Project Acronym: ODS
Grant Agreement number: 297229
Project Title: Open Discovery Space: “A socially powered and multilingual open learning infrastructure to boost the adoption of eLearning resources”

D8.1 - Review of Social Data Requirements

Revision: Draft version 0.8

Authors:
Hendrik Drachsler, Wolfgang Greller, Soude Fazeli (OUNL)
Katja Niemann (FRAUNHOFER)
Salvador Sanchez-Alonso, Enayat Rajabi (UAH)
Matthias Palmér, Hannes Ebner (MetaSolutions)
Bernd Simon, Daniela Nösterer (Knowledge Markets)
Kostas Kastrantas, Nikos Manouselis (AK)
Ilias Hatzakis (GRNET)

Project Co-ordinator

Company name: INTRASOFT International S.A.
Name of representative: Antonis Ramfos
Address: Rue Nicolas Bove 2b
Phone number: + 30 310 6876482
Fax number: + 30 210 6876478
E-mail: Antonis.ramfos@intrasoft-intl.com
Website: www.opendiscoveryspace.eu
# Table of Contents

1. Introduction
2. Social data within ODS
   2.1 Working definitions and scope
   2.2 A social metadata use case - ‘Irma’
   2.3 Validation of the ‘Irma’ use case
   2.4 Requirements derived from the ‘Irma’ use case
   2.5 Inventory matrix of social data used by partner repositories
   2.6 Conclusions for the social metadata infrastructure
3. Existing social data solutions
   3.1 Contextualised Attention Metadata
   3.2 Organic.Edunet
   3.3 Learning Registry
   3.4 NSDL item-level metadata
   3.5 Inventory matrix and conclusions
4. Linked Open Data standards
   4.1 Dublin Core Terms
   4.2 FOAF
   4.3 SIOC
   4.4 REV
   4.5 Schema.org
   4.6 Inventory matrix and conclusions
   4.7 Example how ODS social data will be expressed as Linked Data
5. Privacy, and social data requirements
   5.1 Irma’s privacy requirements
   5.2 Methodology and objectives
   5.3 Cornerstones of the Terms-of-Use
   5.4 Recommendations for deployment
6. Overall conclusions
   6.1 Recommendations from the social data analysis
   6.2 Cross workpackage collaboration
      6.2.1 Coordination with system components chosen by WP9
      6.2.2 Coordination with data harvesting (social and content) from WP7
      6.2.3 Uniform Resource Identifier resolution service
   6.3 Privacy, any social data translated in sample terms-of-use
   6.4 Next steps
References
Appendix - Terms of Use (Draft)
1. Introduction

The main objective of WP8 is to enhance the ODS learning object repository with a social data management layer on top of a social data cloud. Social data based on activities of the user is considered an important information source to provide additional and more personalised access to learning objects. Social data includes publicly accessible information like comments, tags, ratings or bookmarks as well as automatically collected usage data that reveals relevant interactions between users and electronic information. Social data provides the opportunity to filter and recommend tailored information to a target user based on their social data profile.

The deliverable is structured according to four key objectives as described below and reflecting the tasks described in T8.1 of WP8 (DoW, p.37).

Definition of social data
The first objective of deliverable D8.1 is to develop a working definition of social metadata for the ODS project that clearly distinguishes it from other kinds of metadata. In section 2, we provide such a definition and created the ODS use case ‘Irma’ which represents the typical social activities of the intended target users of the ODS platform. Irma is used as a guiding example in the remaining parts of the deliverable to evaluate the suitability of a social metadata format, linked data vocabulary, and the needed aspects of the ‘Terms of Use’ of the ODS platform. Section 2 also covers a survey of available social metadata in the 23 partner repositories of ODS. We conclude this section with an overview diagram of the social metadata infrastructure within ODS which shows how the different data formats will be connected to each other.

Social metadata format
The second objective is the evaluation and selection of the most suitable social metadata format to store the social activities of the users. The ODS project requires a format that captures social metadata as well as paradata of the system according to the definition given in section 2. The following four metadata schemas were considered as relevant for the requirements of the ODS social data layer: 1.CAM, 2.Organic.Edunet, 3.Learning Registry Paradata, 4. NSDL Paradata. In section 3, the different schemata were introduced and compared according to the requirements derived from the Irma use case.

Linked Data vocabulary
The third objective of D8.1 is a review of vocabularies to use when exposing social metadata as Linked Open Data. Since paradata should not be exposed publicly for privacy and integrity reasons, no vocabularies have been reviewed for paradata. Section 4 concludes with a selection of suitable vocabularies to expose ODS social metadata. The basic vocabulary of RDF
itself, the RDF Vocabulary Language (RDFS) and the Web Ontology Language (OWL) are omitted in this review as they are not really optional and should be used when appropriate.

Terms of Use
The fourth objective of the deliverable is an analysis of legal statements that need affirmation from the ODS users in order to use the platform. These legal statements are also created according to the social activities of the Irma use case.

Finally, the deliverable concludes with summative reasoning for the development of the social data layer in line with the DoW. We also point to the next steps and some as yet unsolved issues that need to be addressed in future activities of WP8 and in cooperation with Cluster 1 and Cluster 2 (WP7, WP9).
2. Social data within ODS

2.1 Working definitions and scope

Before embarking on the tasks described above, we need to point at the scope of the study undertaken and its underlying baseline of common understanding created between the project partners. This may in some cases differ from otherwise established notions and definitions, but serves as a shared basis and working specification for the current work in a diverse multinational environment.

First and foremost, we have to specify what we mean by ‘social’ data. Social data can be defined in many ways when seen from different disciplines or perspectives. In this document, and for the purpose of serving the ODS objectives, we are interested in a particular perspective which refers to collecting a memory of information published on the Web that says something explicitly about the quality of a Web resource (artefact or user). This can come from a direct value statement (e.g. rating) or an inference (e.g. most popular resource, derived from usage stats). We consider social data as any piece of information on the Web that describes or says something about a Web resource (identified by a URI) from the perspective of an individual user or an aggregated group of users (clustering). To clarify, this perspective on social data excludes content metadata which typically describe a learning object’s characteristics and attributes (e.g. author, date, licence, document type, etc.).

Let us look further into what kind of direct and indirect/inferred value statements constitute social data. First, there is social metadata which refers to the direct interaction users have with an artefact or with other actors around the artefact (e.g. other people who also like an artefact). Interactions with artefacts can include keyword tags, bookmarking, downloading, ratings, and so forth, whereas interactions around artefacts (such as learning objects) refers to emerging discussions, commentaries, reviews, or recommendations to other users. Second, there is attention (meta)data that correspond to automatic traces of the interaction the user has with various artefacts together with appropriate contextual information. This usage data is often referred to as ‘paradata’. Paradata requires further processing before it can be used effectively.

ODS strives to accomplish interoperability between systems. A common approach is to rely wherever possible on standards for data expressions. More concretely, ODS will focus on Linked Data. Linked Data is not an individual standard but provides guidelines for exposing data
on the web in a way that allows many standards to be combined. The basis of Linked Data is to expose data in a way that is machine readable in a non-proprietary format. Linked Data goes further and suggests that the format used should be RDF and that whenever possible data should be interlinked rather than copied across systems. Linked Data may need authorization before it is viewed, but the biggest advantage of Linked Data is openness that allows interlinking to the “Linked Open Data cloud” (LOD).

Up to now we mentioned five forms of data that we now summarize:

1. **Data** - in the form of digital artefacts such as PDF files or web pages.
2. **Content metadata** - provides more or less objective facts and authoritative statements on artefacts. An example of content metadata are metadata expressed using IEEE LOM.
3. **Social metadata** - provides subjective interactive statements on artefacts. Examples of social metadata include comments, ratings, tags, comments etc. that were intentionally contributed by the users.
4. **Paradata** - provides information about how users have interacted with artefacts and sometimes also the purpose, task at hand and other contextual information. Paradata is data that is automatically tracked by the system.
5. **Linked Data** - is an approach to expose interlinked datasets on the web using the RDF standard. It allows datasets to strengthen and improve the quality of each other by always referring to the source of the data. Note that linked data is a way of exposing data publically, and can be used to make available content metadata and social metadata.

The main difference between ‘social data’ and ‘paradata’ is the provenance of the items whether they have been contributed by the user intentionally or were tracked by the system in the background.

Epistemologically, **social metadata is opinion** rather than knowledge (in the common current scientific understanding of knowledge). From this follows that the expressed quality measure derived from social metadata is “opinionated”, not absolute or benchmarked according to e.g. specific quality assurance criteria. The scope of social metadata in our context (Manouselis & Vuorikari, 2009) includes, for example, a rating of a video in YouTube, a contact in the LinkedIn account of a person, or a tag of a webpage in delicious. Outside the scope of our social data lies, for example, a property relating to some molecular functions in a Linked Data version of the Gene Ontology. A simple web link can be a form of social metadata, but contextual information (e.g. a link description) is required to know something about the valence of that piece of information (e.g. if it is positive, negative, or neutral). The same happens with tweets mentioning a URI. In such cases, techniques like sentiment analysis (Koukourikos et al., 2012) can be used.
to analyse the valence of a piece of social metadata so that something useful can be done with it.

The WP8 project team will not consider learner-centred personal characteristics (*person metadata*) such as competences, proficiency, qualifications, personal experience as social data. Profiling specifications related to person specific attributes like IMS ePortfolio, IMS LIP, UK LEAP, or HR-XML will therefore be disregarded. The metadata schemas relating to ODS content and user profiles will be developed in WP4 T4.2 and used by WP7 and WP9 respectively. Also, content metadata stored in the IEEE LOM standard and the data itself is not considered to be part of social data cloud. For an overview of the objectives of WP8 see figure 2 in the conclusions (2.4) of this section.

Lastly, it is important to note that the project will not create a new standard specification for social metadata, but builds on existing efforts and schemata. It looks at implementations from relevant previous open content RTD projects such as ROLE, MACE, OpenScout, Organic.Edunet that have partially or extensively harvested social data for various purposes. From this, WP8 will formulate comparisons and common grounding which will then be mapped onto the ODS requirements.

### 2.2 A social metadata use case - ‘Irma’

Below, we describe the anticipated user experience of a persona ‘Irma’, that provides a practical and realistic vision for the use of social metadata within ODS. This use case is intended as a basis for the selection of appropriate metadata schemas that describe and cover the most important social activities that should be covered by a social metadata layer of ODS, and which would support the described functionalities in the portal and myODS spaces:

*Irma has just started a career as a maths and physics teacher at a secondary school in the Netherlands. As a newbie teacher, she is still somewhat insecure about how to catch her students’ attention and in finding illustrative supportive learning materials to accompany her teaching and to engage students. Her school has, due to financial austerity measures, abandoned the practice of assigning an official mentor to her who would offer instructional resources and advice. Hence, Irma’s only options are a formal meeting with her head teacher, or an independent approach. As a self-motivated person, she follows the hint by a colleague to check out the ODS resource portal and teacher support network.*
Irma’s first encounter with the ODS portal leads her to a search interface, where she is able to enter search queries of topics as in many other web-based search services she is familiar with. To explore she enters a query term and browses through the result list. She discovers an interesting looking item, a simulation explaining the force of gravity. In order to find out more, she decides to click on it and look at the detailed information of this particular resource.

On the resource page, Irma is pleasantly surprised that this not only contains an overview and a more detailed description of the resource, such as information about the usage licence, the authors and the publisher of the simulation. The resource page also displays information and suggestions based on users similar to her. Irma scrolls through the page. Below an overview of the learning resource, she finds connected resources that are clustered by usage. Here, Irma learns that the simulation she’s looking at also comes with an instruction manual document, two items that are frequently downloaded and used together. Below this section, she sees what resources other users looked at, who had previously visited this resource page. This viewing path is based on the viewing navigation clicks that users take when looking at resources.

Further below, Irma finds a detailed explanation and description of the learning resource, which also includes a section on usage and context description by the collective user base of that simulation. This gives her information on where and how others have used this resource for learning, how often it has been edited and how many derivatives the ODS system knows about. In this section, Irma is informed about when the object was first uploaded, graphical stats on how often it has been viewed, downloaded, and other quantitative data as a popularity measure. It also provides a link to the editing history of the resource and links to different versions.

The next section on the resource page presents the social appreciation and quality assumptions derived from social data. This includes accumulated user ratings and a tag cloud of associated keywords, which leads to new search queries when clicking on a cloud tag. User reviews in this section verbalise the value and experiences with the respective learning object. A link leads to a discussion forum, but Irma sees that no discussions have as yet taken place with respect to this resource, the system offers to her to start a discussion (this would require her to log in).
In the share section below, Irma is able to send the link to the resource page to others via popular social networks (twitter, Google+, Facebook, delicious) or by e-mail. She decides to send an e-mail to herself for reference purposes, so she can pick up the resource later at home.

At the bottom of the resource page, Irma finds that the ODS portal has followed her query and exploration history, showing her quick links to the ten most recent objects she looked at in this session (she is still not logged in). This allows her to quickly revisit the other resource pages she looked at.

Irma is much impressed with what she found, and is curious about the personal space ODS provides for users. She registers at the portal after accepting the “Terms of Use”, which describe in detail how social data will be collected and used. She enters some information about herself in the user profile, such as position “teacher”, areas of interest, preferred language. With this information, the ODS system now is able to provide recommendations to her that match her profile. It includes latest uploaded resources, most used resources, and most highly rated resources in her areas of interest.

In her personal profile page in ODS, Irma is able to bookmark items she likes with a “favourite” marker, so she can quickly find them later again. There is also a personal history available to her now, which shows all searches undertaken and objects she has taken a deeper interest in while logged in. Also her upload library shows her the artefacts she has published and the privacy and licence status she has applied to them. User stats are presented to her in a personal dashboard, showing quantitative information about her activities in ODS, like how many items she rated, reviews she’s written, which groups she is following, votes she gave to other users’ contributions, discussion posts and comments she has made etc. Those activities also contribute to Irma’s ‘karma’ that is part of her user profile. Depending on the amount of activities she is conducting she earns points that improve her level of karma. Users with a higher karma level can be consider as trustworthy ODS members. In the personal dashboard, Irma finds a personal tag cloud of keywords she added to some resources. Irma has the possibility to send updates on her activities in ODS to her Facebook, Twitter and Google+ accounts. This reads like “Irma has just reviewed a new learning object on Open Discovery Space: link”.


In addition to resource related information, her profile page in ODS allows Irma to set her availability and privacy status to indicate whether she wants to be contacted by other portal users. In her ODS profile page, she now sees a section on similar users to herself, i.e., users the system recommends on the basis of Irma’s declared interests and activity profile. The system allows her to “follow” these users and to see objects they have favoured and rated, as well as add karma votes to their contributions which she thinks are of very high quality. When Irma conducts a search from ODS in addition to the information she received before (when she was not registered), she also now gets to see users of the resource in question and is offered a possibility to add them as a connection to follow. On the status page, she sees when logging into ODS, she receives updates on the people she follows, but there is also an option to receive these updates via e-mail when particular learning objects are discussed. But Irma finds this less interesting than the option of sending messages to a variety of groups: “users you follow”, “users that follow you”, and “similar users” (everyone the system considers similar), and there is also an “all users” and “selected users” option. Irma is a bit concerned about messaging spam, but finds a flag button, where she would be able to report inappropriate communication. She can also customise the messages she receives in a control panel.

2.3 Validation of the ‘Irma’ use case

Due to the school year being interrupted by summer holidays, the requirements analysis approach of ODS (WP2) took its initial steps in the first ODS Summer school in July 2012 in Crete. The use case ‘Irma’ was presented to 35 teachers from 14 countries in a visionary workshop. The teachers were mainly secondary level (24), but there was also primary level (2) teachers, museum educators (2), teacher trainers (2), University lecturers (2), educational policy makers (2), curriculum developers (1) and teacher trainees (9) present (some listed to have more than one role). The countries represented were Portugal, Germany, France, Finland, Greece, Austria, Poland, Lithuania, Spain, Hungary, Romania, Cyprus, Ireland, Serbia and the US. Spain (5) and Hungary (4) had more participants than others while the usual number of participants was 1-3 per country. The gender division was almost 50-50% with 15 males and 18 females present. Also the age groups were quite evenly presented with 8 teachers in their 20’s, 11 in their 30’s, 7 in their 40’s and 6 over 50 year-olds present.

The data was gathered from the participants by a questionnaire. During the workshop the participants also had a discussion regarding Irma, the observations during the discussions also supported these views.
Most important social data needs for teachers based on our analysis:

1. Sharing content via social networks such as facebook/twitter (33/33 of teachers use it/think it's useful/important):
2. Recommending content to a friend or colleague (32/33)
3. Commenting (30/33)
4. Control who can see my profile (30/33)
5. Control who I share with (30/33)
6. Creating groups or 'spaces to hang out' (30/33)
7. Recognizing different level users e.g. in forums by how much they have posted/uploaded in order to evaluate whether their content is trustworthy (30/33)
8. Following/participating to existing groups (27/33)
9. Tagging (adding keywords to pictures, text or other media) (26/33)
10. Rating of content with star rating (1-5 stars) (26/33)
11. Following other users (25/33)
12. Notifications on the website on what has 'been happening (25/33)
13. Notifications when someone commented/liked/recommended something that I posted/uploaded (24/33)
14. Notifications emails to what is going on (23/33)
15. Recommendations that are made automatically e.g. by Amazon.com or Ebay for you based on what you have previously looked at or bought (23/33)
16. Creating subgroups of people from ALL 'friends' on social media sites to cluster them (19/33)
17. Reporting inappropriate content or broken links (14/33)

The overall feedback on Irma approach raising from the teachers was promising: They are interested in the social functionalities that ODS can provide for them and they find these issues useful: One teacher from Germany for example noted that in Germany there is already as much learning materials available as one teacher can put into their classes (for Maths for example), but what is still missing is a place to share thoughts with other teachers – support from the community. This is exactly which the socially powered ODS system intends to do with the social data. In other countries like Serbia for example, there is not that much content available (in Serbian), which means that their teachers are more interested in the 'resource repository' facilities of ODS. Needs and requirements of teachers vary in Europe based on the countries' situation (explained further in D2.1. Needs analysis).

2.4 Requirements derived from the ‘Irma’ use case

According to the above Irma use case we modeled a use case diagram (see below figure 1) which summarizes the possible activities of Irma within the ODS environment. It covers social
metadata as well as related paradata events that are tracked by the system. Based on this summary, we create two lists of social events that need to be represented in the social data layer: One list for social metadata requirements, and one for paradata requirements according to our definition in section 2.1 above.

Figure 1: Use case diagram for social data of ODS based on Irma scenario
Social data requirements list:
1 Rate
2 Tag
3 Bookmark
4 Share (FB, twitter, e-mail)
5 Share count (how often has an object been shared)
6 Comment
7 Join groups
8 Posts (discussion or blog posts)
9 Following/follower

Besides these intentional and direct contributions or activities of Irma, there will be also be paradata that is collected by the ODS system. This paradata logs several activities of the user in the environment such as listed below. We continue the numbering of the social metadata as this makes it easier in the comparison tables further below to count the best performing format.

Paradata requirements list:
10 Login / logout
11 Access learning object metadata
12 Navigation history of users
13 Search history of users
14 History of learning object (first upload / edited)
15 IP location of user
16 Language of LO
17 Language of user browser
18 Group metadata to extend user profile with (new interests)

The social metadata list and the paradata list are used in the following sections to evaluate the suitability of different existent social data formats (section 3.5), identify required Linked Data vocabularies to express social data as Linked Data (4.6), and to suggested required legal statements for the Terms of Use of the ODS portal (5.3). Both lists are indicators to judge if a feature is covered by the ODS social layer, linking it to Irma’s expected user experience.

2.5 Inventory matrix of social data used by partner repositories
After the definition of social and paradata and the identification of the required datasets, it is also important to get an overview of the usage of social data schemata in the ODS partner repositories, because the ODS social data cloud should be able to cover all forms of social metadata that are provided at the partner repositories. As it is not possible to merge the
paradata of all partner repositories due to privacy and technical issues, we focused in this survey only on the public social metadata that can be harvested. We reviewed 23 partner repositories by browsing to their website and analysing if and in case what kind of social metadata they provide for their learning objects. But only 13 of the partner repositories could be taken into account for the social metadata analysis, as the other repositories were not reachable or did not provide any social metadata. The outcomes of this survey can be found in Table 1.

<table>
<thead>
<tr>
<th>Name of repository</th>
<th>Rating</th>
<th>Sharing (e.g., g+, facebook, twitter, mail)</th>
<th>Tags</th>
<th>Comments</th>
<th>Usage (e.g., download, bookmark, access)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. OpenScout Federation</td>
<td>x (1-5)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x (not public)</td>
</tr>
<tr>
<td>2. Mace</td>
<td>x (1-5)</td>
<td>-</td>
<td>x</td>
<td>x</td>
<td>x (not public)</td>
</tr>
<tr>
<td>3. Cosmos repository</td>
<td>x (1-5)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>?</td>
</tr>
<tr>
<td>4. Open Science Resources</td>
<td>x (1-5)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>?</td>
</tr>
<tr>
<td>5. GLOBE/ARIADNE federation instance</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>?</td>
</tr>
<tr>
<td>6. eduTubeplus video library</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>?</td>
</tr>
<tr>
<td>7. I2G Intergeo repository</td>
<td>x (1-5)</td>
<td>-</td>
<td>-</td>
<td>x</td>
<td>?</td>
</tr>
<tr>
<td>8. Key2Nature’s</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>9.</td>
<td>FUNecole® Educational Content Repository</td>
<td>?</td>
<td>x</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td><a href="http://funecole.com">http://funecole.com</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>LaProf educational content repository</td>
<td>x (1-5)</td>
<td>-</td>
<td>- selected by the user from an ontology</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td><a href="http://www.language-learning-portal.eu">http://www.language-learning-portal.eu</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>Miksike’s LeFo repository</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td><a href="http://lefo.net">http://lefo.net</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>Bulgarian national educational repository</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>13.</td>
<td>Open Educational Resources Commons</td>
<td>x (1-5)</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td><a href="http://www.oercommons.org">http://www.oercommons.org</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 1. Inventory matrix of social data used by partner repositories**

As shown in Table 1, more than 50% of the repositories provide ratings on a five-star scale. Most of these repositories also provide tagging data and comments created by users. It can be argued that where a learning object repository applies one type of social data, they apply most of the time the common triple rating, tagging, and comments.

For the Cosmos repository, the LOMs contain keywords from an ontology that are represented as tags. A user can only select these keywords, if they upload a document and generate the LOM themselves. Among them, only OpenScout and Mace provide usage data such as bookmarks, downloads, etc. but the data are not publicly available. As a conclusion, OpenScout seems to be the only repository that covers all the social metadata and paradata even though not all the data are openly accessible. In addition to that, we found that the eduTubeplus repository applies a classification taxonomy of tags, but these tags seem to be embedded in the
LOMs and not to be assigned by users. So, these tags cannot be considered as social data by users.

The rest of the repositories listed in Table 1, seem to provide no social data in form of ratings, sharing, tagging, bookmarking, etc. They are:

1. Ariadne Federation
2. Key2Nature’s Dryades repositories
3. Miksike’s LeFo repository
4. Bulgarian national educational repository

In addition to the mentioned repositories, we explored the following nine repositories:

1. SIVECO’s ASPECT repository
2. Virtual Bulgaria educational portal repository
3. Znam.bg collection
4. BMU collection of eLearning content (including Digiš school repository)
5. Croatian national school repository
6. Greek national school repository
7. Bildung.at repository
8. LMS.at repository
9. Ambjörn’s Naeve math repository

However, we couldn’t analyse them properly for one or more of the following reasons:

- The resources are not accessible (sometimes the server seems to be down)
- The platform is in a language not spoken by the investigators that makes it difficult to find out whether they support social data or not.
- The content / metadata seems to be not available publicly and after login. However, it is not possible to register at the time of checking the repository.

We conclude from this survey that most of the existing social metadata is inline with the common data like rating, tagging, comments. Providing paradata like amount of access or amount of bookmarks is rarely presented in the partner repositories. Thus, in principle a relatively simple approach of a social metadata layer can capture the existing data from the partner repositories.

### 2.6 Conclusions for the social metadata infrastructure

The main objective of WP8 of ODS is to create a social data layer around single learning objects within the ODS portal that takes the above mentioned social activities of the use case ‘Irma’ and stores them as social data. The user activities and the social activities will be
aggregated in a so-called social data layer or social data cloud. This metadata will be stored in two different databases:

1. Paradata database (PDb)
2. Social metadata database (SMDb)

The paradata metadata acts as a history of all user activities like viewing, editing, bookmarking a learning object or also contacting another user. The social metadata database will store aggregated information like the average ratings, assigned tags, bookmarks of certain items and not the historical information of a user or a learning object. The aim is that social metadata will be exposed as Linked Data, whereas the paradata won’t. Special care will be taken so that the data expression in the social metadata database are compatible with suitable vocabularies for future exposure as Linked Data to the Linked Open Data cloud. In that way, it can be used by other repositories or projects over APIs. Figure 2 gives an overview for the social data layer illustrating the social metadata layer as resulting from the above section.
As shown in figure 2, there are a some activities that are under responsibility of related WPs like the user profile which will be defined by WP9, who will also select the underlying technical infrastructure. WP7 will take care of exposing the harvested IEEE LOM metadata as Linked Data as this is not considered to be social metadata (see definition above section 2.1). WP8 is mainly responsible for the design and development of the social metadata layer that captures all
events between users as well as between users and learning objects (see section 3). These events are first tracked in the Paradata Database (PDb) and from there a subselection that only deals with *publicly posted social data* is sent to the Social Metadata database (SMDb). The SMDb can be a separate database or a specific view of the PDb. Both databases can be used for the development of navigation or visualization services in T8.4 later in the project.

WP8 also designs the RDF schema to expose the collected information of the Social Metadata database as Linked Data (see section 4 below). Finally, it would be beneficial to harvest existing social metadata from the partner repositories and store it in the ODS social metadata database. However, this task is not explicitly mentioned in the DoW, but would be supportive of the overall project objectives as some repositories already provided a collection of social metadata (see section 2.4 and 6.2.3).

The IPR and licensing task is not modeled in figure 2 as this is a general task that is not directly bonded to the technical architecture of the social metadata layer.
3. Existing social data solutions

The following four metadata schemas (1.CAM, 2.Organic.Edunet, 3.Learning Registry Paradata, 4.NSDL Paradata) were considered relevant for the underlying architecture of the ODS social data layer. Below we shortly introduce each of the schemas and discuss how they could be used within the ODS project (Niemmann et al., 2012).

3.1 Contextualised Attention Metadata

Description:
Contextualized Attention Metadata\(^1\) (CAM, Schmitz et al. 2012) is a format to describe events conducted by a human user, e.g. accessing a document or sending an e-mail. As little information as possible is stored in the CAM instance itself, e.g. the event type and the time stamp. All other information, e.g. metadata describing users or documents involved in the event, are linked. This way, every entity/session can be described in a different and suitable format and no information is duplicated.

\(^1\) https://sites.google.com/site/camschema/

![Diagram of CAM scheme](image)

Figure 3: CAM scheme

The main element of each CAM instance is the event entry which comprises its id, the event type, the timestamp, and a sharing level reference. Examples for event types are “send”,
“update” or “select”. In the EU funded projects that used CAM, a collection of event types already started to evolve, but this is still preliminary, see description below. The sharing level reference points to a description of the privacy level of the event, e.g. “public”, “group”, “personal”. Depending on the event, various entities with different roles can be involved. For example: When sending an e-mail there’s a person with the role sender, at least one person with the role receiver and a document with the role e-mail. Each event can be conducted in a session. A session can e.g. be the time between booting and shutting down a computer or the time between the login and logout of a user in a portal.

Please see the following excerpt of CAM event types and their attributes used in the OpenScout project (see Deliverable D5.2.2) as example of how to use CAM:

- login / logout
- filteredSearch
- socialSearch
- viewMetadata
- goToLearningObject
- addTag
- addRating
- addComment
- publishLOM

For each event type, the following information are stored as entities:

- userid
- timestamp
- source (e.g. Portal, Widget,...)

The following event types store additional event specific entities:

- viewMetadata
  - objectId
- goToLearningObject
  - objectId
- filteredSearch
  - learningResourceType
  - repository
  - authorSearch
  - competency
  - searchTerm
  - target
  - language

---

Each project in which CAM was used (ROLE, OpenScout, NaturalEurope, and MACE) created an own CAM web service. This is due to the fact that CAM is very flexible and in each project only few and sometimes very project specific events were captured. In contrast, the ROLE project developed a CAM service that is able to deal with all possible events which don’t need not to be predefined. The code of the CAM API\(^3\) is publicly available and a detailed description including performance evaluations can be found in the ROLE Deliverable D3.4\(^4\). It must be mentioned, that the CAM API is still in beta stadium.

**Relevance for ODS:**

To capture the described paradata, a format is needed to record user events in ODS, e.g. when a user accesses an object. These stored events can then be used to create behavioural profiles and recommendations. CAM is designed for this purpose, but needs to be further specified in the context of ODS, e.g. by creating a taxonomy of event types and by defining in which format the involved entities will be stored. The flexibility of the CAM scheme allows to store an unlimited amount of distinct events. It can also handle complex events such as multi-criteria ratings where each criterion will be represented by a distinct entity with a respective role that belongs to the event. In the same way, the CAM scheme can be used to store all elements of a filtered/facetted search or a language tag for comments etc.

**Limitations and constraints for ODS:**

The CAM schema can be used to store information about all conducted events. This also includes user-generated social data like tags, ratings and comments that were assigned to learning objects. It’s important to mention that users can e.g. delete tags or change ratings. This information is stored in the CAM database as well, as it can be useful, e.g. for recommender systems that take the usage of learning objects into account. However, when presenting this

---

\(^3\) [http://sourceforge.net/projects/camapi/](http://sourceforge.net/projects/camapi/)

kind of data for a learning object in the ODS portal, we are only interested in the current status of the data and not in deleted tags etc. It’s possible to compute this information using the CAM database, but this is resource intensive. For that reason, we suggest an additional social metadata database. CAM will record all past ratings, tags etc. and relationships between users and LOs. The additional social metadata server will keep track of the latest social data of a LO and also expose this to the linked data cloud.

There are already recommender systems developed that use CAM data as their basis, e.g. in the ROLE project, but they are very application specific. However, ODS adaptations implemented in Mahout are perceived as feasible and straightforward. More sophisticated approaches, using e.g. multi-criteria rating schemas etc., will only be implemented as and if required.

3.2 Organic.Edunet

Description:
The Organic.Edunet portal\(^5\) is a learning portal that provides access to more than 10,000 digital learning resources on organic agriculture and agro-ecology hosted in a federation of external repositories. Regarding social data, Organic.Edunet relies on a representation model detailed in Manouselis and Vourikari (2009) which to some degree is based on CAM, since it stores data about which tags, reviews, ratings and recommendations were assigned to learning resources by which user. This conceptual model, not specific of any portal, context or particular application, was intended as a structured, reusable and interoperable way of representing the different types of user feedback and was used as a basis for the social module in the Organic.Edunet portal. The model by Manouselis and Vourikari (2009) is based on the concept of an **annotation schema**, a formal declaration of the type(s) of feedback (i.e. rating, review, tags, etc.) including the exact structure and value spaces of the collected feedback. For instance, ratings may be collected upon one or more attributes (criteria), and may use different rating scales, particularly in different application areas. The following table shows the schema’s basic information:

\(^5\) [http://portal.organic-edunet.eu](http://portal.organic-edunet.eu)
The following are examples of possible annotation schemas:

- \( \text{myPortal.BasicRatings, 1, rating, 1, 5, 1} \)
- \( \text{myPortal.Ratings, 3, rating, 1, 5, 1} \)
- \( \text{myPortal.Tags, 1, tag, -, -, -} \)

As more than one schema can coexist in the same application, it is mandatory to register all social information schema(s) that will be used in an annotation schema registry (see Figure 4).
Figure 4. An example of an annotation scheme registry (from Manouselis & Vuorikari, 2009)

Once all the schemas are registered, user feedback is collected and stored according to one or more schemas. The following table shows a possible data for storing ratings information:

<table>
<thead>
<tr>
<th>Table: Ratings (annotation schemes registry)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Element</strong></td>
</tr>
<tr>
<td>ratingID</td>
</tr>
<tr>
<td>userID</td>
</tr>
<tr>
<td>schemeID</td>
</tr>
<tr>
<td>itemID</td>
</tr>
<tr>
<td>rating1</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>ratingN</td>
</tr>
<tr>
<td>ratingMean</td>
</tr>
</tbody>
</table>

This conceptual model asks for the creation of several data tables, one per type of annotation, in each application making use of the model. Therefore, a table such as the following would be necessary for storing information about user feedback in the form of tags:

<table>
<thead>
<tr>
<th>Table: Tags (annotation schemes registry)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Element</strong></td>
</tr>
<tr>
<td>tagID</td>
</tr>
<tr>
<td>userID</td>
</tr>
<tr>
<td>schemeID</td>
</tr>
<tr>
<td>itemID</td>
</tr>
<tr>
<td>content</td>
</tr>
</tbody>
</table>

In the Organic.Edunet portal, the conceptual model described was implemented into a concrete practical format that provides registered users of the portal with the ability to tag, rate and review learning resources, as well as to add them to their list of favourites (bookmarking):

- **Tags** (unidimensional). Tags are free and can be any length or any language, so the vocabulary for tagging is not controlled. The system stores information about the tag and the user that provided the tag for each tagging event.
- **Ratings** (tridimensional, 1-5). Ratings in Organic.Edunet provide users with the ability to include information on a rating scale 1-5 for the following 3 categories: Relevance to Organic Agriculture, Educational Usefulness and Quality of Metadata. The system automatically aggregates this information into a unique rating 1-5 star-like information. This is the information that any user can see when accessing the resource e.g. after a search, although the system stores all: user info + 3-dimensional information.

- **Reviews** (unidimensional): In Organic.Edunet users can provide a review of up to 256 characters. One user is not allowed to provide more than one review for the same resource. The system stores information on the review, user, resource and date.

- **Bookmarks** (unidimensional): Users also have the ability to bookmark resources, which is basically a form of indicating interest or quality. The system stores information on the user, resource and date when it was added.

In the Organic.Edunet portal all this social information is managed and stored externally by the so-called “social services”, an external module supported by GRNET (Manouselis et al., 2009). The portal initially was prepared to recommend users with a similar profile as well as resources that might be of interest of a user, but this was never implemented. As for the user profile information, the portal stores information related to the reviews, ratings, tags, bookmarks as well as the following personal information: Username, Full Name, E-mail, Educational level, Password, Front-end Language and Timezone. However, it is important to state that the external social data module uses an anonymous approach with only a user id so that no private information “travels” across systems.

**Relevance for ODS:**
The Organic.Edunet approach is a working solution of a social metadata layer in a partner repository of ODS. It is based on an external service that collects the data in an anonymised way and therefore provides a potential solution to the ODS social metadata that should be exposed as Linked Data in an anonymised way (see figure 1).

**Limitations and constraints for ODS:**
In this section, we report the main limitations of the Organic.Edunet approach for the management of social data and the provision of recommender services.

Architectural
The Organic.Edunet portal uses an external module that is called Social Navigation Module, which allows the storage and retrieval of social metadata. Furthermore, this module provides recommendation services to the Organic.Edunet portal based on the user profile. The Social Navigation Module was designed to support both single and multi-criteria rating schemas. An important characteristic of the module is that more than one annotation schema can coexist.
The main limitation of the architecture is that it is not modular and thus it is not easy to integrate new software components. More specifically, to add a new component that will implement for instance a new recommendation algorithm, a specific PHP development framework should be followed. It is not straightforward to use available implementations of recommender algorithms that are provided by frameworks such as Mahout. These recommender algorithms should be re-developed using the specific PHP framework. Another limitation worth mentioning, is that implementations of mathematical and machine learning functions using PHP are not very efficient in terms of memory handling. This implies limitations for scalability of the module.

The Organic.Edunet portal is connected to the Social Navigation Module through a SOAP API\(^6\). This is a shortcoming since it requires more development effort than a HTTP API. This in some cases can discourage portals to use such an external module even if the advantages of providing recommender services are clear.

Annotation schemas support
It is not possible to support more rich schemas of social data other than the one that is currently implemented in the Organic.Edunet Social Navigation Module. For instance, if a portal would like to store usage information other than tags, reviews and ratings this is not possible.

Recommendation services
The recommendation algorithms currently implemented in the Social Navigation Module are user based. This means that in order to suggest items to users, a unique id for each user is used. Although the module can store social information from different portals it cannot suggest items using data from various portals since the relation of users between different portals is not known. For this reason, the user based recommendation services cannot support a central infrastructure for social data like the one that will be developed in ODS.

Multilingual support
The language of tags and reviews that are populated is not known. This is an important limitation because it cannot support use cases like “Provide all tags for a specific item only in English”.

3.3 Learning Registry

Description:

The Learning Registry\textsuperscript{7} is a joint effort by the US Department of Education and the US Department of Defense, with support of the White House and numerous federal agencies, non-profit organizations, international organizations and private companies. Key members of the collaboration are:

- The Advanced Distributed Learning\textsuperscript{8} Initiative (ADL)
- The Office of Educational Technology\textsuperscript{9} at the US Department of Education

This effort, begun in 2010, is creating a set of technical protocols\textsuperscript{10} as a platform for innovation by content authors and aggregators. Applications built to harness the power of harvesting and analyzing the Learning Registry data will allow educators to quickly find content specific to their unique needs. The Learning Registry will store more than traditional descriptive data (metadata). It will also allow sharing of ratings, comments, downloads, standards alignment, etc.

This effort has been driven by a call for increased openness, sharing and use of digital learning resources as described in both the National Education Technology Plan\textsuperscript{11} and the National Broadband Plan\textsuperscript{12}. The specifications have been developed to support learning organizations from across all education sectors.

Importantly, the Learning Registry is not a specific destination, portal or engine that educators will “go to”. Rather, it is an open technology framework to which any content creator can publish, and any technology vendor (e.g. learning management system, content aggregators, or application developers) can leverage for their applications.

The Learning Registry beta release has been developed by a team funded by the US Department of Education and ADL: SRI International, Lockheed Martin, National Science Digital Library (NSDL), Navigation North, and Butte County Office of Education/CADRE.

**The Learning Registry: how does it work?**

The Learning Registry model collects social data such as tags, comments, ratings, clicked and viewed data, shared data, data aligned to a standard, and any other data about the usage of learning resources and shares this data in a common pool for aggregation, amplification and analysis.

\textsuperscript{7} http://www.learningregistry.org/documents/starter-resources
\textsuperscript{8} http://www.adlnet.gov/
\textsuperscript{9} http://www.ed.gov/technology
\textsuperscript{10} http://www.learningregistry.org/documents
\textsuperscript{11} http://www.ed.gov/technology/netp-2010
\textsuperscript{12} http://www.broadband.gov/plan/
The Learning Registry accepts, stores, and provides access to around 400,000 learning resource descriptions — metadata or social metadata — as documents expressed in JSON notation. The Learning Registry uses a lightweight and open-source document-based database, CouchDB (couchdb.apache.org), as storage that provides data access in the form of views generated by MapReduce functions written in Javascript. The Learning Registry also provides a layer of services (written in Python) as APIs to query and retrieve documents, on top of this database. Developers are encouraged to provide other services in addition to or on top of these services.

By design, a loose format for the submission of metadata is defined without specifying what metadata schema should be used. The Learning Registry uses a Resource Data Description (RDD) document for submitting social metadata as a thin wrapper around the submitted metadata. The services built on top of the Learning Registry can provide extraction or crosswalk services across RDDs that use disparate standards, or can assemble metadata fields from different schemas into custom views.

The following picture illustrates how the Learning Registry works:
Figure 5: Workflow of the Learning Registry

The following picture (fig. 6) has been taken from a demonstration of the Learning Registry in a lecture. It shows a sample learning resource in the middle of the picture with detail information of it below. On the left side, there is a dynamic summary about the resource. This summary includes a full description of what has been happening with this resource, how many times it has been rated, how many times it has been voted or viewed, and so forth. To the right, some of the particular activities and events related to the resource are highlighted.
Technical Specification Overview:
The Learning Registry technical specification is split into several parts:

- **Network**: The description of the resource distribution network and its parts. A fixed multi-level structure of network parts is used to support distributing content and to provide policy-based security and operations.

- **Data Models**: The data models that are used to describe the network and learning resources data.
• **Services and APIs**: The APIs used to publish and consume data and those used to operate the network. The APIs are designed to abstract the logical behaviours of the Learning Registry from any particular implementation tools.

• **General Requirements**: Common behaviours and attributes that apply to all data models and behaviours.

• **Identity and Trust**: Models of trust, authentication, authorization, identity and security. These models are applied to all data models and operations.

• **Operations**: Operational procedures that apply to any implementation.

**Representing Paradata in Learning Registry:**
Paradata capture a user’s browsing activities. The description includes JSON expressions along with English language. There are three main parts to a basic paradata statement: ACTOR, VERB, and OBJECT. An "actor" does "verb" to an "object". For example, "A teacher taught the lesson located at some URL."

- Actor = A teacher
- Verb = Taught
- Object = the lesson located at some URL

The main elements of the paradata schema are described below. All the details are accessible in the formal **Paradata Specification**. Please note, there are no restrictions on adding additional properties to any of the schemata.

---


15 [https://docs.google.com/document/d/1IrOYXd3S0FUwNozaEG5tM7Ki4_AZPrBn-pbyVUz-Bh0/edit?hl=en_US](https://docs.google.com/document/d/1IrOYXd3S0FUwNozaEG5tM7Ki4_AZPrBn-pbyVUz-Bh0/edit?hl=en_US)
"actor" | Refers to the person or group who does something. Alternatively, if you want to talk mainly about individual actions, you should consider using the Activity Streams\(^\text{16}\) format.

| 
| "objectType" | Is used when "actor" is a JSON structure, and refers to the actor itself (the same value you would use if actor was just a string). |
| "description" | Is used to provide a list of attributes that describe the actor. It can also be used to describe the action or object. |

| "verb" | Refers to the action that is taken. |

| 
| "action" | Is used when "verb" is a JSON structure, and refers to the verb itself (the same value you would use if verb was just a string). |
| "measure" | Provides a way to talk about the occurrence of the verb. It is a metric appropriate in the context. It can be a count, average, rating, or other measure. |

| 
| "measureType" | Is the kind of measurement. This can be any value that makes sense, but common ones include "count" and "rating". |
| "value" | Is the value or magnitude of the measurement. "count" has a value that describes how many times a thing happened. "rating" has a value that indicates the average rating. |

| "date" | Is either a point or period of time (aka range of dates). If it's a period of time, it contains two dates separated by a slash. Actually this field is defined by RFC3339 and ISO8601 so look it up if you want more detail. |
| "context" | Provides a way of describing where the activity took place. |

---

\(^\text{16}\) [http://activitystrea.ms/specs/json/1.0/](http://activitystrea.ms/specs/json/1.0/)
"object" | Refers to the thing being acted upon. The important part of an object is the URL (or URI) where you can find out about the object (or get the object, or ideally both).
---

| "id" | Is used when "object" is a JSON structure, and refers to the object itself (the same value you would use if object was just a string). |
|      | |

"related" | Is a collection of things that relate to the paradata (usually the object). It's an array of JSON "objects".
---

| ["object"] | Objects in the list that are related to the main object. |
|            | |

"content" | Is a string that is a human-readable sentence restating the meaning of the paradata expression as a whole.
---

Example:
Let's look at following example that says "Over the course of May 2012, high school physics teachers shared the video 'Vectors and Scalars' 10 times for teaching activities."

"actor" |
---

| "objectType" | Teacher |
| "description" | Secondary School, Physics |

"verb" |
---

| "action" | shared |
And is represented in JSON as follows:

```json
{
    "activity": {
        "actor": {
            "objectType": "Teacher",
            "description": [
                "Secondary School",
                "Physics"
            ]
        },
        "verb": {
            "action": "shared",
            "measure": {
                "measureType": "count",
                "value": 10
            },
            "date": "2012-05-01/2012-05-31"
        }
    }
}
```
Relevance for ODS:
The Learning Registry, as an open source framework and an open community of resource publishers and consumers is relevant for ODS because of its approach to storing and handling social metadata and paradata. One of the interesting points of the Learning Registry is that sets of standards can be imported into the system and users can align them to learning resources. This approach can be used for validity of a learning resource. This feature can also be considered for ODS. For instance, educators and learning resource providers can align their data and resources to imported standards by ODS.

Limitations and constraints for ODS:
Learning Registry is only concerned with para and metadata for learning objects and it defines a set of services and data structures for encapsulating both meta and paradata which are then federated amongst an ad-hoc network. In this way, the data about learning objects can be used to formulate ratings, discover how resources might be used by others, locate similar or copies of the same resource at multiple locations, assess quality and alignments of resources and so forth, but its intention is to be as a infrastructure for sharing learning objects and presents some services and API's to be used by other learning repositories to be harvested. While ODS needs complete solution not only for harvesting social data of repositories, but also for manipulating of metadata and social data for extending user profile so that recommends learning objects based on users interests. Meanwhile, Learning Registry supports many social activities such as comments and reviews indirectly. Only repositories log into system for publishing their collections. There is no user session in the Learning Registry and authentication is provided upon each request.

Learning Registry Paradata was only designed to store aggregated events and not single events. Actually it is possible to store single events, but it is not recommended in the system, while ODS wants to store any activities of single user into a database.
3.4 NSDL item-level metadata

*Description:*
NSDL (National Sience Digital Libraray) is an online portal for education and research on learning in Science, Technology, Engineering, and Mathematics. NSDL's mission is to provide quality digital resources to the science, technology, engineering, and mathematics (STEM) education community, both formal and informal, institutional and individual. The STEM Exchange is a collaboration with a range of education partners that has been initiated for the implementation of an NSDL web service to capture and share social media-generated information and other networked associations about educational resources. Collections and records stored in the NSDL repository are made available through the Search API and the NSDL OAI data provider. In addition, the Strand Map Service APIs provide access to AAAS Benchmarks, Maps and visualizations. Developers can use the Search API, SMS APIs and OAI data provider to build customized search and browse interfaces and other applications. The diagram below (Figure 7) shows an overview of the NSDL technical platform.
In creating the concept of the STEM Exchange, two different kinds of Item-Level Metadata evolved, i.e. the NSDL Annotation and NSDL Paradata. The main purpose of NSDL Annotation is to capture user comments, reviews, and teaching tips. It also allows annotations to include additional information, e.g. the metadata record contributor, annotator, or the subject. NSDL Paradata was defined to capture usage data about a resource, such as downloaded, favorited, rated.

**NSDL Annotation:**

---

**Figure 7: Overview of NSDL’s Technical Platform**

In creating the concept of the STEM Exchange, two different kinds of Item-Level Metadata evolved, i.e. the NSDL Annotation and NSDL Paradata. The main purpose of NSDL Annotation is to capture user comments, reviews, and teaching tips. It also allows annotations to include additional information, e.g. the metadata record contributor, annotator, or the subject. NSDL Paradata was defined to capture usage data about a resource, such as downloaded, favorited, rated.

**NSDL Annotation:**

---

https://wiki.ucar.edu/display/nsdldocs/Technical+Overview
Currently, NSDL Annotation defines four types of annotation data, namely text data, ID/URL references, rating data, and information about used standards. Each type is again divided into subtypes, e.g., tags and comments amongst others for text data.

Each resource description can contain several annotation records. When a user adds a rating and a comment at the same time, this can be described in one record. A separate record needs to be created when annotations are collected on different dates, created by different annotators or belong to the same annotation type (e.g., one annotation record cannot contain descriptions about two comments).

Please see http://ns.nsdl.org/ncs/comm_anno/1.00/schemas/comm_anno.xsd for the complete scheme. The following example shows an exemplary NSDL Annotation record containing a rating, a comment, a tag, and a URL to a related resource which were added to a resource with the title “Global Sun Temperature Project” on the 24th November, 2004 by Dr. Joe Whoever from the ABC University. Please note, that it is possible to change tags, ratings etc. assigned to a resource and in this case the Annotation record will be modified, but if a new tag or rating etc. is added, a new Annotation record will be created and added to the resource description.
NSDL Paradata:
NSDL Paradata captures usage data about a resource (e.g. downloaded, favorited, rated) which is designated by audience, subject or education level. In contrast to NSDL Annotations, there is only one NSDL Paradata record in each resource description.
Each NSDL Paradata record is identified by at least one recordId that can be further specified by a catalog and must contain the URL of the resource to which the paradata applies (usageDataResourceURL). A record can also contain the title of the record (paradataTitle), a description of it (paradataDescription), a URL to a page that has the summary of paradata information (usageDataReferenceURL), identifier(s) of the resource(s) to which the paradata applies (usageDateResourceIdentifier), the title of the resource to which the paradata applies (usageDataProvidedForName), and any additional XML elements (moreInfo).

The most important element is the usageDataSummary which comprises all available usage statistics/information about a resource. The scheme defines five different value kinds depending on the event type to be used as usage statistics. An Integer/Float value represents the number of times certain actions have been performed on the resource (e.g. awarded, cited, commented, downloaded, tagged, viewed). A String value is a textual value that has been associated to the resource (commented, tagged). A RatingType value is the numerical average that represents the judging of a resource on a numerical scale (star, usability). A VoteType value represents the number of positive and negative responses to a resource (accurate, like, useful, useInClassroom). A RankType value represents the standing of a resource in a hierarchy (best, most).

Besides its kind and value, each usageDataSummary element contains the start and end date for the usage data (dateTimeStart, endTimeStart), information about the Audience that conducted the event (educator, student, general public), the Subject of the used resource (computing, engineering, mathematics, science) and at which educational level (EdLevel) the resource was used (e.g., Pre-K-12, Grad1-Grade12, Higher Education).

Some value types can hold additional attributes. A String can also indicate the number of times the textual value has been associated with the resource (e.g. a resource was tagged three times with the term “management” by the same audience etc.). A RatingType must include the minimum and maximum numerical values of the scale used and the total number of ratings performed. A VoteType includes the numerical minimum and maximum ranking values and the name of the ranking (e.g., of 2010, popular, used, easiest).

Please see http://ns.nsdl.org/ncs/comm_para/1.00/schemas/comm_para.xsd for the complete scheme. The following example shows a paradata record containing usage data information such as viewed 36 times by high school educators teaching mathematics on the 17th February, 2011.
<commParadata
    xsi:schemaLocation="http://ns.nsdl.org/ncs/comm_para
http://ns.nsdl.org/ncs/comm_para/1.00/schemas/comm_para.xsd"
xmlns="http://ns.nsdl.org/ncs/comm_para"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">

    <recordId catalog="CTE">CTE-PARA-211067</recordId>

    <paradataTitle>
        <string language="en">Paradata for 'Plants'</string>
    </paradataTitle>

    <usageDataReferenceURL>any url</usageDataReferenceURL>
    <usageDataResourceURL>any url</usageDataResourceURL>
    <usageDataResourceIdentifier catalog="DLESE">SPACE-01-01C</usageDataResourceIdentifier>

    <usageDataSummary>
        <integer type="viewed"
            dateTimeStart="2011-02-17T00:03:52" dateTimeEnd="2011-02-17T00:03:52"
            audience="Educator" edLevel="High School"
            subject="Mathematics">36</integer>

        <string type="tagged" total="2"
            dateTimeStart="2011-02-17T00:03:52Z" dateTimeEnd="2011-02-17T00:03:52"
            audience="Educator" edLevel="High School"
            subject="Mathematics" language="en">climate</string>

        <string type="tagged" total="4"
            dateTimeStart="2011-02-17T00:03:52" dateTimeEnd="2011-02-17T00:03:52"
            audience="Educator" edLevel="High School"
            subject="Mathematics" language="en">global warming</string>

        <rating type="usability" min="1" max="10" total="56"
            dateTimeStart="2011-02-17T00:03:52" dateTimeEnd="2011-02-17T00:03:52"
            audience="Educator" edLevel="High School"
            subject="Mathematics">7.65</rating>

        <vote type="useful" positive="14" negative="9"
            dateTimeStart="2011-02-17T00:03:52" dateTimeEnd="2011-02-17T00:03:52"
            audience="Educator" edLevel="High School"
            subject="Mathematics"/>

        <rank type="most" name="useful" best="1" worst="5"
            dateTimeStart="2011-02-17T00:03:52" dateTimeEnd="2011-02-17T00:03:52"
            audience="Educator" edLevel="High School"
            subject="Mathematics">2</rank>
    </usageDataSummary>

</commParadata>
Why is that relevant for ODS portal and how can it be applied?
The NSDL Item-Level Metadata instances can be used additionally for describing the content (e.g., LOM) to offer information about the social metadata and the usage of the resources in a standardized way, e.g., for re-harvesting.

Limitations and constraints for ODS:
However, by themselves they are insufficient for storing events in ODS, because the NSDL Annotation schema only supports the storage of annotation data, e.g., tags and ratings and the NSDL Paradata schema is designed to store information about the usage of an object only in an aggregated way which means no single events are stored. In ODS we aim to store single events to be able to create personalized recommendations, e.g. based on the history of a user.

3.5 LRE Social Data Manager and Aggregator
A special case of a social data aggregator component is the one of the Learning Resource Exchange (LRE, http://lreforschools.eun.org) that is deployed and maintained by the European Schoolnet (EUN). This component, called the LRE Social Data Manager, is collecting and storing social metadata from the LRE portal together with a user identification. Part of this information is exposed through an LRE Social Data Aggregator node that is implemented using the NSDL paradata format - so it cannot be considered a new social data format. Unfortunately there is no publicly available information about the format used within the LRE Social Data Manager, therefore it was not possible for us to examine the user-related information it stores. Thus we do not consider this as a separate case in the analysis that follows.

3.6 Inventory matrix and conclusions
In order to evaluate the four described candidate schemata and for selecting the best suited format for ODS social metadata and paradata recording, we shall now return to our use case Irma which contains the functional requirements of the ODS social layer. We extracted these requirements for contrastive comparison which we present in the following table. The table illustrates a feature comparison of the four mentioned data formats: CAM, Organic.Edunet, Learning Registry, and NSDL Paradata. Each of the formats are either rated with a (+) to indicate it supports a requirement, or with a (-) meaning it to does not support this requirement. The first nine requirements relate to social metadata, the second nine requirements show support of paradata. These have been abstracted from the use case in section 2.3 above.
<table>
<thead>
<tr>
<th>Nr.</th>
<th>Social Metadata requirements from Irma</th>
<th>CAM</th>
<th>OrganicEdunet format</th>
<th>Learning Registry paradata</th>
<th>NSDL paradata</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Rate</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>Tag</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>Bookmark</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>Share (FB, twitter, e-mail)</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>5</td>
<td>share count (how often has an object been shared)</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>6</td>
<td>Comment</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>7</td>
<td>Join groups</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>Posts (discussion, blog, etc)</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>(Google discussion group)</td>
</tr>
<tr>
<td>9</td>
<td>following/followers</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nr</th>
<th>Paradata requirements from Irma</th>
<th>CAM</th>
<th>Organic Edunet format</th>
<th>Learning Registry paradata</th>
<th>NSDL paradata</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Login / logout</td>
<td>+</td>
<td>+</td>
<td>+ (Google)</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>Access learning object metadata</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>12</td>
<td>Navigation history of users</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>Search history of users</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>History of LO (new upload or edit)</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>IP location of user</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>Language of LO (and of browser of the user)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>17</td>
<td>Language of user browser</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
The comparison expressed in table 2 emphasizes that any of the described formats can store the social metadata from the 23 partner repositories that we analysed in section 2.4. All the mentioned formats can store common social activities of users such as rating, tagging as well as comments, and bookmarking. However, as the use case illustrates, in ODS we aim to go beyond this.

Bigger differences in applicability of the various schemata appear between the formats when we consider the requirements evolving from the Irma use case. The most promising data format here is CAM, as it covers all 18 requirements from the use case, while NSDL supports only 9 of them. Learning Registry and Organic.Edunet have also 13 and 12 points in the table respectively. All the mentioned formats use application specific models and services for implementing social services for their users.

The most important feature in the Irma use case is metadata grouping for extending user profile, since ODS portal can recommend users many learning objects based on their interests. As Table 3.5 illustrates, only Organic.Edunet and CAM group metadata to extend user profile interests.

In addition to the support of the requirements derived from the Irma use case it is also important to consider practical implications of the choice we take within the ODS project. These criteria are related to sustainability, effectiveness and efficiency of the implementation to be carried out. Thus, it is important for ODS to select a format that will continue to exist or even further develop in future, one that matches especially European learning contexts (e.g., multilingualism, multi-culturalism), that can be implemented with less effort and where existing prototypes are in place that can be adapted to the ODS system.

**Table 3. Additional considerations relevant for the ODS social metadata specification**
### Sustainability (A1):

Key members of the Learning Registry are the Advanced Distributed Learning Initiative (ADL) and the Office of Educational Technology at the US Department of Education. It has been developed by a team funded by the US Department of Education and ADL. The NSDL format is developed by National Science Digital Library that receives the majority of its funding through the generous support of the US National Science Foundation. Little is known about the intentions and future direction these US-led formats will take and one of the limiting factors from an EU perspective is the lack of influence that a project like ODS can have on these.

The Organic.Edunet format was developed in the context of the Organic.Edunet project that is funded by the European Union in the eContentplus programme. As such it is very project specific and the limitations mentioned above (section 3.2) indicate potential issues of flexibility and future development. CAM too has been EU funded, but it is supported by an even bigger research community from four different EU projects (MACE, ROLE, OpenScout, and Natural Europe). This also includes a wider stakeholder community of co-funding organisations.

<table>
<thead>
<tr>
<th></th>
<th>Natural Europe</th>
<th>Educational Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>Designed to store aggregated (AE) vs. single event (SE) (Effectiveness)</td>
<td>SE</td>
</tr>
<tr>
<td>A3</td>
<td>Known implementations (Efficiency)</td>
<td>CAM services (developed in ROLE, OpenScout, NaturalEurope and MACE)</td>
</tr>
<tr>
<td>A4</td>
<td>Expertise in the consortium (1-3 stars) (Efficiency)</td>
<td>***</td>
</tr>
</tbody>
</table>
Effectiveness and Efficiency (A2-A4):
Organic.Edunet and CAM were designed for storing single events, while NSDL and the Learning Registry store aggregate events. Both, Organic.Edunet and CAM have an own login/logout system, while the Learning Registry uses Google for this feature and also for making discussion posts. All formats except Organic.Edunet allow to share the learning objects in social communities like Facebook and Twitter but all of them provide access to learning object metadata to users as well.

Organic.Edunet and CAM already provide suitable services for the ODS project that need to be customized and adapted. Because some ODS consortium partners were also involved in the projects that applied Organic.Edunet and CAM, the expertise in the consortium is better developed regarding these two formats than with the NSDL and Learning Registry paradata.

4 Linked Open Data standards
The following section will review and suggest vocabularies to use when exposing social metadata as Linked Open Data. Since paradata should not be exposed publicly for privacy and integrity reasons, no vocabularies will be reviewed for paradata. The section will conclude with a selection of suitable vocabularies to expose the social metadata in. The basic vocabulary of RDF itself, the RDF Vocabulary Language (RDFS) and the Web Ontology Language (OWL) are omitted in this review as they are not really optional and should be used when appropriate.

4.1 Dublin Core Terms
Description:
Dublin Core Terms provides a set of terms useful for describing resources. The metadata terms are intentionally generic in their nature to make them fit for a wide range of situations. Originally, there were only 15 terms, now there are over 100 terms to cover for needs in a wide range of domains, including education. The recommendation is to find a selection of terms that fits a specific need rather than use all of them. In addition, the terms are quite generic, encouraging refinement of the terms.

Relevance for ODS:
Dublin Core Terms provides a common basis that is reused among many other standards. And even if they are not part of another vocabulary that we might choose in ODS, there are terms like creator, date, and title that are certainly relevant for expressing social data.
4.2 FOAF

Description:
FOAF (Friend-of-a-Friend) vocabulary describes user information and their social connections through concepts and properties in form of an ontology using Semantic Web technologies. The FOAF vocabulary describes personal information and social relationships. It reveals basic information of users (FOAF Basics) such as name, surname and also personal information about the people that a user "knows" and their areas of interest (personal info).

A definition from Wikipedia: "FOAF (an acronym of Friend of a friend) is a machine-readable ontology describing persons, their activities and their relations to other people and objects. Anyone can use FOAF to describe themselves. FOAF allows groups of people to describe social networks without the need for a centralised database. FOAF is a descriptive vocabulary expressed using the Resource Description Framework (RDF) and the Web Ontology Language (OWL). Computers may use these FOAF profiles to find, for example, all people living in Europe, or to list all people both you and a friend of yours know. This is accomplished by defining relationships between people. Each profile has a unique identifier (such as the person's e-mail addresses, a Jabber ID, or a URI of the homepage or weblog of the person), which is used when defining these relationships."

FOAF enables users to find documents on the Web according to the documents’ properties and relationships. FOAF can help us to find information about people based on their group memberships and expressed interests. FOAF also helps users to share ratings, tags, bookmarks, etc. through a distributed and decentralized system.

Relevance for ODS:
FOAF is a de-facto standard for describing persons and basic user profile information in the LOD cloud and cannot be ignored. Its importance is shown by how many other initiatives use and build new ontologies on top of FOAF. And even those that start from scratch such as schema.org have acknowledged being inspired by FOAF and a mapping back to FOAF is in progress. Furthermore, describing users by FOAF:agent allows the system to become more distributed (as for example the WebID initiative builds on FOAF) and in a longer perspective more scalable.

Limitations and constraints for ODS:
Unfortunately, there is no mapping between the OpenId user attributes and FOAF, neither is there an established mapping between OpenSocial user profile information and FOAF. Furthermore, even though there are lots of services that provide their users with FOAF profiles, few of the big players in the field do so out of the box. Nevertheless it should be noted that at
the time of writing there is no competing data expression that is better positioned to achieve data interoperability (according to the authors knowledge and judgement).

4.3 SIOC

Description:
The SIOC core ontology consists of the following classes (the descriptions originate from the SIOC specification):

- **Community**: a Community may consist of different types of objects (people, sites, etc.) joined by a common topic, interests or goals.
- **Container**: an area in which content Items are contained.
- **Forum**: a discussion area on which Posts or entries are made.
- **Item**: an Item is something which can be in a Container.
- **Post**: an article or message that can be posted to a Forum.
- **Role**: a Role is a function of a UserAccount within a scope of a particular Forum, Site, etc.
- **Site**: a Site can be the location of an online community or set of communities, with UserAccounts and UserGroups creating Items in a set of Containers. It can be thought of as a web-accessible data Space.
- **Space**: a Space is a place where data resides, e.g. on a website, desktop, fileshare, etc.
- **Thread**: a container for a series of threaded discussion Posts or Items.
- **UserAccount**: a user account in an online community site.
- **UserGroup**: a set of UserAccounts whose owners have a common purpose or interest. Can be used for access control purposes.

More detailed descriptions of the classes and a detailed listing of the classes properties can be found in the SIOC specification at [http://rdfs.org/sioc/spec/](http://rdfs.org/sioc/spec/).

Relevance for ODS:
Some SIOC classes and properties are relevant for ODS, because they provide a potentially good match for the data models used in ODS (to be explored). Can be used with FOAF (above 4.2) and REV (below 4.4). Blogging software such as Wordpress and Drupal already have support for SIOC via extentions which makes it both possible to harvest social metadata as well as build upon existing technology if needed.

4.4 REV

The Review ontology (REV) can be used to express reviews and ratings in RDF. See specification at [http://vocab.org/review/](http://vocab.org/review/).
The following terms are defined in this schema:

- **Comment**: a comment on a Review.
- **Feedback**: feedback on the Review, to express usefulness.
- **Review**: review of a work.
- **Commenter**: commenter on the Review.
- **Has review**: association between work and Review.
- **Has comment**: association between a Review and its Comment(s).
- **Has feedback**: association between a Review and its Feedback.
- **Max Rating**: numeric value.
- **Min Rating**: numeric value.
- **Positive Votes**: number of positive usefulness votes.
- **Rating**: numeric value.
- **Reviewer**: person that has written the Review.
- **Text**: text of the Review.
- **Title**: title of the Review.
- **Total votes**: number of usefulness votes.
- **Type**: type of media of a work under review.

**Relevance for ODS:**
REV complements other initiatives such as SIOC and FOAF.

**4.5 Schema.org**

Schema.org is a recent initiative among the leading search engine companies to provide a common vocabulary for describing webpages. See specifications at [http://schema.org/](http://schema.org/).

There is a mapping to RDF at [http://schema.rdfs.org](http://schema.rdfs.org).

Small selection of types which can be of interest for ODS:

- EducationalOrganization and its subtypes, such as School, HighSchool, CollegeOrUniversity, etc
- Review and Rating: reviews and ratings of an item, such as CreativeWork.
- Audience: intended audience for a creative work.

The Learning Resource Metadata Initiative (LRMI) has released a specification at [http://www.lrmi.net/the-specification](http://www.lrmi.net/the-specification) which uses schema.org terms to describe learning resources. One of the goals of LRMI was to be compatible with schema.org and other
established standards such as DC and LOM. Via schema.rdfs.org there is also a direct mapping into RDF available.

**Relevance for ODS:**
Schema.org is relevant since it is backed by the large search giants. Even if it is not chosen as the primary vocabulary to use in the Linked Data expression it will still be valuable to improve the appearance of ODS data in external search results.

### 4.6 Inventory matrix and conclusions

In the following table (table 4), we analyse which of the presented Linked Data schemas support the social metadata requirements from the Irma use case to exposed it to the Linked Open Data cloud. This table is slightly different to the table 2, as it identifies needed combinations of schemas to expose the data in a sustainable and machine readable way. Its main focus is on combining standards rather than selecting a certain data format over another. Please note that the social media requirement Nr. 8 Posts (discussion, blogs, etc) are needed to be listed here as separate items in the table.

<table>
<thead>
<tr>
<th></th>
<th>DCTerms</th>
<th>FOAF</th>
<th>SiOC</th>
<th>REV</th>
<th>Schema.org</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Rate</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>Tag</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>Bookmark</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>Share (FB, twitter, email)</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>Share count</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>Comment</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>7</td>
<td>Join groups</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>8a</td>
<td>Discussion / posts</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>8b</td>
<td>Blog / posts</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>8c</td>
<td>Microblog / post</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>Follower / Following</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
From this overview it is clear that SIOC in combination with REV covers most forms of social metadata. As SIOC depends on FOAF and DC Terms, those are used indirectly. The schema.org vocabulary is not far away, but at this point in time it lacks expressiveness at least with regard to microblogging, forums, and social sharing/bookmarking.

The vocabulary of the candidate schemas will be mapped with an internal ODS vocabulary server. This vocabulary server will store all the suggested vocabularies and include explicit term-to-term mappings.

4.7 Example how ODS social data will be expressed as Linked Data

The example shows how social metadata produced by Irma from the scenario above can be expressed in RDF using the SIOC, REV, FOAF and DCTerms vocabularies. The Turtle format is used for making the RDF more readable (for simplicity assuming there is an ods.net namespace where the ODS portal is located):

```turtle
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix sioc: <http://rdfs.org/sioc/ns#> .
@prefix siocct: <http://rdfs.org/sioc/types#> .

<http://ods.net/container/15>
  rdf:type sioc:Forum ;
  sioc:topic <http://dbpedia.org/resource/Mathematics> ;
  dcterms:title "A forum centered on math LOs" .

<http://ods.net/container/256>
  rdf:type siocct:FavouriteThings ;
  dcterms:title "Irmas favourite LOs" .

<http://ods.net/item/145>
  rdf:type siocct:MicroBlogPost ;
  sioc:has_container <http://ods.net/container/15> ;
  sioc:about <http://ods.net/lo/345> ;
  dcterms:creator <http://ods.net/account/irma> ;
  dcterms:created "2012-09-02T11:54:49"^^xsd:date ;
  sioc:content """"This learning object is really a
nice example of how you can teach linear algebra in both a fun and pedagogical way!"

<http://ods.net/item/146>
  rdf:type sioc:Item;
  sioc:has_container <http://ods.net/container/256>;
  dcterms:creator <http://ods.net/account/irma>;
  dcterms:created "2012-09-02T11:58:11"^^xsd:date;

<http://ods.net/item/226>
  rdf:type rev:Review;
  sioc:about <http://ods.net/lo/345>;
  dcterms:creator <http://ods.net/account/irma>;
  dcterms:created "2012-09-07T10:34:09"
  rev:rating "7"^^xsd:integer.

<http://ods.net/lo/345>
  dcterms:title “Linear Algebra for newbies”.
  // ... a lot more content metadata here
  rdfs:seeAlso <http://example.com/lo42432>;

<http://ods.net/account/irma>
  rdf:type sioc:UserAccount;
  sioc:email_sha1 "e520b98b2fe8f5f4e8dd27a06b21ca56f71f71a8";

<http://ods.net/person/irma>
  rdf:type foaf:Person;
  foaf:name "Irma";

In summary, the example shows:

2. A bookmark (item/146) of the same learning object (lo/345) in Irma’s personal favourites (container/256).
3 A review/rating (item/226) of the same learning object (lo/345)
4 Irma’s user account (account/irma) and FOAF profile (person/irma).
5 The learning object (lo/345) with content metadata and reference to non-ODS identifiers for the same learning object. Listed shortly for completeness, the LOD compatible expression for learning objects are not within the scope of this report.
5. Privacy, and social data requirements

5.1 Irma’s privacy requirements

As indicated and foreseen in the DoW, the ODS services of WP8 will make extensive use of personal data of users, including the tracking of their online activities and navigation. Since personal data enjoys high protection in the various legal systems via data protection and privacy legislation, these services require a legally compliant way to be operated and a structure that passes ownership and control to the users for their acceptance and agreement - after being properly informed (i.e. requiring informed consent). It also implies that agreement from the users can be withdrawn at some point and the service structure needs to be flexible and transparent enough to react to such demands.

According to the use case description - Irma in section 2.2, ODS will provide a more personalised view on learning objects. As illustrated in the use case above, but also based on the qualitative evidence collected in the context of the ODS summer school, we envisage users to become increasingly sensitive in terms of their privacy, intellectual property rights and other legal issues relevant in an increasingly connected world.

As a consequence, WP8 is charged with the task to investigate the requirements in other domains than data models and APIs. In the spirit of a good service design, we developed a recommendation for terms of use in educational portals concerned with the collection, storage, and use of social and paradata.

The following subsection describes the methodology applied. Then, the main requirements of the terms of use are documented. The actual terms of use developed with the support of a legal professional are documented in the Appendix.

5.2 Methodology and objectives

The mentioned task aimed at the development of a terms-of-use formulation that encapsulates the requirements of this work package. The requirements were collected by looking into the above-mentioned use case. In addition, good practices like amazon.com, lms.at, or mash-up.tv were studied in order to extract and adopt relevant user requirements with respect to providing social services. To add a multicultural perspective, two out of the weekly WP8 flash meetings were used explicitly, to quality assure the requirements collected from the viewpoint of the different EU member countries of the ODS project.

The objective of the development task can be summarized as follows:
Focus on social data: Provide sample terms-of-use for educational social portals (e.g. ODS portal)

Address rights and obligations: identify rights and obligations of the service provider and their users.

Agnostic in terms of (commercial) business models: The terms-of-use developed are agnostic in terms business models. Hence, no business model related aspects are modeled into the terms-of-use

Developing a legal sample document for social data services is a complex task. Given the focus of the workpackage, the terms-of-use developed are not addressing the following aspects of service provision:

- Licensing of Learning Resources: This subject is already successfully dealt by Creative Commons or Commercial Licenses and is covered in WP7.

- Third-party agreements: At the moment, the terms-of-use developed are designed for a single service provider. This statement needs to be viewed from a legal perspective only. Hence, technically speaking of course multiple service providers are possible. However, the terms-of-use do not cover the scenario where social data is collected with a Facebook-like third party plug-in economy. Special attention, therefore, is required for consortium structures and distributed architectures as intended by international collaborations (like for example ODS).

- No repair function for non-existing terms-of-use: Users already registered at an educational portal take of the functionalities of this portal subject to legacy terms-of-use. In the absence of terms-of-use local, national, or international legislations might apply. The terms-of-use presented herein are not designed to “repair” the non-existence of agreements, neither will they unify the terms-of-use of distributed service providers.

5.3 Cornerstones of the Terms-of-Use

From this discussion, the cornerstones of the terms-of-use can be summarized as indicated in the table below. The full version of the terms-of-use are provided in the Appendix.

<table>
<thead>
<tr>
<th>Registration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration is needed</td>
</tr>
<tr>
<td>Users are asked to provide true information on themselves</td>
</tr>
<tr>
<td>Users need to keep their credentials confidentially</td>
</tr>
<tr>
<td>Right to Use</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td><strong>Users need to update their password</strong></td>
</tr>
<tr>
<td><strong>Operators are allowed to restrict access in case of suspicion of third party usage</strong></td>
</tr>
</tbody>
</table>

**Social Data Collection and Protection**

| We collect data that you provide and use tools for storing them |
| Operators have to stick to the applicable data protection regulations |
| Operators have to protect personal data |
| To run an educational social portal successfully, a minimum amount of data has to be visible for other platform users and web based systems such as search engines |
| Personal data is not accessible by search engines |
Users have the right to request the permanent deletion of their data

Operators have the right to log user information

Operators have the right to use cookies and save it on your computer

Operators anonymize collected data for statistical evaluation

Operators are allowed to combine information in order to improve the service

Phishing: Identity theft

Meta (about the terms of service)

Modification of the Terms of Service

In case of dispute the applicable law is ...

Table 3. Term-of-use guidelines

5.4 Recommendations for deployment

In this document, we distinguished our service requirements as social metadata and paradata (above section 1 and 2.1), where the former is intentionally provided information, whereas the latter is the system tracking of users’ activities that happens in the background. In our legal considerations, we need to treat these two datasets differently for implementation. Concretely, we have three layers of control at our disposal:

Layer 1: General terms of use: accepted at the point of registration
This regulates the permission to use user data either anonymously in aggregated form (e.g. ratings, tags), or explicitly (e.g. displaying the username next to a comment or review). It also needs to cover in detail the acceptance to use paradata, explaining the purpose and distribution of these datasets for providing personalised recommendations and services.

Layer 2: Anonymisation and aggregation
Detaching the users’ id from the submitted data, e.g. by means calculation for aggregated ratings. The method of aggregating social information without the user id is well-known in social networks. It allows for the safe provision of average ratings of an item or click counters (e.g. likes, shares) without revealing the person providing this information.
Layer 3: Permission control
This entails putting an item level or group level control system in place that enable users to exercise privacy and access controls. Familiar contexts are e.g. the sharing of e-mail addresses with different networks (groups, logged-in users, the World, or no-one). Although not related to social data, permission controls also allow the application of usage and distribution licences (e.g. creative commons) for various types of content, e.g. photos or uploads.
6. Overall conclusions

In this final section of the report, the WP8 team summarises the outcomes against the main objectives of T8.1 Social Data Requirements in the DoW and their implications for future work. The task at hand was to study existing implementations and requirements related to the current representation, storage, processing and use of social data (social metadata and paradata) that users produce during their interactions with learning resources or other users. T8.1 entails four sub-objectives (cf. DoW p.37), and we can summarise our activities as follows:

**Objective 1-2: Data schemas & specifications & usage/annotation data types & formats**
We addressed this subtask in section 3, where we reviewed existing metadata schemas, conducted an in depth analysis on the suitability of data formats and their support for the Irma use case. Furthermore, we also addressed the second sub-objective of the DoW ‘usage/annotation data types & formats’ in section 3 when we compared the different data formats according to a set of requirements for technically representing and storing social data sets.

**Objective 4: Open & Linked social data for learning**
We addressed this sub-objective from the DoW - Open & Linked social data for learning in section 4 of this report, where we analysed the suitable Linked Data vocabulary according to the requirements again derived from the Irma use case. We put this objective forward, as it is closely related to the identification of the most suitable social metadata format discussed in the previous objective.

**Objective 3: IPR & licensing requirements**
Finally, we elaborated and proposed privacy statements that are required for a socially empowered learning environment such as ODS. These privacy statements are an extension to 'standard' terms-of-use agreements that explicitly address the social and linked data aspects of the ODS project and brings it in line with existing data protection and privacy legislation.

6.1 Recommendations from the social data analysis

From the analysis in D8.1, it appears that Organic.Edunet might be a suitable candidate for the social metadata database that is also intended to be exposed as Linked Data. CAM, on the other hand, clearly turned out to be the most suitable data format for the paradata database (see figure 2). Both formats scored very high according to the requirement lists from the Irma use case. They also got good scores on the practical issues that will support an efficient development process within the ODS project. Both formats have a strong European community
behind them and some ready-to-use services that are well known to the ODS consortium partners.

However, there are also some shortcomings that should be addressed within the ODS project. Social data is usually stored locally in the content management system of a single portal. Harvesting and aggregating such data from various learning object repositories will allow the generation of a social data cloud and will enable the provision of new services to the portals. For instance, more accurate recommendations can be generated by taking into account social data from more than one learning object repository. This overall objective is also expressed in the ODS project and especially within WP8. But collecting heterogeneous social data from different sources is not a trivial task and requires the adoption of efficient technologies and protocols. The main aspects that should be taken into consideration for the social metadata cloud within ODS are therefore:

1. A common data schema
Defining a common schema that will be used by the intended navigation and visualization services in T8.2 - T8.4 to manage and process social data. Potential options for such a schema are CAM for paradata in combination with Organic.Edunet for social metadata. Although these seem to be promising and most feasible, it needs to be analysed how Organic.Edunet can be aligned to events stored in CAM. The Organic.Edunet partners are preparing a new release of their social data schema by the end of the year 2012 that will address this issue and provide required adjustments to link CAM to Organic.Edunet.

2. Transformation of the social data schema
In order to design a common social data schema supported by different repositories a transformation of existing schemas to the ODS schema is required. The analysis was based on the repositories within ODS that have been pro-actively investigating how they can represent and store social data even before the launch of the project. We addressed this issue in section 2.4 by analyzing the used social data in 23 partner repositories. The next step would be to carry out a post-analysis step that will investigate the types of social data already produced in ODS portals, even if they are not explicitly stored in separate databases. This deeper investigation will serve as the basis for a complete mapping of the social schemas used by the partner portals, in order to define and implement a mapping / transformation from their schema to the ODS one (Niemman et al., 2012). This task is not scheduled for the upcoming T8.2, but is rather important to develop an effective social metadata cloud for all repositories. Furthermore, we propose in section 6.2.3 below a possible solution to harmonise the different social metadata schemas.

3. A standard protocol to harvest social metadata
To enable a sustainable collection of social data from different portals, a protocol for the aggregation of the social data needs to be developed. Since social data can be considered as metadata, the use of a standard protocol for metadata harvesting such as OAI-PMH could be applied. An alternative could be to use a REST API to provide access to ODS social metadata since making calls to an HTTP API is significantly easier than making calls to a SOAP API. The latter requires a client library and a higher learning curve. The former is native to all programming languages and simply involves constructing an HTTP request with appropriate parameters appended to it.

4. Uniform resource identifier service
A challenge for the ODS social metadata cloud is a Uniform Resource Identifier (URI) resolution service that will check resources’ URI persistence. Since most of the data in ODS will be metadata that is harvested back and forth it is a challenge to identify a unique object. The URI service is a very important module due to the fact that the navigation and visualisation services will rely on the resource’s URI. This also affects the Linked Data objectives and is further discussed in section 6.3.1 below.

5. A modular infrastructure
In order to facilitate the integration of new services in ODS but also in other repositories, a modular approach for the ODS architecture should be followed. This will enable the implementation of some computational and memory intensive components such as navigation, visualization, recommendation algorithms in a distributed way. Thus, the ODS social metadata cloud should be powered by a cloud infrastructure that consists of a combination of different servers. In case one server would not be available, the social data is mirrored on other servers and could be provided from there. This is crucial for the reliability of the ODS social metadata cloud has it is intended to become the central repository for all schools in Europe and also as core service provider for other repositories.

6. Towards federated recommendation
The modular approach in the infrastructure also calls for services that will be operating on top of the aggregated social data in a horizontal and cross-repository manner. Such services, such as federated recommender systems (Zhou et al., 2012) would be of benefit/use both for the ODS social platform and all other portals in the network. This would require the development of such services in the context of WP9 as modular components that will be exposed through interoperable APIs and be made available to all ODS portals (instead of being specifically binded to the ODS social platform architecture).

7. Multilingual services
Finally, the infrastructure should store information about the language of the resource, tags and review. This will enable the provision of truly multilingual services to the users based on the social data. A complete solution to such a challenge would not be feasible in the context of ODS, still liaisons with relevant projects on multilingual online services (such as the CIP PSP Organic.Lingua) will be sought and exploited.

References


Paradata Specification v1.0, Retrieved from https://docs.google.com/document/d/1IrOYXd3S0FUwNozaEG5tM7Ki4_AZPrBn-pbyVUz-Bh0/edit

Appendix - Terms of Use (Draft)

○ Registration
  ■ Registration is needed.
    ● “To use the service you need to register”
  ■ Users are asked to provide true information on themselves
    ● “You agree to provide true and complete information”
  ■ Users need to keep their credentials confidentially.
    ● “You may not to provide access to your account to a third party”
    ● “You are responsible for keeping your password and access data protected against the access of third parties.”
  ■ Users need to update their password.
    ● “You have to change your password at regular intervals for security reasons. Passwords can be changed online at any time. If you suspect your password may be known to third parties, you have to change it immediately. In case you do not have access to your user account anymore retrieve a new passwort or contact our administrators.”
  ■ Operators are allowed to restrict access in case of suspicion of third party usage.
    ● “If the operators have reasonable grounds to suspect any unauthorized use of your account by a third party, the operators reserve the right to terminate your access. In this case, you will be provided with new access data.”

○ Right to Use
  ■ Users have the right to use the platform only for the intended purpose
    ● “You agree to use the service only for the intended purpose and not in any unusual way. The service may not be misused, e.g. for e-mail advertising material (spam).”
  ■ Users need to be able to represent themselves as legal persons
    ● “You may use the service only if you or your authorized representative can form a binding contract with us, and only in compliance with these terms and all applicable local, state, national, and international laws, rules and regulations.”
  ■ Users are not permitted to publish any content that is against law or accepted principles of morality.
    ● “You agree that you will not publish or create any content which constitutes a public nuisance or is contrary to accepted principles of morality. You agree that you will not create or publish any sexually objectionable, racist, extreme or inhuman content on the service.”
Users are not permitted to publish any harmful content like virus-infected content or spyware.
- “You agree that you will not publish any content which is infected by computer viruses or spyware.”

Operators are allowed to delete content that is against law or accepted ethic principles at the expenses of the author.
- “If you break this obligation, the operators reserve the right to remove the content without prior notification from the internet at your expense and withdraw your right to use the service.”

No exhaustive usage of the service
- “You must make sure that the infrastructure of the service is not overloaded due to excessive use.”

Operators do not guarantee 100% availability of the service.
- “In principle, the operators make the services available without any interruptions. Nevertheless the operators do not warrant that the service will be available all the time and uninterrupted. It cannot be excluded that delays and restrictions of services may occur due to events and conditions.”

Operators will back-up data once a day.
- “Back-up of the platform data is performed once a day.”

Operators reserve the right at its sole discretion to modify, or discontinue or terminate the service.
- “The design including the visual layout, structure and the scale of the functionality of the services are at the operators' sole discretion. The operator of the service reserves the right at its sole discretion to discontinue or terminate the service including individual functions of the service.”

Social Data Collection and Protection
- We collect data that you provide and use tools for storing them.
  - “We collect information that you voluntarily provide to become a registered user of the service, and when you use certain aspects of the service.”

Operators have to stick to the applicable data protection regulations.
- “The operators comply with the Data Protection Directives and all other applicable data protection law of the European Union.”

Operators have to protect personal data.
- “Personal data that you supply will not be given or otherwise disclosed to any third parties without authorization.”
To run an educational social portal successfully, a minimum amount of data has to be visible for other platform users.

- “As part of the service, it is necessary that a minimum amount of personal data such as e-mail address or name is visible to registered users.”

Personal data is not accessible by search engines.

- “As a rule, this data is not accessible through search engines.”

Users have the right to request the permanent deletion of their data

- “If you would like us to delete your record in our system, please contact us with a request that we delete your information from our database. We will use economically justifiable efforts to honor your request. We may retain an archived copy of your records as required by law.”

Operators have the right to log user information

- “When you browse or use the service, technological tools like Web Server Logs and Cookies collect certain information in a server log. The server log includes your Internet Protocol Address (by which computer are using may be uniquely identified), the remote host (name and IP address of the computer which requests the site), time of your request, status, volume of transferred data and the web site from which you visited the requested site (referrer) as well as product and version information of your browser.”

Operators have the right to use cookies and save it on your computer

- “After you have logged on, the service uses cookies to identify you for the time of your visit. A cookie is saved on your computer.”
  - Variation: session cookie (the cookie is deleted when you log off from the service).

Operators anonymize collected data for statistical evaluation

- “The operators anonymize the protocol data (logs) without identification or references to the user) to use it for statistic analyses, particularly for server load prediction.”

Operators are allowed to combine information in order to improve the service.

- “The operators may combine the information collected from you with information collected from other users to attempt to provide you with a better experience”

Phishing.

- “Identity theft and the practice currently known as “phishing” are of great concern to us. Safeguarding information to help protect you from identity theft is a top priority. We do not and will not, at any time, request your credit card information, your account ID, login password, or national
identification numbers in a non-secure or unsolicited e-mail or telephone communication.”

○ **Meta (about the terms of service)**
  - Modification of the Terms of Service
    - “We reserve the right to modify these Terms of Service at any time and without prior notice. In case we modify these Terms of Service we will inform you about the modifications made.”
  - In case of dispute the applicable law is $X
    - “In case of dispute the applicable law is $X”.