

Permission to make digital or hard copies of part or all of this work for personal, professional or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists for profit or commercial advantage requires prior specific permission from the author(s).

**Citation:**

Wu, Paul Horng-Jyh, "Towards Preservation, Description and Arrangement of Web Records: A Case Study of Archiving Professional Seminar eLearning Space," in *Proceedings of the InterPARES 3 International Symposium, 4-5 June 2009, Seoul, South Korea* (Seoul: Sungkyunkwan University, 2009), 393-411.

# Towards the Preservation of Web Records: a Case Study of the Capture, Arrangement and Description of a Professional Seminar eLearning Space

Paul Wu Horng-Jyh; Nanyang Technological University; Singapore

## Abstract

*Professional Seminar is a Master's level course with a blended learning environment that includes online and offline learning spaces. Modeled after constructivist pedagogy, it is organized around student-centered activities with their dialogic and reflective postings on a Web 2.0 platform consisting of blogs, forums and wikis. Upon each speech given by respected industrial veterans or captains in their respective fields of expertise, students organize and moderate break-out discussion on topics such as leadership, communication, ethics and critical thinking. This is followed by online interaction in the Web 2.0 platform that documents students' socializing, dialoging and sense-making efforts. These posting documents are essential as they serve as records required to account for a final reflective report by students to substantiate their learning results. The challenges for Professional Seminar's eLearning records management are how the postings on a Web 2.0 platform can be maintained, arranged and described so to achieve their reliability, authenticity (identity and integrity), accessibility and usability over time. A system prototype will be explained that integrates web "archiving" and web annotation functions. Complete snapshots of web pages of the Web 2.0 platform are captured as records and ingested into a recordkeeping system through a web crawling process. The records are then grouped according to a classification scheme, through a metadata annotation process, based on a sense-making theory that explains how students authenticate, rationalize, norm and position their thoughts for transformative learning. Additional metadata can then be created to aggregate these annotations as a description of the captured eLearning space. Many technological challenges and alternative solutions will be discussed along with the prototype system presentation. A brief state-of-the-art survey on similar efforts will also be discussed to highlight the distinctiveness of the proposed approach.*

## Introduction

This paper describes a preliminary case study on eLearning records management in a blended learning environment. In InterPARES 2, the ideal record-making and record-keeping processes in the government, arts and science sectors are studied. Similarly, eLearning presents yet another domain such study may be conducted. Furthermore, in the past few years, we have witnessed the wide adoption of internet digital technology, such as Web 2.0 and Virtual World, for educational purposes, similar to the cases in e-government, digital arts and e-science. In this paper we shall examine the concept of eLearning records based on a course called Professional Seminar, with the objective to formulate a framework for preserving eLearning records in general. The examination will proceed in two steps<sup>1</sup>: first, the context in which eLearning documents are created will be examined. This is because the concept of record in the eLearning space is inevitably shaped by its underlying pedagogical theories, which confluences with the perspectives of the record authors, preservers, users, and creators in defining the records in such space. Second, the documentary form of the documents in this course mediated by the Web 2.0 technologies will be analyzed for their record-ness and a method to keeping them will be demonstrated.

In the following, we should start by explaining the particular theory – Constructivist pedagogy, after which the object of our case study – Professional Seminar – is modeled. After an exposition of Professional Seminar's structure which provides the context of our study, we then propose a framework of ideal record-keeping and record-keeping processes. As Professional Seminar platform leverages on Web 2.0 technology for its social learning capability, we then discussed the specific challenges towards preserving such Web records produced in blogs, forum and wiki's. A brief critique of the current state-of-affairs in preserving Web records is given, where the predominant methods rely on crawling a website so to retain snapshots of the web documents. Against this critique, a framework is proposed based on the InterPARES 2 study where a generic model of stored and manifested documents can be used to analyze the mechanism for the records creating process of blogs, forum and wiki. A Web Annotation

---

<sup>1</sup> Particularly, we will be following the framework explained in the paper: Durani L. & Thibodeau K., (2006) The concept of record in interactive, experiential and dynamic environments: The view of InterPARES. *Archival Science*, Vol. 5, No. 2, pp. 13-68.

system under development will be demonstrated to show the ways these records may be classified and described based on metadata created from the annotation process.

## **Constructivist Pedagogy**

In most current contexts, academic records typically consists of exam scripts, Grade Point Average (GPA), final year report, thesis (if required), and documents alike<sup>2</sup>. However, pedagogical practice changes as new technologies advance. With the emergence of social media enabled by Web 2.0, “pedagogy through participation” rather than “pedagogy through instruction” has become the predominant paradigm of pedagogical theorizing under the “participatory revolution”<sup>3</sup>. Under such pedagogical thinking, students are “freed” from teachers’ deterministic instructional design while at the same time, “burdened” with the need to construct their own understanding based on their backgrounds through socializing with other students. Teachers’ responsibility shifted from a molder to a facilitator that nurture a collaborative learning space where “learning to be” rather than “learning about” becomes the main activities enacted in the learning process. These emerging pedagogical practices inevitably change the ways records are defined in a learning space. More than summative assessment such as GPA and thesis, the emphasis should be put on formative assessments and narrative accounts of the learning process – a learning portfolio, for instance; the latter should be regarded as the main evidence of students’ learning achievements.

The above peer-to-peer learning process describes the so-called Constructivist Pedagogy. Constructivist Pedagogy, by itself, is not a new concept; as advocated by Vygotsky<sup>4</sup>, it is very much focused on the relationship between the individuals and their social surrounding; the most compelling theoretical rationale for cooperative learning comes from Vygotsky, who claimed that human mental functions and accomplishments have their origins in social relationships. Mental functioning is the internalized and transformed version of the accomplishment of a group. The theory therefore assigns value to a group’s common perspectives and solutions to problems as they are arrived at, through debate, argument, negotiation, discussion, compromise and dialectic.

## **Evidence-Based Practice: Record-making and Record-keeping in Constructivist Pedagogy**

Drawing the insights obtained in InterPARES 2 from the arts sector, the methodology of teaching after Constructivist Pedagogy can be likened to that of an arts performance, particularly as performance of a piece of Jazz’s music. Jazz is known for its ingenuity and improvisation. In Constructivist Pedagogy, the instructor’s role can be likened to that of a scores author and the lead musician, who adhere to a pre-agreed guideline (the scores), ensuring sustained “performance,” including conversation, dialogue and consensus building among the students, and removing impasses or conflicts through dynamic and in-situ interventions only when required. Like learning Jazz, we recognize the better way to learn soft skills is in the sustained practice of them (learning to be), rather than the knowledge about them (learning about).

To account for the pedagogical functions in such educational paradigm, the questions for the stakeholders (including teachers and students) and record-keepers alike are: how are the learning process to be accounted for? Following the InterPARES 2 framework<sup>5</sup>, this question can be readily formulated as: where are the potential *prospective* records – like the score of a piece of music – and *retrospective* records – like the recordings of music performance? The purpose of these records is to afford reflection and assessment between iteration of action so that the “performance” can be fine-tuned. Indeed, this process is well-known in educational research as *Action Research*<sup>6</sup>. Thus, in the mode of Action Research, our working hypothesis became that the

---

<sup>2</sup> For instance, a sample records retention schedule for student files can be found at <http://www.lib.iastate.edu/spcl/arch/Schedule/Retention=Student%20Files.pdf>

<sup>3</sup> The social learning principle advocated by Web2.0 (1) adjusts the focus from the content of the subject to *the learning activities* and *human interactions* around which the content is situated, in line with the gap identified by Light cited above. It also (2) redefines the concept of ‘mastering’ a subject as “*learning to be a full participant*” in the subject area which involves “acquiring the practices and norms of established practitioners in the field or acculturating into a community of practice”- “*learning to be*” -rather than “*learning about*” the subject matter. Reference: Brown, J. S., & Adler, R. P. (2008). Minds on Fire: Open Education, the Long Tail, and Learning 2.0. *EDUCAUSE Review*. 43 (1), 16-20,22,24,26,28,30,32.

<sup>4</sup> Vygotsky believed that the “*relation between education and life, and between the school and the social order must serve as a starting point for pedagogics*” which in turn became one of the basic tenets of constructivism. See Section on Professional Seminar for reasons why this is the relevant pedagogy to adopt. Cf.

Vygotsky,L.(1926). *Educational Psychology*. St. Lucie Press. Florida, 1992.

<sup>5</sup> “Interactions between humans and computer systems, experiences enabled or mediated by experiential systems, and processes which are composed and carried out with at least some degree of spontaneity by dynamic systems are not the residues of action. They are not means of remembering either what was done or what is to be done. In short, they are not records. But, they can be *captured in documentary form* and some of these documents could be treated and used as records of interactions, experiences, or dynamic processes, that is, they may become *records of those activities*. In addition, interactions, experiences and processes are enabled by documents within such systems and these documents can serve as *prospective* records.” Durani L. & Thibodeau K., (2006) p 59.

<sup>6</sup> Action research is defined by Halsey (1972) as a “small scale intervention in the functioning of the real world . . . and the close examination of the effects of such interventions.” See HALSEY, A. H. (Ed.) (1972). *Educational Priority Volume 1: educational priority area problems and policies* (London, HMSO). See Also, Newman, Judith M. (2000, January). Action research: A brief overview. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(1). Available at: <http://qualitative-research.net/fqs/2007>

successful practice of these soft skills by the students in the Professional Seminar manifest itself in the formation of a vibrant collaborative learning community where students engage each other in serious sense-making processes. Certainly, these records need to be determined to be of a fixed form and of stable content, which will be dealt with in detail in the last section on capturing Web 2.0 records. We are now ready to examine a particular case of Constructivist Pedagogy – Professional Seminar – by identifying the prospective and retrospective records in the case, and how these records can account for Action Research using a particular iteration as an example.

## Professional Seminar – A Case of Constructivist Pedagogy

Since 2005, Division of Information Studies of the Wee Kim Wee School of Communication and Information at Nanyang Technological University has been offering Professional Seminar as a core course for its incoming batch of Master of Science students, with majors in Information Studies, Information Systems, and Knowledge Management. In 2008 we saw the number of students in the program reach 230. The students came from very diverse backgrounds in several aspects including age, culture, language and specialty. They also aimed at achieving different goals through the program, ranging from interest of new career in information, career advancement in existing work, exposing themselves in international learning environment, to self-enrichment and life-long learning. To take advantage of this complexity, rather than to be handicapped by it, one of the essential tasks to be achieved, then, is to orient and guide the students in forming a coherent and vibrant learning community. Such a community will then be conducive for students to forge collaboration and share knowledge by cross-fertilizing their diversity that naturally complements each others'. However, to engage this complex mix of students in one course is not as straightforward. In the past, foundational courses, such as Information and Society, had been offered with limited success, as content-specific courses will tend to suit one group but disfranchise another. Thus, we adapted by setting the goal of this core course to be cultivating students' soft skills, such as leadership, communication, critical/creative thinking, and entrepreneurship, which shall be relevant across the complex mix of motives and backgrounds.

### Overall Course Structure

The seminars were spaced out so that there was one scheduled each month from 9.00 am to 5.00 pm on a selected Saturday. As shown in the Figure 1.a, each seminar session was identical, one each in the morning and afternoon session, each of which was followed by panel discussion. With breaks in between, the seminar was followed by student moderated break-out discussion sessions where the students in their groups moved to separate rooms to discuss the key take-aways of the seminar. During any one session, five sub groups were assigned moderation duty. The groups on moderation duty were required to facilitate and moderate the discussions of the other groups as well as consolidate the points that were brought up during the discussions in a 300 word summary report.

Seminar 2 : <span style="float: right;">Date: September 27th 2008 (Saturday)</span>		
Venue: Lecture Theatre 1, North Spine, Nanyang Technology University		
Time	Agenda	Speaker/Moderator
0900 – 1000	Topic: Leadership	Mr. Heng Chiang Gnee
1000 – 1030	Panel Discussion	Mr. Heng and Dr. Wu
1030 – 1100	Tea Break	
1100 – 1230	Morning Break-out Session	Moderating Subgroups & Groups
1230 – 1330	Lunch Break	
1330 – 1430	Topic: Ethics	Prof Cheong Hee Kiat
1430 – 1500	Panel Discussion	Prof Cheong and Dr. Wu
1500 – 1530	Tea Break	
1530 – 1700	Afternoon Break-Out Session	Moderating Subgroups & Groups

Figure 1.a: Sample One Day Seminar, including Speeches and Break-out Sessions

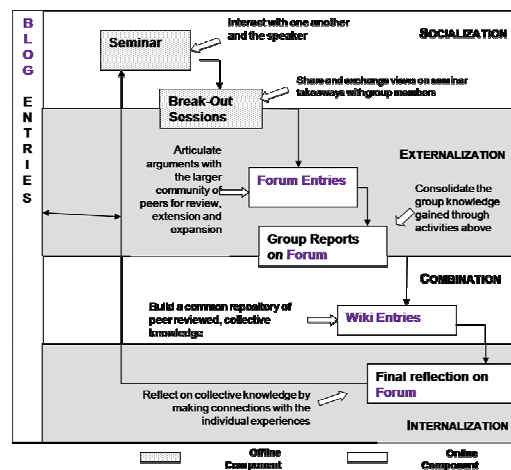


Figure 1.b: Online and Offline Spaces & SECI Knowledge Sharing Spiral on Web 2.0

Mandatory participation in all three seminars was required of all the students enrolled in the unit. Absentees were required to submit official reasons for being absent as well as prepare summary reports on the affected seminar presentations. The interaction carries on beyond the seminar conducted in the physical (offline) space. As shown in Figure 1.b, in between the seminars, the students were required to extend these discussions using *blogs, forums* and eventually consolidate their view points

under the *wikis* on the e-learning platform designed to support these activities on the edveNTUre. The participation on edveNTUre was strongly encouraged but not mandated.

**Prospective records – design parameters and scaffolding**

Beyond the course structure, design parameters and scaffolding<sup>7</sup> can be seen as prospective records that constitute a framework for forming successful learning community. Among them, one important design parameter of professional seminar is that there are minimal requirements for students to pass the course: they are only required to attend all three seminars and submit a 500 word long final report. The intended objective of this is to free them of any burdens so that they concentrate solely on interacting with speakers and with one another. In invoking the Golden Rule, the lecturer explains to the students the extent of the efforts and resources required in organizing the professional seminar in an open letter to the students as a way to initiate them for the course (see Fig. 2.a). The letter also markets the quality of the invited speakers, the ambience combining fun and bite, and the facilitation and logistic work done by the TAs and division officers. Similarly, rewards and gifts are put in place, on a case-by-case manner, to encourage participants who demonstrate initiative in participation as the seminar progresses (see Fig. 2.b, for certificates of Best Team in Collaborative Team Challenge). There are other scaffolding steps taken depending on the circumstances such as advice in organizing the group in the blog and initiating discussion in the forum.

Dear Students,

Welcome to a new chapter in your life where you decide to take on the task of graduate study. Presented with many challenging courses and interesting topics to choose from, I know you are excited about the opportunities to understand, analyze, write reports, and perform well in exams and I congratulate you for having such enthusiasm, since an old Chinese saying teaches that "a good start is the beginning of success".

Hopefully, success will begin with this mandatory course for all students in the three MSc programs called "Professional Seminar." This course will be held over three weekends and will allow you to grab some free food while you network and learn. Which gurus can magically transform us into professionals in a couple of sittings you ask? Well, much thought and hard work has gone into the planning of this seminar. Special speakers were selected both from academia and industry to share their experience and knowledge with us. Food and venue were also prepared in advance to ensure that you have a great time while you learn. Two teaching assistants have also been assigned to help you along the way. All these were done with the hope that you would have a fun time enjoying the intellectual discourse while you socialize with each other over food and drinks. To pass this course, what you need to do is just attend all three sessions and turn in a 500 word report reflecting your experience. Simple isn't it?

Yet, you are expected to put in your best effort and share your thoughts and reflections openly with your classmates on your experience of the seminars, the topics and the speakers. Most importantly, you are expected to take charge and facilitate the discussion sessions, so that the sharing among yourselves can be facilitated. In other words, you are encouraged to learn to serve one another during the course of your studies.

Being at the forefront of technology, we have also taken advantage of the opportunity to allow you to extend your learning online. We are fortunate to have one of the most advanced eLearning platforms here in NTU called edveNTUre (edventure.ntu.edu.sg), equipped with Web 2.0 technologies like discussion forums, blogs and wiki. You will have the chance to experience this social media revolution by plugging yourself into cyberspace and continue your learning there. Expect an exciting ride with our "participatory community of eLearning."

Attached is a write up on the course objectives and further details to help you understand more about the course. Please read it carefully, after which you should start contacting your course mates to brainstorm, organise yourselves and get ready to run with the seminar! If you have any queries please contact the following teaching assistants: Mr. \*TA 1\* - \*\*\*\*@ntu.edu.sg and Mr. \*TA 2\* - \*\*\*\*\*@ntu.edu.sg

Best of luck!

Sincerely,

Course Co-ordinator of Professional Seminar  
Email: \*\*\*\*\*@ntu.edu.sg

**Figure 2.a: Instructor's Initiation Letter**

 **WKK School of Communication & Information  
Division of Information Studies**



*Awards*

Name of Student

*for*

**Best Team In  
Collaborative Action Challenge**

**Dr Paul Wu Horng-Jyh**  
Course Co-ordinator, Professional Seminar

**Assoc Prof Abdus Sattar Chaudhry**  
Division Head (Information Studies)



**WKK School of Communication & Information  
Division of Information Studies**



*Awards*

Name of Student

*for*

**Best Team  
Collaborative Action Challenge**

**Dr Paul Wu Horng-Jyh**  
Course Co-ordinator, Professional Seminar

**Assoc Prof Abdus Sattar Chaudhry**  
Division Head (Information Studies)

**Figure 2.b: Certificate of Best Team Recognition**

<sup>7</sup> Vygotsky's notion of ZPD leads to the concept of scaffolding which Bruner views as assistance in the zone of proximal development. The term itself is a metaphor for a support system that compresses lengthy explanation of the term into a meaningful image that gives a glimpse of what lies beneath it. The term was initially used by Wood, Bruner, and Ross (1976) in an educational sense when they described scaffolding as a "process that enables a child or novice to solve a problem, carry out a task or achieve a goal which would be beyond his unassisted efforts". A research on Scaffolding for Professional Seminar will be published in Ravindran, T., Wu, H-J, Chia, Y. B., Nourbakhsh, A. "Scaffolding for Knowledge Creation in a Constructivist learning Environment – The case of Professional Seminar."

By sensitizing students' sense of reciprocity, which is to counter what the course offers as well as the efforts put in by their fellow students, most of them will be initiated to demonstrate their share of the participation. For students stronger in certain soft skills, the course provides an opportunity for them to exercise skills that benefits their fellow students. For those weaker in certain soft skills, they may learn from the stronger ones or they may experience the need to invest effort in learning such skills. Most critically, students may deploy their stronger skills to complement one another's; for example, we have witnessed how students who are good at critical thinking are bridged by their fellow students in communicating their thoughts to the rest (see Fig. 4.d for a sample reflective report on knowledge sharing).

Nonetheless, risks exist when students are not sufficiently skilled to start the above process of reciprocity and team work. In such cases, most of them will experience a certain dilemma: either they failed to moderate a discussion, or they are unable to be engaged in some circumstances. Nonetheless, they are encouraged to reflect on such experiences in their reports and blog postings, and engage each other in discussing and problem-solving the dilemma in forums. In such cases, lecturers may take a more active role of intervention, which, however, does not entail imparting them with explicit advice, but making recommendations and helping the students in seeking a *minimal common ground* so that they can function together as a group. The objective is to help students find the most appropriate ways out for themselves, and by bootstrapping, they grow more complex skills through recursively leveraging upon more basic ones. Ultimately, in reaching praxis in the students' journey of learning, they need to go beyond the practice of soft skills, and be entrusted with the responsibilities of constructing their own practices based on self-reflection. Now, in the mode of Action Research, how do we become informed so to allow us to enact change that helps the students? It's time to examine the issue of the retrospective records as how well the students have been responding to the constructivist pedagogy as configured in Professional Seminar.

### Retrospective Records and its Classification - SECI spiral & knowledge sharing functions

As shown in Fig 1.b, students' participation is extended in cyberspace in a cascade of Web 2.0 technologies: blogs, forums and wikis, which allows students to find an appropriate space for Socialization, Externalization, Combination and Internalization – the so called SECI knowledge sharing functions<sup>8</sup>. As shown in Fig 3.a, the functions of SECI is to guide students through a process of authenticating self-understanding, rationalizing and articulating thoughts, and norming and connecting on consensus, and then positioning and embodying their actionable knowledge through critical reflection, achieving a spiral of truly internalized knowledge. Along these processes, blog, forum and wiki online spaces are being populated with entries and reports which can then be classified according to the above SECI process (see Fig. 3.b).

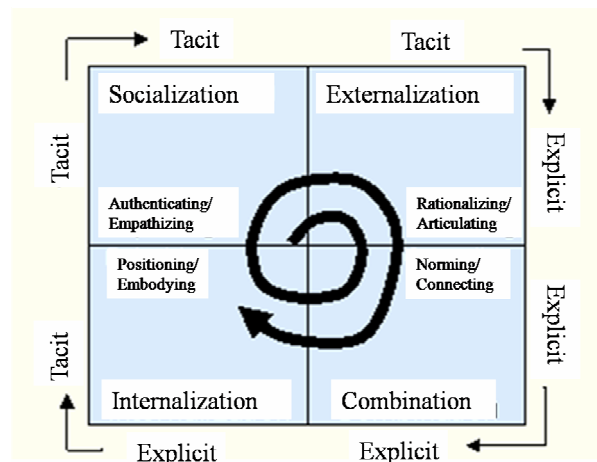


Figure 3.a: SECI spiral and knowledge sharing functions

#### Blog entries – Socialization

- Interact with one another and the speaker
- Share and exchange views on seminar takeaways with group members

#### Forum entries – Externalization

- Articulate arguments with the larger community of peers for review, extension and expansion
- Consolidate the group knowledge gained through activities above

#### Forum & wiki entries – Combination

- Build a common repository of peer reviewed, collective knowledge

#### Final report & forum – Internalization

- Reflect on collective knowledge by making connections with the individual experiences

Figure 3.b: Web 2.0 record creating spaces

In the following, we shall discuss sample retrospective records according to their classification assigned by SECI functions performed by students in their learning process. By identifying these records, they form the basis for the formative assessment of the students' learning process, an emphasis over the conventional summative assessments. Observation is also made as which

<sup>8</sup> Nonaka, I., Toyama, R., and Konno, N. (2000). Seci, ba and leadership: a unified model of dynamic knowledge creation. *Long Range Planning*, 33(1):5-34.

space these records may be identified. Before the discussion, the following Fig. 4a to 4.d shows the postings captured from blogs, forums and wikis.

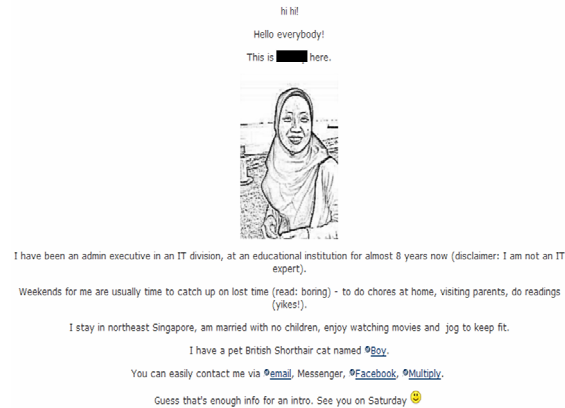


Figure 4.a: A sample blog post

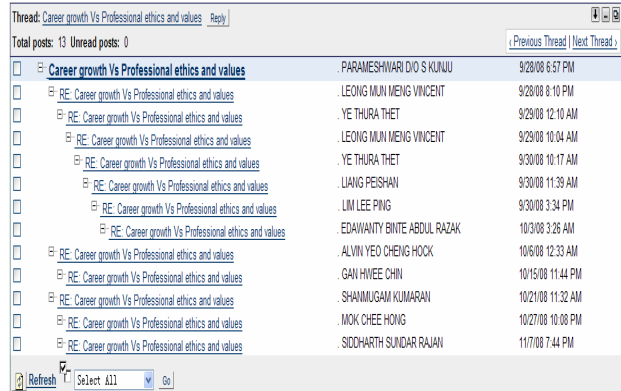


Figure 4.b: A sample forum thread

**Group 4.1**

- 1) Reflection on the Issues
  - Loyalty to co., friend?
  - Conflict of interest
  - Unfair advantage
- 2) What are the values?
  - Serve e.co. to the best of my ability
  - Impartiality (fairness)
  - Friendship
- 3) Impact of the action
  - May not really be the best candidate deprive my suitable candidate
  - Friendship affected
- 4) Is it ethical?
  - No

**Group 4.3**

- 1) Reflection on the Issues
  - Whether people who are alerted to the loop hole report to the management
  - Should Kathi report the problem to her supervisor?
  - Whether Andre use the loop holes to be benefit himself
- 2) What are the values? Andre
  - Integrity
  - Trust
  - Richard
  - Trust
  - Loyalty
- 3) Impact of the action Not Report
  - Financial loss to the company
  - Well being of Andre's mother
  - Conflict of interest
- Report
  - Disciplinary actions against Andre
  - Review the employee benefit program
  - Richard is betraying the trust of the Andre
  - Andre losses confident in Richard
- 4) Is it ethical?
  - Yes – From company's point of view
  - No – From employees' relationship

**Group 4.4**

- 1) Reflection on the Issues
    - Using office copier for personal use
  - 2) What are the values?
    - Integrity
    - Honesty
    - Fairness
    - Responsibility
    - Loyalty
  - 3) Impact of the action
    - Using up limited resources
    - Indiscriminate misuse
    - Setting bad examples
    - Reduce productivity
    - Create unhappiness
    - Betraying company's trust
  - 4) Is it ethical?
    - No
    - Possible solution
- Pay for usage of copier, after seeking permission

Figure 4.c: Sample wiki entry with consolidated summary reports

It's true that knowledge sharing is power. No matter the cultural, lingual or ethical difference between groups of people knowledge sharing still fosters growth and so things are done faster and effectively than when knowledge is hoard. I guess that's why we are prospective knowledge Managers which is what we preach and are expected to practice in future.

After the professional seminar that day, I and my group decided we were going to find answers to the quiz given. We shared the work between us in the group in such a way that division of labor will be ensued. \*Studente A\* decided to collate the story while \*Studente B\* and others decided to find answers to the puzzles via the internet and books so to say.

Before researches became an option, we thought of using a very funny technique as we decided to act the drama posed in the questions given. We decided to tie \*Studente C\*'s, \*Studente D\* and \*Studente E\*'s face and from there we coined out something related to the answer. Likewise we got coins and started placing our coins but this time it wasn't really a round table that we got (lets called a spherical one where we sat for the brief minutes before we all left for home). Although we were unable to fill up the table when someone said she got the logic to the question and decided to conduct a detailed finding about it:

After the drama, we all left for our houses worn out and tired but with deep sense of anticipation that we can get the answers to the questions with our eyes gazed on the certificates promised. \*Studente F\* came up with the answer to one of the question but unfortunately, it came in Mandarin and it was sent to my email; I requested for the translation and thereafter \*Studente B\* translated it.

As for the Second question about the red caps, it likewise did come in Mandarin and \*Studente B\* did me lots of favour to translate it into English and after meeting on the Monday thereafter we were all able to get what it meant. It was such an interesting.

This is just that story of how things worked out; The answer is not here..... lol

Figure 4.d: Sample final report which reflects on the importance of sharing and complementing

**Socialization**

This process focuses on mutual transmission of tacit knowledge. Tacit knowledge goes across the private space through self-disclosure (see Fig. 4.a for an example) thus enables personal knowledge to be exchanged during the process of interactions, observing, discussing, analyzing, spending time together or living in the same environment. Socialization also functions to create new knowledge through shared experiences. The manner of production is more like in traditional environments where, for example, a novice learn the technique of wood craft from his master by working with him, rather than from reading from books or manuals.

Socialization in Professional Seminar may occur any time and any place: participating in the lectures, interacting with the speaker, introduce yourself and know each other in the group, chatting with others during the tea break, and so on. All these activities foster a kind of social network. Also, Professional Seminar enables students to keep in contact with one another through an online space, they gain new knowledge outside of their confine of physical environment, for example, by participating in online discussion, and browsing others' blogs, or viewing the wiki entries produced by other students

## ***Externalization***

This process focuses on the transmission of tacit to explicit knowledge. It helps to create new knowledge as tacit knowledge comes out of its boundary and became explicit knowledge. Externalization is often driven by metaphor, analogy and models, and during the process knowledge gets crystallized. It's like the way that an engineer taught his apprentices how to operate the machine. In the process, he articulated his knowledge and experience in a methodical manner so to be understood more readily.

In Professional Seminar, externalization mainly takes place in the offline break-out discussion first, and then, carries onto the online discussing forum. As students speak or write down their opinions, they present what they think, something that may have already embedded in their minds or been inspired by the speeches and by other students' statements. But if there is no platform like forum for them to externalize the views, it's possible that all these thoughts are merely "thoughts," and in no way, they would be able to be distilled into conceptual knowledge through a dialogical process. When someone wants to convince others, he/she should firstly be able to put the views in a reasonable frame, so images, symbols and language are used to make a clear statement. In the process of argumentation in the forum, students usually obtain a deeper level of understanding of their and others' opinions.

## ***Combination***

Combination is a process in which knowledge transforms from explicit to more systematic knowledge. Combine the existing explicit knowledge of different sources by collecting, understanding and modification can be a way of generating new knowledge. It's a process of packing knowledge and making them easily accessible by members. An example is that of a finance department that collects all financial reports from each department and produces a consolidated annual financial performance report. Creative use of database to make business report, sorting, adding, categorizing are also examples of combination process.

The process of combination is represented by the wiki and forum in Professional Seminar. Groups in charge of the moderation will collect other group's point of view, and by discussing and editing in group blog, the group members then sum up the opinions and work out the final report, and post it on wiki. During the combination work, students can examine the rationalities of knowledge sources and arrive at a common ground which may satisfy all parties' interests.

## ***Internalization***

By internalization explicit knowledge is transformed into tacit knowledge and is shared across the organization. It's also regarded as "learning by doing". Like in the case in externalization, remembering the academic steps to run a machine doesn't ensure the apprentices have mastered the skills. They have to practice by themselves, and thus internalization occurs.

Internalization in Professional Seminar is designed by reflective reports, in which they document what PS is meant for them and what they have learnt from this course. Through the experience of all sections in Professional Seminar, students absorb the explicit knowledge and generate their own views, which may be represented in actions and future practice. At the end, the students' personal views may become consistent with the mainstream as the norm being agreed upon; on the other hand, they may diverge from it despite a certain degree of change. Hopefully this dissonance will become the starting point of a new round of engagement and SECI spiral, and the learning process continues.

In sum, PS provides students a way to learn new knowledge and share what they know through interaction, thus helps them to internalize the valuable knowledge. As such, the dual goals of soft skills practice and community formation are optimally integrated and to be achieved simultaneously.

## ***Preliminary assessment of retrospective records for action research***

Several studies are underway to triangulate with the retrospective records produced by the students in assessing the effectiveness of Professional Seminar. One quantitative study uses questionnaire to survey students' perception of the senses of community and efficacy in the learning process, and its effect on that of the quality of knowledge shared. The initial finding of this study shows that the factors of Trust and Shared Vision have influenced students' assessment of the quality of knowledge shared. A second study uses qualitative interviews to investigate how students experienced the seminar in response to Constructivist pedagogy. Most students interviewed have related that they have gained much practical knowledge in communicating with fellow students. Meanwhile, based on the initial assessment retrospective records preserved, a complete SECI cycle is rarely achieved by students. Nonetheless students' reflection shows most have actively engaged with one another and been impacted to re-examine their life and resolve to commit to put into practice lessons learnt from the seminar. At the point of this paper, however, we are yet able to present a systemic analysis, which shall be covered in a future paper.



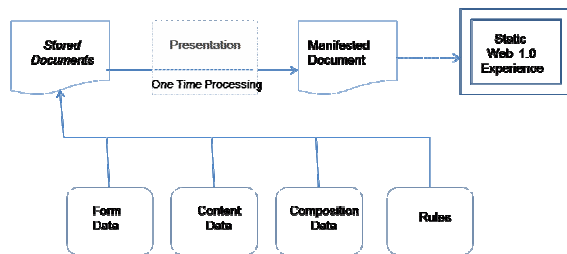
## Capturing Web 2.0 Records

We now turn to the discussion on the record-keeping process and the technological context in Professional Seminar. We do this by first examining the change from Web 1.0 to Web 2.0: most Web 1.0 websites contain static web pages. By taking snapshots of these web pages it will be sufficient to fix the form and content of these documents that are potential records. Even if the web pages are not static but are generated by retrieving content from querying database, they can become fixed in the form and content by themselves being taken snapshots of. Snapshot taking can be performed by typical web crawling methods. Most current web archiving efforts adopt this method for web capturing. As it assumes static HTML web pages, it does not, thus, distinguish between static documents, on the one hand, and interactive, experiential or dynamic web documents, on the other. This assumption proved to be too simplistic and several challenges have been faced when preserving Web 2.0 documents.

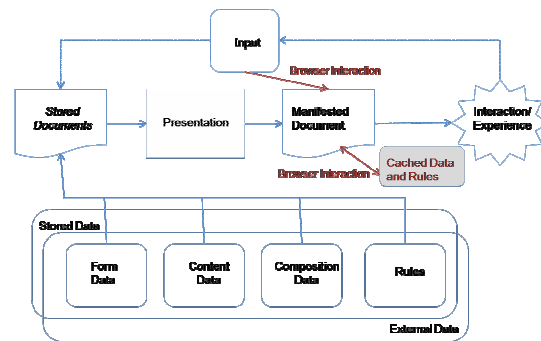
### ***Record-ness of blog, forum and wiki documents in professional seminar***

For Professional Seminar, blog, forum and wiki are the critical social media as it activates the interactivity via the documents. A closer inspection – as explained below revealed that they are basically stored dynamic documents, as they have a fixed form – certain variety of posting-comment structures and version tracking, and they stored input from users in generating the next manifested documents. The inputs are actually recorded experiences or responses resulting from the users being triggered in the engagement with other users. As such, Web 2.0 has also become an experiential system, as the users are responding and interacting with each other by expressing their thoughts and experiences on other people’s thoughts. These documents are indeed records of Professional Seminar. This is because (1) they have fixed form and content (refer to the next section on capturing dynamic records in blog, forum and wiki)<sup>9</sup>; (2) they have explicit linkages to other records – for instance, comments are to a post in blog and replies are to a thread in a forum discussion, within or outside of the digital system, through a classification code associated with SECI sense-making process: Socialization, Externalization, Combination, and Internalization. (3) The documents are kept in a eLearning platform with clear administrative context (4) Each posting in blog, forum, wiki has a unambiguous writer and addressee, while the author is the instructor who prescribed the SECI learning spaces (5) The action associated with these records are retrospective records of a collective “performance” that are designed according to a Constructivist Pedagogy and “conducted” by the instructor of the course. After establishing the record-ness of blog, forum and wiki entries, it is time to examine how exactly these dynamic records can be captured to have stable content.

### ***Capturing Dynamic Records in Blog, Forum and Wiki***



**Figure 5.a: Generic Web 1.0 Model of Stored and Manifested Documents Adapted from InterPARES 2**



**Figure 5.b: Generic Web 2.0 Model of Stored and Manifested Documents Adapted from InterPARES 2**

<sup>9</sup> Here we cite an excerpt as support for blog, forum, and wiki postings as records from Durani L. & Thibodeau K., (2006) p 62. “A stored dynamic document, such as one which stores user inputs and uses them in subsequent manifestations or one which processes and presents, but does not store, data from users or from other external sources, might be said to be always in the process of creating, but never completing a manifested record. The manifested document might be a record if the processes which cause it to be forever in progress were terminated or if it were removed from the dynamic environment and kept in some frozen form. However, the final state of a document somehow isolated from or rendered immune to dynamic processes would be static. But producing a static document in either way would amount to creating a different document. It might serve as a record of the dynamic process or its state at the moment it was frozen, but it would not be able to serve the dynamic purpose of producing variable output in response to a variety of different inputs or stimuli. In some cases, the stored dynamic document might be kept as a record, but that would not be possible in the case of documents covered by class 2.2.4 of Table 1.”

Compare Fig. 5.a with Fig. 5.b, we noted two main differences. First, there is no input feedback to the application server in Web 1.0, where there is in Web 2.0. Hence, there is no interactivity in Web 1.0 documents, although there exists one-time system processing to “integrate” various types of data and present the manifested document – that is, the web page as seen in the browser application. Second, there is no distinction between stored and external data in Web 1.0, while there is in Web 2.0. This is because Web 2.0 documents are dynamic and accommodate multiple sources of data that are updated in real-time, while Web 1.0 has a single source of content data. Furthermore, with the adoption of JavaScript and AJAX (that is, Asynchronous JavaScript + XML) programming paradigm, the server resources, such as content data and rules, may be transmitted and cached in the browser application to have interactivity either directly through browser application, or remotely through server application, as shown in Fig. 5.b.

With the models technically distinguished, we are able to examine the current Web preservation technological framework. As mentioned, most of the current methods of Web preservation are based on snapshot taking. This is equivalent to say that Web preservation system can be seen as a transformation of Web 2.0 modeled system into a Web 1.0 modeled system<sup>10</sup>, as the prevailing aims are at *encapsulating* the records, as well as its manifested results, instead of *emulating* the software applications<sup>11</sup>. However, this state-of-the-art approach to web documents capturing has three potential weaknesses. First, during the crawling processes, the Web 2.0 modeled system is captured when the URL’s of a website is being recursively enumerated. The captured resources – content, form, composition data and rules – of the manifested web pages are stored away as static HTML files. When these static HTML files are accessed through the Web preservation servers, the interactivity is lost as the input is no longer able to direct messages to the “correct” application server. This will result in broken links or the browser receiving data from live, instead of the preservation – the correct source, Web servers. Second, the crawler (or Robot) has a pre-determined mechanistic ways of enumerating the URL’s of a website. Many of the inputs that can be enacted by AJAX in Web 2.0 environment are outside of such scope, as the inputs were designed for personalized choices and random in nature; thus, the resulting manifested documents are not enumerated and captured, which should have been captured as an intrinsic part of the user experience interacting with the Web system. Third, as reflected by the “External Data” in Fig. 5.b, content may be added, deleted, or updated by parties simultaneously when the current session of interaction the crawler is simulating – for example, a comment is being added to a blog. However, these changes are not captured by the crawler as the update is not “pushed” to the browser simultaneously; as a result, such changes will only be captured in the next cycle of crawling process, while they could have been part of the user experience if so desired; here lies a subtle distinction between the 2D and 3D Web environments, which we now turn our discussion to.

From the overall perspectives of the Web, the above describes Web 2.0 as a *Synchronous Experiential* environment supported by interactive and dynamic documents, which is quite the same to a 3D Web environment supported by Virtual World or Computer Game, where “continuous” experience of interaction is essential. However, a closer examination of most Web 2.0 systems reveals that they are more *Asynchronous* than Synchronous *Experiential* environment where the real-time continuous experiences are not that critical – except perhaps for Twitter or Facebook type of social networking sites where degree of real-time-ness is one of important attractions. On the other hand, blog, forum and wiki systems do not require real-time, continuous interaction to be an essential part of the experience to be captured; rather, it is more about the logical flow of the information exchange where only the chronological order is important to be kept intact. If real-time interactivity is not critical, it’s possible to adopt an alternative mechanism to preserve Web pages by the creators to “push” their content and structure through syndication format such as ATOM. This provides an alternative to the prevailing crawling method, which shall be discussed in a later paper.

### **Developing Web annotation system to describe and arrange captured records**

In this section, we briefly describe the Web Annotation for Web Intelligence (WAWI) system<sup>12</sup>; two technical design principles are recognized to achieve the objectives of preserving the context in cataloguing and arranging web archives:

- Relate Web content to the semantic content in the metadata
- Relate metadata to ontology in relational metadata.

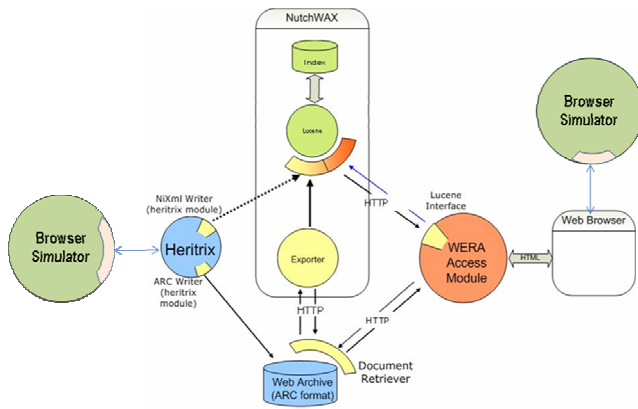
The WAWI annotation system is integrated with the web archiving platform developed by International Internet Preservation Consortium (IIPC)<sup>13</sup>, which comprises web harvesting and access components<sup>14</sup>: Heritrix, Nutchwax, and Wera.

---

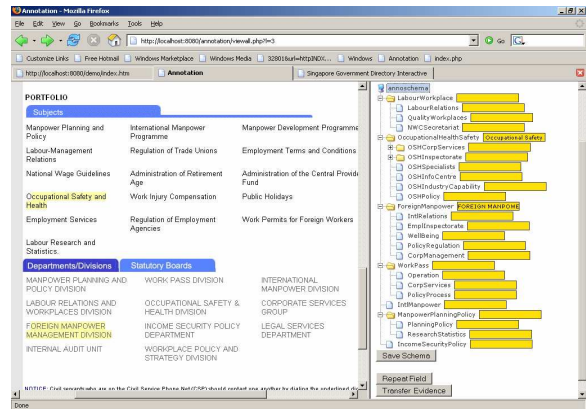
<sup>10</sup> For example, the International Internet Preservation Consortium (IIPC) employs Heritrix developed by Internet Archive.

<sup>11</sup> In general, there are four digital preservation strategies: encapsulation, migration, emulation, and universal virtual machine. Among many similar efforts, see National Library of Australia’s (NLA’s) Preserving Access to Digital Information (PADI) at <http://www.nla.gov.au/padi/topics/18.html>

<sup>12</sup> The content in this section is largely adapted from a published article: Wu, J., P. H., Heok, H., A. K., Tamsir, and P., I. (2007). Annotating web archives - structure, provenance, and context through archival cataloguing. *New Review in Hypermedia and Multimedia*, 13(1):55-75.



**Figure 6.a: WAWI annotation and cataloguing system integrated with IIPC Web Archive platform**



**Figure 6.b: Annotation Schema, an ontology reflecting an organization chart in Singapore Government Directory interactive Website**

As shown in Fig. 6.a, the IIPC platform of Web crawling and access components is integrated with an annotation component. The annotation component is to retrieve captured web pages and allow users to annotate directly on the captured web pages. As to overcome limitations explained above to account for AJAX programming, server-side and client-side browser simulators need to be incorporated to enact the interactivity during both the capturing as well as the accessing time.

The annotation process is essentially a process of linking web content to a metadata scheme that was pre-determined by the preservers. In Fig. 6.b, it is demonstrated using a preserved Singapore government website as an example, where the right-hand panel shows the semantic model of the organizational chart, which is linked with the content on the left-hand-side, which is the captured web page published to the public about the government organization. We are not able to demonstrate using Professional Seminar captured web records as we are still overcoming the difficulty faced in crawling the blogs, forums and wikis. Ultimately, for Professional Seminar, the SECI knowledge sharing schema will be used as schema to annotate the captured web pages; that is, blog, forum, and wiki posting will be classified as Socialization, Externalization, Combination, and Internalization. These records will be arranged as a SECI spiral according to each Internalization records, mostly expressed in the final reflection report. These aggregated records, pertaining to SECI spirals achieved by individuals or groups, can be seen as “mini-series” that represent the process-bound information on the learning experience – retrospective records – of the students’ performance according to Constructivist Pedagogy. Encoded Archival Description (EAD) can be applied to further organize these mini-series, in which case, an ontology corresponding to EAD shall be configured to “link” metadata of mini-series. These metadata is stored in a XML-database and thus, can be searched through specifying various search parameters. For readers who are interested in the further detail of WAWI, please refer to the reference in Footnote 11, entitled “Annotating web archives - structure, provenance, and context through archival cataloguing.”

In an e-learning environment, it’s possible to capture the records in parallel to its creation. The timely capture of records certainly helps establish the reliability of these records at the first place. By capturing all the necessary digital components, and fix their forms, contents and structures, the integrity of the records will also be established. By describing these captured documents using annotation metadata and classification code upon crawling, the authenticity and reliability of the records can be further enhanced. In WAWI system, the metadata, including the ontology, are stored as instances of XML-documents and are indexed accordingly. The accessibility to these captured documents is ensured. In sum, we believe the Web annotation system, together with improved Web capturing systems, will be an essential tool to ensure the high quality of Web records preservation.

<sup>13</sup> URL: <http://www.netpreserve.org/about/index.php>

<sup>14</sup> Heritrix URL: <http://crawler.archive.org/>; NutchWax URL: <http://archive-access.sourceforge.net/projects/nutch/>; Wera URL: <http://archive-access.sourceforge.net/projects/wera/>

## Concluding Remarks

In this paper, we considered the issues involved in identifying the records in Constructivist Pedagogy. Using Professional Seminar as a case, we explained how it can be seen as an experiential learning environment supported by Web 2.0 technologies; by likening Professional Seminar to a Jazz performance, we identified the potential prospective and retrospective records of Professional Seminar. We argued how SECI knowledge sharing process can be used to classify documents captured in blog, forum and wiki, as the essence of their performances. Research, with other triangulation measures, is underway to examine these records for assessing the students' learning achievement in the current iteration of action research. Initial finding suggests that although there are few complete SECI knowledge sharing cycles achieved by students, they have gained much practical knowledge in communicating with one another and in forming an active learning community. In the second half of the paper, we turned our attention to the development of Web preservation efforts, and highlighted several limits of the current approach to capturing dynamic records. Based on the framework established in InterPARES 2, we are able to isolate the problems and suggest solutions to overcome such limits. Lastly, we briefly showed a Web annotation system – WAWI – which is under development to classify and describe captured Professional Seminar e-Learning records. We argued Web annotation system like WAWI is an essential tool in a Web preservation system as they are able to maintain the quality of preserved records. Our immediate task is to complete the development of WAWI and conduct a full pilot study on a Professional Seminar course. Besides resolving the technical issue, one important success factor is the collaboration needed from the technical support unit – Center of Educational Development (CED), which like typical technical organizations, may not readily appreciate the complexity of digital preservation and need to help them overcome the initial hesitation before involving them.

Two potential future works may extend the current project. First, it may be timely to consider how records and their classification can be defined for e-Learning systems in general, even though they may not endorse Constructivist Pedagogy. We do suspect Constructivist Pedagogy marked the boundary for such consideration as it's the most further away from the traditional norms, which already have well-established record-keeping practices. Second, although the prevailing Web records capturing tool is based on the crawling method. However, it is noted for Web 2.0 environment, there may exist an alternative to the traditional method. Given that Web 2.0 is essentially asynchronously experiential, the content syndication framework such as ATOM may well be appropriate for preserving Web 2.0 systems as well. However, besides blog syndication, little has been attempted for forum and wiki along the line of standardized content packaging. Nonetheless, most industrial players do recognize "migration" (from one server to another server) is critical to Web 2.0 systems, and promoted ATOM and its equivalents for such purposes. In a similar way, we shall argue: for digital preservation ATOM can be promoted as a standard content format not only

## Speaker's Brief Biography:



*Dr. Paul Wu Horng-Jyh is a Senior Fellow with the Wee Kim Wee School of Communication and Information at Nanyang Technological University (NTU). Presently, he researches and directs the academic areas of Digital Preservation and Archival Informatics, particularly in the e-Learning and e-Social Science domains. Before joining NTU in 2004, he had been involved in technology start-ups for almost a decade and was a co-founder and CEO of Mustard Technology, which specializes in multilingual search and data quality technologies for businesses and government agencies in the region and beyond. After obtaining PhD in Artificial Intelligence from University of Michigan, he served as a Senior R&D Member of Kent Ridge Digital Labs in the 90's, during which he invented two US Patents on search and language processing technologies. He has two decades of experiences in theoretical and practical consultancy in information science and technology. The most recent project he consulted was the Web Archives of Singapore (WAS) project – a collaborative effort between National Library Board (NLB) and NTU, which created the first Web Archives in this region. Currently, he is conducting a general study on Website Preservation and a case study on preserving e-Learning records, under the InterPARES 3 framework.*