



## Datasets for Technology Enhanced Learning

Hendrik Drachsler, Katrien Verbert, Miguel-Angel Sicilia, Martin Wolpers,  
Nikos Manouselis, Riina Vuorikari, Stefanie Lindstaedt

► **To cite this version:**

Hendrik Drachsler, Katrien Verbert, Miguel-Angel Sicilia, Martin Wolpers, Nikos Manouselis, et al..  
Datasets for Technology Enhanced Learning. 2012. <hal-00722845>


**HAL Id: hal-00722845**

**<https://telearn.archives-ouvertes.fr/hal-00722845>**

Submitted on 5 Aug 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Alpine Rendez-Vous 2011

## **STELLAR Alpine Rendez-Vous White Paper**

### **Workshop 6: dataTEL - Datasets for Technology Enhanced Learning**

#### **Workshop Organisers**

Hendrik Drachsler, (Open University of the Netherlands, NL)

Katrien Verbert (K.U. Leuven, BE)

Miguel-Angel Sicilia (University of Alcalá, ES)

Martin Wolpers (Fraunhofer Institute for Applied Information Technology, DE)

Nikos Manouselis (Greek Research and Technology Network, GR)

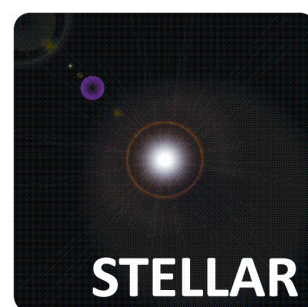
Riina Vuorikari (European Schoolnet, BE)

Stefanie Lindstaedt (KnowCenter, AT)

#### **Workshop Provocateur:**

Frank Fischer (LMU Munich, DE)

powered by



## 1. Introduction and motivation

This workshop was originally proposed by the STELLAR Theme Team dataTEL in cooperation with the MAVSEL project. The dataTEL Theme Team consists of Riina Vuorikari, Stefanie Lindstaedt, Katrien Verbert, Nikos Manouselies, Martin Wolpers and Hendrik Drachler. The MAVSEL project was represented by Miguel-Angel Sicilia.

The workshop was motivated by the issue that very less educational datasets are publicly available in TEL, so that the outcomes of different TEL adaptive applications and recommender systems that support personalised learning are hardly comparable. In other domains like in e-commerce it is a common practise to use different datasets as benchmarks to evaluate recommender systems algorithms to make the results comparable (MovieLens, Book-Crossing, EachMovie dataset).

So far, no universally valid knowledge exists in TEL on algorithm that can be successfully applied in a certain learning setting to personalise learning. Having a collection of datasets could be a first major step towards a theory of personalisation within TEL that can be based on empirical experiments with verifiable and valid results.

Therefore, the main objective of the dataTEL workshop was to explore suitable datasets for TEL with a specific focus on recommender and adaptive information systems that can take advantage of these datasets. In this context, new challenges emerge like unclear legal protection rights and privacy issues, suitable policies and formats to share data, required pre-processing procedures and rules to create sharable datasets, common evaluation criteria for recommender systems in TEL and how a dataset driven future in TEL could look like.

The workshop aimed to bring together TEL researchers, data scientists, and privacy and legal protection experts to:

- identify the most pressing topics on educational datasets
- come-up with achievable objectives to overcome the current issues on educational datasets and potential data applications

The relevant topic of the workshop were:

- publicly available datasets for educational systems
- dealing with legal protection rights towards datasets on a European level
- privacy preservation for educational datasets
- methods of effective anonymisation of educational datasets
- management and pre-processing procedures for educational datasets
- future scenarios for educational datasets
- impact of educational datasets for learners and teachers
- mash-ups based on educational datasets
- recommender approaches that are based on educational data
- evaluation methodologies and metrics for educational recommender systems

## 2. Workshop description

The participants were invited to submit original unpublished research as papers (4-8 pages) to the workshop. Demonstrations and Hands-on sessions were explicitly encouraged. All submitted papers have been peer-reviewed by two members of the program committee. Based

on the contributions of the participants the organisers identified 4 most pressing topics of the workshop and clustered the workshop contributions accordingly.

The most pressing topics were:

- Evaluation of recommender systems in TEL
- Data supported learning examples
- Datasets from learning object repositories and web content
- Privacy and data protection for educational datasets

Next to the contributions of the participants the organisers invited two keynote speakers from related research fields to share their view on the dataTEL topics. The keynote speakers were *Shlomo Berkovsky* (AUS) and *John Stamper* (USA).

### **Shlomo Berkovsky:**

Shlomo Berkovsky is a Senior Research Scientist and Research Team Leader at the TLI project (CSIRO – Commonwealth Scientific and Industrial Research Organisation, Tasmanian ICT Centre). His current research project aims to provide individual users and their families with a personalized dietary and health information to help them to maintain a healthier lifestyle.

His research interests include user modelling and personalisation of information. In particular, he is interested in recommender systems, collaborative and content-based filtering, mediation of user models, ubiquitous user modelling, context-aware personalisation, personalised content generation, and use of machine learning and data mining techniques in user modelling and personalisation.

On the 1st day Shlomo gave a keynote about: *Setting Up a Data Contest*

Research contests have attracted attention in many areas, mainly due to their potential to boost research on a specific problem. Contests also facilitate a fair and objective evaluation means, as all the participants share the same data and task. His talk focused on the details of organising a research contest. Initially, he gave an overview about several past contests: KDD Cup competition series, Netflix prize competition, and CAMRa challenge on context-aware recommendations. Then, he presented some of the essential components of a successful data contest: selection of appropriate tasks, data processing and preparation, publicity and attraction of participants, and the logistics of carrying out the contest. Finally, Shlomo showed the implications and constraints for a data competition in TEL on predicting the performance of students with intelligent tutoring systems.

### **John Stamper:**

John Stamper is the Technical Director of the Pittsburgh Science of Learning Center DataShop. He is also a member of the research faculty at the Human-Computer Interaction Institute at Carnegie Mellon University. His primary areas of research include Educational Data Mining and Intelligent Tutoring Systems. John received his PhD in Information Technology from the University of North Carolina at Charlotte, holds an MBA from the University of Cincinnati, and a BS in Systems Analysis from Miami University. Prior to returning to academia, John spent over ten years in the software industry. John is a Microsoft Certified Systems Engineer (MCSE) and a Microsoft Certified Database Administrator (MCDBA). John was the co-chair of the 2010 KDD Cup Competition, titled “Educational Data Mining Challenge,” which centred on improving assessment of student learning via data mining.

On the second day John gave a keynote about: *DataShop: An Educational Data Mining Platform for the Learning Science Community*

In his talk he discussed the vision of creating a true platform for conducting educational data mining research. The talk focused on DataShop, part of the Pittsburgh Science of Learning Center, which is an open data repository and set of associated visualization and analysis tools. DataShop has data from thousands of students deriving from interactions with on-line course materials and intelligent tutoring systems. The data is fine-grained, with student actions recorded roughly every 20 seconds, and it is longitudinal, spanning semester or yearlong courses. As of February 2011, over 245 datasets are stored including over 51 million student actions which equates to over 150,000 student hours of data. Most student actions are “coded” meaning they are not only graded as correct or incorrect, but are categorized in terms of the hypothesized competencies or knowledge components needed to perform that action. John focused his talk to workshop related key issues like the developing of an open data repository, security, privacy, and data diversity.

Based on the 4 pressing topics and the two keynote speakers we created the following workshop programme.

### Programme and workshop participants

| ARV2011 | 30.03.2011       | Presenter  | Title   |   |
|---------|------------------|--|---|---|
| 8:30    | Organisers       | Hendrik Drachsler,<br>(Open University of the Netherlands, NL) | Welcome, Introduction   |   |
| 8:50    | Keynote          | Shlomo Berkovsky,<br>(Tasmanian ICT Centre, AU)                | Setting Up a Research Contest for TEL   |   |
| 9:30    | Presenter 1      | Katrien Verbert,<br>(K.U.Leuven, BE)                           | Evaluating Collaborative Filtering Algorithms on TEL Data Sets  | <b>Topic 1: Evaluation of TEL recommender systems</b> |
| 10:00   | Presenter 2      | Peter Kraker,<br>(KnowCenter, AT)                              | Personalized Services supporting Work-Integrated Learning: An Evaluation of applicable Recommendation Mechanisms for open accessible Datasets |   |
| 10:30   | Coffee break     |  |   |   |
| 11:00   | Presenter 3      | Catherine Mulwa,<br>(Trinity College Dublin, IE)               | A Recommender Framework for End User Experience In Adaptive Technology-Enhanced Learning Systems  |   |
| 11:30   | Organisers       | All  | Discussion of topic 1   |   |
| 12:15   | Post-its session | ARV2011 Activities   |   |   |
| 12:30   | Lunch            |  |   |   |
| 13:30   | Snow activities  |  |   |   |
| 16:30   | Presenter 4      | Raquel Crespo-García,<br>(University of Madrid, ES)            | Peeking into the black box: visualizing student activities  | <b>Topic 2: Data supported learning</b>               |

|       |              |  |  |   |
|-------|--------------|--|--|---|
| 17:00 | Presenter 5  | Maren Scheffel,<br>(Fraunhofer Institute for Applied Information Technology, DE) | From Keyword Extraction To Key Action Extraction   | <b>Topic 2: Data supported learning</b> |
| 17:30 | Coffee break |  |  |   |
| 18:00 | Presenter 6  | Rory Sie,<br>(Open University of the Netherlands, NL)                            | Why should I connect and to whom should I connect? Recommendation of Knowledgeable Peers in a Co-authorship Network to Foster Innovation |   |
| 18:30 | Presenter 7  | Felix Mödritscher,<br>(Vienna University of Economics and Business, AT)          | On reconstructing and analyzing personal learning environments of scientific artifacts   |   |
| 19:00 | Organisers   | All  | Discussion of topic 2  |   |
| 19:45 | End of day 1 |  |  |   |

**Second workshop day**

| ARV2011 | 31.03.2011       | Presenter  | Title  |  |
|---------|------------------|--|--|--|
| 8:30    | Organisers       | Hendrik Drachsler,<br>(Open University of the Netherlands, NL)         | Introduction day 2   |  |
| 8:45    | Keynote          | John Stamper,<br>(Pittsburgh Science of Learning Center DataShop, USA) | The PSLC DataShop: A Data Repository for the TEL community                             |  |
| 9:30    | Presenter 8      | Joris Klerkx,<br>(K.U.Leuven, BE)                                      | Contextual Open Educational Resources for Future Recommender Scenarios                 | <b>Topic 3: Datasets from learning object repositories and web content</b> |
| 10:00   | Presenter 9      | Miguel-Angel Sicilia,<br>(University of Alcalá, ES)                    | Recommenders inside learning object repositories: requirements for meaningful datasets |  |
| 10:30   | Coffee break     |  |  |  |
| 10:45   | Presenter 10     | Eelco Herder, L3S Hannover, DE   | Experiences in Building the Public Web History Repository                              | <b>Topic 4: Privacy and data protection for educational datasets</b>       |
| 11:15   | Presenter 11     | Seda Gurses,<br>(K.U.Leuven, BE)                                       | Privacy issues and data protection in Technology Enhanced Learning                     |  |
| 11:45   | Organisers       | All  | Discussion of topic 3 and 4  |  |
| 12:15   | Post-its session | ARV 2011 activities  |  |  |
| 12:30   | Lunch            |  |  |  |

|       |                     |   |  |
|-------|---------------------|---|--|
| 13:30 | Organisers          | Hendrik Drachsler,<br>(Open University of the Netherlands, NL)<br><br>Peter Kraker,<br>(KnowCenter, AT) | Wrap up round, What are the low-hanging fruits? Writing down the Grand challenges. Reflection round supported by Research2.0 tools from STELLAR. |
| 15:00 | End of the workshop |   |  |

### 3. Emerging research questions

The workshop focused as much as possible on group discussions and group work rather than individual presentations. Before the actual workshop the participants were asked to submit their extended abstracts to the dataTEL group space at TELEurope.org and prepare the questions and statements for the workshop.

Every participant had 15 minutes to present the main message of his/her research followed by 15 minutes questions and discussions.

At the end of each topic session we had an overall discussion that took into account the latest presentations. We finalised every topic session with a speed statement round, where each participant were asked to write down his/her main challenges and ideas in one sentence. Afterwards we collected these statements and clustered those according to the 4 main topics. In that way we collected a couple of emerging research questions during the workshop for each topic:

1. **Topic:** Evaluation of recommender systems in TEL
  - Does a common data format for evaluation also require a common format for TEL recommender systems?
  - Learning is a collaborative process, how can we translate that into evaluation measures for TEL recommender systems?
  - Do we have to converge to a common evaluation framework or should we diverge to a wide range of (accepted) evaluation methods to choose from?
  - We need metrics that can be applied to every standardized dataset!
2. **Topic:** Data supported learning examples
  - How can we get richer representations of the social context in learning?
  - How important are visualisations to reflect and learn from a dataset?
  - Can teachers or students deal with data visualisations to reflect their learning process?
  - Which new competences requires data supported learning?
  - How can we integrate the context of learning into the support systems?
3. **Topic:** Datasets from Learning Object Repositories and Web content
  - Based on the experiences with SCORM and IMS-LD that should have created a European Learning Object market, will a common dataset format really lead to more datasets in TEL?
  - How can we create datasets that capture real-life learner data?
  - How can we overcome the lack of data sharing opportunities?
  - Do we need a dataset format or rather well documented datasets? or both?
  - How can we deal with the diversity of data from various TEL systems? What are appropriate levels of granularity?



- We need to create a representative association that requests datasets from the big players in LMS (e.g., Moodle, Blackboard) and learning object repositories (e.g., MERLOT, OERCommons)

#### 4. Topic: Privacy and data protection for dataTEL

- Does dataTEL require an ethical discussion on privacy, data protection and surveillance?
- We are able to develop a new generation of support tools when we are allowed to track the context and behaviour of learners. How can we deal with privacy issues in a practical way?
- Data driven research will make unveil information visible that will challenge the way we learn, teach and conducting research.

At the second day of the workshop we focused on the description of dataTEL Grand Challenges emerging out of the 4 pressing topics and the research questions. Therefore, we split the workshop into 4 smaller groups that developed a Grand Challenge for every pressing topic.

### 4. Grand Challenge Problems

#### Grand Challenge 1 / Topic 1:

***Reduce the drop-out rate in online learning environments by 10% through applying well evaluated and tested recommender systems for learning.***

*What problems of the European education system are addressed, and what are the long term benefits for society?*

A challenging problem for educational institutes and lifelong learning in general are the high drop-out rates esp. in online and distance education settings. The isolation and confusion of students may cause them to withdraw from their studies. These groups of students are called 'drop-outs'. The research on TEL recommender systems can contribute to decrease the drop-out rate by disseminating its research outcomes for the development of different support systems for teachers and students to offer relevant information at the right time. Regarding the drop-out problem it is thinkable to develop a drop-out analyzer that informs the tutor of a (Moodle) course which learners are likely to drop-out. This could be done by training a certain recommender technology on the drop-out patterns of previous (Moodle) courses. The trained analyzer could then be applied on follow-up (Moodle) courses and mark students in a list that show similar drop-out patterns. The tutor of the course could then make an intervention and contact those students personally to offer additional support for their studies.

*What are the main activities to address this Grand Challenge Problem?*

- Customize existing recommendation algorithms for learning
- Employ recommender systems in real-life scenarios
- Develop suitable evaluation criteria for different kind of recommender systems

*What is the timeframe for the Grand Challenge Problem?*



First implantations recommender systems are already available and can be implemented within a year (Manouselis et al. 2011; Manouselis et al., 2010). More challenging is the evaluation of the recommender systems that will take up to 2 – 3 years. For the further development of such systems publicly available educational datasets are needed to evaluate and compare different recommendation approaches to gain a solid body of knowledge (5-8 years).

*What are measurable progress and success indicators?*

Measurable progress and success indicators are depending on the applied type of recommender system (curriculum recommender system, drop-out analyzer etc.) (Drachler, Hummel, Koper, 2009). For the this Grand Challenge a significant decrease of the drop-out rate within an educational institution would be an promising measure to value the impact of such a system. A challenging issue will be to isolate the effect of decreasing drop-out rates only to the recommender system as most educational institutes permanently improve their educational services.

*How can funding be attracted?*

Next to European and national funding such a research project could be funded by single Universities (Innovation funds) and LMS providers like Blackboard or IMC AG. Next to the commercial providers recommender projects can be initiated as open source project in the Moodle or SAKAI community for instance.

### **Grand Challenge 2 / Topic 2:**

***ACTUALLY, help students and teachers in TEL to use data supported information systems.***

*What problems of the European education system are addressed, and what are the long term benefits for society?*

In order to make data supported information systems an effective tool for educational practice, various limitations and hurdles in technology, privacy and education need to be addressed. It is important to realise that data supported tools work with computational results that are not easy to understand and need to be presented in an easy way (e.g., by visualizations) to address the daily practice of the educational stakeholders. It is crucial to interpret the presented outcomes in a correct manner to take the right follow-up activities that can lead to improved learning. Therefore, the interpretation of educational data and its related tools requires new competences to deal with the outcomes (statistical knowledge, critical thinking, privacy awareness and ethical competences).

*What are the main activities to address this Grand Challenge Problem?*

- Developing new data driven tools that are easy to understand
- Make new real time data tools available as test applications
- Identify suitable algorithms and map them to certain datasets and learning purposes

- Integrate statistical, critical thinking, privacy awareness and ethical competences into the teacher education programs

*What is the timeframe for the Grand Challenge Problem?*

The development of the data tools has already started, on every conference new data driven tools are presented (Zhang & Almeroth, 2010). Systems like Mendley and open access journals show us already the future of academic work. The training of the new competences for teachers and students in the primary and secondary education level will take more time (5 to 10 years) with having many different levels in the EU partner countries.

*What are measurable progress and success indicators?*

- An increased effectiveness, efficiency or satisfaction of the learning process
- Courses at educational providers that train competences to handle data products
- An increasing amount of data mashup systems for different educational stakeholders (students, teachers, parents and educational providers)

*How can funding be attracted?*

For the competence training the Lifelong Learning Programme of the EU is suitable. The development of the new data driven tools can be funded by FP7 and national calls.

### **Grand Challenge 3 / Topic 3:**

***Create a generic infrastructure for sharing, analyzing and reusing learning resources and learning activity logs (educational datasets) and related research findings.***

*What problems of the European education system are addressed, and what are the long term benefits for society?*

The increased application of LMS, e-portfolio systems, and PLEs in schools and higher education institutions produces large amounts of educational data. But, although these e-learning environments store educational data automatically, exploitation of this data for learning and teaching is still very limited. These educational datasets offer an unused opportunity for the evaluation of learning theories, student support, learning technology, and the development of future learning applications. Furthermore, educational datasets can be supportive to advance research on TEL towards a basic theory for TEL (Verbert et al. 2011) by offering the recorded and observed behavior of the stakeholders (students, teachers, parents, lifelong learners, educational institutes) in different learning settings (formal – informal learning). In that way, the educational datasets extend the methodological and empirical approaches to analyze TEL that is dominated by design-based research approaches, simulations, and field studies (Gray, 2009).

*What are the main activities to address this Grand Challenge Problem?*

- Data ownership and access rights are challenging because the LMS and PLE systems are collecting educational data and the current assumption is that this data belongs to them. However, who exactly holds the ownership of the data created by the students and what can be done with it is still unresolved.
- Data policies (licences) that regulate how different users can use, share, and reference certain datasets. Until now there are very limited data policies available in educational institutes. It could be considered to apply the Creative Commons licensing rights as a standard way to grant permissions to datasets.
- Common dataset formats like from the CEN PT Social data group (<https://sites.google.com/site/camschema/home>) and a standardised documentation of datasets so that others can make proper use of it.
- Methods to anonymise and pre-process data according to privacy and legal protection rights.

*What is the timeframe for the Grand Challenge Problem?*

Anything between 5 and to 8 years. For learning resources there are already standards like LOM and Dublin Core. For learning activities it's more complicated (apart from very generic formats such as XML – which does not guarantee that data can be reused).

*What are measurable progress and success indicators?*

- An increasing amount of publicly available datasets and research articles that are based on shared datasets
- The availability of data or privacy policies at educational providers
- More data-driven tools at educational providers
- A common dataset format

*How can funding be attracted?*

Funding can be attracted from governmental funding bodies like FP7, national funding, or funding by companies like Microsoft, Google, or IBM.

**Grand Challenge 4 / Topic 4:**

*Reduce delivery costs and create more effective learning environments by applying advanced information retrieval technologies on educational data sets.*

Europe's education systems suffer from decreasing amount of teachers and the request to increase the amount of high-educated students in a short time period. As a consequence there is less time available for the individual support of students, thus the teaching quality decreases. On the other hand, the education systems are increasingly based on electronic systems like LMS and e-portfolios. With the increase in available educational data, the application of information retrieval technologies becomes valuable to create new services for the educational stakeholders (students, parents, teachers, and educational institutes). The combination of educational data and information retrieval techniques also known as Learning Analytics (LA) will become a powerful means in educational practice and student guidance (Johnson et al., 2011). LA promises the educational field to reduce delivery costs, create more effective learning environments and experiences, accelerate competence development, and increase collaboration between students and teachers.

But LA also have barriers and limitations among these are issues of privacy and data protection that need to be addressed by policy guidelines. Additional, challenges arise with respect to data surveillance<sup>1</sup> (social sorting, cumulative disadvantages) and its ethical implications.

*What are the main activities to address this Grand Challenge Problem?*

- In order to discuss and improve the above-mentioned situation a new vocabulary needs to be accomplished in order to discuss privacy, data protection and surveillance issues. For instance, what are better terms to express concepts like ownership and access control, when in digital systems replication and distribution is so easy that the concepts have no traction.
- Research is needed on how existing privacy and transparency solutions can be integrated in dataTEL practice. Further, research is desirable on how state of the art security solutions can be used to secure large educational datasets.
- There is a need for data awareness education for society. Such an educational program should not be limited to teaching individuals when to reveal or conceal their data, but

---

<sup>1</sup> Data surveillance refers to the process which individualizes each member of the population (or a group), and permits the observation and recording of each individual's activities, then collates these individual observations across the population. From these conglomerated observations, statistical norms are produced relating to any of a multitude of characteristics. These norms are then applied back to the subjected individuals, who are categorized and perhaps acted upon, either with gratification or punishment, according to their relation to the produced norm. (Phillips, Privacy Policy and PETs, 2004)

also to increase their awareness with respect to large datasets, surveillance practices, and related problems.

- User and stakeholder studies (case studies) are necessary to understand the complex requirements with respect to privacy, data protection, surveillance in dataTEL.
- The issues around privacy, data protection, and surveillance need to be addressed from the beginning of the research and not as an add-on. Methodologies and guidelines that support this vision need to be developed to support privacy and ethical practices.
- There needs to be research on how to bridge between dataTEL researchers and ethical boards with respect to advances in technologies and research and the related privacy, data protection, and surveillance concerns that arise with them.
- Policies have to be defined to avoid unethical data mining research.

*What is the timeframe for the Grand Challenge Problem?*

The first four activities can be addressed in a time frame from 2 to 3 years because they mainly require the application or translation of existing examples or solution from other domains to the educational field. The activities 5 to 7 will require a longer timeframe (3-5 years) as they can only be developed out of the experiences with the activities 1 to 4.

*What are measurable progress and success indicators?*

Measurable progress and success indicators are an increasing amount of ethical boards in LA units at educational organisations. The integration of privacy and data protection statements in research projects as well as between educational providers and the students. The integration of data and privacy competence in job profiles at the educational providers.

*How can funding be attracted?*

Funding for these challenges could be attracted from EU FP7 projects and the Lifelong Learning Programmes (Erasmus, Leonardo or Comenius).

## **5. Researchers and Communities**

In the last 3 to 6 years a couple of new research communities emerged around the dataTEL topics like Educational Data Mining, Recommender Systems in TEL, and Learning Analytics. These research communities are interdisciplinary and populated partly from Psychology Science, Educational Science, Computer Science, Data Science, Ethical science and Jurisprudence.

For the Grand Challenge 1 and 3 an interdisciplinary team should consist of computer scientists, educational experts and lawyers. Before any development can take place lawyers are needed to create suitable privacy and data sharing agreements that secure individual data on the one side and enable research to use the data on the other side. In a next step the educational experts can cooperate with the computer scientists to develop the systems required by the Grand Challenge 1 and 3.

For Grand Challenge 2 an interdisciplinary team out of educational experts, policy makers and data scientists is needed to define the new competences required to understand the outcomes of dataTEL tools and create a training program for educational stakeholders esp. teachers.

In order to address Grand Challenge 4 it is essential to work with educational scientists, policy makers, lawyers and computer scientists together. Furthermore, an European association or a Special Interest Group is needed that moderates the public discussion on ethics and privacy in TEL and offers guidelines for the work and use of educational data and related data tools.

## 6. References

- Drachler, H., Hummel, H. G. K., & Koper, R. (2009). Identifying the Goal, User model and Conditions of Recommender Systems for Formal and Informal Learning. *Journal of Digital Information*, 10(2), 4-24.
- Drachler, H., Bogers, T., Vuorikari, R., Verbert, K., Duval, E., Manouselis, N., Beham, G., Lindstaedt, S., Stern, H., Friedrich, M., Wolpers, M., (2010) Issues and Considerations regarding Sharable Data Sets for Recommender Systems in Technology Enhanced Learning. *Proceedings of 1<sup>st</sup> Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010)*, Elsevier Computer Science.
- Gray, J., (2009). A Transformed Scientific Method, In: *The 4th Paradigm: Data Intensive Scientific Discovery*.
- Johnson, L., Smith, R. Willis, H. Levine, A. Haywood, K. (2011). *The 2011 Horizon Report*. Austin, Texas: The New Media Consortium.
- Manouselis, N., Drachler, H., Vuorikari, R., Hummel, H. and Koper, R. (2010). Recommender Systems in Technology Enhanced Learning. In Kantor, P.B.; Ricci, F.; Rokach, L.; Shapira, B. (Eds.) *1st Recommender Systems Handbook*, Berlin: Springer (2011).
- Manouselis, N., Drachler, H., Verbert, K., & Santos, C. S. (Eds.) (2010). *Recommender System in Technology Enhanced Learning*. Elsevier Procedia Computer Science: Volume 1, Issue 2. *Proceedings of the 1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL)*. September, 29-30, 2010, Barcelona, Spain. doi: 10.1016/S1877-0509(10)00329-7.
- Verbert, K., Drachler, H., Manouselis, N., Wolpers, M., Vuorikari, R., & Duval, E. (2011). *Dataset-driven Research for Improving Recommender Systems for Learning*. 1st International Conference Learning Analytics & Knowledge. February, 27 - March, 1, 2011, Banff, Alberta, Canada.
- Zhang, H. and Almeroth, K. (2010). Moodog: Tracking Student Activity in Online Course Management Systems. *Journal of Interactive Learning Research*, 21(3), 407-429. Chesapeake, VA: AACE. Retrieved May 1, 2011 from <http://0-www.editlib.org.aupac.lib.athabascau.ca/p/32307>.