

# Issues and considerations regarding sharable data sets for recommender systems in technology enhanced learning

Citation for published version (APA):

Drachsler, 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. *Procedia Computer Science*, 1(2), 2849-2858. <https://doi.org/10.1016/j.procs.2010.08.010>

**DOI:**

[10.1016/j.procs.2010.08.010](https://doi.org/10.1016/j.procs.2010.08.010)

**Document status and date:**

Published: 15/12/2010

**Document Version:**

Peer reviewed version

**Document license:**

CC BY-NC-ND

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

<https://www.ou.nl/taverne-agreement>

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[pure-support@ou.nl](mailto:pure-support@ou.nl)

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 03 Dec. 2023

**Open Universiteit**  
[www.ou.nl](http://www.ou.nl)



# dataTEL – Issues and Considerations regarding Sharable Data Sets for Recommender Systems in Technology Enhanced Learning

Hendrik Drachsler<sup>1</sup>, Toine Bogers<sup>2</sup>, Riina Vuorikari<sup>3</sup>, Katrien Verbert<sup>4</sup>, Erik Duval<sup>4</sup>, Nikos Manouselis<sup>5</sup>, Guenter Beham<sup>6</sup>, Stephanie Lindstaedt<sup>6</sup>, Hermann Stern<sup>6</sup>, Martin Friedrich<sup>7</sup>, Martin Wolpers<sup>7</sup>

<sup>1</sup>Open University of the Netherlands (OUNL),  
P.O. Box 2960, 6401 DL Heerlen, The Netherlands  
[hendrik.drachsler@ou.nl](mailto:hendrik.drachsler@ou.nl)

<sup>2</sup>Royal School of Library & Information Science (RSLIS),  
Birketinget 6, 2300 Copenhagen S, Denmark  
[tb@iva.dk](mailto:tb@iva.dk)

<sup>3</sup>European Schoolnet (EUN),  
24, Rue Paul Emile Janson, 1050 Brussels, Belgium  
[riina.vuorikari@eun.org](mailto:riina.vuorikari@eun.org)

<sup>4</sup>Katholieke Universiteit Leuven (K.U.Leuven)  
Celestijnenlaan 200A, 3001 Leuven, Belgium  
{erik.duval, katrien.verbert} [@cs.kuleuven.be](mailto:@cs.kuleuven.be)

<sup>5</sup>Greek Research and Technology Network (GRNET)  
56 Messogeion Av., 115 27, Athens, Greece  
[nikosm@ieee.org](mailto:nikosm@ieee.org)

<sup>6</sup>Know-Center GmbH,  
Inffeldgasse 21a, A-8010 Graz, Austria  
{gbeham, hstern, slind} [@know-center.at](mailto:@know-center.at)

<sup>7</sup>Fraunhofer Institute for Applied Information Technology (FIT)  
Schloss Birlinghoven, 53754 Sankt Augustin, Germany  
{martin.wolpers, martin.friedrich} [@fit.fraunhofer.de](mailto:@fit.fraunhofer.de)

**Abstract.** This paper raises the issue of missing standardised data sets for recommender systems in Technology Enhanced Learning (TEL) that can be used as benchmarks to compare different recommendation approaches. It discusses how suitable data sets could be created according to some initial suggestions, and investigates a number of steps that may be followed in order to develop reference data sets that will be adopted and reused within a scientific community. In addition, policies are discussed that are needed to enhance sharing of data sets by taking into account legal protection rights. Finally, an initial elaboration of a representation and exchange format for sharable TEL data sets is carried out. The paper concludes with future research needs.

**Keywords:** data sharing, recommender systems, data sets, format, technology-enhanced learning, legal protection rights

## 1. Introduction

Personalisation is a key approach to overcome the plethora of information in the Knowledge Society. It is expected that personalised learning has the potential to reduce delivery costs, to create more effective learning environments and experiences, to accelerate time to competence development, and to increase collaboration between learners. Recommender Systems are one of the promising technologies to support people in finding most suitable information and peer learners. They are increasingly applied in Technology-Enhanced Learning (TEL) in various European projects (TENCompetence, OpenScout, APOSDLE, ROLE, STELLAR, Organic.Edunet, VOA3R) in order to personalise learning content and connect suitable peer learners according to their context (individual needs, preferences, and learning goals). In the world of consumer recommender systems, it is a common practise to use different data sets with specific characteristics to evaluate new recommender system algorithms (MovieLens, Book-Crossing, EachMovie data set). These data sets are used as common benchmarks to evaluate new kinds of recommendation algorithms [1][2][3].

Data sets for TEL are multi dimensional as TEL takes place in the whole spectrum of learning roughly distinguished between formal and non-formal learning settings. Both settings offer a rather different context that has to be taken into account by recommender systems in order to offer personalized information to individual learners. Formal learning is usually organized according to some curriculum, traditionally occurs in teacher-directed environments with person-to-person interactions. Non-formal learning mainly takes place in learning phases of lifelong learners who are not participating in any formal learning context. They are acting more self-directed and they are responsible for their own learning pace and path. The learning content for their learning comes from different Web 2.0 sources like blogs, social bookmarking tools, or document sharing platforms like Script or Slideshare.

Between both ends of the spectrum of learning various learning settings like workplace learning [4] and informal learning are located and overlap with each other. From the perspective of a recommender system developer each of them has their own particular data sets, user models, recommendation tasks, and most suitable recommendation algorithms.

So far, there are no standardised data sets publicly available neither for formal learning nor non-formal learning or something in between. Therefore, the performance results of different recommender systems research efforts are hardly comparable [5]. There is no universally valid knowledge which recommender system algorithm can be successfully applied in a certain learning setting. Drawing conclusions about the validity and generalisability of scientific experiments depends on the possibility of verification, repeatability, and comparisons of results. Thus, a data set framework is needed offering different data sets for different learning settings. This would enable researchers to create repeatable experiments to gain valid

and comprehensible knowledge on how certain recommender system algorithms performed on a certain data set (learning settings).

Therefore, we propose in this paper to create such a data set framework for recommender systems in TEL. Based on the above-described spectrum of learning settings, we suggest the first steps towards a general data set framework consisting out of different data sets in order to address the specific learning settings. The two main objectives of a data set framework will be first to collect data sets from different TEL research fields, and second to track outcomes of applied recommender system algorithms on these data sets. Various additional challenges have to be solved in order to create such a data set framework. We will tackle with the first four challenges (i.e. creating good data sets, creating reference data sets, considering legal/privacy issues, representing/sharing data sets) in the remainder of this paper.

In Section 2, we present some initial suggestions that could help developers create good TEL data sets for testing recommender systems in particular learning settings. Section 3 elaborates a bit more on the process, by focusing on how these data sets can be created in a way that will make them reference ones for the TEL scientific community. In Section 4, we will focus on policies that are needed to enhance sharing of data sets for TEL by taking into account legal protection rights of individuals. Afterwards, we suggest in Section 5 a first format for exchanging data sets that includes a common metadata structure to enable the access to and interpretation of data sets. Finally, we conclude the paper with future research questions that are not sufficiently answered so far.

## 2. Suggestions for creating suitable data sets

The *GroupLens* research group explicitly discussed the challenges of creating and using data sets specifically for testing recommender algorithms [6]. They emphasized that it is very important that the recommendation task the algorithm is designed for needs to be similar to the tasks supported by the system from which the data was collected. An example is the MovieLens recommender system, which offers support for the *Find Good items* user task. This means that the recommender system tends to show those items it has more ratings about very popular movies. As a consequence, the MovieLens data set has fewer ratings for unpopular movies. Therefore, using this data set to evaluate a different task such as *Recommend Sequence* or *Annotation in Context* would be inappropriate. In [7], a comprehensive overview of recommendation tasks for TEL recommender systems is given. These recommendation tasks could be further allocated to different learning settings that are possible for TEL recommender systems.

Thus, data sets with realistic representations of the recommendation task and the learning setting are needed in order to evaluate recommendation algorithms with particular recommendation tasks. Learning settings could be for instance data sets from Moodle instances running in schools or universities (formal learning), data sets from professional communities like meeting recordings (Flashmeeting data [8], repositories of research articles (Google scholar, EC-TEL proceedings, Mendeley,

CiteUlike, ResearchIndex, collected Web 2.0 contributions of Mashup environments like ReMashed [9] or TeamSpace [10].

We identified three initial suggestions for creating suitable data sets for TEL recommender systems. Following these rules may support system developers and researchers estimate how generalisable results of single experiments can be for related learning settings.

1. *Try to create a data set that realistically reflects the variables of a learning setting.* It is important to use a data set that is representative to the target learning setting. In case a recommender system should support ad hoc online communities, the data set also needs to present these ad hoc communities in terms of the amount of users, items, and type of resources. A good data set includes and stores information that corresponds to the actual content and user interaction in a given learning setting.
2. *Try to use a sufficiently large set of user profiles.* The success of recommender systems often relies on the availability of a sufficiently large set of users providing a sufficiently large set of ratings. If real data are not available (e.g. if a new system is being developed) then try to use data from similar application contexts or synthetic/simulated data sets that mimic the properties of the targeted one [11].
3. *Try to develop data sets comparable to others.* Using a data set from a particular TEL project does not mean that the results are generalisable to a learning setting. Therefore, it makes sense to develop data sets that are structured in a way similar to the one of other data sets (especially from similar learning settings) in order for developers and/or researchers to be able to test their system components (e.g. algorithms) with different data sets that have a similar structure/format.

### **3. General approach for creating sharable data sets**

Several steps have to be taken to move from the simple notion of using Web 2.0 data in recommendation experiments to producing and sharing actual benchmark data sets with the scientific community. In this section, we list what we believe are some of the most important steps to ensure the quality of a data set. We divide these steps into three different phases: (1) collection, (2) processing, and (3) sharing.

While the *collection phase* is often rushed, gathering Web 2.0 data from a website for use in recommendation experiments is not simply a matter of pointing a crawler to the website in question and pressing a button. The first decision to be made is which source(s) to select. This naturally depends on the purpose and application domain of the recommender system, but even websites within the same domain can provide different functionality and information. For example, many different social bookmarking websites exist, such as Delicious, Furl, Simpy, and Diigo just to name a few. While they all share the same goal—making it easier for users to bookmark valuable websites and information—they all differ in how they exactly provide this service. It is therefore important to carefully analyse the available websites to

determine what information is available, if it is suitable for the learning setting, and whether it could be used for generating recommendations.

In addition, Web 2.0 websites can also differ wildly in the size and characteristics of their user base and in their demographic makeup. Web 2.0 websites in particular tend to attract a younger and technology-minded people, whereas a data set of a distance education institute tends to represent older, less technology minded people. But different websites might attract different subcultures. While these characteristics are perhaps not always public knowledge and therefore not always within the experimental control of the researcher, they can influence different aspects of the recommendation experiments and are therefore important to keep in mind. It is therefore advisable to collect multiple data sets from different sources to improve the generalisability of the results.

After these decisions have been made, it is essential to think about how to ensure the representativeness of the data to be collected. Web 2.0 websites, especially the popular ones, are a massive, moving target, which makes it practically impossible to collect all information available on such a website. It is therefore important that the data collected is a representative sample of the user population of the website and of its general usage. For instance, when crawling a social bookmarking website, care has to be taken to not simply start crawling the list of most popular tags, and then collecting all users and items that have been used or have been annotated with those tags. Popular tags are more likely to be linked to popular items and users with large profiles, which then make up a disproportionately large part of a data set. If the goal is instead to collect a representative sample of all types of users, instead of the most active ones, it is often better to guide the collection process in a different way. For instance, one could rank all users by the profile size, take the logarithm of that size, and divide the users into ten equally-sized groups. By randomly selecting the same number of users from each group, new or less active users have a greater chance of being selected. Here, techniques from other fields such as index size estimation for search engines could also be used to ensure the quality of the sampled data [12].

Finally, before the final collection process takes place, it is advisable to contact the Web 2.0 website directly and see if they are willing to share their data. This saves time for the researcher and bandwidth for the website. Examples of Web 2.0 websites that freely share their data set include CiteULike, LibraryThing, Last.FM, and Filmtipset.se.

After the initial data sample has been collected, it will have to be *processed*. This can involve a number of different pre-processing, cleaning, and filtering steps. One likely problem with collected data from non-formal learning settings is that it contains spam. Any system that relies on user-generated content is vulnerable to spam, and Web 2.0 sources are no exception. One of the simplest possible solutions to dealing with spam could be to rank all users by their profile size and automatically disregard the top 5-10% as spam users. However, a more fine-grained approach might be better. Krause et al. [13] present an overview of many different approaches to dealing with spam in social bookmarking websites and can serve as a good starting point for spam detection.

Another issue with collecting data from a Web 2.0 website are duplicates. The ideal scenario would be that every time a user adds some content to a Web 2.0 website, the system can detect whether or not that resource has already been added to

the system. While most Web 2.0 services will attempt to match identical content, no matching system can be perfect. Compared to spam, this is obviously less an issue of malicious intent, and more one of carelessness or lack of awareness. However, the presence of duplicates can still affect the algorithms that operate on the data set. For instance, a recommender system might recommend two versions of the same resource in the same session, or would perhaps not be able to locate all interesting content when the popularity counts are diluted and fragmented. Duplicate detection has mostly been studied as the problem of ‘record linkage’ in databases [14], with the only known application to Web 2.0 websites by Bogers [15].

A third issue related to processing the data set is that of noise reduction. It is common practice in research on recommender systems to filter data sets on noise, as spurious connections between users and items could make it difficult for a recommender system to generate meaningful suggestions. A popular filtering threshold, for instance, is to filter out all users with less than twenty items in their profile. Another possibility could be to view the user-item patterns as a bipartite network and take the  $k$ -core of this network, i.e., retain only those users that have at least  $k$  items and only those items that were added by at least  $k$  users [16]. Whichever filtering regime is used, it is advisable to retain both the unfiltered version of the data set as well as the filtered version, so future users can implement their own filtering scheme.

The third and final phase involves *sharing* the experimental results and the Web 2.0 data set. Besides privacy rules (see section 4) some attention should also be paid to the way the data is formatted to be shared within a scientific community. In principle, the choice of representation format is completely free, apart from being bound by the specific input requirements of the recommender system. However, for sharing purposes it is prudent to stick to a common representation format that is accepted by open-source recommendation software. As of yet there are no widely accepted representation formats available, but an inspection of the available data sets, such as MovieLens, BookCrossing, and Netflix, could serve as inspiration. Large amounts of item metadata could be represented in XML format, while another option is make a database dump directly available, as was done in the 2008 ECML/PKDD Discovery Challenge [17].

#### **4. Policies for privacy and legal protection rights**

Privacy and legal protection rights are a very challenging and also important topic when talking over data set sharing. Finding the right balance between not accessible (private) data and public available data can easily fill many books. The most prominent incident with this respect was the release of an AOL data set with over 20 million web queries from 650,000 AOL users. The data includes all searches from those users for over three months, as well as whether they clicked on a search result and what that search result was. An inspiring data set for researchers and a scare full source for people that believes in data protection rights [18].

Privacy protection rules are surely one of the greatest challenges for sharing data sets in TEL as well. Especially, in the formal and workplace learning settings privacy protection rights hinder the exchange of these data sets.

Thus, before a data set can be shared, care must be taken to anonymise the data as much as possible. Individual user names should be anonymised so they cannot be traced back to their profile on the original Web 2.0 website, for instance by calculating a salted MD5 hash of the user name. It is important to keep in mind that with some dedicated investigation, even anonymised users and their items in the data set can often be matched to the data on the Web 2.0 website. These privacy concerns are also important when sharing results based on experiments with data collected from the Web. We therefore suggest to report only about general or aggregated conclusions over all or a group of users. Reporting about results of individual users and especially using their data as examples in publications is usually an attack of privacy. The APOSDLE project followed this approach by using unique user IDs which could not be tracked back to the individual user.

In order to overcome these challenges policies have to be created that protect the individual privacy and legal rights but enable also access to the data for research purposes and to develop recommender systems. For non-formal or informal data sets the privacy rights might be less of an issue as they are mainly based on Web 2.0 data or open educational resources that are already freely available on the Internet. But still the data might only be shared under certain licensing conditions. While data published on the Web is in the public domain, collecting it and sharing it by third parties is currently still a grey area with many different solutions. Addressing this issue is best done in collaboration with the data set provider that owns the original data.

In the field of workplace-learning privacy issues could be addressed by following best practices which have already been used in systems such as APOSDLE. Additionally, privacy guidelines should be designed and comply with privacy directives such as the European Directive on data protection (95/46/EC). An effortless classification of data sets into certain categories is needed to enhance the sharing of data. It is roughly thinkable to have public and private data sets. For public data sets no permission to view or use the data should be required. It would be possible to apply the creative commons protection rights as a standardized way to grant copyright permissions to the data sets. For private data sets users would need the permission of the creator or an organization to get access and use a data set for their own research.

In addition, every user or organization that contributes data sets to a data set framework should have full control over the data. Thus, the owner of a data set should be able to add or remove access of third parties to their own data.

Finally, the policies should also give a clear rule how and when a reference to a data sets needs to be done. Thus, in case a person used a data set for his/her own research acknowledgements should be added to the article that mentions the data set creator.



## 5. Formats to exchange data sets

A representation of implicit or explicit feedback from the users regarding the candidate items is required by a recommender system to produce a recommendation. This feedback can be in several forms. For example, in the case of collaborative filtering systems it can be ratings or votes (i.e. if an item has been purchased, viewed or bookmarked). In the case of content-based recommenders, it can be product reviews or simple tags (keywords) that users provide for items. Additional information is also required such a unique way to identify who provides this feedback (user ID) and upon which item (item ID). The user-rating matrix used in collaborative filtering systems is a well-known example.

An important issue related to the representation and export of user feedback data sets from a given application environment is the declaration of the type(s) of feedback that is being collected and its format. For example, such an exported data set has to declare if the collected information is in the forms of ratings, reviews, or tags. In addition, it has to declare the exact structure and value spaces of the collected feedback. For instance, ratings may be collected upon one or more attributes (criteria), and using different rating scales. This can be particularly evident when examining evaluation or quality models for different application areas. In such a previous survey for TEL, it has been found [19] that rating of learning resources takes place using a variety of review schemes or instruments.

To this end, an initial suggestion would be representing a data set using some plain and commonly acceptable format like comma-separated file formats (like CSV) or simple XML. To facilitate processing of similar information, it is strongly recommended that in data sets that store multiple types of social information/annotations (e.g. ratings, tags, textual reviews, bookmarks), different/separate data set files are produced - e.g. one file for the tags on the items, another file for the ratings, etc. Therefore, individual records in submitted files would be expected to at least include the following info:

- *userID*: ID of user providing the data (e.g. rating or tag). See note below on privacy.
- *itemID*: ID of the item that the data concerns (e.g. rating or tag)
- *content/value*: the actual data/info concerning the item (e.g. the text of the tag or the value of the rating)

In the case of multi-attribute data sets, files should include individual records that follow a similar format:

- *userID*: ID of user providing the data (e.g. rating or tag). See note below on privacy.
- *itemID*: ID of the item that the data concerns (e.g. rating or tag)
- *content/value of attr1*: the actual data/info concerning dimension/attribute #1 of the item (e.g. rating on dimension #1)
- *content/value of attrN*: the actual data/info concerning dimension/attribute #N of the item (e.g. rating on dimension #N)

As general notes/considerations, someone could add that: additional information like timestamps could be included at the record level; no data compression/encoding/reduction should take place (apart from the elimination of blank/invalid values); and user information should be anonymized (IDs should allow for distinguishing among users but should not reveal the true/real identify of the users or any other personal information).

In order to facilitate interoperability between different systems and potential reusability of exported user feedback in other application environments, it has been argued that it is important to define the particular social information scheme used [20]. This means that apart from the data set files to be produced and shared, an accompanying description of the properties of the data sets is required,

The explanation of the attributes/dimensions should include a number of important attributes that will give enough information to a potential user of the data set for both the structure/format of the data set and the context in which it has been created/collected. To this end, such a description could include the following information about the submitted data set files; one table describing the data scheme used in each file (see Table 1)

**Table 1.** Example of a description of a rating scheme used when collecting a data set.

<b>Example of Table: Rating Scheme used</b>	
<i>Element</i>	<i>Description</i>
<b>schemeID</b>	example.com ratings
<b>dimNum</b>	1 (ratings are provided as total satisfaction)
<b>Type</b>	Rating
<b>minVal</b>	1
<b>maxVal</b>	5
<b>stepVal</b>	1 (rating is collected as integer)

- *schemeID*: title of data scheme used
- *dimNum*: number of dimensions that the scheme uses (useful when e.g. ratings use multiple dimensions/attributes)
- *Type*: type of data scheme in file (i.e. tag, rating values, textual reviews, bookmarks, ...other)
- *minVal*: if applicable, minimum allowed value
- *maxVal*: if applicable, maximum allowed value
- *stepVal* : if applicable, rating/assessment scales intervals - for instance, step by which rating scale increases (e.g. if rating is an integer between 1..5, stepVal is 1)

For the overall data set (all files), information on the context in which the data set has been collected

- Application/environment: short description of the application in which the data set has been collected (e.g. Mendeley.com) and the relevant context (e.g. service/application provider, targeted users, educational setting, etc.).
- Data set contact: Currently responsible contact person with contact details (incl. institution and e-mail).
- Legal protection / open access policy: Is the data set available for public or does it require certain protection rules. In case it requires protection rules when and how it could be used by other researchers?
- Data set collection/processing method: How was the data set collected, which techniques have been applied to collect (e.g. logging users, storing user-provided info in system's DB), was the raw data set extended with additional information.
- Data set coverage (overall/instance): is the data set a complete reflection of the data in the system or just one instance of specific users or a given time period? Also define sampling period/time (when and in which time period was the data set collected).
- Application/environment stats: number of overall users, number of overall items, number of overall tags/ratings/reviews/... in the application (not the submitted data set but in total until date of reporting)
- Data set stats: number of submitted files, statistics per file (how many users, items, data entries per file), other meta information.
- Raw or cleaned data set: Is the data set a raw data set or was already certain noise reeducation or stemming methods applied. In case yes what kind of methods have been applied and on what part of the data.

Initial work on the development of a common metadata structure that will enable the representation, storage, sharing and reuse of such data sets has been carried out in the past [19][20]. Currently, this is a topic explored in the context of the STELLAR Theme Team on dataTEL (<http://www.teleurope.eu/pg/groups/9405/datatel/>), as well as a Project Team on “An information model and its XML binding for capturing information about the perceived quality and (re)-usability of learning objects” of the Workshop on Learning Technologies (WS-LT) of the European Standardisation Committee (CEN). The two groups are in close communication and collaboration, and are expected to produce results of joint interest and agreement.

## 6. Conclusions

This paper raised the issue of missing standardised data sets in TEL that can be used as benchmarks to compare different recommendation systems. We discussed how suitable data sets could be created according to some initial suggestions, and investigated a number of steps that may be followed in order to develop reference data sets that can contribute to the scientific community. We discussed policies that are needed to enhance sharing of data sets by taking into account legal protection rights. Finally, an initial elaboration of a representation and exchange format for

sharable TEL data sets was presented. In the following section we discuss related challenges and future research needs.

Besides the sharing of data sets within TEL, another challenging issue is the creation of common evaluation criteria to compare the effects of TEL recommender systems. Consumer product recommendation systems are evaluated based on common technical measures like Accuracy, Recall, F1 and other performance indicator like execution time [21]. But focusing only on technical measures for recommender systems in TEL without considering the actual needs and characteristics of the learners is questionable. In the literature there are two evaluation approaches to measure the impact of TEL recommender systems: 1. A combined evaluation approach by [22] and a layered evaluation approach by [7]. They need to be further explored and maybe combined to enable a standardised way to compare TEL recommender systems.

Besides the selection of most suitable evaluation criteria another challenge is the creation of an overview of the performance of different TEL recommender systems. Therefore, already available data sets known from the TEL literature need to be combined in a framework and the applied recommender system and their outcomes should be described in relation to them. In that way, it could be monitored which and how certain recommender systems have been applied and what were their outcomes according to the standardised evaluation criteria. This would be a very helpful instrument for researchers to get an overview about applied recommendation approaches for different learning settings. The researchers could judge which recommendation approach performed better than others for a certain data set. It could be a first step forward to an evidence driven theory for recommender systems in TEL.

## Acknowledgements

This work is kindly supported by the dataTEL Theme Team and the STELLAR Network of Excellence. Katrien Verbert is a Postdoctoral Fellow of the Research Foundation - Flanders (FWO). The work on Nikos Manouselis has been funded with support by the EU project ECP-2006-EDU-410012 'Organic.Edunet' of the *eContentplus* Programme.

## References

1. Goldberg, K., Roeder, T., Gupta, D., Perkins, C.: Eigentaste: A Constant Time Collaborative Filtering Algorithm. *Information Retrieval*, vol. 4, no. 2, pp. 133–151 (2001).
2. O'Sullivan, D., Wilson, D., Smyth, B.: Improving Case-Based Recommendation A Collaborative Filtering Approach. In: Craw, S., Preece, A. D. (eds), *Proceedings of the 6th European Conference on Advances in Case-Based Reasoning*, Lecture Notes of Computer Science, vol. 2416, pp. 278--291. Springer-Verlag London, UK (2002)
3. Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Analysis of recommendation algorithms for e-commerce. In: Jhingran, A., Mason, J. M., Tygar, D. (eds) *Proceedings of the 2nd ACM conference on Electronic commerce*, pp. 158-167. ACM, New York, NY (2000).

4. Lindstaedt, S. N., Beham, G., Kump, B., Ley, T.: Getting to Know Your User – Unobtrusive User Model Maintenance within Work-Integrated Learning Environments. In: In Proceedings of the 4th European Conference on Technology Enhanced Learning: Learning in the Synergy of Multiple Disciplines (Nice, France, September 29 - October 02, 2009). U. Cress, V. Dimitrova, and M. Specht (eds.) Lecture Notes In Computer Science, vol. 5794, pp. 73–87. Springer-Verlag, Berlin, Heidelberg (2009)
5. Manouselis, N., Vuorikari, R., Assche, F.V.: Collaborative Recommendation of e-Learning Resources: An Experimental Investigation. *Journal of Computer Assisted Learning*, Blackwell, August 2010.
6. Herlocker, J. L., Konstan, J. A., Borchers, A., Riedl, J.: Evaluating Collaborative Filtering Recommender Systems. *ACM Transactions on Information Systems*, vol. 22, no. 1, pp. 5--53 (2004).
7. Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H. G. K., Koper, R.: Recommender Systems in Technology Enhanced Learning. In: P. B. Kantor, F. Ricci, L. Rokach & B. Shapira (Eds.), *Recommender Systems Handbook*. Berlin: Springer (2010).
8. Wild, F., Ochoa, X., Heinze, N., Crespo, R. M., Quick, K.: Bringing together what belongs together: A recommender-system to foster academic collaboration. In: Verbert, K., Duval, E., Lindstaedt, S., Gillet, D., Scott, P. (eds), *Proceedings of the second Alpine Rendez-Vous*, November 30 – December 2, 2009.
9. Drachsler, H., Rutledge, L., Van Rosmalen, P., Hummel, H. G. K., Pecceu, D., Arts, T., Hutten, E., Koper, R.: ReMashed - An Usability Study of a Recommender System for Mash-Ups for Learning. *International Journal of Emerging Technologies in Learning (iJET)*, vol. 5, pp. 7–11 (2010).
10. Glahn, C., Specht, M., Koper, R.: Visualisation of interaction footprints for engagement in online communities. *Educational Technology & Society*, vol. 12 no. 3, pp. 44–57 (2009).
11. Manouselis, N., Costopoulou, C.: Experimental Analysis of Multiattribute Utility Collaborative Filtering on a Synthetic Data Set. In: M. Ma & G. Uchyigit (Eds.), *Personalization Techniques and Recommender Systems*, Series in Machine Perception and Artificial Intelligence (V.70), World Scientific Publishing, pp. 111--134, May 2008.
12. Henzinger, M. R., Heydon, A., Mitzenmacher, M., and Najork, M.: On near-uniform URL sampling. *Computer Networks*, vol. 33, no. 1-6, pp. 295--308 (2000)
13. Krause, B., Schmitz, C., Hotho, A., and Stumme, G.: The anti-social tagger: detecting spam in social bookmarking systems. In Castillo, C., Chellapilla, K., Fetterly, D. (eds) *Proceedings of the 4th international Workshop on Adversarial information Retrieval on the Web (Beijing, China, April 22 - 22, 2008)*, vol. 295, pp. 61–68, ACM, New York, NY (2008)
14. Elmagarmid, A., Panagiotis G. I., Verykios, V.: Duplicate Record Detection: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 1, pp. 1–16 (2007)
15. Bogers, T.: *Recommender Systems for Social Bookmarking*. Ph.D. thesis, Tilburg University, December 2009
16. Seidman, S.B.: Network structure and minimum degree. *Social Networks*, vol. 5, pp. 269–287 (1983)
17. Hotho, A., Benz, D., Jäschke, R., Krause, B. (eds): *ECML PKDD Discovery Challenge 2008 (RSDC'08)*. Workshop at 18th European Conference on Machine Learning (ECML'08) / 11th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'08), 2008.
18. AOL search data scandal. Accessed online on 30 June 2010 at [http://en.wikipedia.org/wiki/AOL\\_search\\_data\\_scandal#External\\_links](http://en.wikipedia.org/wiki/AOL_search_data_scandal#External_links)
19. Vuorikari, R., Manouselis, N., Duval, E.: Using Metadata for Storing, Sharing, and Reusing Evaluations in Social Recommendation: the Case of Learning Resources. In: Go D.H. & Foo S. (Eds.) *Social Information Retrieval Systems: Emerging Technologies and*

Applications for Searching the Web Effectively, Hershey, PA: Idea Group Publishing, pp. 87-107, January 2008.

20. Manouselis, N., Vuorikari, R.: What if annotations were reusable: a preliminary discussion. In: M. Spaniol, Q. Li, R. Klamma & R. W. H. Lau (eds.), Proceedings of the 8th International Conference on Web-based Learning (ICWL 2009), Aachen, Germany, pp. 255–264, August 2009.
21. Adomavicius, G., Sankaranarayanan, R., Sen S., Tuzhilin, A.: Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems*, vol. 23, no. 1, pp. 103–145 (2005)
22. Drachsler, H., Hummel, H. K. G., Koper, R.: Identifying the Goal, User model and Conditions of Recommender Systems for Formal and Informal Learning. *Journal of Digital Information*, Vol 10, No 2 (2009)
23. Burke, R.: Hybrid recommender systems: survey and experiments. *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331-370 (2002)
24. Kirkpatrick, D. L.: Techniques for evaluating training programs. *Journal of the American Society of Training Directors*, vol. 13, pp. 3–26 (1959)