second sub-function addresses the problem of categorizing and organizing the terms and vocabulary generated into meaningful groupings, and facets, by drawing on expertise in

Digital Archiving as Information Production:
Using Experts and Learners in the Design of Subject Access

classification.

Relevance. In the absence of a literary warrant for canonical terms, the first sub-task was to generate terms and a vocabulary based on user relevance and warrant. This took the form of an exercise devised such that the significance of the work would be grounded in the interpretations of a learner community; and a selection of materials from the archive tagged and annotated by a group of students studying information management and information systems.

The materials to be tagged consisted of four videos lasting between 40 mins. and 1 hr. The broad subject of each video relates to different perspectives or ‘trajectories’, e.g. technical, artistic, through the *Riders Have Spoken* archive (Benford et al., 2009). Each video involves an ethnographer shadowing the player as they participate in the game; and is further accompanied by textual commentary written by the trajectory’s designer. Seventy-four postgraduate students from the Information School at the University of Sheffield were invited to tag the four trajectories from the archive as a voluntary exercise. Learners were evenly distributed across the four videos and each participant invited to generate up to ten tags. Each tag or annotation represented a point of interest in the video for the learner. Forty-six students completed the exercise and 356 tags were generated. As a collection these tags and annotations represent the terms on which a subsequent classical content analysis was conducted (Bauer, 2000) and a basic subject language and classification scheme developed.

Accreditation. When a critical mass of terms and a vocabulary have been developed a key function of any subject language and classification scheme is its capacity to group and categorize the terms into categories of terms that share similar characteristics. Drawing on a ‘bottom-up’ approach information tools such as tag clouds and folksonomies aggregate users’ contributions in this regard. Although more democratic and participative in their approach, these tools are by definition limited in their ability to provide consistent subject descriptions; hence Benkler’s (2006) distinction between subjective relevance and objective credibility. For a more objective knowledge mapping to occur a scheme needs to be developed that credibly organizes the terms into a classification scheme based on more objective criteria.

Once the significance or point of view (s) on the work have been determined, the designer of a faceted scheme needs to identify the facets that are relevant to the area. The designer of a faceted scheme for subject access will need to “separate out any particular aspect of interest” and “it is these separately listed aspects that are known as “facets” (Vickery, 2008). For a number of reasons ‘Agents’ as creators and performers of action, was chosen as the starting-point for the scheme. First and foremost agents are key to the work in that it is the rider or cyclist who experiences and performs the game of *Rider Spoke*; and whose perspective acts as a key source of coherence in journeying through the game. In contrast to interacting with objects through interfaces

“the purpose of cultural user experiences is not to reach a destination, solve a problem or complete a task, but rather to enjoy an engaging journey [...] while these journeys may pass through different places, times, roles and interfaces...they maintain an overall sense of coherence; of being part of a connected whole. These journeys are steered by the participants, but are also shaped by narratives that are embedded into spatial, temporal,

Digital Archiving as Information Production:
Using Experts and Learners in the Design of Subject Access

and performative structures by authors. They are also influenced by the dynamic process of orchestration... finally they may be undertaken by groups and/or involve encounters among participants” (Benford et. al., 2009).

Second, the notion of a cultural user experience as a journey or ‘trajectory’ is key to the design of the archive in its current videographic form; and the four different trajectories or films through the *Rider Spoke* work. Third the rider, and other agents including the work’s creator Blast Theory, is in keeping with the first ‘personality’ facet of Ranganathan’s universally applicable and fundamental categories of personality, matter, energy, space and time. Finally the ‘Agents’ facet is one of the main categories in the AAT. The principle of division is by role; and the order within array is by the sequential contribution to the production of *Rider Spoke*. The facet accounted for 13.76% of the tagged content. Once the key facet has been established other candidates for facets can more easily be determined: “Which of the facets would users regard as the one that is key to the area? What is the relation to the key facet, of each of the facets that you have identified? This procedure may give you fresh insights into the facets you need” (Vickery, 2008). The second main facet is ‘Technology’. The principal actor on the stage, the rider, performs *Rider Spoke* through the use of different types of technology e.g. tablet, wifi. Learners also made observations about the recording technology used. This is also an aspect of Ranganathan’s Materials facet, and therefore in keeping with PMEST’s principle of relevant succession (Spiteri, 1998). The principle of division is by type; and the order within array is by relevance of the technology to a *Rider Spoke* event. The facet accounted for 9.30% of the tagged content. The third main facet is ‘Activities’ as riders perform distinct actions with the technology. The principle of division is by kind of activity e.g. cycling, listening, recording; and the order within array is by sequence of first activity. This was the most populated facet in terms of tags and accounted for 37.5% of the tagged content. The fourth main facet is ‘Space’ i.e. the rider uses technology to perform distinct actions with the technology as part of a location-based game. The principle of division is by location, and the order in array by geographical location, and then by form and by function within that location. The facet accounted for 6.65% of the tagged content. The fifth main facet is ‘Objects’ in the sense of objects situated in space. The principle of division is by type of object, and the order in array the frequency with which the user observed the different types of object. The facet accounted for 11.44% of the tagged content. The fifth facet is Time and as been added as a basis for subject access to the different stagings of *Rider Spoke*. The principle of division is by date; with the order in array chronological time. Drawing on the AAT thesaurus the final facet is named “Associated Concepts” and currently carries two sub-facets: aesthetic concepts, specified by form; and psychological concepts, specified by emotion. Together these associated concepts accounted for 17.70% of the tagged content.

In summary the facets follow a principle of relevant succession broadly consistent with Ranganathan’s fundamental categories of PMEST i.e. Agents (Personality), Technology (Materials), Activities (Energy), Space (Space), Objects (Materials), Time (time). While ‘technology’ and ‘objects’ are both sub-facets of Materials; given that technology is integral to the performing of *Rider Spoke*, a decision was taken to have ‘Technology’ as a mutually exclusive facet. Ranganathan’s view of objects also relates to the making of objects rather than their appreciation; and therefore a decision

Digital Archiving as Information Production: Using Experts and Learners in the Design of Subject Access

was also taken to have 'Objects' as a mutually exclusive facet independent of any material association. Mention should also be made of two further resources for exploring possible facets and arrays: the AAT that informed the use of an 'Activities' and an 'Associated Concepts' facet; along with Anselm Strauss's (1993) 'sensitizing concept' of trajectory and related concepts. Strauss defines a trajectory as "*any experienced phenomenon as it evolves over time [...] and the actions and interactions contributing to its evolution*" (Strauss, 1993: 53) and cites 'an engineering project', a 'chronic illness', 'dying' and a 'national revolution' as examples of such trajectories. The course of such trajectories may be anticipated but not entirely predictable as the actions and interactions of participants contribute to its evolution. Strauss further explicates the idea of trajectory with reference to 'orders' that he defines as "analytic abstractions that summarize what, with regard to an evolving phenomenon, the actions and interactions, are directed at shaping". These orders include: spatial, technological, informational, sentimental and aesthetic orders that are shaped by the unfolding actions and interactions of the *Rider Spoke* participants. The proximity of the concept to that of the *Rider Spoke* trajectory and the manner in which Strauss provides a range of concepts associated with the action and interactions relevant to a trajectory, also informed the development of this scheme, and in particular the 'associated concepts' of aesthetics and emotion.

5.0 Discussion

The nascent and interdisciplinary field of electronic art presents the subject indexer with considerable challenges. To date the conventional solution to the development of any scheme has been to draw on existing information science expertise in the construction of the scheme and the identification of facets; and to rely on a curator's informal observations of the electronic artwork(s) rather than on literary warrant for the development of terminology. The approach taken to the construction of a basic subject language for the *Riders Have Spoken* archive is one informed by a general information production and exchange framework (Benkler, 2006), and the relevance/accreditation function of this framework in particular. The virtue of adopting this framework from a subject indexing perspective has been a) to provide a systematic underpinning for combining the sub-functions of subjective relevance and objective credibility as part of the same logical framework, and b) to demonstrate the potential that a networked information environment affords for modularizing and distributing responsibility for the performance of these sub-functions. In common with previous approaches the strategy has been develop terms and a vocabulary from the ground up. The distinctive contribution here however has been to draw democratically on the contributions of a wider pool of users (see also Winget, 2009; Rafferty & Hilderley, 2005; Svenonius, 2000). This has the virtue of developing terms and a vocabulary that are relevant to a particular user audience and therefore have 'user warrant'. Like previous approaches the construction of the main organizing categories has drawn on more objective criteria however and the knowledge and expertise embedded in previous classification scheme; in particular Ranganathan's fundamental categories of PMEST, the main organizing facets of the AAT, and Anselm Strauss's concept of trajectory. The test of any subject language however will remain its usefulness for accessing the artistic work and its materials; and while the emphasis of the study has been on the innovative development of a subject schema that is both relevant to a particular work, and consistent in its construction. In this respect a necessary next step will be to conduct an evaluation of the performance of

Digital Archiving as Information Production:
Using Experts and Learners in the Design of Subject Access

that subject language, and whether it meets the requirements of its current and future users.

6.0 Conclusion and Implications

This study has focused on the construction of a relevant yet consistent subject language that aids in providing access to a digital archive of materials for the an electronic artwork *Rider Spoke*. The approach that has been taken to the development of this subject language is one that has been informed by an information production and exchange framework that considers digital archiving as a form of information production. The implications of this framework relate both to the methods used in the production of the digital archive, and to the character of the digital archive produced. With respect to methods, the framework enables us to perceive information organization both as a form of relevance/accreditation production, and as one component among others that contribute to the production and exchange of information. The use of the framework also draws attention to the character of a digital archive. Unlike a traditional archive, a digital archive can be considered to be—from initial utterance through relevance/accreditation to value-added distribution—less a final product than an open work.

References

- Bauer, M.W. (2000). "Classical content analysis: a review". In: Bauer, M.W. & Gaskell, G. eds., *Qualitative Researching with Text, Image, and Sound*, 131-151. London: Sage.
- Benford, S., Giannachi, G., Koleva, B., Rodden, T. (2009). "From interaction to trajectories: designing coherent journeys through user experiences", *CHI 2009*, April 4-9, 2009, Boston, Massachusetts, USA.
- Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. London: Yale University Press.
- Broughton, V. (2004). *Essential Classification*. London: Facet Publishing.
- La Barre, K. (2010). "Facet analysis", *Annual Review of Information Science and Technology*, 44, 243-284.
- Langridge, D.W. (1976). *Classification and Indexing in the Humanities*. London: Butterworth.
- Maron, M.E. (1977). "On indexing, retrieval and the meaning of about", *Journal of the American Society for Information Science*, 28, 38-43.
- Rafferty, P. & Hilderley, R. (2005). *Indexing Multimedia and Creative Works: The Problems of Meaning and Interpretation*. Aldershot: Ashgate Publishing.
- Ranganathan, S.R. (1963). *Colon Classification: Basic Classification*. 6th edition. Bombay: Asia Publishing House.
- Spiteri, L. (1998). "A simplified model for facet analysis", *Canadian Journal of Information and Library Science*, 23, 1-30.

Digital Archiving as Information Production:
Using Experts and Learners in the Design of Subject Access

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Strauss, A.L. (1993). *Continual Permutations of Action*. Chicago: Aldine Press.

Svenonius, E. (1994). "Access to nonbook materials: the limits of subject indexing for visual and aural languages", *Journal of the American Society for Information Science*, **45** (8), 600-606.

Svenonius, E. (2000). *The Intellectual Foundation of Information Organization*. London: MIT Press.

V2_ (2004). "Capturing Unstable Media", available at:
<http://capturing.projects.v2.nl/index.html>

Vickery, B.C. (1960). *Faceted Classification: A Guide to Construction and Use of Special Schemes*. London: Aslib.

Vickery, B. (2008). "Faceted classification for the Web", *Axiomathes*, 18, 145-160.

Winget, M. (2009). "Describing art: an alternative approach to subject access and interpretation". *Journal of Documentation*, **65** (6), 958-976.

Digital Archiving as Information Production:
Using Experts and Learners in the Design of Subject Access

Appendix 1

FACET	SUB-FACETS	EXAMPLES
A. AGENTS	<i>Specified by role</i>	
	Creator	Blast Theory
	Participant	George
	Bystander	Pedestrians, diners,
	Videographer	
	Viewer	Learner, researcher
B. TECHNOLOGY	<i>Specified by type</i>	
	Game technology	Bicycle, GPS, Nokia tablet, wifi
	Recording technology	Video camera
	Non-game technology	Mobile phone
	Production technology	Audio, video
C. ACTIVITIES	<i>Specified by kind</i>	
	Cycling	
	Listening	Riders' utterances
	Recording	Riders' utterances
D. SPACE	<i>Specified by composition</i>	
	Location	Linz
	Form	Bridge, street, tunnel
	Function	Crossroads, park, townsquare
E. OBJECTS	<i>Specified by type</i>	
	Built environment	Bank, church, clock, gate, tower, fountain, hotel, shop.
	Transport	Car, tram; street lamp, traffic light
	Signs	Street sign, graffiti
F. TIME	<i>Specified by date</i>	2009
G. ASSOCIATED CONCEPTS	Aesthetic concepts	
	<i>Specified by form</i>	Colour, music, landscape
	Psychological concepts (45)	

Digital Archiving as Information Production:
Using Experts and Learners in the Design of Subject Access

	<i>Specified by emotion (18)</i>	Fear, horror

For Peer Review