



# Supporting the discoverability of open educational resources

Renato Cortinovis<sup>1</sup>  · Alexander Mikroyannidis<sup>1</sup> · John Domingue<sup>1</sup> · Paul Mulholland<sup>1</sup> · Robert Farrow<sup>1</sup>

Received: 18 December 2018 / Accepted: 23 April 2019 / Published online: 30 April 2019  
© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

Open Educational Resources (OERs), now available in large numbers, have a considerable potential to improve many aspects of society, yet one of the factors limiting this positive impact is the difficulty to discover them. This study investigates and proposes strategies to better support educators in discovering OERs, mainly focusing on secondary education. The literature suggests that the effectiveness of existing search systems could be improved by supporting high-level and domain-oriented tasks. Hence a preliminary taxonomy of discovery-related tasks was developed, based on the analysis of the literature, interpreted through Information Foraging Theory. This taxonomy was empirically evaluated with a few experienced educators, to preliminary identify an interesting class of Query By Examples (QBE) expansion by similarity tasks, which avoids the need to decompose natural high-level tasks in a complex sequence of sub-tasks. Following the Design Science Research methodology, three prototypes to support as well as to refine those tasks were iteratively designed, implemented, and evaluated involving an increasing number of educators in usability oriented studies. The resulting high-level and domain-oriented blended search/recommendation strategy, transparently replicates Google searches in specialized networks, and identifies similar resources with a QBE strategy. It makes use of a domain-oriented similarity metric based on shared [schema.org/LRMI](https://schema.org/LRMI) alignments to educational frameworks, and clusters results in expandable classes of comparable degree of similarity. The summative evaluation shows that educators appreciate this exploratory-oriented strategy because – balancing similarity and diversity – it supports their high-level tasks, such as lesson planning and personalization of education.

**Keywords** Discoverability · Exploratory search · OER · [schema.org](https://schema.org) · Educational alignments

---

✉ Renato Cortinovis  
renato.cortinovis@open.ac.uk; rmcortinovis@gmail.com

<sup>1</sup> The Open University (UK), Walton Hall, Milton Keynes MK7 6AA, UK

# 1 Introduction: Background and motivation

## 1.1 The problem: A hidden treasure

Open Educational Resources (OERs) can be defined as any educational resource that can be freely used as well as repurposed (Atkins et al. 2007). Examples of OERs include interactive exercises, virtual laboratories, lesson plans, open textbooks or Massive Open Online Courses. In the last years, millions of potentially useful OERs have been developed and made available on the Internet. This huge number of educational resources openly available to educators, students, and self-learners all over the world, could have a large positive impact on society. UNESCO (2012, p. 1) mentions, for example, that OERs can foster “access to education at all levels, both formal and non-formal”, can “contribute to social inclusion, gender equity and special needs education”, and “improve both cost-efficiency and quality of teaching”. This is “increasingly being recognized as one of the most significant educational movements thus far in the 21st century” (Shear et al. 2015, p. 1). Yet, this enormous potential is far from being fully realized, for example because of the lack of awareness, reliable quality indicators, or even equitable access. Among many barriers, a frequently mentioned one is the challenge of discoverability: UNESCO (2012), LRMI (2013a), Barker and Campbell (2016a) among many others.

The challenge of OER discoverability can be understood in terms of several complex, interrelated aspects. In particular, there is a scarcity of quality metadata which describe these resources, and there are many incompatible standards to specify these metadata. As a result, there are a plethora of isolated search platforms, and users rarely wish to spend their time searching repositories individually. The large majority of educators looking for OER, therefore, make use of basic Google search (LRMI 2013b; Abeywardena et al. 2013). This is hardly surprising: search systems producing ranked results starting from simple keywords introduced in their textbox, have been so successful to frame our mental model of the Web (Schraefel 2009). Yet, despite its ubiquity, this popular search mechanism has severe limitations. Educators in particular lament that, when used to search for OER, it generates too many irrelevant hits, and it is too time consuming to be useful (Abeywardena et al. 2013). Yet, these search engines can be improved in many directions (Schraefel 2009). This research focuses on two of these: supporting discovery-oriented exploratory-search, and supporting users in their domain-oriented tasks.

## 1.2 Main issues and research gaps

This section briefly summarizes the fundamental aspects related to the challenge of OERs discoverability, highlighting the main research gaps.

### 1.2.1 Lookup versus discovery: From information retrieval to exploratory search

The term lookup, which is defined by Marchionini (2006, p. 42) as “the most basic kind of search task”, aims to produce very precise results starting from precisely formulated queries, and is related to the traditional field of “Information Retrieval”. On the contrary, the term “discoverability” emphasizes that the existence of the objects to be

discovered is not previously known, and is more related to the recent field of Exploratory Search (Marchionini 2006).

Researchers recognized the importance of shifting focus from lookup search, to supporting interactive search activities in a “more continuous exploratory process” (White et al. 2007, p. 2877). Indeed, many scenarios “require much more diverse searching strategies” from Google’s keywords-oriented search “elegant paradigm”, “including when the users are unfamiliar” with the domain, its terminology, or even with “the full detail of their task or goal” (Wilson et al. 2010, p. 9). The precise objectives of a discovery oriented activity, for example an educator searching for OERs to motivate his/her students, might not even be known in advance. On the contrary, precisely identifying these objectives is part of the activity itself. This situation calls for a far more articulated concept of search that is well beyond simple and ubiquitous keyword search, and represents an opportunity for improvement: as Wilson et al. (2010, p.4) claim, “there is substantial room for improving the support provided to users who are exhibiting more exploratory forms of search”.

### 1.2.2 From isolated search to high-level tasks in their broader context

In the context of the previous trend, search is not seen any more as an isolated activity, but as a sub-activity of wider high-level and domain-oriented tasks, dubbed “Work Context” (WC) tasks by Wilson et al. (2010). These tasks are recognizable “parts of a person’s duties towards his/her employer” (Byström and Hansen 2002, p. 242), such as organizing an in-depth educational activity, or planning a remediation activity. These tasks provide the context for lower-level, more context-independent tasks, such as finding resources related to a list of keywords. Qu and Furnas (2008, p. 534) argue that “there has been a paradigm shift in the design of search systems to support the larger task rather than” simply providing information matching the user-query keywords. Wilson et al. (2010) observe that traditional information retrieval tasks are the elementary steps to achieve higher level goals. Kabel et al. (2004, Section 2) claim that “the performance of the information retrieval task is inextricably bound to the work task”.

The application of such an approach in the case of open educational resources is to facilitate the exploration and discovery of OER. Educators and instructional designers regularly look online for a range of materials that they can use in their teaching activities. Many learners also regularly search online for resources that can help them. The central challenge, therefore, is to ground search and discovery in authentic educational workflow and activities.

### 1.2.3 From traditional metadata to linked data

The main traditional strategy to solving the problem of OER search (including discovery), consists in exploiting suitable metadata. Metadata are data describing meaningful educational characteristics of the resources, for example the educational audience, the type of educational resource, or its formal learning objectives.

Attempts to standardise these metadata have met with mixed success, resulting in a landscape of many incompatible standards (Riley 2010). An additional major challenge is the well-known unwillingness by authors to provide metadata (Doctorow 2001). Consequently, even more recent standards such as the IEEE (2002) Standard for

Learning Object Metadata (LOM), the Instructional Management Systems (IMS) standards (IMS 2015), and the Dublin Core Metadata Initiative (DCMI) Education Application Profile (Sutton and Mason 2001; DCMI 2012) could not fundamentally improve this situation (Barker and Campbell 2016a).

As discussed by Downes (2003) already, a single universal standard might not necessarily be the best solution – given that it is arguable even whether a resource is “educational” or not (particularly when it was not originally created for an educational purpose). Hence, alternative strategies attempted to design explicitly for diversity and fully support the heterogeneity of the Internet (Dietze et al. 2013). With this objective, many initiatives shifted their focus from traditional metadata towards semantic / Linked Data (LD) technologies (Al-Khalifa and Davis 2006) which emphasise that meaningfulness in metadata must be understood to be contextual.

Associating a formal semantic model that can be understood and processed by computers to the resources on the Web, is the grand vision of the Semantic Web (Berners-Lee et al. 2001). This is driving the evolution of the World Wide Web to a Global Giant Graph (Berners-Lee 2007) extending the Web of documents to a Web of meaningful interconnected Linked Data (Bizer et al. 2009). LD technologies make it possible to perform searches that are not simply based on keywords-matching but on their semantics. Nilsson (2010) claim that they could represent a possible solution to support the heterogeneity of the Web, facilitating the “harmonization” of many different vocabularies. However, existing datasets are typically quite isolated (D’Aquin et al. 2013).

#### 1.2.4 Current focus on [Schema.org](https://schema.org/)/LRMI

Following years of experimentation in the academic environment, Linked Data (LD) technologies have been adopted by major commercial search engines. Google in particular, has evolved its traditional search based on word statistics and structural links analysis, to a semantic search based on its knowledge base called “Knowledge Graph” (Singhal 2012). It is possible to contribute to this knowledge graph via “[schema.org](https://schema.org/)” (2013), an initiative launched by Google, Bing and Yahoo! in 2011, aiming at improving search results by providing a standardized simple mechanism to add semantics to Web documents. [Schema.org](https://schema.org/) defines an ontology to describe resources on the Web, which can be annotated by embedding metadata in Web pages.

[Schema.org](https://schema.org/) has potential to be widely adopted, because developers are motivated to use it knowing that major search engines recognize it, and because, by design, it is relatively easy to use, reducing one of the obstacles of traditional LD (Guha et al. 2016). Hence, it has been recently extended with the vocabulary developed by the Learning Resource Metadata Initiative (LRMI) aiming to support end-users in searching and discovering educational resources (LRMI 2014). The LRMI specification represents the latest attempt to describe educational resources, taking full advantage from previous experiences in metadata standards as well as LD. Dietze et al. (2017) contend, considering its increasing adoption (51% from 2014, 139% from 2015) despite the recent introduction, that it does provide potential to power search related applications.

Particularly relevant for this research is the so-called “killer” feature of LRMI (2013c): the alignment of a resource to a standard in an existing educational framework. This type of metadata can be used, for example, to express statements such as “this

educational resource teaches X”, where X is a specific learning objective (or competency standard) in an existing educational framework (Barker 2014). A notable example of an educational framework is the Common Core State Standards (CCSS) (Porter et al. 2011) in the United States of America, which defines detailed learning objectives in Maths and English at K-12 level. Indeed, such frameworks and descriptions are quite ubiquitous in the case of educational materials (such as textbooks) that have been written for a specific audience in formal education.

The importance of the alignment to educational frameworks is confirmed by the fact that a similar feature was already foreseen in previous metadata standards, such as LOM and IMS. Sutton (2008) argues that educators commonly use alignments of resources to standards to improve their efficiency in searching for resources, as well as to certify the compliance of their teaching activities to the standard curriculum. And yet, alignments are still rarely used or misused in [schema.org](http://schema.org) (Dietze et al. 2017), and more research is needed to fully exploit their potential (Barker and Campbell 2016b). In the case of “little” OER (Weller 2010) produced by educators in their own time, these alignments are often omitted altogether.

### 1.2.5 Blended search/recommendation systems under user control

Evidently, recommendation systems have an important role to play for the discovery of educational resources (Manouselis et al. 2011), complementing traditional search (lookup) systems by suggesting related resources. In the past decade, in parallel to the extensive efforts aiming to support users searching for (“pulling”) information in a wider domain-oriented context, there have been considerable efforts on developing various types of recommendation systems suggesting (“pushing”) personalized items of potential interest to users (Dietze et al. 2014). More recently, recommendation systems are increasingly seen as a fundamental component of modern interactive search systems: a new area of research is in blending these technologies, so that search engines become more personalized, and recommendation systems increasingly search-like and under user control (Chi 2015). Of course, the algorithms successfully used in the commercial domain must be adapted to the pedagogical domain, where the relationships must be based on pedagogical aspects (Verbert et al. 2011).

### 1.2.6 Exploratory search evaluation

In the field of Information Retrieval there are well established metrics that can be used to evaluate search systems. These are based on the traditional Cranfield model (Cleverdon 1960), which predefines the corpus of resources, the collection of queries, and the collection of relevance assessments. Yet, in the case of Exploratory Search, queries might not even be precisely known in advance, hence traditional metrics such as precision (fraction of relevant instances among the results), recall (fraction of relevant instances among the relevant ones in the whole dataset), completion time, or number of errors, are no longer appropriate because they are based on a precise pre-classification of resources as relevant and non-relevant. This makes evaluation of exploratory systems a research area on its own (Kules and Shneiderman 2008; Wilson et al. 2010), that needs to focus on users and their context.

Wildemuth and Freund (2009) argue that the evaluation of exploratory search systems needs to focus on tasks. Belkin (1995, p. 4) also claims that “evaluation begins with studies of users in their tasks, in order to identify the criteria which they apply in evaluating success”. Hence, an essential step to evaluate solutions aiming to support OER discoverability, is to identify the user and domain-oriented tasks that need to be supported. This makes the evaluation of Exploratory Search systems strongly related to the research area of usability evaluation (Madan and Dubey 2012). The term usability is intended here to mean “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use”, as reported in International Standards Organization standard 9241–11 (ISO 1998). Usability is therefore firmly grounded in the domain-oriented user-tasks that the applications are supposed to support.

### 1.2.7 Synthesis of main issues

In summary, there is an opportunity to complement the ubiquitous keyword-oriented search with more exploratory-oriented solutions, better suited for ill-defined problems. Search systems should address high-level user’s tasks and their specific context. They should integrate pro-active features of recommenders, suitably adapted to the pedagogical domain. Linked Data technologies, and more recently [schema.org/LRMI](https://schema.org/LRMI), are considered a potential solution to overcome some limitations of traditional metadata. Its educational alignments, in particular, are considered a “killer feature” but are still relatively unexplored. Finally, the evaluation of exploratory search systems requires a new focus on users and their tasks, with usability oriented studies.

### 1.3 Goal and research question

Given the importance of the discoverability of OERs, there have been, and there are, many attempts to address it. Relevant initiatives include hundreds OER repositories with search facilities, federations of repositories, even federations of federations (Globe 2016). Developing yet another OER search portal or engine would worsen, fragmenting it further, the current situation. Indeed, educators are not willing to hop from one platform to the other, and end-up by using the inadequate standard facilities of Google. Hence, the goal of this research was rather to propose innovative solutions to be integrated in existing and future search applications. The first aim was to identify new requirements to support educators looking for educational resources, by analysing their domain-oriented tasks and their relative importance. The second – fundamental – aim was to suggest suitable strategies to satisfy the requirements identified. The overall research question was therefore twofold:

What are the main tasks associated with OER discovery, and how can educators be supported in performing (some of) these tasks?

The anticipated outcome of this research – strongly focused on users and their tasks – was to improve the effectiveness of existing and future OER search platforms, improving the discoverability of OER, which was identified as one of the obstacles to reap the benefits of the OER movement. The present study focuses on secondary education, but

also suggests some basic extrapolations which are relevant to other education levels and OER scenarios (such as non-formal learners).

## 2 Research methodology: Enhanced DSR

Following the identification of the problem, and the research gaps identified from the literature, the objective of the research was twofold: identify (1) high-level and domain-oriented educators' tasks that need to be supported, and (2) suitable strategies to support them. The research targeted these objectives mainly by designing and experimenting with software prototypes, hence it followed the Design Science Research (DSR) methodology (Hevner et al. 2004), as illustrated in Fig. 1. This figure extends Fig. 3 by Vaishnavi and Kuechler (2015), by articulating the path to the first DSR cycle. This path represents the preluding activities leading to the DSR iterations, when these are

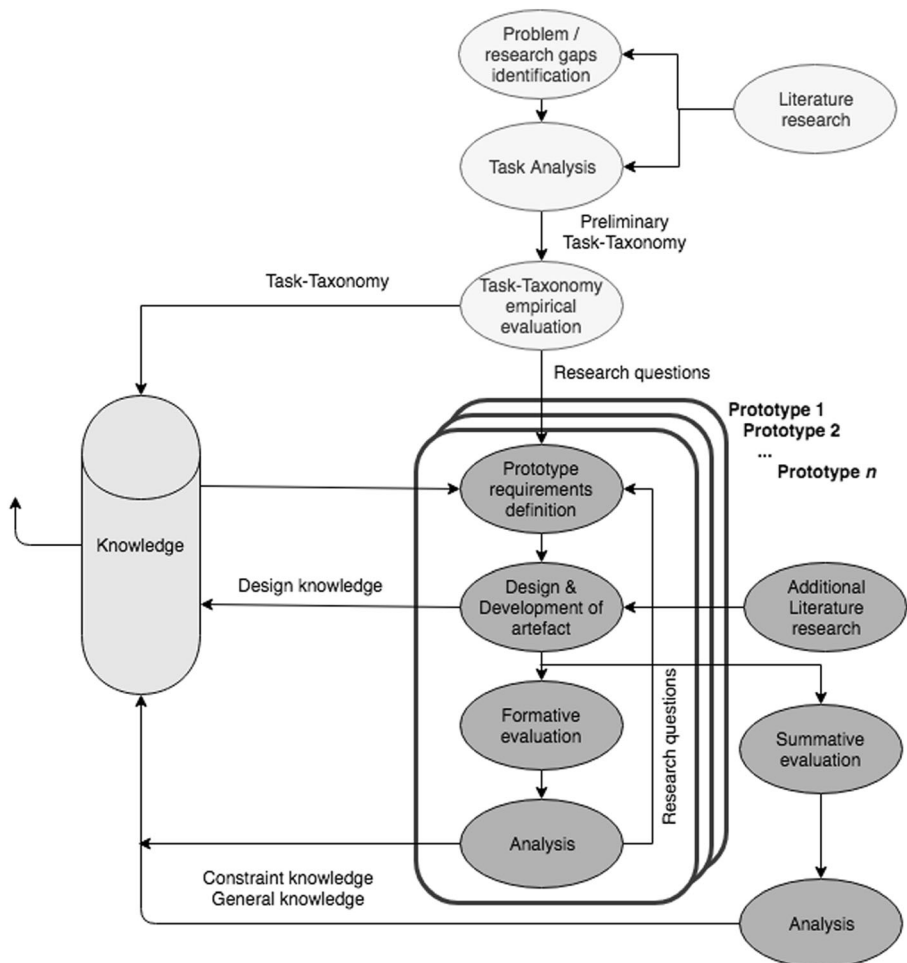


Fig. 1 The adopted research methodology



driven by a specific problem to be solved (Peffer et al. 2007). Accordingly, a task analysis attempted a preliminary identification of requirements and challenges. Its evaluation provided the initial input to the iterative development and evaluation of a sequence of prototypes: Injector, RepExp, and Discoverer. These prototypes, represented in the figure by stacked round rectangles, made it possible to experiment with new solutions, confirming, refining, or extending the requirements preliminarily identified, discovering further challenges and research questions, and generating new knowledge of the requirements of an effective design.

## 2.1 Preliminary task analysis and related empirical evaluation

A preliminary identification of the educators' tasks, that OER search/discovery applications should support, was based on a review of the research literature and existing OERs search portals, interpreted through Information Foraging Theory (IFT). IFT was developed by Piroli and Card (1999) noticing similarities between the behaviour of users ("informavores") looking for information, and foragers hunting for food. This behavioural model is widely used to explain and predict users' behaviour in different information search circumstances. This model-driven task analysis was meant to produce a preliminary but well-founded general domain-oriented taxonomy open to modifications and extensions.

A first empirical study collected quantitative and qualitative data about the taxonomy, to obtain a preliminary understanding of priorities, habits, as well as thinking strategies of educators when looking for educational resources. The data collection was carried out through a detailed survey and follow-up interviews, submitted to a small sample of experienced teachers in secondary education. The survey collected quantitative data, first with the objective to engage respondents in critical thinking, in order to elicit highly valued qualitative information. Hence, respondents were asked to rate first the importance of tasks and categories identified by the previous task analysis, with single-item constant-sum questions (CSQs), allocating a total of 100 points to groups of related tasks or categories. The main disadvantage of CSQs is the high cognitive load imposed on respondents (Sue and Ritter 2007). However, by removing the simplistic possibility to score every item as "very important" as on standard rating scales, CSQs forced respondents to reflect on the precise relative importance of every category and task, increasing discrimination power and engaging them in critical thinking (Timpany 2015). It is only following this activity that respondents were invited, with open questions following each CSQ, to offer suggestions for additional tasks / categories, their modification, reorganization, or any other comment. A final section included a question on the overall perceived completeness of the OER task-taxonomy with a 7 point Likert scale, a few open questions to collect additional qualitative feedback on possible tasks not covered by the OER task-taxonomy, as well as any additional comments considered relevant.

The bulk of questions collecting quantitative and qualitative data for the various task categories, were preceded by basic demographic and general questions related to country, experience, subject, and level of teaching, search portals employed and frequency of use. The whole questionnaire was prepended by an introduction with goals, background information about the task analysis, instructions, privacy, data management, and optional contact information. While no sensitive data were collected



from participants, they were anonymized once the necessary clarifications could be obtained.

The questionnaire was implemented as a Web application, by extending Google Forms to support CSQ type of questions, in order to have the possibility to collect anonymous feedback.

The reliability of the data collected was checked by a hidden redundant question. Additionally, outliers in the scores obtained, were double-checked in follow-up interviews, in order to eliminate mistakes and fully understand motivations. Every comment collected via open questions was followed-up, to understand the underlying motivations and to elicit additional information.

A triangulated quantitative/qualitative analysis was carried out. Quantitative data were mainly analysed with non-parametric techniques suitable for small samples, especially when there were doubts about their normal distribution. The relative importance of different tasks, for example, was investigated by charting basic descriptive statistics, and by using the non-parametric Wilcoxon Signed Rank Test. The qualitative data collected via open questions in the survey, as well as in structured and follow-up interviews, were analysed with qualitative content analysis (Cho and Lee 2014). Following an inductive approach, the text was first subdivided in to sections expressing single concepts. Mutually exclusive categories were iteratively extracted from the key concepts, and organized hierarchically. Finally, the original extracts were encoded with the categories extracted.

## **2.2 Prototype design & evaluation iterations**

The preliminary identification of requirements in the previous study, was followed by a series of DSR design and evaluation cycles.

A prototype was developed in each iteration. The development of prototypes already fostered a better understanding of the initial ideas (Winston 1984), but – most important – each prototype was evaluated, with the objective to identify possible shortcomings, as well as to improve the preliminary understanding of the requirements, prune the solution space, and generate new research questions. These were used to drive the design and evaluation of an enhanced prototype in the subsequent iteration.

In each cycle evaluation, a small group of participants were encouraged to use the prototype. Specific search tasks were suggested in the first cycle, but they were increasingly unconstrained in the following more realistic prototypes. Participants were exposed to the prototypes for a variable amount of time according to their needs and interest, ranging from 20 min for the first prototype, to 2 h for the last one. They were encouraged to think aloud, and allowed to request clarifications or discuss any concern with the researcher. When a direct interaction with the evaluators was not possible, the activity was carried out remotely. In these cases, a representative demo screencast was provided, in addition to a remote demonstration and the possibility to interact with the researcher. Following this activity, evidence about the relevance of the needs preliminarily identified and the suitability of the proposed solution was collected through questionnaires, administered as survey and structured interviews, and field notes resulting from the observation of test-users.

Participants could not be expected to have expertise in evaluation. Hence they were supported by heuristics: specialized evaluation knowledge in the form of check-lists, derived and adapted from evaluation heuristics available in the literature. These include the widely adopted System Usability Scale (Brooke 1986), the heuristics by Molich and Nielsen (1990), and by Gerhardt-Powals (1996). The heuristics addressed usability and user experience, integrated with more specific aspects related to the functionalities supported by the various prototypes (McNamara and Kirakowski 2006).

The formative evaluations in the first two cycles, to maximize efficiency, involved the minimum number of participants sufficient to reach saturation, when major shortcomings to be addressed in the following cycle were clearly identified. It was considered convenient to have multiple inexpensive formative evaluations, that Nielsen (1995) calls “discounted”, distributed along the design process. Using more samples than needed in the early studies would have been an unwise waste of resources, which were better spent in subsequent design iterations, to improve the overall quality of the activity. Given the exploratory nature of the activity and the limited number of test-users, the primary form of analysis for user feedback was qualitative content analysis.

Once the design had stabilized, a summative evaluation with a larger number of participants was carried out with the last prototype. The aim of this evaluation was not so much to identify weaknesses and areas for improvements as in previous iterations, but was more oriented to further confirm the relevance of the addressed tasks / scenarios, and gathering additional supporting evidence about the effectiveness of the discovery solution proposed. In this case, given the more realistic implementation of the prototype, participants were encouraged to freely use it to carry out unconstrained search tasks of their interest, increasing the ecological validity of the evaluation (Prat et al. 2014). The larger number of participants involved, made it possible to carry out a triangulated quantitative and qualitative analysis.

The questionnaire used in the previous evaluation, to be administered again as a structured interview or survey, was slightly adapted to the new goals, improved, and translated also to Italian, letting participants respond in English, Italian, French, or Spanish.

Quantitative data were collected mainly via Likert-type scales and were treated as ordinal data. Even considering the data as ordinal, the mean was considered an appropriate measure of central tendency, as data did not contain outliers. The measure of dispersion adopted was the range across quartiles (IQR). Results were analysed across different profiles, attempting for example to identify possible differences among educators teaching different subjects, or having a different experience in the use of OERs. To this scope, the non-parametric independent-samples Kruskal-Wallis test was used. The effect size was estimated through correlation analysis with Kendall’s tau. Finally, the correlation among Likert-type scale variables was analysed with the Spearman’s rank-order test.

Qualitative data were collected via open questions in the questionnaire, structured interviews, and observation notes while participants were using the system. Data were processed with qualitative content analysis, using a simple Computer Assisted Qualitative Data Analysis Software. The frequency of the resulting codes was used as a measure of the importance of corresponding concept. Some of the concepts were then correlated to educators’ profile characteristics.

### 3 Results & discussion

#### 3.1 Preliminary task analysis

The initial review of the research literature and existing applications, interpreted through IFT, produced a preliminary taxonomy of educators' tasks that OER search/discovery applications should support (Fig. 2).

Tasks and categories were derived from the analysis of the functionalities of existing systems, and from task models, use cases (including from LRMI), and search-tasks used in evaluation studies, available in the literature. The top level categories, for example, are the four “themes” adopted by Atenas and Havemann (2013) in their evaluation of OERs portals – Search, Share, Reuse, and Collaborate – merged with the similar “steps” of the OER life cycle discussed by Gurell and Wiley (2008) in the OER Handbook, Find, Compose, Adapt, Use, and Share. The top-level categories were iteratively specialized in subcategories as much domain-oriented as possible. Notably, the category “Using” (denoting the use of OER for teaching) was specialized in the subcategories “Lesson planning” and “Lesson delivery”, which should be the ultimate reasons for educators to use OERs, hence a target of this research.

IFT was used to suggest user tasks by transposing foragers' behaviour to corresponding informavores' behaviour. For example, the forager behaviour “discover potentially interesting prey following the footprints of other foragers”, can be transposed to the informavore behaviour: “discover potentially interesting resources by identifying those previously used by other users”. Inversely, IFT helped validating the soundness of previously identified user tasks, by considering the corresponding foragers behaviour.

Many IFT tasks exploit similarities based on examples, which correspond to the key sub-category Expansion. This category includes tasks to discover resources, related to a sample resource previously identified, by a relatedness or proximity metric, such as likedness (liked by the same users who liked the current resource) or togetherness (used

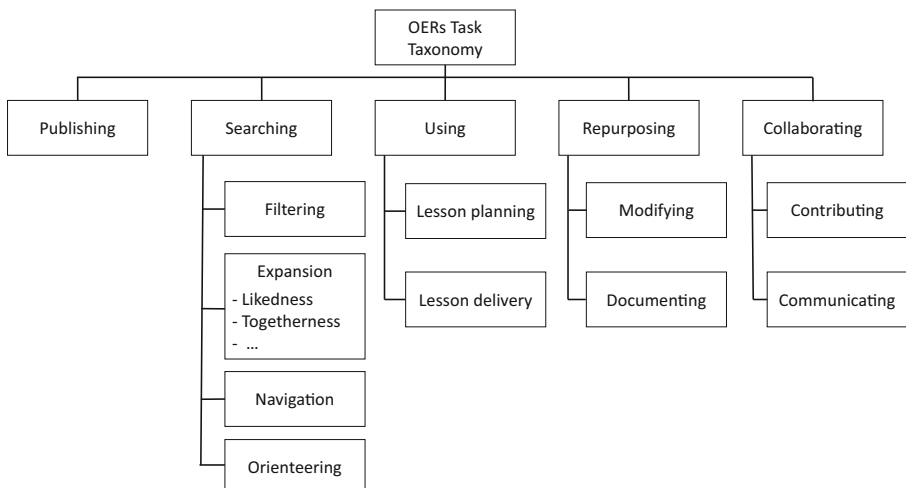


Fig. 2 Preliminary domain-oriented task-taxonomy

together). The corresponding queries can be seen as Query By Examples (QBE), where the example is the resource this process starts from. This is the fundamental class of exploratory-search and discovery-oriented tasks, based on relationships among resources (Knoth 2015). This sub-category is particularly promising (Wilson et al. 2010), blending discovery oriented exploratory search, query by examples, and recommendation features under full user control – in line with latest trends in search systems (Chi 2015).

Filtering refers to the frequently available functionalities allowing users to specialize (or generalize) search results by adding (or removing) constraints (filters). Filtering could, for example, restrict current results to OERs targeting students with a given age range.

### 3.2 Task analysis empirical evaluation – Main results

The OER task-taxonomy previously developed was empirically evaluated, to gain a preliminary understanding of priorities, possible novel tasks, and initial requirements, to drive the first DSR cycle.

Nine educators participated in the survey. Six had more than 20 years of teaching experience; eight had experience at secondary level, seven in Italy, while two in the UK and at international level; six searched for OERs more than 20 times per year. Five participants opted for the email solution, two for a face to face supported structured interview, two for a telephone supported structured interview; none selected the Web-based solution.

To support the reliability of results, the participants were all experienced educators, but as an additional check, two different variables measured the same construct – the importance of adding non authoritative metadata – in different contexts. The Spearman's correlation coefficient indicated a strong positive monotonic relationships ( $r_s = 0.674$ ) with a high significance level ( $p = 0.047$ ) – showing high internal consistency.

The survey collected detailed weights indicating the relative importance of each category and task. Figure 3 shows the weights for the top-level categories, and the sub-category Searching.

Considering the suggestion in the literature for further research in expansion operations, a comparative analysis attempted to identify whether participants attributed importance to expansion in addition to the ubiquitous filtering. The mean difference is about 17%, identified as significant by a Wilcoxon Signed Rank test ( $p = 0.013$ ). However, the higher importance attributed to filtering was mainly due to the higher

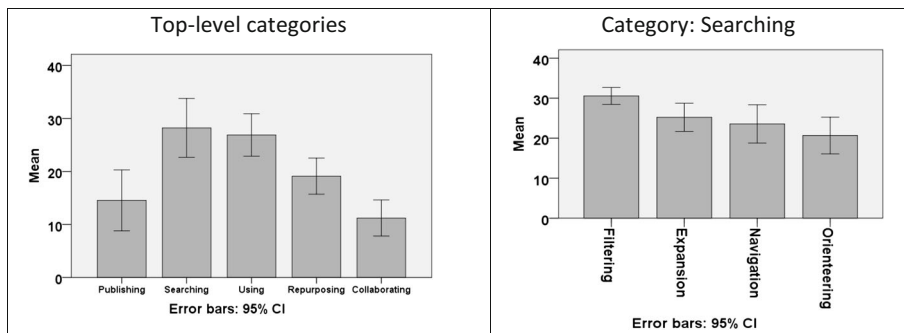


Fig. 3 Weights with 95% confidence intervals

familiarity of users with filtering, and to the consideration that expansion is generally to be used following filtering:

“I am more familiar with filtering conceptually, but I fully recognize the importance of having the possibility of expanding to similar resources”.

“I think you have to filter first, then, once you have found something, you may expand your search”.

The overall OER task-taxonomy completeness, measured on a Likert scale anchored from 1 (very low) to 7 (very high), obtained a mean of 6.8. The three scores that were not the maximum possible, were followed-up: the respondents motivated their score with their own lack of confidence due to limited personal knowledge, but could not pinpoint any shortcoming in the analysis. However, five respondents, while asking for clarifications and during post interviews, suggested to include additional “expansions”, such as on same topics, same educational standards, and even same authors.

Qualitative data showed that participants were not too keen to explicitly use educational standards, rather preferring subject taxonomies, even if a Paired Samples T-Test failed to indicate a statistically significant difference. Here too, the main reason reported was higher familiarity of educators with ubiquitous topics, compared to educational standards:

“The possibility to target a specific educational standard is quite interesting in principle. But we don’t use formal educational standards!”

Yet, participants considered educational alignments very useful to precisely target educational resources, for example:

“I feel I am more in control by using topics taxonomies, but educational alignments would allow a more precise targeting”

“I think that filtering by educational alignments could be very powerful [...]”.

Studies (e.g. LRMI [2013b](#)), have suggested that Google would be the search engine most frequently used to search for OER. Google proved to be the search engine used by every educator in the sample to look for OERs, even if everybody lamented its limitations for this particular task: irrelevant results, uninformative snippets, etc. Despite this awareness, however, just one educator out of nine complemented its use with other specific OER search engines.

The study generated other useful results, which are not described in this paper. One example is the weighting ascribed for each task and category by educators which could be used as a metric for the analytic evaluation of OER search portals (Agarwal and Venkatesh [2002](#)).

### **3.3 Task analysis empirical evaluation – Discussion**

The analysis of respondents’ feedback supports the proposed OER task-taxonomy, as demonstrated by the high rating on the overall completeness scale, the positive final comments, and the lack of suggestions for modifications. These results are fully in line

with the research literature, as the taxonomy proposed represents by construction a synthesis of the research community's understanding.

Participants were mainly interested in searching and using the resources – confirming the importance of high-level domain-oriented tasks. However, results indicated that respondents included in the expansion category, tasks that could be alternatively carried out with a suitable combination of operations already foreseen in the taxonomy proposed. For example, the task “find the resources aligned to the same educational standards of a given resource”, could be carried out by the following sequence of lower level tasks:

1. select a resource;
2. get the educational standards the resource is aligned to;
3. for each standard X, get the resources aligned to X;
4. rank the resulting resources.

While these expansion tasks could be unwisely dismissed because technically redundant, they represent useful short-cut operations that are close to the natural task-oriented thinking strategy of educators. Forcing educators to decompose these “natural” tasks in sub-tasks, obliges them to think in procedural terms and to take into account complex underlying data structures, imposing an unnecessary cognitive overload. This was a key finding of this first study, in line with the need to identify domain and user oriented tasks to advance the research in the field, advocated in the scientific literature (Marchionini 2006), (Wilson et al. 2010). These unforeseen tasks became the focus of the following Design Science Research iterations.

While filtering emerged as the most important operation, expansion was considered highly important too, especially considering the higher familiarity of educators with filtering, and its temporal precedence over expansion. Marchionini (2006) indeed argues that Information Retrieval and filtering oriented operations, mainly serve the purpose to bring searchers in a position from where to start exploratory search, that is from where more discovery (expansion style) oriented functionalities can be used.

Respondents expressed a preference for the use of more familiar topic taxonomies. Yet qualitative data indicated that they fully realized the advantage to target precisely resources with standard alignments. More precisely, they considered that educational standards would be appropriate for domain-oriented expansion operations, provided that users were not required to operate on them explicitly.

Figure 4 summarizes the findings of this study, which drove the design of the first prototype. It is worthwhile remarking that while these preliminary insights, derived from the empirical evaluation of the task-taxonomy, were used as input to the first prototype, their refined formulation, their relevance for educators, as well as the strategy to support them, were going to be further tested and improved or refined over the following DSR cycles.

### 3.4 Injector

The prototype Injector, developed in the first design & evaluation DSR cycle, identified educational resources and injected related educational metadata as well as expansion/discovery functionalities, in the original Google Search Engine Results Pages (SERPs).

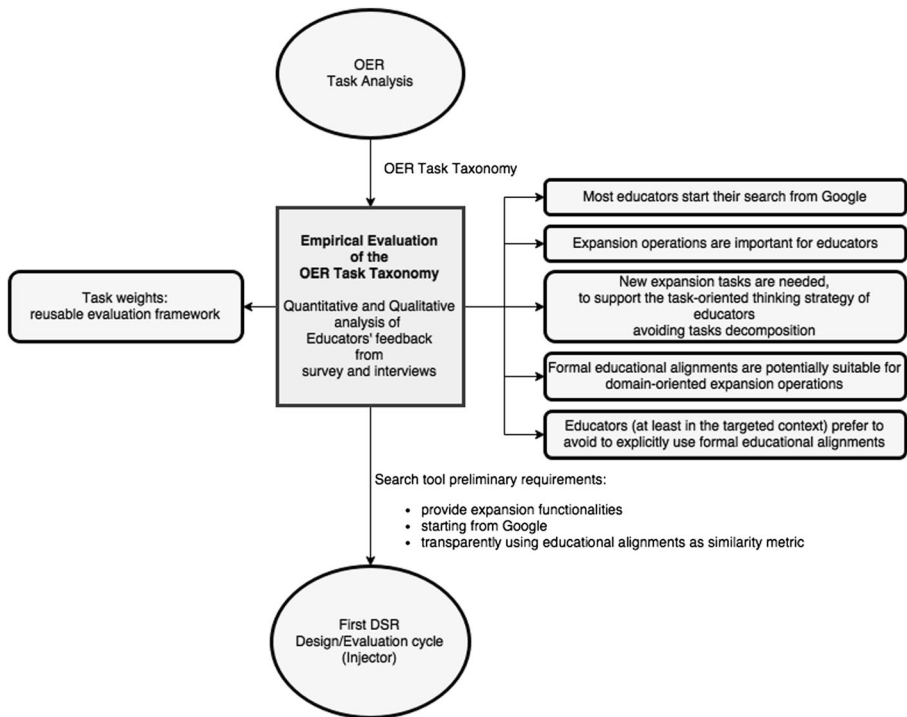


Fig. 4 OER Task-taxonomy Empirical Evaluation: an overall map

The development of Injector was driven by requirements and insights that emerged from the task analysis. Expansion by similarity was indicated in the literature as fundamental in exploratory-search, and its importance was confirmed by the IFT. Additionally, the task-analysis empirical evaluation suggested that expansions could conveniently support task-oriented thinking strategies of educators. The empirical evaluation also showed that expansion was considered somewhat less important than filtering, but this was mainly due to its lower familiarity, and its temporal precedence. While filtering is already widely investigated and commonly available in existing search platforms, expansion represents a significant research gap, hence it was decided to focus on expansion. Accordingly, Injector identified and ranked similar resources via a novel similarity metric, defined as the number of educational alignments that  $Res_i$  and  $Res_j$  have in common:

$$\text{Similarity}(Res_i, Res_j) =_{\text{def}} \frac{|\{\text{educational alignments of } Res_i\} \cap \{\text{educational alignments of } Res_j\}|}{|\{\text{educational alignments of } Res_i\}| + |\{\text{educational alignments of } Res_j\}|}$$

Coherently with the results of the task-analysis, this domain-oriented metric exploited the acknowledged power of educational alignments, without requiring users to be directly aware of them.

Considering the habits and preferences of educators, Injector identified the educational resources directly in Google SERPs, by parsing them and looking for corresponding entries in the Learning Registry (2016), a large repository of learning resources' metadata. This way, it could then retrieve domain-oriented educational metadata from the Learning Registry and enrich the original Google SERP, by injecting



the metadata in the snippets corresponding to the identified resources. Similar basic techniques were previously demonstrated by the “Browser Plugin” (Lockley 2011) and “AMPS” (Klo 2011) prototypes. These prototypes, which are no longer supported, identified educational resources in Google SERPs too, but they only injected static descriptive metadata. The distinctive feature of Injector was to inject expansion/discovery functionalities, that is, active links to additional similar resources.

Figure 5 reports a representative screenshot of the prototype, which shows a window with the enriched (highlighted) Google SERP, overlapped by a window with the similar (expanded) resources.

Injector was evaluated in this first cycle with a discounted heuristic evaluation, by a small sample of four secondary education teachers, from Italy and the UK. The evaluation questionnaire was administered as a structured interview in two cases, and as a self-administered survey in the two other.

The results of this evaluation were very coherent with the findings from the empirical evaluation of the task-taxonomy. For example the discovery oriented functionality based on educational alignments was unanimously considered very useful, with mean 7 on a scale from 1 (totally useless) to 7 (very useful). The “transparency” of the tool, exploiting educational alignments without the need to manipulate them explicitly, obtained a mean of 6.8 on a scale from 1 (totally useless) to 7 (very useful). A participant reported:

“The transparent use of educational alignments is very much appreciated, as I am not familiar with existing standards”.

The relevance / similarity of the suggested resources was also considered very positively, obtaining mean 6.5 on a scale from 1 (very weak) to 7 (very strong).

The general comments were very positive:

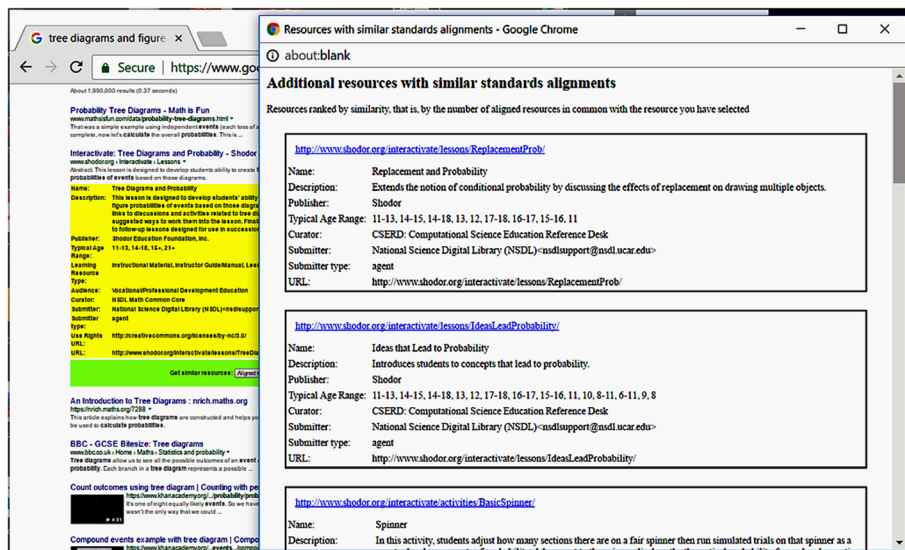


Fig. 5 Injector: enriched Google SERP and expanded resources

“I think this is the potentially perfect ‘all-in-one’ instrument for us educators”,  
 “Opens up, literally, a whole new dimension in knowledge and content searching”,  
 “Educational metadata are much more useful than the traditional snippets provided by Google”.

However, test-users consistently lamented the modest number of educational resources it could identify in Google SERPs (intrinsic sparsity), and the modest number of resources that could be actually expanded (alignments sparsity). Intrinsic sparsity is a structural limitation of Google search since Google results contain “false positives”: heterogeneous resources that are related just because they share some search keywords, and hence contain many items which are not educational resources.

Participants did appreciate the possibility to start directly from Google SERPs, yet two of them expressed the wish to see results restricted to exclusively educational resources:

“I would prefer to see just the educational resources in [Google] results pages”,  
 “It would be better if it could offer only educational resources”.

This concern was addressed in the following activity, by retaining the strategy to start the search from a standard Google SERP, yet relaxing the constraint of identifying the scarce number of educational resources exclusively among the original Google results.

At this stage, there was no more need to engage further participants: sparsity was the major challenge to address in the next prototype.

### 3.5 RepExp

A new prototype, RepExp, was developed in the second DSR cycle, to address the challenges identified in the previous cycle: intrinsic and alignments sparsity. In order to reduce intrinsic sparsity, this new prototype transparently replicated the initial Google search, using the same search keywords automatically intercepted from a user’s Google query, in the Learning Registry again. Hence, it returned a custom SERP with solely educational resources. In the background frame in Fig. 6, for example, obtained by replicating a Google query with the keyword “biology”, 399 educational resources were identified. In order to reduce alignments sparsity, RepExp restricted results in its SERP to educational resources *having* educational alignments: this way any resource included is always expandable. Expanding the first resource in the previous SERP, for example, produced the foreground frame in Fig. 6, where we can notice that 1345 similar educational resources were identified. This is incomparable to what is possible to achieve with a traditional Google search. Furthermore, as additional flexibility, similar educational resources could be identified also starting from any resource being explored, while navigating on the Internet.

The prototype was evaluated with a heuristic evaluation, this time with a sample of six educators, four males and two females, five teaching at level of secondary education, one at tertiary level, four from Italy, one from the UK, and one from Brazil, teaching technical or scientific subjects. Three educators had more than 11 years of

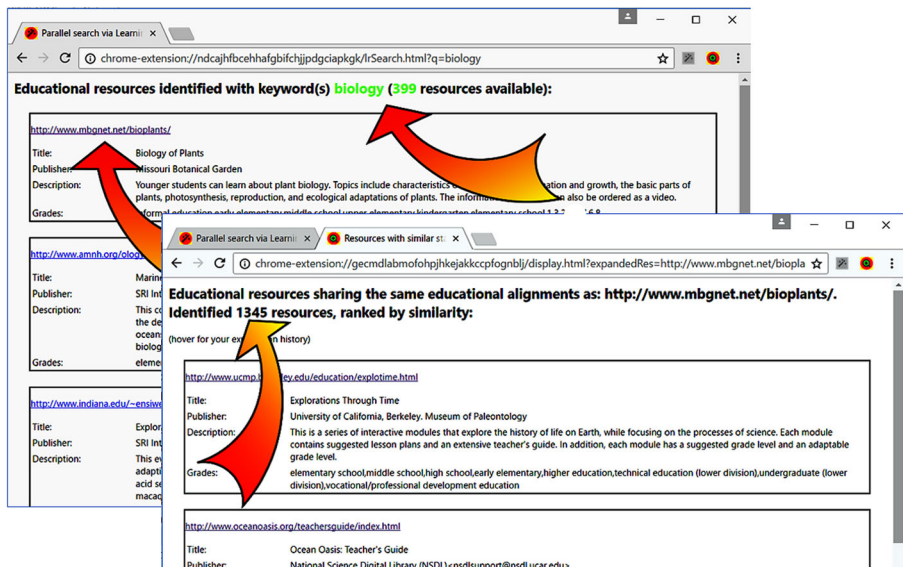


Fig. 6 RepExp: replicated SERP and expanded SERP

experience, two had more than 21 years. Four of them were selected opportunistically among educators personally known, two of them were recruited with a snowball strategy. Six participants were sufficient to identify the new challenges to be addressed in the following DSR cycle.

The evaluation questionnaire was administered to three educators as a remote survey, supported by email and Skype, and as structured interview to the remaining three educators. In this last case, field notes were annotated directly on the survey and approved by interviewees.

The results of this evaluation were very consistent with the results of the previous studies. First, the strategy to offer discovery functionalities directly from Google pages, in comparison with the alternative to use specialized portals, was again very much appreciated, obtaining a mean of 1.2 measured on a Likert Scale anchored from 1 (much better) to 7 (much worst). Explanations were quite emphatic:

“I always use Google, I ignore other portals”

“Most of us educators start, and stop, in Google”.

Concerning User Experience, participants expressed, as in the previous evaluation, willingness to use the system frequently: mean 6.7 on a scale from 1 (fully disagree) to 7 (fully agree). One of them even indicated:

“I would like to use such a system not only for my work as educator, but for self-development activities too”.

Participants also appreciated the solution to use a similarity metric based on educational alignments, in comparison to the use of keywords, obtaining a mean of 1.3 on a scale

from 1 (much better) to 7 (much worst). Motivations explicitly mentioned, in five over six cases, its strongly domain oriented characteristics:

“Precisely focused on the educational domain”,  
“Very appropriate in education”.

Instrumental to the objectives of this formative evaluation, participants expressed some critical observations about core features of the prototype, which pinpointed areas of concerns and potential improvements, to be addressed in the next prototype. The first important concern was related to the degree of similarity of the presented resources:

“Resources should not be too similar; that is, they should be somewhat similar, but not equals”.

Indeed, the prototype ranked and presented the identified resources in order of similarity, starting with the most similar. Consequently, when there were many similar resources available, the first few resources presented were characterized by the highest degree of similarity. These first few resources were usually the only ones examined by participants. The resulting unintentional effect, was that the prototype frequently ended-up by presenting to test-users exclusively resources, which in some cases could be too similar to be useful. Indeed, Smyth and McClave (2001, p. 348) claim that the “standard pure similarity-based retrieval strategy is flawed in some application domains”, and “recommenders are often faulted for the limited diversity of their recommendations” (p. 360). RepExp indeed, can be considered a recommender (under user control) that suffers from the limited diversity of its recommendations: in addition to similarity, it should take into account diversity too.

Another participant pointed out that:

“Maximum similarity is not necessarily what one looks for in every opportunity”.

Indeed, maximum similarity is not necessarily the best solution to maximize utility. Consistently with the goal of supporting users in their high-level tasks, it is necessary to consider more precisely for what purpose, an educator might need to look for similar educational resources. An educator searching for educational resources to be used in a remediation activity, would require resources with a high degree of similarity (in terms of learning objectives) with the resources previously used in the main classroom activity. Indeed, the goal in this case is to offer students another chance to achieve the same learning objectives that could not be achieved before. On the contrary, an educator looking for educational resources for in-depth activities, would need resources with a lower degree of similarity, that is with a more limited overlap of learning objectives.

Another class of remarks concerned the difficulty to make sense of the large number of results produced by this new prototype. While in the previous prototype, Injector, the major concern was about sparsity, that is the *limited* number of resources identified, here participants expressed concerns for the opposite reason, that is, for the excessively *large* number of resources identified:

“Sometimes there are too many [resources]”,

Identifying a large number of resources was a direct goal of this second prototype, which was indeed successful in this regard: why is this now seen as a concern? While the identification of a large volume of resources is a positive aspect indeed, it uncovered a new challenge that was previously masked: how to make sense of large result-sets. This was explicitly revealed by the following remark:

“It would be useful to get a quicker global picture of the available similar resources”.

Finally, a participant noted that a sequence of expansions (requesting resources similar to the current one of interest, selecting one of these, and then repeating the process multiple times), after a while, was repeatedly producing mostly the same results:

“If we keep expanding, we end up getting the same resources over and over”.

Indeed, the repetition of resources following repeated expansions, is just the visible symptom of a larger problem, which we dubbed “*lock-in*”. When users select a resource from a group of very similar ones and expand it, they tend to obtain again the same group of resources they started from. This makes it difficult or impossible for users to navigate from the original group of resources to other groups (“re-patching”, in terms of the patch model in IFT).

The problem of “*lock-in*” can be explained in terms of the characteristics of the similarity relationship adopted, and the strategy of ranking resources by similarity. The relation of highest similarity is an approximate equivalence relation, hence it partitions the set of resources in approximate equivalence classes. Consequently, the most similar resources belonging to the same class, are the only ones that keep showing up in repeated expansions.

The identified challenges of sense-making, *lock-in*, and the need to provide users with some control on the degree of similarity, were addressed in the subsequent prototype.

### 3.6 Discoverer

A third prototype, Discoverer, was developed in the third DSR cycle. Its new additional key feature was to present users with a representative set of resources, grouped in three expandable clusters of different degrees of similarity. This way, educators could quickly make sense of the usually large set of resources available, supporting sense-making and reducing information overload. Additionally, users could then select, explore, and iteratively expand resources of the desired similarity, at will, eliminating the problem of *lock-in* and further supporting an exploratory-search oriented approach.

The prototype was also improved in other aspects, such as displaying additional metadata to flag OERs, and including a thumbnail image for each resource to provide a more visually attractive presentation as well as to improve its effectiveness (Dziodosz and Chandrasekar 2002). In order to provide an engaging user experience, the system conformed to the Google user-centric RAILS (Response, Animation, Idle, Load) performance model (Kearney 2017): loading data incrementally, allowed the essential information to be delivered within a window of 300–1000 ms. Figure 7 shows a

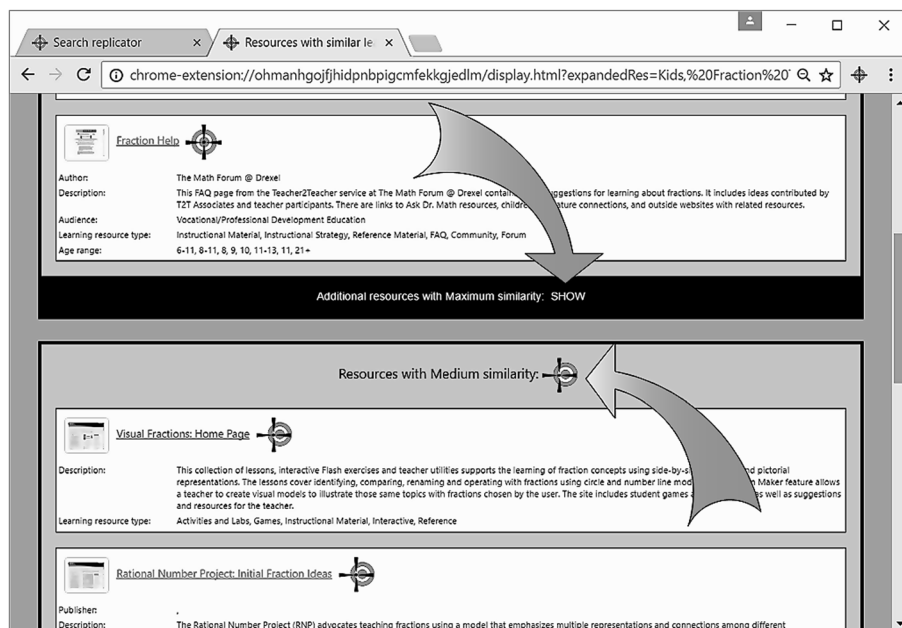


Fig. 7 Discoverer: resources clustered by similarity

screenshot of the prototype: the upper frame with resources of maximum similarity and a button to request additional ones, and a lower frame with representative resources of medium similarity.

Discoverer was evaluated with a larger summative evaluation. The invitation to participate in the evaluation was sent to about 50 educators in three different countries, identified via a snowball strategy. Twenty-nine educators from three countries, teaching different subjects mostly (79.3%) at secondary level, with dissimilar experience about OERs, accepted to participate. This sample size was considered adequate to the qualitative/exploratory oriented character of this research, in line with the recommendations from Marshall et al. (2013), of a sample size between 15 and 30 for qualitative oriented studies. Quantitative data collected via Likert-type scales, were not normally distributed, because most values were skewed towards the maximum possible. Hence they were analysed with non-parametric statistics, which is also suitable to analyse samples of this size.

### 3.6.1 Quantitative analysis (attitudinal data)

Table 1 reports some of the data collected with basic descriptive statistics. The groups in the table relate to three main interrelated aspects – functionalities (concerning the application), usability (concerning the interaction), and user experience (concerning more holistic aspects) – plus a final one about overall relevance.

Participants feedback was very positive: the mode corresponded, in all but one case, to the maximum possible positive value. Likewise, IQR was zero in most cases. One question asked about the usefulness of the expansion by similarity in general, considering also other metrics such as togetherness or likedness. These additional metrics

**Table 1** Basic descriptive statistics of Likert-type scales

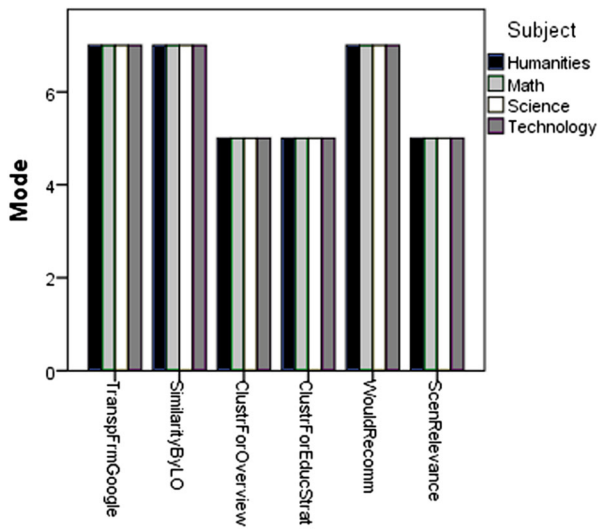
Variable name	Description and possible range	Mode	Mean	IQR
TranspFrmGoogle	Usefulness of transparently starting a specialized search directly from Google, compared to dedicated portals. [1 (much worse) .. 7 (much better)]	7	6.8	0.5
SimilarityByLO	Usefulness of the expansion by similarity based on Learning Objectives, compared to traditional metrics based on shared words. [1..7]	7	6.7	0.5
SimilarityGeneral	Usefulness of the expansion by similarity in general, considering also other metrics (such as togetherness, likedness). [1 (not at all useful) .. 5 (very useful)]	4.5	4.0	1.5
DiffcltyAltTechn	Difficulty to find resources that share the same learning objectives with alternative tools currently used. [1..7]	7	6.4	1
ClustrForOverview	Usefulness of clustering to help educators making sense of large volumes of hits. [1..5]	5	4.8	0
ClustrForEducStrat	Usefulness of clustering to support search of resources targeting specific educational strategies. [1..5]	5	4.8	0
TranspUseLO	Usefulness of the tool transparency, which avoids explicit handling of formal learning objectives. [1..5]	5	4.8	0
WdLikeUsing	Willingness to use the tool. [1..7]	7	6.8	0
WorkloadReduct	Effectiveness in reducing workload. [1..7]	7	6.6	0
WouldRecomm	Willingness to recommend the tool to a colleague. [1..7]	7	6.9	0
ScenRelevance	Relevance of the scenarios proposed in the evaluation. [1..5]	5	4.8	0

were not implemented in the prototype, because they were considered of limited priority by most educators who participated in the OERs tasks taxonomy empirical evaluation. However, they were mentioned in the questionnaire because frequently considered in the literature. This was the only case where participants did not give the highest favourable scoring.

The distribution analysis across different profiles did not show significant differences among educators, in most cases (teaching subject, teaching level, gender, age). The clustered bar charts in Fig. 8, in particular, shows the mode for some meaningful variables (names available in Table 1) across educators teaching different subjects. The mode was used in this case, to stress the visual effect: the values are all perfectly aligned to the maximum possible value (5 for unipolar scales, 7 for bipolar). This looks reasonable, given that most data consist anyway in the highest possible value.

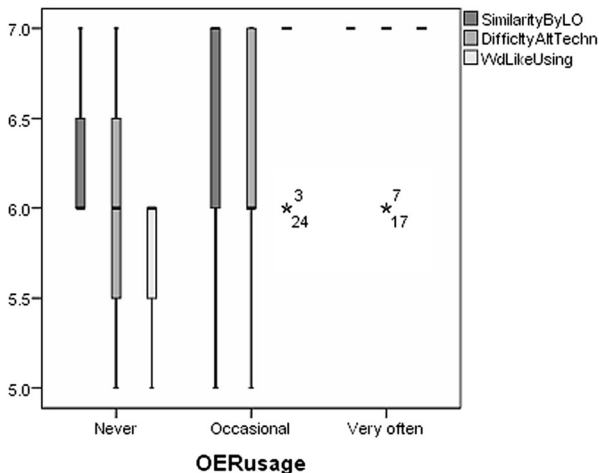
Interestingly, the characteristic that most differentiated educators' responses, was their experience in using OERs. It is evident from the boxplots in Fig. 9, that the more users were experienced with OERs, the more they were likely to (1) appreciate similarity by learning objectives, (2) think that it is difficult to obtain the same results with existing alternatives, and (3) wish to use Discoverer. Indeed, the picture shows that the median (as well as first and third quartiles) of the scores indicating the level of appreciation increases monotonically, and the degree of dispersion (incertitude) decreases decisively, when the self-reported use of OERs in their teaching activities increases from "never", to "occasional", to "very often".





**Fig. 8** Distribution across teaching subjects

This visual analysis was confirmed analytically by the independent-samples Kruskal-Wallis test, which rejected the null hypothesis that the distributions of SimilarityByLO ( $H(2) = 6.055$ ,  $p = 0.048$ ), DifficultyAltTechn ( $H(2) = 6.212$ ,  $p = 0.045$ ), and WdLikeUsing ( $H(2) = 16.891$ ,  $p < 0.01$ ), were the same across categories of educators with different experience in the use of OERs. In addition to significance, the effect size was also important: for example, the Kendall's tau correlation between OERUsageCode and SimilarityByLO was  $r_\tau = 0.436$ .



**Fig. 9** Higher experience in using OERs, leading to more positive feedback

### 3.6.2 Qualitative analysis

Qualitative data were analysed again with inductive qualitative content analysis. Data were quite homogeneous: a few codes were sufficient to label all concepts expressed. While this process was separately carried out for the text related to each question, the core concepts addressed by this research, such as “exploratory search”, “domain orientation”, “WC tasks”, and “personalization”, were consistently repeated in the different contexts, and used to justify positive feedback. For example, the very positive scores assigned to the similarity metric based on learning objectives, were justified by three reasons: domain-orientation (15 times), efficiency (10) and precision (9). Domain-orientation was clearly the most appreciated aspect. As another example, the very positive rating on the usefulness of clustering to support educational strategies, was largely justified (17) by its support to WC tasks. Users mentioned in particular specific activities such as reinforcement, remediation, and in-depth activities, as well as, more in general, personalization of education:

“Very useful to personalize educational activities”,  
 “It simplifies personalized educational activities”,  
 “Useful for both remediation and in-depth activities, as students require personalized support”.

Educators were unanimously positive about the relevance of the scenario proposed in the evaluation: an unsatisfactory Google search, followed by a specialized keyword search with Discoverer, followed in turn by the iterative exploration and expansion of additional similar resources. They (15) appreciated again, in particular, the domain-oriented support for WC tasks:

“They correspond exactly to my search patterns”,  
 “The scenarios correspond to my activities as a teacher”,  
 “Because it fits the search workflow of teachers”,  
 “It corresponds directly to my strategy when looking for resources”.

Interestingly, seven respondents were only moderately positive about the use of other similarity metrics (such as togetherness and likedness), while four respondents even mentioned explicitly that they had no interest in relying on the opinion of their peers:

“I would trust other people’s judgement only partially”,  
 “I am not interested in other people opinions and behaviour”.

These opinions explain the higher dispersion (IQR) of the corresponding scores, and are consistent with the outcomes from the task-taxonomy evaluation. Yet, this result is not always consistent with the literature, which frequently emphasizes collaborative aspects. The research by Drachsler et al. (2012), for example, might give the impression that all educators are extremely keen on “social” functionalities, while the outcomes of this study indicate that they are among the least important. This contrast can be explained because the study mentioned focused

exclusively on collaborative aspects, without considering them in a broader context, and probably on the inevitable bias in the selection of their sample of educators highly conscious about collaboration, the core topic of that project. The present study, that might well be itself biased, demonstrates anyway the importance of obtaining a global picture of the educators' needs.

Participants were also asked about what they disliked most about the prototype. Six of them answered none:

“Really nothing”,

“Nothing, from what I can see it would be very useful in all its aspects”.

Others indicated some minor aspects of the interface, such as the small fonts, the exclusive use of the English language, and limitations in the amount of metadata available.

Qualitative data were also examined across different participant profiles. While quantitative analysis with Kruskal-Wallis and Jonckheere-Terpstra did not reveal significant differences concerning users teaching experience, qualitative analysis spotted some differences. Notably, experienced educators were more specifically interested in exploiting the prototype for personalization:

“Because it perfectly supports the need for personalization of the educational activities” [experience >20],

“To organize personalized educational activities in the same class” [experience >20].

Less experienced users were more interested in using the tool for initial generic lesson planning:

“[...] most important points when you have to prepare a lesson” [experience <6],

“To find the best strategy to present a topic” [experience <6].

All these comments confirmed that the prototype was definitely appreciated for its support to high-level domain-oriented tasks.

### **3.6.3 Discoverer versus Google**

Basic Google keywords search is the most frequent technique used by educators looking for OERs, as widely reported in the literature and confirmed by participants in this research. Hence Google was always at least implicitly the reference engine against which Discoverer was compared. For example, participants explicitly considered that search by similarity based on educational alignments was a much better solution than the traditional search based on keywords (variable *SimilarityByLO* in Table 1). But when participants reported that Discoverer was very effective to reduce their workload (variable *WorkloadReduct*), or when they expressed appreciation for the key characteristics of Discoverer such as its domain orientation or its exploratory character, they were comparing it to their current practices, hence again, this time implicitly, to Google.

The objective of this research was to identify educators' needs and strategies to support them, hence emphasising functionalities in order to inform the design of an OER exploration prototype. The actual performances of Discoverer against Google, in terms of number of resources identified, were less relevant, also because these mainly depend on the metadata available in the dataset used by Discoverer. Yet these aspects too, helped users to appreciate the advantages of the proposed strategy. The performances of Discoverer, which are identical to those of RepExp in this regard, were previously remarked, commenting on the large number of educational resources identified in Fig. 6: 399 resources identified from the initial search with the keyword "biology", and 1345 resources identified by the expansion of a single one of them. Indeed, it was exactly the large number of resources identified by RepExp, which motivated the design of Discoverer. In comparison, Google just returned a single clearly identifiable educational resource in its first page, in addition to relevant but generic material such as definitions or descriptions of the term.

As an even more meaningful example, reported in Fig. 10, searching with the generic keyword "add" did not even return any educational resource in the first 10 snippets in Google, while Discoverer, which returned 30 educational resources in the first search-replication step, returned 617 resources following the expansion of the first one of them, and further 2854 resources following a second expansion. As a final example, searching with the more specific keyword "probability", returned two educational resource in the first 10 snippets in Google, while Discoverer returned 46 educational resources in the first search-replication step, 218 ones following a first expansion, and 1440 ones following a second expansion.

The most important difference, however, is that the resources reported by Discoverer are documented with rich educational metadata, and conveniently organized in clusters of different similarity (Fig. 10 again), which can be explored and further expanded

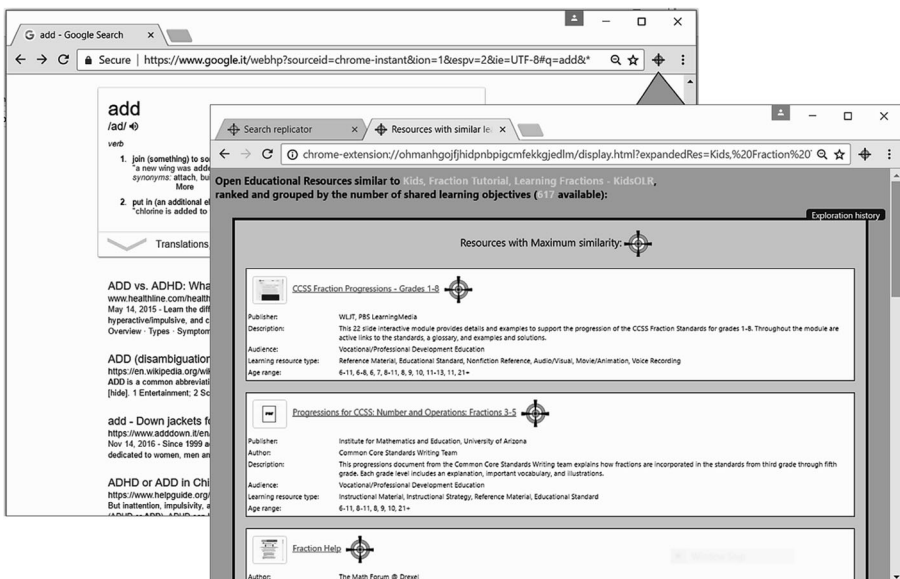


Fig. 10 Comparing Google with Discoverer

interactively at will, based on precisely defined learning objectives. When better datasets and search engines become available, Discoverer, a reusable component on top of them, will just produce even better results.

### 3.7 Research limitations

The relatively limited number and representativeness of participants involved in this research is a limitation which may rise concerns about the generalizability of results, and needs to be further discussed. The research involved a total of 49 participants in the various studies. The first study had the main scope to preliminarily identify areas worth of further investigation. The nine educators involved, thinking critically on the results of the analysis of the literature and existing search platforms in the light of their own experience, were sufficient to preliminarily identify the opportunity to investigate QBE search by similarity to support high-level and domain oriented tasks, in the following DSR cycles. The limited number of participants involved in the first two DSR studies (4 and 6) were in line with the recommendations from the literature (Nielsen 1995), and were sufficient to spot critical aspects of the prototypes, to be addressed in the following studies. Involving a larger number of participants at these stages would have been a waste of resources. The 29 participants involved in the summative evaluation are on the higher end of the recommendations from the literature for a qualitative analysis (Marshall et al. 2013). The quantitative analysis in the first and last studies made use of non-parametric statistic, which is applicable to samples of that size. Despite weighing-in the actual samples size, it yielded statistically significant results, as well as considerable effects size, suggesting strength of statistical claims.

The data, collectively obtained from the 49 participants, were very consistent among the different studies, and quantitative and qualitative data were mutually reinforcing within the studies. The positive feedback from these studies can possibly be appreciated by comparing the results obtainable from the prototype (Figs. 10 and 7), versus the results obtainable from the alternative used by the majority of educators, that is the keywords based mechanism provided by Google. Indeed, Discoverer: (1) finds exclusively educational resources, (2) in large numbers, (3) precisely targeted using formal learning objectives, (4) conveniently clustered by degree of similarity, and (5) documented with rich educational metadata. Last but not least, Discoverer has (6) a simple and exploratory oriented QBE strategy which allows users to interactively reissue new queries from previous results, which is the main feature which made a participant exclaim “it is a Stargate to knowledge!”

Another limitation concerns the degree of universality of needs among educators with different profiles. The analysis tested and did not identify any difference among educators teaching different topics, such as technical, maths or humanities. Minor differences were identified among educators from different countries (Italy and Brazil versus UK), attributed to their different familiarity with formal learning objectives. Differences among educators with different level of experience were also identified. However, while a few test-users had experience in higher education or professional training, their limited number did not make it possible to test potential differences among these different domains. Here, a major potential source of differences does exist indeed, and is due to Discoverer being powered by educational alignments. These are frequently available in secondary education, sometimes because their use is

mandatorily required, and they have an even longer tradition in professional training – where precisely formulated learning objectives or alignments to skill standards are more easily available. Certainly Discoverer, as it is, is dependent on contexts where alignments are not available. However, some of its strategies could be useful anyway, for example the integration of exploratory oriented QBE facilities on top of Google, the iterative interactive search by similarity starting from any resource being explored, or the clustering of results in degrees of similarity. However, different domains, or even different organizational and cultural contexts in the same domain, might rise partially different requirements. Notably, they might require the adoption of different metrics, for example based on the combination of reading difficulty and topic, or on crowdsourced tags.

Finally, the evaluation mainly collected subjective measures from educators, and their reaction was based on a relatively brief exposure to the prototype. While educators reported that the prototype is very relevant and useful to support their high-level tasks, a more thorough evaluation could measure its actual impact on their work. This would require to integrate the prototype in an operational OER search platform, and devise a challenging (Kules and Shneiderman 2008) evaluation strategy at work context level.

## 4 Conclusions

The overall goal of this research was to contribute to support educators in discovering OERs, focusing on secondary education. The strategy proposed corresponds to a modern, mixed search / recommendation system under user control. Notably, it includes: (1) discovery functionalities for educational resources, transparently starting from Google SERPs, (2) an interactive QBE expansion operation letting users identify similar resources starting from any educational resource as a sample, (3) a domain-oriented similarity metric based on educational alignments, and (4) a strategy to present summarized results in expandable clusters of different degrees of similarity.

Evaluation results show that Discoverer was strongly appreciated especially for its support to exploratory search, its domain orientation, and its support to WC tasks, such as lesson planning and personalized education. Some interesting differences emerged across educators with different profiles. First, the more experienced educators were with OERs, the more they (1) appreciated the proposed metric based on educational alignments, (2) were aware of the complexity of using existing alternative techniques to reach the same results, (3) expressed eagerness to use the prototype. Second, less experienced educators considered the prototype very useful to explore different educational strategies for initial generic lesson planning, while more experienced teachers tended to appreciate it most, for the possibility to personalize education. All test-users were extremely positive, and would enthusiastically recommend Discoverer to their colleagues.

The main limitation of this study is the number and representativeness of the evaluators involved: most respondents from Italian high schools. While this specific context was the object of interest of one of the researchers, the

generalization of the results would benefit from a sample more representative of a wider population. Yet, considering the exploratory character of the research, its focus on secondary education, the use of quantitative techniques appropriate for small samples, and relevant arguments from the literature – about saturation, efficiency, sample sizes used by other authors, and recommendations by methodologists – the samples used in the various studies can be considered acceptable. Additional aspects supporting the validity of the findings, include: (1) the engagement of test users with plenty of experience in education, (2) the significant statistical results, (3) the consistency of results from the different studies, (4) the mutually reinforcing results from the triangulation of qualitative and quantitative analysis, (5) the theoretical grounding in the IFT. While we do not believe a reasonable statistical generalisation is warranted by the present data, there is an encouraging pattern of support for the idea that OER exploration benefits from the approach used in the prototypes (especially Discoverer).

This research could be extended in various directions. First, it could collect further evidence supporting the generalizability of the identified requirements and strategies to support them. Integrating Discoverer (which was designed as a reusable component) in an existing search platform and automatically collect analytics, would provide the opportunity for evaluations which could involve a more representative sample of users, target different stakeholders such as learners, and even measure its impact on actual educator performances.

The research reinforced the claim that existing search systems can be improved by supporting user-tasks in their specific domain: further research could open-up again the design space, identifying additional domain-oriented tasks, possibly in different domains, and investigating strategies to support them. For example, Discoverer made the use of alignments totally transparent to its users, which was very much appreciated by Italian teachers, slightly less by British teachers. Yet educators in different domain, or possibly the same educators after gaining experience with the use of Discoverer, might well wish to understand the reasons of the results provided, in order to refine their search strategies. Further research could therefore investigate the possibility to provide explanations (Daher et al. 2017).

Search personalization is another technique that could potentially help to improve search effectiveness. Search results could be customized based on user characteristics such as location or language, or users special interests – either declared or automatically inferred for example from search history (Dou et al. 2007). For example, knowing from the profile of an educator that she is interested in resources for a certain age range, her searches could be personalized by filtering or prioritizing, among the results, educational resources with those characteristics. The initial search based on traditional keywords could take advantage from information about previous searches: if it is known that an educator is frequently looking for resources in a certain context, potentially ambiguous keywords such as “sum” could be associated to that specific context (sum of integrals in calculus versus sum of coefficients in a chemical reaction), contributing to provide more relevant results.

Finally, while this research focused on some aspects of metadata consumption, the challenge of discoverability as a whole is double faced: the other side being metadata production. Dietze et al. (2017) observe that, at the moment, there is not yet much evidence on the proper use of [schema.org/LRMI](https://schema.org/LRMI) educational alignments, which limits



the commitment to make these metadata available. This research shows that educational alignments can be very useful to support educators in their high-level tasks, and should help motivating efforts to make them available. However, there is a need for further research focusing directly on metadata production.

In conclusion, the overall goal of this research was to contribute to understand how educators could be supported in discovering OERs, one of the main challenges identified in the literature. The research community advocates to target an inclusive view of search which is exploratory-oriented (Marchionini 2006), domain-oriented, and supports high (Work Context) level tasks (Wilson et al. 2010). The experimentation with the prototypes developed in this research, whose potential impact is going to grow with the increased availability of metadata, provides evidence that some steps were made in those directions, hopefully contributing to reap the benefits of the OER movement.

**Acknowledgments** We would like to thank Prof. Marian Petre for her advice and feedback.

## Compliance with ethical standards

**Ethical approval** The research was reviewed by, and received a favourable opinion from, the Open University Human Research Ethics Committee.

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Abeywardena, I. S., Chan, C. S., & Tham, C.Y. (2013). OERScout technology framework: A novel approach to open educational resources search. *The International Review of Research in Open and Distance Learning*, 14(4), 214–237.
- Agarwal, R., & Venkatesh, V. (2002). Assessing a firm's web presence: A heuristic evaluation procedure for the measurement of usability. *Information Systems Research*, 13(2), 168–186.
- Al-Khalifa, H. S., & Davis, H. C. (2006). The evolution of metadata from standards to semantics in E-learning applications. In *Proceedings of the seventeenth conference on Hypertext and hypermedia* (pp. 69–72). New York: ACM.
- Atenas, J., & Havemann, L. (2013). Quality assurance in the open: An evaluation of OER repositories. *INNOQUAL: International Journal for Innovation and Quality in Learning*, 1(2), 22–34.
- Atkins, D. E., Brown, J. S., & Hammond, A. L. (2007). *A review of the open educational resources (OER) movement: Achievements, challenges, and new opportunities*. Menlo Park: The William and Flora Hewlett Foundation.
- Barker, P. (2014). Explaining the LRMI alignment object. *Sharing and learning*. Resource document. <http://blogs.pjnk.net/phil/explaining-the-lrmi-alignment-object/>. Accessed 2 June 2018.
- Barker, P., & Campbell, L. M. (2016a). Technology strategies for open educational resource dissemination. In *Open education: international perspectives in higher education* (pp. 51–70). Cambridge: Openbook Publishers.
- Barker, P., & Campbell, L. (2016b). Learning resource metadata on the web. In *International conference on world wide web (WWW '15 companion)* (pp. 687–688). <https://doi.org/10.1145/2740908.2741745>.
- Belkin, N. J. (1995). Strategies for evaluation of interactive multimedia information retrieval systems. *Proceedings of the final workshop on multimedia information retrieval (MIRO 95)*, Glasgow, London: Springer.
- Berners-Lee, T. (2007). Giant global graph. *timbl's blog*. Resource document. <http://dig.csail.mit.edu/breadcrumbs/node/215>. Accessed 10 February 2014.
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, 284(5), 28–37.
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data – The story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22.

- Brooke, J. (1986). *System usability scale (SUS): A quick-and-dirty method of system evaluation user information*. Reading UK: Digital Equipment Co Ltd..
- Byström, K., & Hansen, P. (2002). Work tasks as unit for analysis in information seeking and retrieval studies. In *Proceedings of the 4th international conference on conceptions of library and information science: emerging frameworks and methods (CoLIS4)* (pp. 239–252). Greenwood Village: Libraries Unlimited.
- Chi, E. H. (2015). Blurring of the boundary between interactive search and recommendation. *Proceedings of the 20th International Conference on Intelligent User Interfaces*. <https://doi.org/10.1145/2678025.2700998>.
- Cho, J. Y., & Lee, E. (2014). Reducing confusion about grounded theory and qualitative content analysis: Similarities and differences. *The Qualitative Report*, 19(32), 1–20.
- Cleverdon, C. W. (1960). ASLIB Cranfield research project on the comparative efficiency of indexing systems. *ASLIB Proceedings*, XII, 421–431.
- D'Aquin, M., Adamou, A., & Dietze, S. (2013). Assessing the educational linked data landscape. *Proceedings of the 5th annual ACM Web science conference* (pp. 43–46). New York: ACM.
- Daher, J., Brun, A., & Boyer, A. (2017). *A review on explanations in recommender systems*. Technical Report LORIA - Université de Lorraine.
- DCMI (2012). *DC-education application profile*. Resource document. [http://dublincore.org/moinmoin-wiki-archive/education/wiki/pages/DC\\_2dEducation\\_20Application\\_20Profile.html](http://dublincore.org/moinmoin-wiki-archive/education/wiki/pages/DC_2dEducation_20Application_20Profile.html). Accessed 2 May 2016.
- Dietze, S., SanchezAlonso, S., Ebner, H., Yu, H. Q., Giordano, G., Marenzi, I., & Pereira Nunes, B. (2013). Interlinking educational resources and the web of data: A survey of challenges and approaches. *Program*, 47(1), 60–91.
- Dietze, S., Drachsler, H., & Giordano, D. (2014). A survey on linked data and the social web as facilitators for TEL recommender systems. In *Recommender systems for technology enhanced learning* (pp. 47–75). New York: Springer.
- Dietze, S., Taibi, D., Yu, R., Barker, P., & d'Aquin, M. (2017). Analysing and improving embedded markup of learning resources on the web. In *Proceedings of the 26th international conference on world wide web companion* (pp. 283–292). Republic and Canton of Geneva: International World Wide Web Conferences Steering Committee.
- Doctorow, C. (2001). *Metacrap: Putting the torch to seven straw-men of the meta-utopia*. Resource document. <http://www.well.com/doctorow/metacrap.htm>. Accessed 15 April 2015.
- Dou, Z., Song, R., & Wen, J. R. (2007, May). A large-scale evaluation and analysis of personalized search strategies. In *Proceedings of the 16th international conference on World Wide Web*, 581–590. ACM.
- Downes, S. (2003). One standard for all: Why We Don't Want It, Why We Don't Need It. *National Research Council*. Resource document. [http://zope.cetis.ac.uk/lib/media/one\\_standard.pdf](http://zope.cetis.ac.uk/lib/media/one_standard.pdf). Accessed 2 May 2016.
- Drachsler, H., Grellier, W., Fazeli, S., Niemann, K., Sanchez-Alonso, S., Rajabi, E., Palmér, M., Ebner, H., Simon, B., Nösterer, D., Kastrantas, K., Manouselis, N., Hatzakis, I., & Clements, K. (2012). D8.1 review of social data requirements. *Open Discovery Space project*.
- Dziadosz, S., & Chandrasekar, R. (2002). Do thumbnail previews help users make better relevance decisions about web search results? In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 365–366). New York: ACM.
- Gerhardt-Powals, J. (1996). Cognitive engineering principles for enhancing human-computer performance. *International Journal of Human Computer Interaction*, 8(2), 189–211.
- GLOBE (2016). *GLOBE – Connecting the world and unlocking the deep web*. Resource document. <http://globe-info.org>. Accessed 30 July 2016.
- Guha, R. V., Brickley, D., & Macbeth, S. (2016). Schema.org: Evolution of structured data on the web. *Communications of the ACM*, 59(2), 44–51.
- Gurell, S., & Wiley, D. (2008). *OER handbook for educators*. Resource document. [http://wikieducator.org/OER\\_Handbook/educator\\_version\\_one](http://wikieducator.org/OER_Handbook/educator_version_one). Accessed 11 Aug 2017.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
- IEEE (2002). *Draft standard for learning object metadata*. Resource document. [http://ltsc.ieee.org/wg12/files/LOM\\_1484\\_12\\_1\\_v1\\_Final\\_Draft.pdf](http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf). Accessed 10 April 2014.
- IMS (2015). *Interoperability standards helping to transform the digital curriculum*. Resource document. [www.imsglobal.org/sites/default/files/articles/sli17-051815.pdf](http://www.imsglobal.org/sites/default/files/articles/sli17-051815.pdf). Accessed 10 June 2015.
- ISO. (1998). *ISO 9241-11:1998: Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability*. Geneva: International Organization for Standardization.
- Kabel, S., De Hoog, R., Wielinga, B. J., & Anjewierden, A. (2004). The added value of task and ontology-based markup for information retrieval. *Journal of the American Society for Information Science and Technology*, 55(4), 348–362.

- Kearney, M. (2017). *Measure performance with the RAIL model*. Resource document. <https://developers.google.com/web/fundamentals/performance/rail>. Accessed 13 August 2017.
- Klo, J. (2011). *AMPlified search extension for chrome*. Resource document. <https://github.com/jimklo/AMPS-Chrome>. Accessed 01 May 2016.
- Knonth, P. (2015). *Linking Textual Resources to Support Information Discovery*. Ph.D. thesis. Milton Keynes: The Open University.
- Kules, B., & Shneiderman, B. (2008). Users can change their web search tactics: Design guidelines for categorized overviews. *Information Processing & Management*, 44(2), 463–484.
- Learning Registry. (2016). *Learning Registry*. Resource document. [www.learningregistry.org](http://www.learningregistry.org). Accessed 1 May 2016.
- Lockley, P. (2011). *Pgogy/learning\_registry\_chrome*. Resource document. [https://github.com/pgogy/learning\\_registry\\_chrome](https://github.com/pgogy/learning_registry_chrome). Accessed 1 May 2016.
- LRMI (2013a). *The smart Publisher's guide to LRMI tagging*. Resource document. [http://www.lrmi.net/wp-content/uploads/2013/03/LRMI\\_tagging\\_Guide.pdf](http://www.lrmi.net/wp-content/uploads/2013/03/LRMI_tagging_Guide.pdf). Accessed 5 March 2014.
- LRMI (2013b). *LRMI survey report*. Resource document. <http://www.lrmi.net/wp-content/uploads/2013/08/LRMI-Survey-Report-August-2013-Update.pdf>. Accessed 5 April 2014.
- LRMI (2013c). *LRMI's "killer feature": educationalAlignment*. Resource document. <http://lrmi.dublincore.net/2013/06/12/lrmis-killer-feature-educationalalignment/>. Accessed 2 May 2016.
- LRMI (2014). *LRMI Project Description*. Resource document. <http://wiki.lrmi.net/LRMI+Project+Description>. Accessed 5 April 2014.
- Madan, A., & Dubey, S.K. (2012). Usability evaluation methods: A literature review. *International Journal of Engineering Science and Technology*, 4(2), 590–599.
- Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., & Koper, R. (2011). Recommender systems in technology enhanced learning. In *Recommender systems handbook* (pp. 387–415). New York: Springer.
- Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communications of the ACM*, 49(4), 41–46.
- Marshall, B., Cardon, P., Poddar, A., & Fontenot, R. (2013). Does sample size matter in qualitative research? A review of qualitative interviews in IS research. *Journal of Computer Information Systems*, 54(1), 11–22.
- McNamara, N., & Kirakowski, J. (2006). Functionality, usability, and user experience: Three areas of concern. *Interactions*, 13(6), 26–28.
- Molich, B., & Nielsen, J. (1990). Improving a human - computer dialogue. *Communications of the ACM*, 33(3), 338–348.
- Nielsen, J. (1995). Applying discount usability engineering. *Software, IEEE*, 12(1), 98–100.
- Nilsson, M. (2010). *From interoperability to harmonization in metadata standardization: designing an evolvable framework for metadata harmonization*. Ph.D. thesis. Stockholm: KTH.
- Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77.
- Pirolli, P., & Card, S. K. (1999). Information foraging. *Psychological Review*, 106, 643–675.
- Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011). Common core standards the new US intended curriculum. *Educational Researcher*, 40(3), 103–116.
- Prat, N., Comyn-Wattiau, I., & Akoka, J. (2014). Artifact evaluation in information systems design-science research-a holistic view. In *PACIS 2014 Proceedings* (vol. Paper 23, pp. 1–16).
- Qu, Y., & Furnas, G. W. (2008). Model-driven formative evaluation of exploratory search: A study under a sensemaking framework. *Information Processing & Management*, 44(2), 534–555.
- Riley, J. (2010). *Glossary of metadata standards*. Resource document. [http://www.dlib.indiana.edu/~jenlrile/metadatamap/seeingstandards\\_glossary\\_pamphlet.pdf](http://www.dlib.indiana.edu/~jenlrile/metadatamap/seeingstandards_glossary_pamphlet.pdf). Accessed 02 May 2016.
- Schema.org. (2013). *Frequently asked questions*. Resource document. <http://schema.org/docs/faq.html>. Accessed 10 June 2015.
- Schraefel, M. C. (2009). Building knowledge: What's beyond keyword search? *Computer*, 42(3), 52–59.
- Shear, L., Means, B., & Lundh, P. (2015). *Research on open: OER research hub review and futures for research on OER*. Menlo Park: SRI International.
- Singhal, A. (2012). Introducing the knowledge graph: Things, not strings. *Official Google Blog*. Resource document. <http://googleblog.blogspot.co.uk/2012/05/introducing-knowledge-graph-things-not.html>. Accessed 13 April 2015.
- Smyth, B., & McClave, P. (2001). Similarity vs. diversity. In *Case-based reasoning research and development* (pp. 347–361). Berlin, Heidelberg: Springer.
- Sue, V. M., & Ritter, L. A. (2007). Designing and developing the survey instrument. In *Conducting online surveys* (pp. 59–88). Thousand Oaks: Sage.
- Sutton, S. A. (2008). Metadata quality, utility and the semantic web: The case of learning resources and achievement standards. *Cataloging and Classification Quarterly*, 46(1), 81–107.

- Sutton, S. A., & Mason, J. (2001). The Dublin Core and metadata for educational resources. In *International conference on Dublin Core and metadata applications*, Wroclaw, Poland (pp. 25–31).
- Timpany, G. (2015). *Reaching the constant sum*. Resource document. <http://survey.event.com/blog/customer-insights-2/using-constant-sum-questions>. Accessed 2 May 2016.
- UNESCO (2012). *UNESCO Paris OER declaration*. Resource document. [http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/Events/Paris%20OER%20Declaration\\_01.pdf](http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/Events/Paris%20OER%20Declaration_01.pdf). Accessed 8 February 2014.
- Vaishnavi, V., & Kuechler, W. (2015). *Design science research in information systems*. Resource document. <http://www.desrist.org/design-research-in-information-systems/>. Accessed 5 May 2017.
- Verbert, K., Drachsler, H., Manouselis, N., Wolpers, M., Vuorikari, R., & Duval, E. (2011). Dataset-driven research for improving recommender systems for learning. In *Proceedings of the 1st international conference on learning analytics and knowledge* (pp. 44–53). New York: ACM.
- Weller, M. (2010). Big and little OER. *OpenED2010: Seventh Annual Open Education Conference*, 2–4 Nov 2010, Barcelona, Spain. <http://oro.open.ac.uk/24702/>.
- White, R. W., Drucker, S. M., Marchionini, G., & Hearst, M. (2007). Exploratory search and HCI: Designing and evaluating interfaces to support exploratory search interaction. In *CHI'07 extended abstracts on human factors in computing systems* (pp. 2877–2880). New York: ACM.
- Wildemuth, B. M. & Freund, L. (2009). Search tasks and their role in studies of search behaviors. In *Third annual workshop on human computer interaction and information retrieval*, Washington DC.
- Wilson, M. L., Kules, B., Schraefel, M. C., & Shneiderman, B. (2010). From keyword search to exploration: Designing future search interfaces for the web. *Foundations and Trends in Web Science*, 2(1), 1–97.
- Winston, P. H. (1984). *Artificial intelligence*. Reading: Addison-Wesley.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.