

Open Research Online

The Open University's repository of research publications
and other research outputs

Supporting the Discoverability of Open Educational Resources: on the Scent of a Hidden Treasury

Thesis

How to cite:

Cortinovis, Renato Mario (2020). Supporting the Discoverability of Open Educational Resources: on the Scent of a Hidden Treasury. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2019 The Author

Version: Version of Record

Link(s) to article on publisher's website:
<http://dx.doi.org/doi:10.21954/ou.ro.00011608>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Supporting the Discoverability of Open Educational Resources: on the Scent of a Hidden Treasury

Renato Mario Cortinovis, MSc.

Thesis submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

Knowledge Media Institute,
Faculty of Science, Technology, Engineering & Mathematics,
The Open University (UK)

August, 2019

This page is intentionally left blank

Abstract

Supporting the discoverability of Open Educational Resources: on the Scent of a Hidden Treasury

by

Renato Mario Cortinovis

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 thesis investigates and proposes strategies to better support educators in discovering OERs.

The literature suggests that the effectiveness of existing search systems, including for OER discovery, could be improved by supporting users, such as teachers, in carrying out more exploratory search activities closer to their existing methods of working. Hence, a preliminary taxonomy of OER-related search tasks was produced, based on an analysis of the literature, interpreted through Information Foraging Theory. This taxonomy was empirically evaluated to preliminarily identify a set of search tasks that involve finding other OERs similar to one that has already been identified, a process that is generally referred to as Query By Example (QBE). 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 alignments to educational standards, and clusters results in expandable classes of comparable degrees of similarity. The summative evaluation shows that educators do appreciate this strategy because it is exploratory and – balancing similarity and diversity – it supports their high-level tasks, such as lesson planning and personalization of education. Finally, potential barriers and opportunities for the uptake of OER discovery tools were investigated in a structured interview study with experts from the OER field. Identified issues included how to work across multiple OER portals, variability in the use of metadata and how to align with the working practices of teachers.

The findings of the thesis can be used to inform the research and development of methods and tools for OER discovery as well as their deployment to serve the needs of educators.

Acknowledgments

To my family – Chiara (my wife), Alice (doctoral student in Math), Irene (doctoral student in Physics), and Enrico (young programmer and digital artist) – and to the participants in the Open Education movements.

With a professional background in industrial R&D (in the context of the European Space Agency and other organizations) and Technology Transfer (in the context of the United Nations), I always cultivated the wish to better understand academic research. Accordingly, this PhD has been a fruitful journey in scientific research, yet paralleled by an even more interesting mirrored journey in human nature.

Concerning scientific research, I could satisfy my quest for understanding the differences between industrial R&D and academic research. In my personal experience, the key aspect is the deliberate intention to carefully avoid any “surprise” in the industrial setting, versus the deliberate intention to focus on, harness, exploit, and especially enjoy precisely the unexpected, in the academic sector. For this, I would like to thank especially Prof. Marian Petre.

Concerning human nature, I had the privilege to interact with beautiful minds capable of fascinating firm grasping of ideas, apparently limitless creativity, and a generous attitude to help. For this, I’d certainly like to thank first my supervisors Dr. Paul Mulholland and Dr. Rob Farrow. Additionally, I would like to thank my interviewees, whom I am not supposed to name for privacy reasons, but who were so engaged in the open movement to look themselves like “living open resources”. Finally I would like to thank the members of my examination panel, Prof. Andy Lane and Dr. Eamon Costello, in particular for their openness and lack of bias in following my thinking and my research, rather than their own views and interests.

Last but not least, I would like to thank my wife Chiara for her strategic support in enrolling test-users for me, and the many test-users for the precious time and effort they devoted to my experiments.

Contents

Publications of the candidate	1
List of Acronyms	2
1. Introduction	5
1.1 The problem – OERs: a hidden treasury	5
1.2 Discoverability	7
1.3 Goal and motivation of the research	7
1.4 Research questions	9
1.5 Dissertation structure	9
2. Preliminary literature review	13
2.1 Introduction	13
2.2 Traditional educational metadata	13
2.3 Linked data and semantic search.....	14
2.4 Schema.org/LRMI.....	15
2.4.1 LRMI alignment to educational frameworks	17
2.5 Social network based approaches.....	19
2.6 Recommendation and blended search/recommendation systems	20
2.7 From keywords-matching to Exploratory Search	21
2.8 From isolated search to high-level tasks in their broader context.....	22
2.8.1 Organizational and cultural aspects: educational standards in Italy.....	23
2.9 Exploratory search evaluation: focus on users and their tasks	24
2.10 Conclusions: user and domain-oriented tasks	25
3. Core research framework: Design Science Research.....	27
3.1 Introduction	27
3.2 Research and design: tensions and roles	27
3.3 Distinguishing research from design practice: knowledge production, rigour, evaluation	28
3.4 The adopted core framework: Design Science Research	29
3.5 The research process in detail	29
3.5.1 Evaluation in the DSR iterations	31
3.6 Conclusions	32
4. Focused literature review: model-driven task-analysis	33

4.1 Introduction: a domain-oriented taxonomy of OER search-related tasks.....	33
4.2 Sources of tasks: evaluation studies, use cases, and functionalities	34
4.3 Search-task models in the literature.....	35
4.3.1 A context model of search: focus on Information Seeking and Work Context	35
4.3.2 Information Foraging Theory: focus on QBE expansions by similarity.....	36
4.3.3 Strategic and episodic models	37
4.4 Literature related to metadata standards	37
4.4.1 LRMI / schema.org.....	38
4.4.2 Open Discovery Space (Irma case): social aspects.....	38
4.5 Literature on specific search tasks	38
4.5.1 Search tasks used in experimental search studies: RepAST	38
4.5.2 Integrating search with support for WC level tasks.....	39
4.5.3 Search interfaces	40
4.6 Notable learning resources discovery projects	40
4.6.1 MACE: domain-oriented visual interfaces	40
4.6.2 OpenScout: access via a taxonomy of problems.....	41
4.6.3 National Science Digital Library: exploring resource relationships.....	42
4.6.4 mEducator: repurposing.....	42
4.6.5 OERCommons: approximate alignments to educational standards.....	43
4.6.6 SocialLearn: multiple recommender systems	44
4.6.7 OpenEd: personalized remediation activities	44
4.6.8 Gooru: supporting different delivery strategies	45
4.6.9 Achievement Standards Network: educational frameworks	46
4.6.10 Summary of the projects reviewed	47
4.7 Task analysis: OER domain-oriented taxonomy and sample tasks	47
4.7.1 Top level categories of the domain-oriented taxonomy	48
4.7.2 Searching: subcategories and related tasks	49
4.7.3 Using: subcategories and related tasks	53
4.7.4 Summary of the domain-oriented taxonomy	54
4.8 Conclusions: an open, domain-oriented taxonomy.....	55
5. OER task-taxonomy empirical evaluation.....	56

5.1 Overview: empirical evaluation of the OER task-taxonomy with actual educators	56
5.2 Relationship to research questions	56
5.3 Methods: triangulated quantitative and qualitative analysis	56
5.3.1 Quality of the data: validity and reliability	57
5.3.2 Data collection protocol	59
5.4 Data analysis methodology	59
5.4.1 Quantitative data analysis	59
5.4.2 Qualitative data analysis	61
5.5 Results	61
5.5.1 Participants demography	61
5.5.2 Internal consistency reliability check	62
5.5.3 OER task-taxonomy completeness.....	63
5.5.4 Normal distribution of task weights	64
5.5.5 Task weights	65
5.5.6 Planned comparative analysis: relative importance of specific tasks.....	67
5.5.7 Explorative comparative analysis among educators with different profile	69
5.5.8 Explorative correlation analysis	69
5.5.9 Additional qualitative data: use of Google.....	70
5.6 Discussion	70
5.6.1 Limitations: sample size and representativeness.....	72
5.6.2 Findings: new domain-oriented tasks.....	73
5.7 Conclusions: new expansion task transparently based on educational alignments	73
6. Injector: first design and evaluation DSR cycle.....	76
6.1 Injector research questions: design and knowledge oriented	76
6.2 Preliminary requirements: resources with similar educational alignments from Google SERPs	77
6.3 Injector Design and Development.....	77
6.3.1 Overview: custom rich snippets injected in Google SERPs.....	77
6.3.2 Injector sample session	78
6.3.3 Similarity metric based on shared educational alignments	80
6.3.4 Selection of datasets and services providing alignments metadata	81

6.3.5 Injector architecture	81
6.4 Injector Discounted Heuristic evaluation	84
6.4.1 Goals: identify Injector major drawbacks and get preliminary feedback	84
6.4.2 Methods: discounted heuristic evaluation.....	84
6.4.3 Injector evaluation: results and discussion	85
6.4.4 Evaluation conclusions: the challenge of sparsity	88
6.5 Injector design and evaluation: conclusions	89
7. RepExp: Replicator-Expander prototype.....	92
7.1 RepExp goal and research questions: mitigate intrinsic and alignments sparsity.....	92
7.2 RepExp Design.....	92
7.2.1 Sparsity: problems and possible solutions	92
7.2.2 Separating Replicator and Expander: RepExp.....	94
7.2.3 RepExp sample session.....	94
7.2.4 RepExp architecture.....	97
7.3 RepExp Heuristic evaluation	99
7.3.1 Goals: identify RepExp further drawbacks and possible areas for improvements	99
7.3.2 Methods: a second formative heuristic evaluation with content analysis.....	99
7.3.3 RepExp evaluation results and discussion	101
7.3.4 RepExp evaluation conclusions: degree of similarity, sense-making, lock-in, interface	109
7.4 RepExp design and evaluation: reflection and further work	110
8. The prototype Discoverer	112
8.1 Discoverer research questions: sense-making, lock-in, domain-orientation	112
8.2 Discoverer Design	113
8.2.1 Clustering results in classes of comparable similarity.....	114
8.2.2 Additional metadata including licensing information to highlight open resources	117
8.2.3 Improving the presentation: resource thumbnail images	117
8.2.4 Discoverer sample session.....	118
8.2.5 Discoverer architecture	124
8.3 Discoverer Summative Evaluation	126
8.3.1 Updating the evaluation questionnaire	126

8.3.2 Discoverer Evaluation – results and discussion	127
8.3.3 Discoverer versus Google	142
8.3.4 Evaluation conclusions: exploratory search and domain-oriented WC-level tasks	143
8.4 Discoverer design and evaluation: conclusions	145
9. Feedback from representative experts in the OER ecosystem	147
9.1 Method: Semi-structured interviews and thematic analysis.....	147
9.1.1 Semi-structured interviews.....	148
9.1.2 Thematic qualitative analysis.....	149
9.2 Results.....	151
9.2.1 Final coding map.....	151
9.2.2 Coding and data extracts examples	152
9.3 Discussion	153
9.3.1 Theme 1: Discoverer (micro).....	154
9.3.2 Theme 2: Metadata (meso).....	159
9.3.3 Theme 3: OERs challenges (macro).....	167
9.3.4 Limitations of this study.....	170
9.4 Conclusions	170
10. Roadmap to the future	172
10.1 Further evaluations.....	172
10.1.1 Blind comparison of Discoverer with alternative baseline solutions	172
10.1.2 Broader quantitative oriented evaluation by instrumenting the prototype	175
10.1.3 Evaluation in different domains and cultural contexts	176
10.1.4 Evaluation of the impact on productivity and quality	177
10.1.5 Evaluation with different stakeholders.....	178
10.2 Micro-level enhancements to the current prototype.....	178
10.2.1 Recognizing different encodings of the same educational standard	178
10.2.2 Controlling the number of clusters.....	178
10.2.3 Parsing in-page schema.org markup	179
10.2.4 Avoiding queries producing no results, without decreasing efficiency	179
10.2.5 Server-side metasearch engine for a service-oriented architecture	180
10.2.6 Optimizing screen real estate with fisheye techniques.....	180

10.3 Extensions to the current similarity metric	180
10.3.1 Normalizing generalizations and specializations in the educational framework.....	180
10.3.2 Jaccard similarity: normalizing the number of alignments	181
10.3.3 Factoring-in the alignments' level of abstraction	182
10.3.4 Standards cross-walking: degree of equivalence and degree of alignment	182
10.3.5 Similarity vector versus similarity matrix	183
10.3.6 Providing explanations, letting users prioritize educational alignments.....	185
10.3.7 Integrating quality indicators based on social data and paradata.....	186
10.4 Meso-level extensions to the overall strategy	186
10.4.1 Additional similarity metrics	186
10.4.2 Exploiting different types of alignments	187
10.4.3 Clustering results of the initial keywords search	188
10.4.4 Search personalization	188
10.4.5 Resource identifiers	188
10.4.6 Educational alignments as OER genes in an OER Genome Project.....	189
10.5 Identification of additional requirements.....	189
10.5.1 Discoverer as a design probe	190
10.5.2 Identification of prerequisite resources.....	190
10.5.3 Identification of interdisciplinary resources	190
10.6 Broadening the research areas	191
10.6.1 From metadata consumption to metadata production.....	191
10.6.2 New metadata standards	192
10.6.3 In-page authoritative metadata versus centralized open metadata.....	192
10.6.4 OER-enabled pedagogy	193
10.7 Conclusions	193
11. Conclusions	194
11.1 Overall aim and research questions	194
11.2 Overview of the research	195
11.2.1 Preliminary literature analysis: the research gaps.....	196
11.2.2 OER-search Task Analysis	196
11.2.3 Empirical evaluation of the taxonomy of OER-search tasks	196

11.2.4 Injector: design and evaluation	196
11.2.5 RepExp: design and evaluation	196
11.2.6 Discoverer: design and evaluation	197
11.2.7 Feedback from representative experts in the OER ecosystem	197
11.3 Summary of findings.....	198
11.3.1 RQ1: What tasks do teachers need to carry out in relation to OER discovery, in the framework of the OER life cycle?	199
11.3.2 RQ2: What is the relative importance of tasks and categories identified, and which aspects deserve further investigation?.....	200
11.3.3 RQ3: What are the general requirements for tools supporting the aspects of interest identified?.....	200
11.3.4 RQ4: Which tools could support the identified requirements, and how do teachers respond to them?	201
11.3.5 RQ4.1: How can an OER discovery tool be embedded into web search?	201
11.3.6 RQ4.2: How can Query By Example be added to an OER discovery tool?	201
11.3.7 RQ4.3: How can similarity based on educational alignments be added to an OER discovery tool?	201
11.3.8 RQ4.4: How can the intrinsic and alignments sparsity of web resources be addressed by an OER discovery tool?	202
11.3.9 RQ4.5: How can a large number of similar educational resources be presented by an OER discovery tool to support work context level tasks?	203
11.3.10 RQ5: What barriers are restricting the deployment and uptake of OER discoverability tools and how can they be overcome?.....	204
11.4 Implications and relevance of findings	205
11.5 Research limitations	206
11.5.1 Generalizability of results	206
11.5.2 Universality of needs among educators with different profiles	207
11.5.3 Breadth of functionalities explored	207
11.5.4 Depth of evaluation.....	208
11.5.5 Data-set representativeness	208
11.5.6 Open versus blind evaluation	209
11.6 Future research	209
11.7 Concluding remarks	210

12. References	211
Appendix A: Task analysis for the sub-categories Publishing, Repurposing, and Collaborating ..	232
Appendix B: Task Analysis Survey.....	237
Appendix C: Task-Taxonomy Survey: weights for tasks and categories	259
Appendix D : Injector Heuristic Evaluation Form.....	266
Appendix E : RepExp Heuristic Evaluation Form.....	271
Appendix F: Discoverer Heuristic Evaluation Form	278

Publications of the candidate

Part of the work presented in this thesis was used to produce the following research paper:

Cortinovis, R., Mikroyannidis, A., Domingue, J., Mulholland, P. and Farrow, R. (2019) 'Supporting the discoverability of open educational resources', *Education and Information Technologies*. Available at <http://link.springer.com/article/10.1007/s10639-019-09921-3>. DOI: 10.1007/s10639-019-09921-3.

List of Acronyms

AI	Artificial Intelligence
AMPS	AMPlied Search extension for Chrome
ANOVA	ANalysis Of VAriance
API	Application Programming Interface
ARIADNE	Alliance of Remote Instructional Authoring and Distribution Networks for Europe
ASN	Achievement Standards Network
CAQDAS	Computer Assisted Qualitative Data Analysis
CAT	Coding Analysis Toolkit
CC	Creative Commons
CCSS	Common Core State Standards
CEMARC	Curriculum-Enhanced MARC
CSE	Custom Search Engine
CSQ	Constant-Sum Question
DCMI	Dublin Core Metadata Initiative
DSR	Design Science Research
GCSE	Google Custom Search Engine
GLOBE	Global Learning Objects Brokered Exchange
HARD	High Accuracy Retrieval from Documents
HCI	Human-Computer interaction
HREC	Human Research Ethics Committee
HTML	Hypertext Markup Language
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IFT	Information Foraging Theory
IMS	Instructional Management System
IQR	InterQuartile Range

IR	Information Retrieval
IS	Information Seeking
ISKME	Institute for the Study of Knowledge Management in Education
ISO	International Organization for Standardization
ISS	Information Seeking Strategies
JSON	JavaScript Object Notation
JSON_LD	JavaScript Object Notation for Linked Data
K12	from kindergarten to 12th grade
LD	Linked Data
LMS	Learning Management System
LOM	Learning Object Metadata
LR	Learning Registry
LRMI	Learning Resource Metadata Initiative
LTI	Learning Tools Interoperability
MACE	Metadata for Architectural Contents in Europe
MARC	MACHine Readable Cataloguing
MIUR	Ministero dell'Istruzione, dell'Università e della Ricerca
MOOC	Massive Open Online Course
NGSS	Next Generation Science Standards
NSDL	National Science Digital Library
OA	Open Access
ODS	Open Discovery Space
OER	Open Educational Resource
QBE	Query By Examples
RAIL	Response Animation Idle Load
RDF	Resource Description Framework
RDFa	Resource Description Framework in Attributes
RepAST	Repository of Assigned Search Tasks

RepExp	REplicator EXPander (defined in this dissertation)
REST	REpresentational State Transfer
RQ	Research Question
SDG	Sustainable Development Goal
SERP	Search Engine Result Page
SPARQL	SPARQL Protocol and RDF Query Language
TA	Task Analysis
TREC	Text REtrieval Conference
UKOER	UK Open Educational Resources
UNESCO	United Nations Educational, Scientific and Cultural Organization
URI	Universal Resource Identifier
URL	Uniform Resource Locator
XML	eXtensible Markup Language
W3C	World Wide Web Consortium
WC	Work Context
WWW	World Wide Web

1. Introduction

1.1 The problem – OERs: a hidden treasury

Open Educational Resources (OERs) are defined by Atkins et al. (2007, p. 4), as “teaching, learning, and research resources that reside in the public domain or have been released under an intellectual property license that permits their free use and re-purposing by others”. Many other definitions exist, which underline the possibility to freely modify or repurpose them, in addition to simply reuse them as they are. For example, Wiley (n.d.) focuses on the meaning of “Open”, which should allow users to engage in the “5R” activities: Retain, Reuse, Revise, Remix, Redistribute. Examples of OERs include (Butcher, 2015) interactive exercises, virtual laboratories, video lessons, e-books, assessment tests, lesson plans, open textbooks or, in some cases, Massive Open Online Courses (MOOCs). MOOCs in particular – free online courses that can be attended by a massive number of students – certainly contributed to the recent visibility of OERs. Yet they are sometimes mistakenly considered OERs because their openness is frequently severely restricted (Silveira, 2016). They are generally open as far as enrolment is concerned (no up-front fees for students) and prior qualifications required (Lane, 2013), but their content does not usually satisfy the 5Rs, which makes them unsuitable to be adapted and reused (Atenas, 2015).

The goal of OERs “is to improve access to learning opportunities by sharing knowledge and learning resources” (Gurell and Wiley, 2008, p.5), with a belief in “the simple and powerful idea that the world’s knowledge is a public good and that technology in general and the World Wide Web in particular provide an extraordinary opportunity for everyone to share, use, and reuse knowledge” (Atkins et al., 2007, p. 5).

In the last years, millions of potentially useful OERs have been developed and made available on the Internet. For example, more than 10,000 resources were developed in the context of the UK Open Educational Resources (UKOER) Programme from 2009 to 2012 (Barker and Campbell, 2016a), and the number of resources in large federations such as the Global Learning Objects Brokered Exchange (GLOBE, 2016) is of the order of one million. On top of this, there is a growing interest in MOOCs, which promised a revolution in the educational sector (Kanwar and Mishra, 2015).

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) states, 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”. Gurell and Wiley (2008), in their OER handbook for educators, indicate, in addition, that some types of OERs can even “compensate for teacher shortages”. Indeed, OERs represent a component which could contribute to support the Sustainable Development Goals (SDGs), in particular SDG 4 related to inclusive and equitable education (Lane, 2017a).

There are a few criticisms levelled at OERs too. Possibly the most severe one (Baker, 2018), is that OERs could potentially become an instrument of neo-colonialism, promoting the views of western and mostly Anglophone countries producing the majority of resources, to countries which are mostly consumers of these resources. While Daniel (2012, p. 4) argued optimistically that identifying this “potential problem helped the world to avoid it”, the same concern was expressed again recently for example by Cobb (2018, p.22): “OERs can inadvertently carry and transmit dominant beliefs and values, which may not reflect local cultural values, knowledges, aspirations, and identities”. Luckily, the salient characteristic of OERs that permit repurposing, helps to reduce this risk, allowing educators to localize and contextualize them, as recommended in the Ljubljana OER Action Plan (UNESCO, 2017).

Clearly, OERs represent an important dimension of what is referred to as “open education”, which is “increasingly being recognized as one of the most significant educational movements thus far in the 21st century” (Shear et al, 2015, p. 1). And yet, the enormous potential of these resources 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. This challenge is obstinately persistent in time: Yergler (2010), Piedra et al. (2010), UNESCO (2012), Abeywardena et al. (2013), LRMI (2013a), Barker and Campbell (2016a), and again, more recently, the Ljubljana OER Action Plan (UNESCO, 2017), Anderson and Leachman (2019), Luo et al. (2020). Tang et al. (2020), for example, claim that there is still a need to simplify the search process in order to improve the acceptance of OERs by K-12 teachers. This challenge is ubiquitously perceived beyond the western world, including for example Africa (Zaid and Alabi, 2020) and Arab States (Tlili et al., 2020). And despite the optimistic plans of Creative Commons to quickly develop a comprehensive search engine for different “media” including OERs (Creative Commons, 2017), the current version is still focusing, in the middle of 2020, just on images (Creative Commons, 2020).

The challenge of OER discoverability can be understood in terms of several complex, interrelated aspects. A major one concerns metadata, that is data describing meaningful educational characteristics of the resources, such as the educational audience (e.g. teachers or high school students), the type of educational resource (e.g. a lesson plan or a test), or the formal learning objectives of a resource. In particular, there is a scarcity of quality metadata which describe these resources, and there are many incompatible standards to specify these metadata. Partly because of this, and because of lack of coordination at global scale, there are a plethora of isolated search platforms, and users rarely wish to spend their time searching repositories individually (Drabkin, 2016a). The large majority of educators looking for OERs, 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). Additionally, not all the resources available are visible through generic search engines like Google, because they fail to index parts of the Web (so called “Deep Web”), for example pages that are dynamically generated (He et al., 2007).

Yet, search engines, hence discoverability, can be improved in many directions (Schraefel, 2009). As it will be introduced in the next sections, this research focuses on two of these: supporting discovery-oriented exploratory-search, and supporting users in their domain-oriented tasks.

1.2 Discoverability

Before proceeding, it is convenient to define the key term “discoverability” with some more precision. Indeed, there is a certain ambiguity in the common use of terms such as search, find and discover. This is evident, for example, from the ambiguous definitions available in Wikipedia. Wikipedia (2016a) defines discoverability as the “ability of something [...] *to be found*”; yet in the entry for findability it specifies that “*findability is different* from discoverability”, because “the word ‘find’ refers to locating something in a known space while ‘search’ is in an unknown space or not in an expected location” (Wikipedia, 2016b).

While there is a considerable overlap among the terms, it is useful – to better understand the meaning of discoverability – to clarify first the distinction between lookup and discovery. Marchionini (2006, p. 42) indicates lookup as “the most basic kind of search task”, which in library / information science “is called ‘known item search’, to distinguish it from subject or topical searches”. Lookup, starting from precisely formulated queries, produces very precise results. He observes that many search use cases, however, including learning search tasks, require an analytical lookup mainly to bring users in a suitable position to start exploratory browsing. Similarly, Knoth (2015, p. 29) writes that discovery is more “concerned with detecting links or relationships between resources” to support exploratory search.

Discoverability, therefore, emphasizes that the objectives of a discovery oriented activity might not even be known precisely in advance. On the contrary, precisely identifying these objectives is part of the discovery activity itself, which can be addressed with an interactive exploratory-oriented browsing. An educator, for example, might like to explore different alternatives, in order to refine his/her preliminary idea for the objectives of an in-depth activity. The term search is more generic and can refer to either lookup or discovery, or more frequently both of them at the same time.

1.3 Goal and motivation of the research

While information search as a “fundamental life activity” (Marchionini, 2006, p.42) might be an overstatement, there is no doubt that search is increasingly important in the context of the Internet and our knowledge oriented society in general (Wilson et al, 2010). The discoverability of OERs in particular, as discussed, is widely recognized as a challenge, instrumental in reaping the potential of the OER movement (UNESCO, 2017).

There are many factors contributing to this challenge, including, as mentioned in the first section, the lack of educational metadata at Web scale. Even just standardizing these metadata to describe the educational resources so that they can be found, proved to be a very complex problem. Indeed, a number of these largely incompatible educational metadata standards were developed (Riley, 2010), including the IEEE (2002) standard for Learning Object Metadata (LOM) and the Instructional Management System project (IMS, 2015), the Dublin Core Education Application Profile (DCMI, 2012), and the Learning Resource Metadata Initiative (LRMI, 2014). Obtaining metadata of the necessary quality also proved to be quite challenging, as Doctorow (2001) remarked already, and which is a persisting problem (Dietze et al., 2017). Beyond technical challenges, another obstacle is the potential incompatibility of sharing metadata with the business model or organizational strategies of the existing actors (Campbell and Barker, 2014). For example, there were cases, such as a commercial oriented OER search portal (OpenEd, 2016a), where the educational resources were “free”, but the metadata to describe and discover them were considered intellectual property. Or, on the contrary, cases where the metadata were “free”, but were used inappropriately to setup a storefront to attract customers to buy commercial resources (Petrides, L., 2013). Finally, when these metadata are available, users do not necessarily care about using them. Thomas et al. (2012) report indeed that users are not that keen to use available metadata, and that they usually attempt to make use of very simple keyword searches to reach their objectives.

Another challenge is that discoverability is also affected by the context. While the search engines most frequently used by educators support generic functionalities, discoverability, as it will be extensively discussed, should not be considered as an isolated self-standing activity, but in the context of broader domain-oriented tasks (Wilson et al., 2010). These tasks, in turn, are inevitably influenced by organizational and cultural aspects, that need to be taken into account. In this regard, the activities of this research first address the context of Italian high-school educators, because of personal interest, and because, pragmatically, I have more easily access to representative users for my studies. Yet, some efforts are devoted, within the limits of the available resources, to generalize the initial findings to additional contexts.

Given its importance, there have been, and there are, many attempts to address the discoverability of OERs. Relevant initiatives include plenty of projects (more than 80 just by UKOER (Barker and Campbell, 2016a)), hundreds of OER repositories with search facilities, federations of repositories (ARIADNE, 2016), even federations of federations (Globe, 2016). And yet, the problem is far from being solved. Developing yet another generic OER search portal or application would only worsen, fragmenting further the current situation. Therefore, there is no intention to develop in this study an additional OER search portal or application, but rather to propose innovative reusable solutions to be integrated in existing and future OER search applications. A positive outcome of this research would contribute to enhance their effectiveness, improving the discoverability of OER, which was identified as one of the obstacles to reap the benefits of the OER movement.

1.4 Research questions

The first aim of this research is to identify requirements – which may have not yet received sufficient attention from the research community – to support educators looking for educational resources. Having identified the requirements, the second – fundamental – aim is to suggest suitable strategies to satisfy them, and evaluate their usefulness. Following the identification of the functionalities/tasks, hierarchically organized in categories, the objectives will be achieved by prototyping and evaluating user-oriented (Karger, 2014) innovative reusable building-blocks (Geisler, 2008) supporting them, aiming to complement and enhance existing and future search applications. The overall research question is therefore twofold:

What are the main tasks associated with OER discovery, and how can educators be supported in performing them?

The overall research question is organized in the following sub-questions:

RQ1: What tasks do teachers need to carry out in relation to OER discovery, in the framework of the OER life cycle?

RQ2: What is the relative importance of tasks and categories identified, and which aspects deserve further investigation?

RQ3: What are the general requirements for tools supporting the aspects of interest identified?

RQ4: Which tools could support the identified requirements, and how do teachers respond to them?

RQ5: What barriers are restricting the deployment and uptake of OER discoverability tools and how can they be overcome?

These research questions are further subdivided and operationalized in the specific research activities.

1.5 Dissertation structure

This dissertation is organized in eleven chapters, as shown in Figure 1.

This introductory chapter, Chapter 1, identifies the goal of the research, motivates its relevance, clarifies the basic terminology, outlines the research questions, and provides an overview of the dissertation structure.

Chapter 2 briefly reviews the main areas of research involved, identifies the research gaps, and establishes the need for a domain-oriented task analysis. It shows in particular the need to focus on exploratory forms of searching, to shift the attention from traditional search to the context where lower-level search problems are originated, and to support high-level tasks in the specific user domain.

Chapter 3 presents and justifies the main research methodology adopted, Design Science Research (Hevner et al., 2004), and describes how it applies to this doctoral research. This methodology foresees, in particular, the identification of a challenge and the design and evaluation of a series of prototypes to address it incrementally.

Chapter 4 describes the task analysis carried out to preliminarily identify major educators' main tasks that OER search / discovery applications should support. This analysis was based on a focused review of the research literature and existing applications, interpreted through Information Foraging Theory (Pirolli and Card, 1999). It produced a domain-oriented task-taxonomy, providing a general framework open to modifications and extensions.

Chapter 5 presents the first empirical study: the evaluation of the task-taxonomy previously produced. The evaluation was based on a quantitative and qualitative analysis of data collected from a restricted but representative sample of experienced educators reviewing the task-taxonomy. This study identified the opportunity to support directly, at a declarative level, a new set of task/domain-oriented expansion by similarity operations.

Chapter 6 presents the design and evaluation of the first prototype Injector, based on requirements and insights that emerged from the previous study. Injector identifies educational resources directly in Google SERPs (Search Engine Result Pages), the platform used by most educators to search for OERs. It extends the original Google snippets, by injecting domain-oriented descriptive metadata and expansion/discovery functionalities to identify similar resources; the similarity metric is based on the transparent use of alignments to educational frameworks. The feedback from its formative heuristic evaluation identified the main challenge of sparsity, that is, the intrinsically limited number of educational resources identified within Google results pages, and the limited number of expandable resources.

Chapter 7 presents the design and evaluation of a second prototype RepExp, developed to address the challenges identified in the previous cycle. The challenge of sparsity was addressed by increasing the relevance of the starting SERP, transparently replicating the initial Google search in a large hub of educational resources. Additionally, RepExp offers the possibility of obtaining similar resources, starting from any existing educational resource, with a Query By Examples approach. RepExp was evaluated again with a discounted heuristic evaluation. Results indicated that the problem of sparsity had been effectively addressed. Yet, now that plenty of resources could be returned by the prototype, new, unexpected challenges emerged. In particular, participants showed difficulties in dealing with frequently large result sets.

Chapter 8 describes the design and summative evaluation of the last prototype, Discoverer, which once again addressed the challenges previously identified, as well as the opportunity to better support educators in their high-level domain-oriented tasks. The new key feature of this prototype was to present users with similar resources grouped in three expandable clusters of different degrees of similarity, offering educators a quick overall view of the resources available, and letting them balance

similarity versus diversity, according to their specific educational goal. A summative evaluation was carried out involving twenty-nine educators from three different Countries. It provided evidence that educators appreciate the proposed strategy, mainly because it is strongly exploratory oriented and domain-oriented. Its domain orientation, in particular, supports educators in their high-level tasks, such as the personalization of education.

While the previous research activities and related chapters focused exclusively on educators, Chapter 9 describes a new activity which collected further feedback from expert stakeholders more representative of the wider OER ecosystem. Feedback, mainly collected through semi-structured interviews, concerned pros and cons of the proposed strategy, barriers and suggestions for its adoption in an operational environment, as well as broader aspects related to the sustainability of the OER ecosystem more in general.

Chapter 10 discusses in detail possible future research activities. These mainly concern improvements and extensions to the prototypes previously developed, in particular to the similarity metric adopted, as well as further evaluation activities to better support the generalizability of the results obtained. It suggests, in particular, to integrate the prototype in the future OER search engine of Creative Commons (Creative Commons, 2017), exploiting a service oriented open architecture. Additional suggestions concern the symmetrical challenge of metadata production and standardization, and the identification of further user requirements.

Finally, Chapter 11 reports the conclusions of this research: it summarizes the main findings in relation to the research questions, discusses its limitations, and summarizes further research activities.

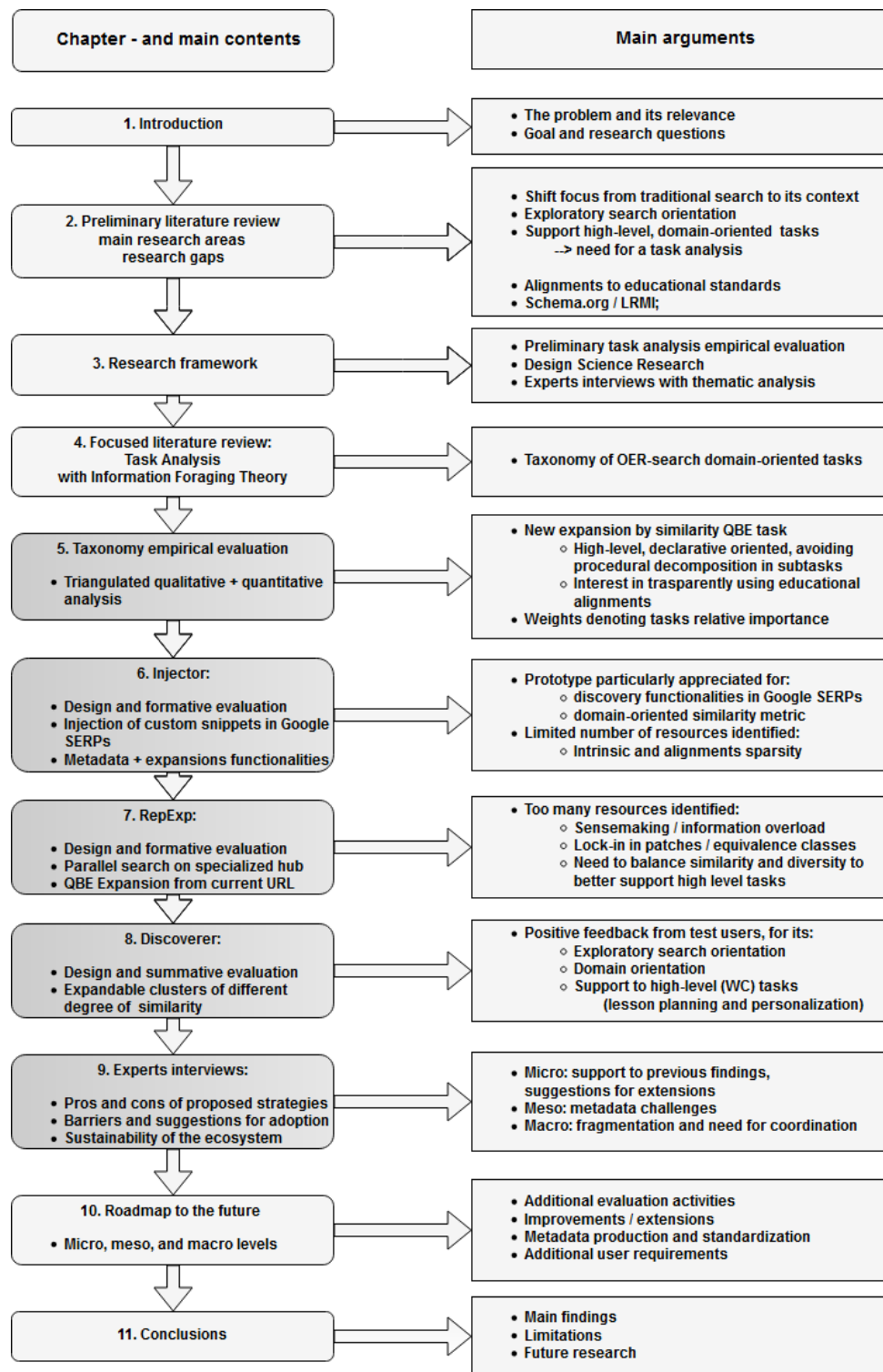


Figure 1.1 – Organization of the dissertation.

2. Preliminary literature review

2.1 Introduction

The aim of this research is to support OERs discoverability: there is a vast literature on the subject, encompassing multiple research areas (Kules and Shneiderman, 2008). This first chapter presents a preliminary literature review, following the historical evolution when possible, outlining the main research areas involved. The goal is to identify the most relevant aspects for the objectives of this thesis, and the opportunities for further research (research gaps). Once the scope of the problem is identified, Chapter 4 provides a more focused literature review. In particular, this chapter establishes the primary need to conduct a domain-oriented task analysis. The task analysis, based on a focused review of the literature and existing systems, interpreted with a behavioural model, is carried out in Chapter 4. Further analysis of the literature about specific aspects arising from the experimental activities, is discussed in the relevant chapters as needed. For example, the necessary literature concerning the need to balance similarity and diversity, or about mono-dimensional clustering techniques, is discussed in Chapter 8.

2.2 Traditional educational metadata

The main traditional strategy to try solving the problem of OER search (including discovery), consists in specifying suitable “metadata” to describe the resources from the educational point of view. Decades of attempts to even just standardise these metadata have met with mixed success, resulting in a landscape of many incompatible standards (Riley, 2010). There were also efforts to make different standards “interoperable”, so that different systems could exchange and interpret the data consistently, but they substantially failed (Nilsson, 2010). Nilsson argues that the major obstacle is not linked to syntactical aspects (format), but to the semantics or interpretation of the metadata. Offersgaard and Hansen (2016, p. 2510) too, for example, agree that according to their experience, semantics is the major problem, and suggest that “agreeing upon standards, making clear definitions of the semantics” is necessary to make interoperability successful. An additional major challenge is the well-known unwillingness by authors to provide metadata (Doctorow, 2001; Atenas et al, 2014). This may be due to a fundamental usability issue of metadata standards: authors of OERs do not have the required knowledge about metadata nor about metadata standards (Ushakova, 2015). Consequently, even later standards such as the IEEE (2002) Standard for Learning Object Metadata (LOM), IMS (2015), and the Dublin Core Education Application Profile (Sutton and Mason, 2001; DCMI 2012) could not fundamentally improve this situation (Barker and Campbell, 2016b). On the contrary, developing additional standards increases competition, fragments the solution space, and exacerbates the problem.

Figure 2.1 shows, as an example, a schematic representation of a few elements of the LOM data model, which is adopted, with some variations, in many modern search platforms, for example OERCommons (discussed in Chapter 4).

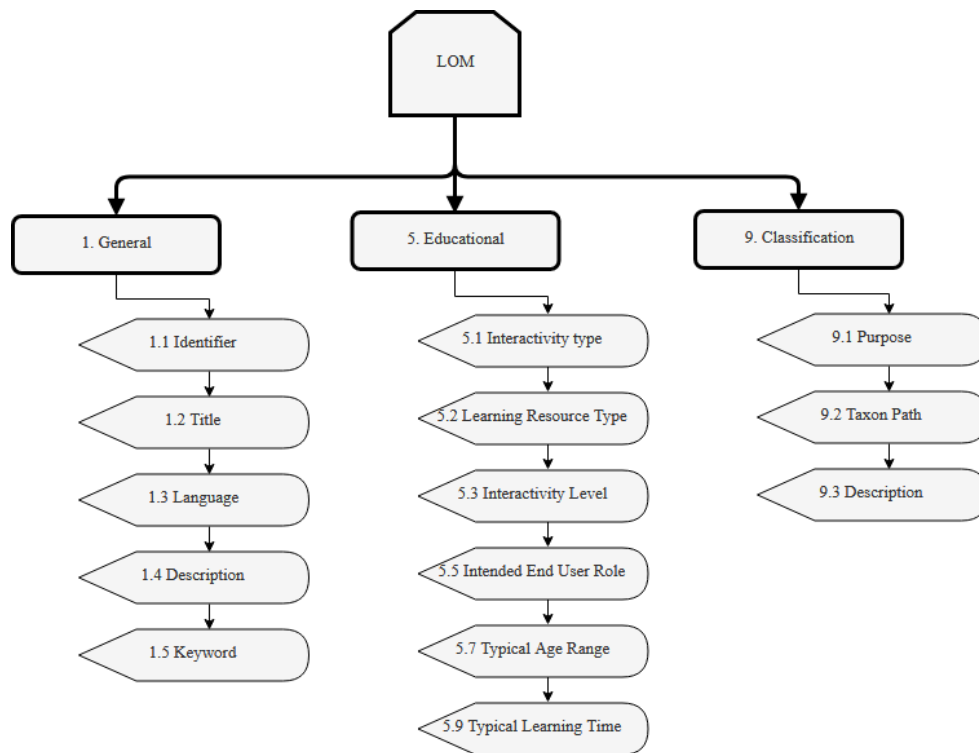


Figure 2.1 – A subset of the LOM data model.

As it can be seen, LOM envisages the specification of educational characteristics with properties such as Interactivity Type or Typical Age Range. The property Interactivity Type can assume the possible values active, expositive or mixed. “Active”, as opposed to “expositive”, signals that the learner is required to provide meaningful input, as in an assessment test. The property Typical Age Range indicates the age of the intended target audience. The element Classification makes it possible to associate a resource, for example, to specific learning objectives in an existing taxonomy. These properties would be very helpful to support the search for educational material, because users could target specific resources of their interest by precisely specifying their desired characteristics, for example suitable for a certain age, covering specific learning objectives, and requiring a maximum amount of learning time.

While history shows that a single common standard has been, so far, wishful thinking, a single standard might not even be the best solution (Downes, 2003), given that it is arguable even whether a resource is “educational” or not, for example in the case of primary source materials that can be used also to support learning. Hence, a better strategy could be 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).

2.3 Linked data and semantic search

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 WWW 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). Linking documents and data on the Web with typed links, for example via RDF triples (Cyganiak et al., 2014), makes it possible to perform searches that are not simply based on keywords-matching but on their semantics.

Concerning search for educational resources, Nilsson (2010) claims that Linked Data (LD) technologies could represent a possible solution to support the heterogeneity of the Web previously discussed, facilitating the interoperability (harmonization) of the many different vocabularies. However, D'Aquin et al. (2013) too, agree that LD makes educational information easier to aggregate and consume in principle, yet they acknowledge that so far the existing datasets are quite isolated. An additional challenge, raised by Karger (2014) in the more general LD research context, also applies here: he argues convincingly that most initiatives address the underlying technologies, while there is an urgent need for an increased focus on user applications and their evaluation. This supports the need for usability studies, as further discussed in Section 2.9.

The Web of Data is a strong motivator to explore new search strategies, “that go well beyond finding documents by way of keyword search” (Wilson et al., 2010, p. 6). Early attempts to search OERs using these technologies, include the OER-CC ontology (Piedra et al., 2010) and DiscoverEd (Yergler, 2010). OER-CC is one of the attempts to solve the problem by developing one more additional ad-hoc ontology. As noted before, different ontologies could in principle coexist in the semantic Web, yet their proliferation contribute to increase the complexity of the solution space, fostering the creation of different, isolated datasets. DiscoverEd, a search prototype developed by Creative Commons, avoided to develop any new ad-hoc ontology, and aimed explicitly to be as open as possible supporting different metadata formats and avoiding to strictly normalize the vocabularies used by different curators. However, while Yergler (2011) claimed that DiscoverEd demonstrated the potential of using structured data to improve searching, he argued that the “many attempts to describe educational resources and how they relate together in a complete, rigorous manner” have failed, because none of them gained widespread adoption.

There are more recent search-related initiatives strongly oriented to Linked Data, such as mEducator (Yu et al., 2011), or activities within LinkedUp (2014). However, major players in the area, such as GLOBE (2011) or ARIADNE (2016), mostly LOM oriented, do not employ at the moment these technologies. The Learning Registry (2016a), that had an approach pursuing a “schema agnostic” strategy by standardizing only the envelope containing the payload with the actual metadata (Jesukiewicz and Rehak, 2011), is a notable exception. Later it started to make use of Linked Data technologies, in particular JSON_LD (W3C, 2014), the Linked Data oriented version of JSON, and schema.org – discussed below.

2.4 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, is evolving 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” (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 defines an ontology to describe resources on the Web, which can be annotated embedding metadata in Web pages. This can be flexibly carried out with various custom or more general LD technologies, such as RDFa (Adida et al., 2015) or, more recently, JSON_LD (W3C, 2014). Rather than directly reusing the wealth (and complexity) of existing semantic standards and ontologies, schema.org opted to start ex-novo. The pragmatic intention was to keep the mechanism as simple as possible, aiming to have it adopted at scale (Douglas and Brickley, 2013).

Schema.org has been later extended with the vocabulary developed by the Learning Resource Metadata Initiative (LRMI) (LRMI, 2014), aiming to support end-users in searching and discovering educational resources. LRMI was originally led by the Association of Educational Publishers and Creative Commons; since then it was moved under the responsibility of the Dublin Core Metadata Initiative. LRMI, of course, has been heavily inspired by previous experiences and standards, in particular the IEEE (2002) Standard for Learning Object Metadata, IMS (2015), which are essentially guidelines for using LOM, and the Dublin Core Education Application Profile (Sutton and Mason, 2001; DCMI, 2012).

The LRMI specification is a collection of properties to describe educational resources, which in schema.org belong to the class `CreativeWork`. Additionally, it specifies the two supporting classes `AlignmentObject` and `EducationalAudience`. The LRMI properties that apply to a schema.org `CreativeWork` class are:

- `educationalUse` – for example group work, assignment;
- `timeRequired` – for example 'P1H25M';
- `typicalAgeRange` – for example 12-16;
- `interactivityType` – active, expositive or mixed;
- `learningResourceType` – for example presentation, handout;
- `isBasedOn` – the resource it derived from;
- `educationalAlignment` – an alignment to an established educational framework, discussed in detail in the next session.

Being part of schema.org, additional characteristics of the educational resources can be specified with other generic schema.org properties of `CreativeWork` (including those inherited from its super-class `Thing`), such as `license`, `title`, `author`, `publisher`, and `dateCreated`.

Schema.org/LRMI has a lot of potential to improve the effectiveness of OER search. First, it is exploited by Google, which is the engine used by most educators to search for educational resources (LRMI, 2013b). Second, it should be widely adopted, because Web developers are motivated to use it knowing that it is recognized by major search engines. Third, the strategy to adopt a simple (and tightly controlled) ontology should further contribute to its wide adoption, because it facilitates its use, reducing one of the obstacles of traditional LD (Guha et al., 2016). It is possible, indeed, to build a Google Custom Search Engine (Google, 2017) that can exploit schema.org structured data (metadata) for filtering, as it was done in Kritikos (Bullough et al., 2013). With this technology, users could search, for example, for all the educational resources (CreativeWork) having an “active” InteractivityType, for students of a given TypicalAgeRange, which contain the free keyword “evaporation”. Finally, the visualization of the results could also be enhanced, compared to the traditional Google generic snippets, that is the limited amount of information shown in its result pages for each resource identified (usually the headline, URL, and a tiny text fragment where the search keywords occur). Indeed, it would be possible to include more specific structured data such as, in the case of OERs, their educational metadata, thus obtaining more informative, so-called “rich snippets”.

However, while the lightweight metadata of LRMI/schema.org can potentially help improving searching with generic engines, metadata need to be produced first, to spur the development of innovative applications using them. For this reason, there are initiatives pushing major actors to map their proprietary metadata schema to LRMI and make them public, notably OERCommons (Campbell, 2014a), Gooru (Campbell, 2014b), Merlot (Campbell, 2014c), and Curriki (Campbell, 2014d). And yet, at the time of writing, there is still a surprisingly limited volume of LRMI tagged educational resources, even if it is improving: 51% from 2014, 139% from 2015 (Dietze et al., 2017).

2.4.1 LRMI alignment to educational frameworks

Particularly relevant for this research is the so-called “killer” feature of LRMI (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) describes clearly the rationale behind the specification of this metadata and how it should be used. Figure 2.2, in particular, represents graphically the alignment of an educational resource (a CreativeWork in schema.org parlance) to a specific entry (node 5) of an educational framework, via an “AlignmentObject”.

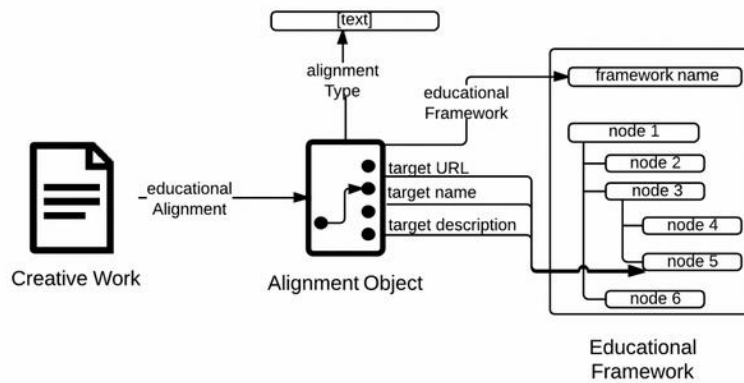


Figure 2.2– Alignment of a resource (Barker, 2014).

A notable example of an educational framework is the Common Core State Standards (CCSS) (Porter et al., 2011) in the United States, which defines detailed learning objectives in Maths and English at High School level. Another relevant example, also at High School level, is the Next Generation Science Standards (NGSS) (Pruitt, 2014). Schema.org does not support the specification of these frameworks: it just lets users define associations between an educational resource and a given “standard” (entry) in an existing educational framework. As represented in the figure, this association is indirect through an AlignmentObject. The AlignmentObject reference to a given standard can be specified in different ways; as an example, a node in the CCSS educational framework could be indicated by its URL (targetURL) “<http://www.corestandards.org/Math/Content/HSS/ID/A/1/>”, its formal name (targetName) “CCSS.Math.Content.HSS.ID.A.1”, or even its extended description (targetDescription): “Represent data with plots on the real number line (dot plots, histograms, and box plots)”.

The concept of “educational framework” is sufficiently broad to make it possible to specify alignments to learning objectives or outcomes, but also professional competency standards, subject taxonomies, or even the target academic level or reading difficulty schemas. Indeed, the alignmentType property of an AlignmentObject makes it possible to specify whether the alignment specifies the learning objective of the resource, the prerequisites necessary to use it effectively, and also other cases indicated by its recommended possible values: “assesses”, “teaches”, “requires”, “textComplexity”, “readingLevel”, “educationalSubject”, and “educationLevel”.

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 IEEE 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 for resource compliance, that is to certify the compliance of their teaching activities to a pre-specified standard curriculum. Yet, despite being so important, alignments are still rarely used in schema.org (Cortinovic, 2016), sometimes are used incorrectly, and more research is needed to fully exploit their potential (Dietze et al., 2017). This includes dealing with long-recognised problems such as the existence of many different (again)

educational frameworks, different ways to express the same alignment, the equivalence among standards in different frameworks, the degree of alignment to a given standard (Sutton, 2008).

2.5 Social network based approaches

There are other techniques to support search and discovery by exploiting crowdsourced social data and paradata left by users. Drachsler et al. (2012) define social data as data generated by people *intentionally*, contributing information about a resource, for example with tags, ratings, bookmarks, and comments. They contrast social data with paradata, which correspond to *unintentionally* contributed and automatically tracked traces of users' interactions with the resources, such as login, navigation history, search history, accessing content or metadata, or memorization of resources in a personal list. An alternative classification (Thomas et al., 2012), distinguishes between *user oriented* data (social data) versus *resource oriented* data (paradata). Cechinel et al. (2013) argue that as most repositories and federations have community dimensions, there is an increasing volume of these data available.

Social tagging, initially popularized by Flickr and Delicious, is a widely adopted technique in the context of the so called Web 2.0, to support the discoverability of online content such as pictures, video, recipes, comments, etc. A large number of users, rather than a few specialists, tag online resources by freely associating one or more keywords; the keywords generated can then be exploited to classify, organize and discover the resources. That is, users can ask a search platform supporting these functionalities, for the resources that were tagged by other users with one or more given tags. This technique, due to its simplicity and effectiveness in leveraging the “wisdom of crowds”, is now adopted also for classifying and discovering OERs in most OER repositories.

The problem with this approach, however, is that people may, for example, use the same tag with different meanings, different tags having the same meaning, or use tags at different levels of abstraction (Andrews et al., 2012). Indeed, the tags just indicate a generic, subjective and unspecified association between a resource and a keyword, failing to capture its meaning or semantics. For example: does “simulation” refer to the interaction style of an OER or is it its subject?

Similarly, easily available paradata such as the number of downloads, could be taken as indicators of the level of interest in a resource, hence of its quality. These information, therefore, could be used to rank resources by their level of quality, improving search effectiveness. Yet, a resource could be more frequently downloaded because it is well-advertised or presented, rather than because of appropriate educational qualities. Alternative metrics, such as the frequency of use of a resource, or even better the actual performances obtained by students using the resource, could be more appropriate indicators of quality. These data, however, are frequently confined in local learning management systems and rarely shared at global scale.

Despite some limitations, social data and paradata have a lot of potential to support advanced searching and discovery applications (Okada et al., 2012; Thomas et al., 2012; Ferguson and

Buckingham Shum, 2012), complementing descriptive (“authoritative”) metadata. This is particularly evident in the area of recommendation systems, discussed below.

2.6 Recommendation and blended search/recommendation systems

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 in developing various types of recommendation systems suggesting (“pushing”) personalized items of potential interest to users (Dietze et al., 2014). The term pushing refers here to the system taking a proactive role, providing information not directly requested by the user, but likely of her interest. 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).

Tang and McCalla (2009, p. 10) argue that recommendation systems in the area of education, however, are quite different from those in the commercial domain, “where user likes are all that matters”. They argue that the majority of recommenders are based on user-user or user-item correlations, without taking into account the context where those correlations were originated. Hence, Verbert et al. (2011) claim that the challenge here is to adapt the algorithms successfully used in the commercial domain, to the pedagogical domain, where the relationships must be based instead on pedagogical aspects.

Recommendation algorithms can be generally classified as collaborative or content-based filtering. Collaborative filtering algorithms (user-based or item-based (Sarwar et al., 2001)) provide recommendations on the basis of similarities about user preferences. Content-based filtering algorithms recommend items on the basis of their specific characteristics. Recommendation systems can be classified according to other aspects, for example Heitmann and Hayes (2010, p. 78) emphasize the difference between open (Internet wide) or closed (limited to a single portal), and identify an additional category of knowledge-based systems, which suggest “items based on inferences about user needs and preferences”. One of the main challenges indeed is to provide recommendations relying on information/knowledge at Web-scale, which is tentatively tackled, in many projects, with Linked Data oriented techniques (Dietze et al., 2014). ODS (Drachsler et al., 2012), a potentially promising project aiming to support the discoverability of educational resources, planned to develop a social metadata cloud with this kind of information about their 1.5 million resources. Another major actor in the area, the Learning Registry (2011), collected and made available a rich set of paradata at Web scale, using a schema-less strategy, but with an increasing orientation to LD technologies.

This research area shares some of its challenges with the research area of metadata standards, in particular the problem of cross-repository harvesting and aggregation, common data schemas, a uniform resource identifier service (because the same data are harvested and re-harvested), modular and distributed architecture (Drachsler et al., 2012). There are also more specific challenges, such as those related to the limited amount of data available (sparsity). Collaborative systems, indeed, provide recommendations on the basis of user preferences, which they collect incrementally: hence, there is a limited amount of data they can exploit at the start of their activities (“cold start”).

Concerning evaluation, as clearly argued by Herlocker et al. (2004), and reiterated by Manouselis et al. (2011), the evaluation of these systems should move from a traditionally oriented focus on technical measures typical of Information Retrieval (IR) research, such as accuracy, to more exploratory and user oriented metrics. The starting point for the design and evaluation of these systems, indeed, is a clear identification of the user-oriented tasks to be supported.

2.7 From keywords-matching to Exploratory Search

Google, according to the literature (LRMI, 2013b; Abeywardena et al., 2013), is the most popular tool used by most educators looking for educational resources. This is hardly surprising, because search systems producing ranked results starting from simple keywords introduced in their textbox, have been so successful in framing our mental model of the Web. Indeed, as argued by Schraefel (2009, p. 52), “the success of the search engine may be our Newtonian paradigm for the Web”. Abstracting from many details at this stage, Google and similar keyword-based search engines basically identify related documents by matching the keywords introduced by the user and the keywords contained in the documents available in the Web. These aspects are mainly addressed in the well-established research field of IR.

IR mainly focuses on retrieving as many relevant documents as possible (recall), including as few non-relevant documents as possible (precision), in relation to a user query. Hence the notion of relevance is at the center of information retrieval. The relevance of a document can be estimated with various techniques, starting from a simple frequency count of meaningful words in the query and in the documents (Van Rijsbergen, 1979). These fundamental techniques have been improved in many ways, for example considering morphological variants, co-occurrence, and proximity (Singhal, 2001). Google, for example, adopts PageRank, anchor text, and proximity information (Brin and Page, 1998) among many other techniques. PageRank, in particular, is a very successful algorithm that takes into account the number of incoming links to a web page, as well as the importance of the incoming pages, to produce a value which is an indication of the importance of a given page. This value is then used to rank the results of the keywords-matching process, in order to present to users the most important pages first. And yet, while this ubiquitous, basic keywords-matching mechanism, as Wilson et al. (2010, p.5) put it, “is really good”, they immediately add: “For what it does”. Indeed, as introduced in Chapter 1, there is a need to move from this kind of “look-up” type of search, to support a more “discovery/exploratory” oriented one.

While lookup and discovery are lumped together under the umbrella-term “search”, as discussed in the introduction, lookup is more precisely related to the research field of IR, while discovery is related to the more recent field of Exploratory Search (Marchionini, 2006). White et al. (2007, p. 2877) already stated that there is a need to move from look-up search, domain of traditional IR, to exploratory search, “to support search-driven information exploration activities” in a “more continuous exploratory process”. Wilson et al. (2010, p. 9) too argue that certain search 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”. This situation calls for a far more articulated concept of search that is well beyond simple keyword search, and represents a considerable research opportunity: 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”.

2.8 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 domain-oriented tasks, that is, the specific high-level problems which users (educators in this case) need to address. For example, 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) too, observe that traditional IR tasks are the elementary steps to achieve higher level goals, that is higher level Information Seeking (exploratory type of search) or Work Context tasks (Byström and Hansen, 2002). Indeed, they define search “as the set of activities that take users from identifying a problem all the way to achieving their goals” (p. 12). Again, they claim that “in Information Seeking and Information Retrieval studies tasks are coming to play a central role” (p. 240), and that work tasks, in particular, are the context for most information seeking activities. Kabel et al. (2004, Section 2) also emphasize the importance to consider both the tasks and the domain: in particular, they claim that “the performance of the information retrieval task is inextricably bound to the work task. Consequently the performance of the retrieval task will depend on taking the context of the work task into account”.

This shift of focus from supporting traditional search to supporting tasks in the application context, is apparent also in the evolution of the Text REtrieval Conference (TREC, 2016). TREC is actively engaged in the evaluation of text retrieval methodologies, proposing tracks with specific objectives, reference queries, test dataset, and evaluation methodologies. They underlined the need for a commonly accepted list of reference tasks, and identified a set of these tasks to make possible the direct comparison of different systems. Unsurprisingly, the tasks initially defined addressed the restricted context of IR (Qu and Furnas, 2008). However, in the following years, following the conceptual expansion of the meaning of “search” from look-up to more exploratory oriented activities, the Interactive Track and the HARD Track of the Text Retrieval Conference broadened their scope to include user-system interaction, and grounded evaluation on users and context. This

was an important step, even if Wilson et al. (2010, p. 67) noted that “the information needs in these tracks are still narrowly expressed in terms of documents to be retrieved, without reference to a higher-level information need.” Yet, later tracks from TREC did take into account higher level domain-oriented tasks, such as the “Clinical Decision Support Track” in 2014. The evolution path of TREC clearly shows a shift of interest from traditional search tasks to tasks in the wider domain-oriented context.

All the above considerations show an important trend in moving from traditional restricted information look-up, addressing lower-level tasks, towards the wider context where those search tasks are originated, that is, towards high-level tasks in the users’ domain. This shows an increased attention to the *use* of search results, reflected in the Information Seeking model proposed by Abraham (2013), which foresees the following four key strategies: Search, Evaluation for Selection (selecting resources to examine), Evaluation for Use (further analysis of potential useful resources), and the last prominent level *Use*.

2.8.1 Organizational and cultural aspects: educational standards in Italy

Considering search in the broader context of domain-oriented work tasks has important implications for this research. Wilson et al. (2010, p. 14) write that “the key benefit of this holistic view is that it maintains an understanding of the situational nature of information and that search activities are performed in the context of a larger work task, and even cultural influences”. Indeed, search is influenced by organizational and cultural aspects, which need to be taken into account: “work tasks are situated in the context of work organization and reflect organizational culture and social norms, as well as organizational resources and constraints” (Wilson et al., 2010, p. 19).

In this context, it is important to remark that, while there is a strong attention to the use of educational standards / frameworks such as CCSS or NGSS in the USA, and Anglophone countries in general – as previously discussed in Section 2.4.1 – there are other countries that do not use formal educational standards at a similar level of detail. The relevant aspects of the Italian case are discussed here, because it is the target context of most interest to the author and is addressed by this thesis.

In Italy there are no detailed formal educational standards at national level, but just general indications (“guidelines”). The Decreto Interministeriale 211 (MIUR, 2010), as an example, contains the national indications with “specific learning objectives” for Italian Licei (high schools) which are, despite their name, very generic. Rather than detailing the contents in terms of educational standards, the guidelines actually establish the general topics and quantify the associated resources, that is the corresponding amount of lesson hours. These guidelines can then be modified by every educational institute in the context of their so called “autonomy”, which gives them the flexibility to adapt those contents and up to 30 % of the specified resources, to fit the educational needs of the local context (Decreto del Presidente della Repubblica n. 275, 1999). The resulting indications need then to be detailed by every teacher, in the context of their “teaching freedom”, which gives them further

flexibility (not specified quantitatively) in organizing and defining the actual detailed activities (including the curriculum) in each class.

This flexibility and uncertainty, compared to Anglophone countries, is a significant cultural and organizational aspect which is expected to affect the way educators in these different contexts search and discover educational resources. It is also important to consider the possible evolution of the situation, to anticipate the possible corresponding evolution of users' needs. In this regard, there is a slow but ongoing effort, in Italy, to define more precisely the learning outcomes: "increasingly, competency-based approaches and learning outcomes are being introduced as a guiding mechanism to inform general education reforms" (European Commission, 2011, p. 9). This is driven, mainly, by the need to harmonize qualifications (European Qualification Framework) in a European-wide educational system (Adam, 2004; Davies, 2017).

2.9 Exploratory search evaluation: focus on users and their tasks

In the IR field there are well established metrics that can be used to evaluate a search system. 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, these methods are system-oriented (Borlund, 2003), while in the case of Exploratory Search, queries might not even be precisely known in advance, and evaluation needs to focus on users and their context. This makes evaluation of exploratory systems a research area on its own (Kules and Shneiderman, 2008; Wilson et al., 2010), where the traditional metrics of precision and recall, completion time or number of errors, are no longer appropriate. Lee (2017) for example, designed and evaluated an exploratory oriented search system for Java code. He established experimentally that in many cases, especially when the search task is less structured and focused, the time to carry it out was not correlated, or in some cases it was even negatively correlated, to the quality of the user experience. This is because a longer completion time can indicate, for example, deeper critical thinking and a more rewarding learning experience.

The first essential step to properly evaluate solutions aiming to support OER discoverability is, therefore, to identify the user and domain oriented tasks they are intended to support. Wildemuth and Freund (2009) argue that tasks are fundamental in evaluating search applications. 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".

The evaluation of Exploratory Search systems needs to take into account the critical role of users in their context, hence it is strongly related to the research area of usability evaluation (Madan and Dubey, 2012). Nielsen (2012) considers that "usability" should not include "utility". He argues that while both attributes contribute to the usefulness of a system, utility assesses at what extent an application supports the functionalities required by the users, while "usability" is only concerned about how they are easy and pleasant to use. On the contrary, this thesis considers "utility" as a fundamental component of "usability", coherently with the ISO standard 9241-11 (ISO, 1998), which

defines usability as “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”. Therefore, “usability” as intended here, is firmly grounded in the domain-oriented user-tasks that the applications are supposed to support.

2.10 Conclusions: user and domain-oriented tasks

Figure 2.3 represents a conceptual map of this chapter. Ovals represent the broad research areas related to this thesis, arrows indicate an evolution path from one area to the other.

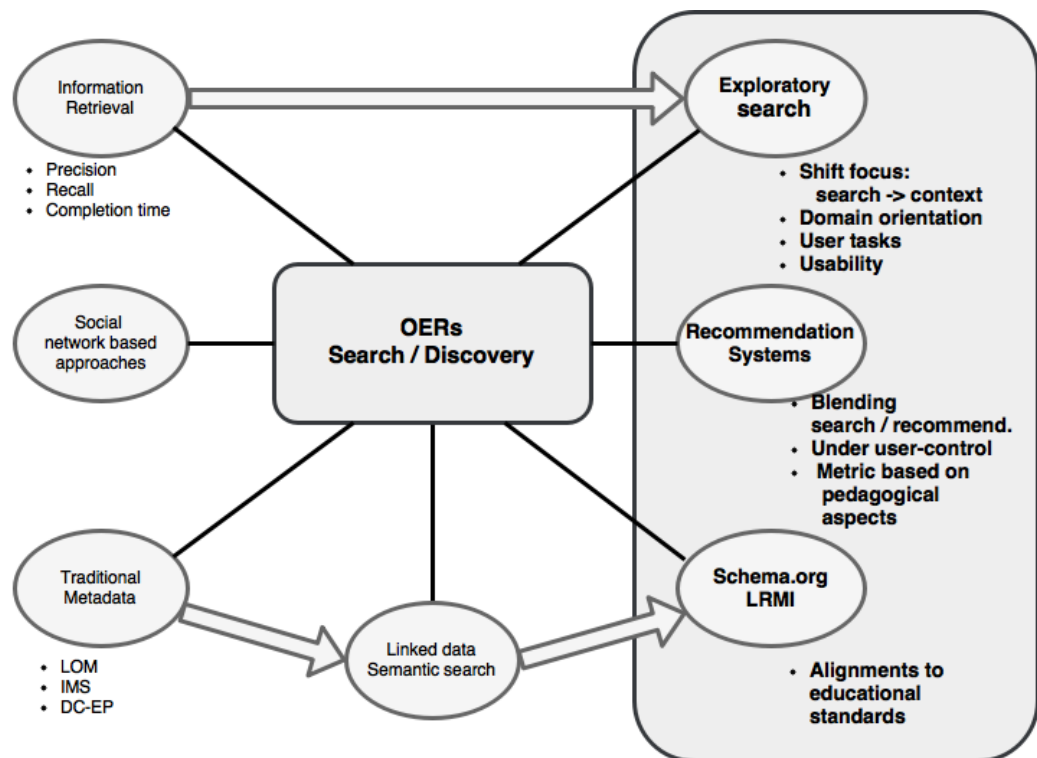


Figure 2.3 – Main research areas and aspects deserving further research.

The rectangle on the right summarizes the main aspects that are considered, in the literature, worth of further research, and that are addressed in this doctoral research:

- Need to shift focus from search to the broader application context, hence supporting domain-oriented high-level tasks;
- Need to focus on exploratory oriented forms of search;
- Need to focus on user-oriented usability studies;
- Opportunity to blend search and recommendation systems, where metrics are based on pedagogical aspects;
- Emerging importance of schema.org / LRMI;
- Killer feature of alignments to educational frameworks;

This chapter has therefore established the situational nature of search, which requires to focus on domain-oriented user-tasks. Hence, Chapter 4 produces, through a focused model-driven analysis of the literature, a broad domain-oriented OER search task-taxonomy. This taxonomy represents the

starting point for the experimental activities of this research. Before that, however, now that the broad research gaps have been identified, the next chapter presents the overall methodology of this research.

3. Core research framework: Design Science Research

3.1 Introduction

In research, it is advisable to identify a coherent framework in which to ground properly the research design, making explicit the assumptions about ontology, epistemology, and methodology (Creswell, 2014). This is what can be referred to as “research paradigm”, defined by Kuhn (1962, p. 45) as “*the set of common beliefs and agreements shared between scientists about how problems should be understood and addressed*”. Research methodologies in particular can be defined as “*specific aggregations of methods or techniques that might be termed ‘approaches to investigation’*” (Petre and Rugg, 2010, p. 98), which are accepted as valid by a research community.

The overall goal of this *research* was to propose a solution to a specific challenge, to be identified experimentally in the context of a broad problem. The objective of the research was not so much concerned with the study of an independent reality, but rather oriented to fully understand a problem, generate possible solutions creating innovative artefacts, and evaluate their relevance and usefulness. This calls for the flexible use of a variety of approaches, adopting the most suitable one for each specific purpose (Creswell, 2014), and is very much in line with pragmatism: not rigidly committed to any specific philosophy (Mackenzie and Knipe, 2006), less concerned with truth and reality, but rather more concerned with “what works”.

3.2 Research and design: tensions and roles

This *research* aimed to propose a solution to a specific challenge by *designing* a software prototype. There seems to be a potential conflict here, as research and design are apparently oriented in opposite directions (Zimmerman and Forlizzi, 2014): natural versus artificial phenomena of interest, descriptive and explanatory versus prescriptive types of theory, answers to questions / knowledge versus solutions to problems, understanding the world versus shaping the future, novelty versus usefulness, generalizability versus applicability, universal versus particular, existing versus non-existing, past and present versus future, and even values such as truth and understanding versus control of the context. As a result, Becker and Niehaves (2007) remarked that there is no unique clear-cut common view on a research paradigm to frame the design of software artefacts. Yet, March and Smith (1995) already, in their influential paper, indicated *build* and *evaluate* in design science, paralleled by *discover* and *justify* in natural sciences, as the two key design activities to reconcile the potentially conflicting points of view of knowledge-production and knowledge-using oriented disciplines.

While a certain tension between research and design is unquestionable, design plays multiple roles in research, which have been extensively discussed during the last 50 years (Offermann et al., 2009). A number of methodologies integrating the two aspects have been recently proposed: these methodologies are much younger than the more established methodologies in natural sciences or even social sciences. While slightly different expressions are used in different disciplines, they all share the similar purpose of *designing* solutions to problems, through a rigorous *research* process as

discussed in the next section. “Research through design” was used in arts and design by Frayling (1993). “Design science in information science research” was used in the widely cited paper by Hevner et al. (2004). This was later renamed as “design science research in information science” (Hevner and Chatterjee, 2010), to better distinguish research *on* from research *through* design. Finally, “design-based research” is used in the educational sector, where it similarly focuses on the design – in this case of interventions – aiming to have a strong practical impact while advancing theory (Barab and Squire, 2004).

3.3 Distinguishing research from design practice: knowledge production, rigour, evaluation

The above expressions include both the terms “research” and “design”, and are clearly distinguished from design practice. Hevner et al. (2004, p. 98), in their foundational paper, specify that “*evaluation, contributions [to knowledge], and rigor are especially important in providing this distinction*”, and include them explicitly in the guidelines they propose.

First, indeed, research aims to produce knowledge, rather than successful products (Zimmerman and Forlizzi, 2014): “*design science research involves the creation of new knowledge through design of novel or innovative artifacts*” (Vaishnavi and Kuechler, 2015, p. 1). Consequently, these methodologies attempt to make explicit the potential contributions to knowledge that can be expected by applying them, keeping into account that the aim is not so much about finding a universal truth, but mainly to (potentially) improve a particular situation, producing knowledge that can help professionals to design solutions for their field problems. Hevner and Chatterjee (2010), in particular, distinguish between *design knowledge*, the overall output supporting practitioners designing suitable solutions, and *constraint knowledge*. Constraint knowledge, derived from things that did not work as previously assumed, narrows down the space of possible solutions, and is generally used to modify the initial hypothesis and guide a new improved design. Zimmerman and Forlizzi (2014) indicate additional contributions to knowledge that can be potentially generated, such as an improved understanding of the problem area, and the anticipated understanding of opportunities that near future technological progress will open-up – which are particularly relevant in this doctoral research.

Iivari (2007), writing about “constructive research”, again aiming to build artifacts to create knowledge about how to solve specific problems, pinpoints a second key difference between academic researchers and practitioners designing innovative solutions, in the level of rigor of the whole process. This is reflected in particular in the evaluation, which is emphasized in every design research methodology model. Hevner et al. (2004, p. 85) claim that “evaluation is a crucial component of the research process”, and identify evaluation aspects useful in this research such as functionality and usability, as well as a comprehensive set of evaluation methods. Peffers et al (2007, p. 72) argue that in Design Science “research, design and the proof of its usefulness is the central component”. Venable et al. (2012) offer more prescriptive oriented guidelines, synthetizing the

literature in a useful evaluation design method and framework, while Prat et al. (2014) propose a detailed and comprehensive taxonomy of specific evaluation criteria and evaluation methods.

3.4 The adopted core framework: Design Science Research

Design Science Research (DSR), which Gregor and Hevner (2013, p. 337) consider “an important and legitimate Information System research paradigm”, was a natural choice as the core methodology of this doctoral research. The methodology is compatible with my aim to experiment with possible solutions to a real world problem, by designing and evaluating software prototypes. The methodology is also suitable to the need to interactively explore the solutions space together with the requirements space, given that while the broad challenge has been identified, further exploration is needed to identify the requirements more precisely. The methodology is also compatible with my “formamentis” and personal background, deeply rooted in industrial R&D – where I used to develop prototypes in various sectors, from Human Computer Interaction to Artificial Intelligence and Technology Based Training, even if without the rigour of academic research. In this regard, adopting a scrutinized and recognized methodological framework, helped to avoid the risk that this design oriented research is “mistaken for poor quality empirical research or for practice case study” (Peffer et al., 2007, p. 73).

3.5 The research process in detail

The research process proposed in DSR-oriented models is organized around a core *iterative* sequence of design and evaluation cycles (Peffer et al., 2007). Zimmerman et al. (2010, p. 313), for example, define “research through design” in the context of HCI, as the “process of *iteratively* designing artefacts as a creative way of investigating what a potential future might be”. Hevner et al. (2004, p. 78) write that the “build-and-evaluate loop is typically iterated a number of times before the final design artifact is generated”. An iterative sequence of design and evaluation cycles is also suggested by the design science research methodology process model by Peffer et al. (2007), the design science research process by Offermann et al. (2009), the guidelines proposed by Hevner and Chatterjee (2010), as well as the design science research process proposed by Vaishnavi and Kuechler (2015).

The overall research process adopted in this doctoral research was therefore mainly organized as an interactive sequence of design and evaluate studies, as illustrated in Figure 3.1. This figure extends Figure 3 by Vaishnavi and Kuechler (2015), mainly by making explicit the path to the first DSR cycle, and indicating a final activity collecting feedback from a broader population. The path to the first DSR cycle articulates more rigorously the “problem centered initiation” discussed by Peffer et al. (2007), which is one of the precluding activities leading to the DSR iterations, when these are driven by a specific problem to be solved. Indeed, the input to the first DSR cycle in this research did not rely on uncontrolled intuition or creativity (Vaishnavi and Kuechler, 2015), but was firmly grounded on the research literature, and on the results of a first empirical study. The last activity collected further feedback from experts more representative of the broader ecosystem.

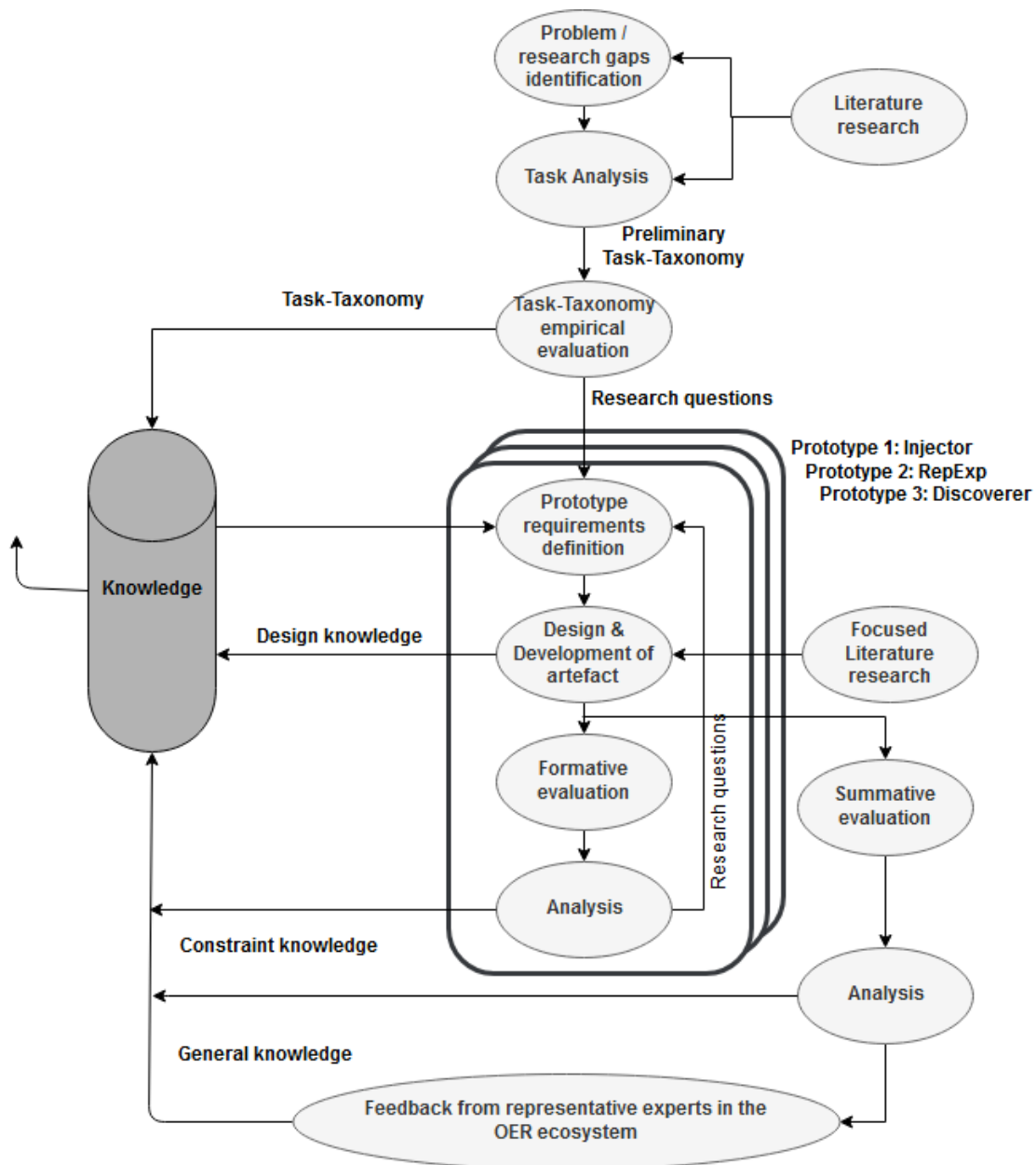


Figure 3.1 – The research process adopted in this research.

More in detail, the activities previously described in the first two chapters identified the general problem, the main research areas involved, and the research gaps (first oval from the top), establishing in particular the need to carry out a domain-oriented task analysis. Hence, in the following step (second oval), an OER search-related task analysis based on the literature was carried out, to capture the current view of the research community, and produce a preliminary task-taxonomy. In the following study, this taxonomy was empirically evaluated with actual educators (third oval), with the main objective to identify priorities, tasks and related aspects, to be further investigated.

These preliminary results could drive the first cycle of the iterative design, development, and evaluation of prototypes (stacked rounded boxes). The development of prototypes fostered a better understanding of the initial ideas: *“Computer models force precision. Implementing a theory uncovers conceptual mistakes and oversights that ordinarily escape even the most meticulous*

researchers. Major roadblocks often appear that were not recognized as problems at all before the cycle of thinking and experimenting began” (Winston, 1984, p. 7). This sequence of prototypes development and evaluations, made it possible to experiment with various solutions to support the identified tasks.

The formative evaluation in each iteration could identify new challenges, contributing “valuable constraint knowledge” (Vaishnavi and Kuechler, 2015, p. 11), useful to improve the understanding of the requirements, prune the solution space, and generate new research questions. These drove the design and evaluation of an enhanced prototype in the subsequent iteration. Once the design had stabilized, a final more comprehensive summative evaluation was carried out, involving a larger number of educators. These came, as far as possible, from different contexts (countries and teaching levels), given the expected influence of organizational and cultural aspects. Following the last DSR evaluation cycle, an additional activity collected further feedback from experts more representative of the broader OER ecosystem, to improve the generalizability of results and situate them in their wider context.

3.5.1 Evaluation in the DSR iterations

Evaluation is a key component of each DSR cycle. Its importance is emphasized in every design research methodology model, and is a fundamental aspect distinguishing design research from design practice (Hevner et al., 2004).

As discussed in Chapter 2, the evaluation of search applications at higher Information Seeking and especially Work Context levels is challenging because it “requires measuring the success of work-context style problems” (Wilson et al., 2010, p.71), and there are many correlated variables and uncontrollable factors (Kabel et al., 2004). Evaluation at these higher levels is therefore frequently carried out with usability oriented studies, collecting subjective measures such as usefulness, usability, and satisfaction (Wilson et al., 2010). Similar considerations apply to the evaluation of recommendation systems, which needs to move from traditional system-oriented to more user-oriented measures (Herlocker et al., 2004). Complying with the results of the preliminary literature review, the evaluations of the prototypes in this research were therefore carried out by means of usability oriented studies, focusing on relevant user tasks, and their context.

To improve external validity, evaluations in this thesis engaged actual educators in participatory-oriented studies (Muller et al., 1998), proposing domain-oriented and realistic search-scenarios to them, consistent with the development status of the prototypes. Educators could not be expected to have expertise in evaluation, hence they were supported in their activity by “heuristics”: specialized evaluation knowledge in the form of check-lists, which were derived and adapted case-by-case from evaluation heuristics available in the literature. These include the widely adopted System Usability Scale (Brooke, 1986), the heuristics provided 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).

It is also necessary to consider the different goals of formative and summative evaluations, which affect both the methods to be used and the number of test-users that need to be involved.

The general goal of formative evaluations is mainly to identify major deficiencies in the earlier phases of a design, in order to improve it. In this research, formative evaluations have also the objective to produce new knowledge about possible solutions to a challenge, in particular constraint knowledge. For efficiency reasons, evaluation activities should engage the minimum number of test-users sufficient to reach their goal. Nielsen (1995) argues that it is preferable to have multiple inexpensive formative evaluations along the design process, that he calls “discounted” because they involve a minimum number of users, rather than a single final summative evaluation involving many users. In this case, the activity is exploratory in nature, and qualitative oriented evaluations are more suitable to the scope.

The main goal of a summative evaluation, instead, is to collect additional supporting evidence about the relevance of the problem and the effectiveness of the proposed solutions (Venable et al., 2012). In this case a larger number of educators were involved as test-users: this made it possible to conduct a triangulated qualitative-quantitative analysis, to provide additional supporting evidence to the claims of this thesis.

3.6 Conclusions

Design Science Research was selected as the core research paradigm for this work, which was mainly based on the iterative Design and Evaluation of a series of prototypes.

The first DSR iteration was driven by an analysis of the literature and the results of a first empirical study, to preliminarily identify the specific task(s) and related requirements to be addressed. Each DSR iteration generated new research questions, addressed in the following cycle. Concerning the fundamental evaluation step in each DSR iteration, this research relied on discounted, heuristic, qualitative oriented formative evaluations of the early prototypes, and a triangulated qualitative / quantitative heuristic summative evaluation in the final cycle. The last DSR cycle was followed by a final activity to collect further feedback from a few experts in the OER domain, concerning potential barriers and suggestions for the uptake of the last prototype in an operational environment.

Now that the overall research strategy has been outlined, the next chapter derives, according to the plan, a first taxonomy of the tasks that educators need to carry out in relation to OER-search, based on a focused literature review.

4. Focused literature review: model-driven task-analysis

4.1 Introduction: a domain-oriented taxonomy of OER search-related tasks

As discussed in the preliminary literature review in Chapter 2, search, in particular exploratory search and discovery, should focus on domain-oriented user-tasks. The main goal of this chapter is to identify and classify the user tasks teachers need to carry out in relation to OERs discovery, drawing on the analysis of the literature and numerous existing initiatives. To this end, a focused review of the literature and some representative portals was carried out, suitably interpreted with a behavioural model (Information Foraging Theory), to produce a domain-oriented task taxonomy.

Figure 4.1 represents a conceptual map of this chapter. The task analysis aims to capture the current shared view of the research community, hence its input are the scientific literature on search tasks, discovery projects, and metadata standards. Information Foraging Theory (Pirolli and Card, 1999), a widely adopted behavioural model, is used to aid in interpreting and organizing the tasks. The task classification is based on, in addition to an established but general purpose context model of search (Byström and Hansen, 2002), a domain specific taxonomy purposely designed in the context of the OER life cycle. The main result of this activity is a taxonomy of OER search-related tasks, including in particular a promising class of Query By Examples (QBE) expansion by similarity operations.

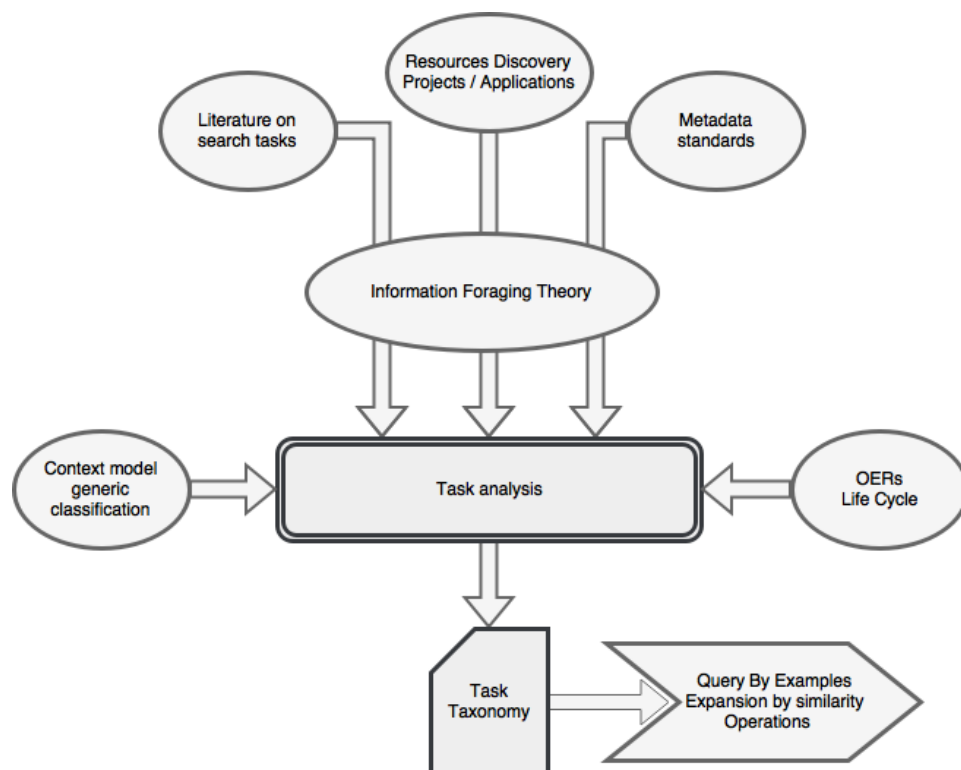


Figure 4.1 – Task Analysis conceptual map.

The main scope of the taxonomy was to provide the necessary input to the following study, where it was assessed by actual educators, with the main objective to answer RQ1 (“What tasks do teachers

need to carry out in relation to OER discovery, in the framework of the OER life cycle?”), and to identify the more specific tasks to be addressed in this research. Yet, this broad taxonomy is also a first potentially useful reusable component (Geisler, 2008) produced by this research.

The first sections of this chapter aim to identify the potential sources of tasks in the literature, and discuss their relevance. The tasks are then organized, in Section 4.7, in a domain-oriented taxonomy.

4.2 Sources of tasks: evaluation studies, use cases, and functionalities

Generic search tasks can be obtained directly from the literature, in particular from studies that make use of reference tasks for their evaluation. At the lower levels, such as Information Retrieval, tasks are context independent, so that there are strong commonalities among tasks in different domains. This makes it relatively easy to identify generic tasks in the literature. Yet, tasks at higher levels, such as Work Context, are more entangled in the specific domain – and are therefore much less easily available in the literature. Use cases, describing with examples the interactions between users and the application, are more common than tasks. Even more common are functionalities (or features), that can also be obtained from the analysis of existing applications. Tasks, use cases, and functionalities, are all useful for the scope of this activity, but their relationships in the context of educational resources need to be clarified.

The literature frequently reports use cases, rather than tasks – for example the Irma use case (Drachsler et al., 2012), Learning Registry use cases, or LRMI (n.d.) use cases. Use cases and task descriptions share the similar fundamental scope of reflecting requirements. However, while tasks are purely in the problem domain, use cases are already a step in the solution domain. Indeed, as Lauesen and Kuhail (2012, p. 17) argue, use cases “force the analyst to design a dialogue at a very early stage, in this way designing a solution rather than specifying the needs”. On the contrary, “task descriptions do not specify a dialogue but only what user and system need to do together”. Let us take as an example, the following sentence extracted from the Irma case (Drachsler et al., 2012, p. 9): “at the bottom of the resource page, Irma finds that the ODS portal has followed her query and exploration history, showing her quick links to the ten most recent objects she looked at in this session”. The core idea of query and exploration history is certainly useful, but it may well be premature to over specify that it should be at the bottom of the resource page, that there should be exactly ten objects, or that they should be the most recently looked up objects tracked by the portal.

Tasks are considered more appropriate in this context, because they describe in a more abstract way what needs to be accomplished, without constraining any specific dialogue in advance, and without the need to allocate prematurely any subtask to users or system. Yet, with the objective to collect as much useful information as possible, the use cases available in the literature are taken into full account – trying to abstract from over specified details.

Other potentially relevant studies do not make explicit reference to tasks, but rather discuss functionalities. This is especially common in papers analysing existing applications, such as Wilson et al. (2010), or Morato et al. (2014). Tasks and functionalities are obviously interrelated:

functionalities are intended to support tasks, and tasks need to be supported by suitable functionalities. However, while tasks belong to the problem domain, functionalities are already in the solution domain and represent “a top-level solution to the problem” identified (Leffingwell, 2001, p. 6). On the contrary, tasks do not imply any particular solution, but can be seen as an abstract specification of requirements, that can be satisfied by different features.

Therefore, this analysis takes into account the functionalities, but attempts to identify more solution-independent tasks, by reverse engineering functionalities to the tasks they are intended to support.

4.3 Search-task models in the literature

A number of search-task models are available in the literature, which inspired this domain-oriented task analysis. While they are generic models, they are useful to think and discuss about tasks. The development of a more specific domain-oriented taxonomy is the aim of this chapter, and is elaborated in the final sections.

4.3.1 A context model of search: focus on Information Seeking and Work Context

Tasks can be classified at different levels of abstraction and hierarchically organized, with higher-level complex tasks generally composed of a number of simpler lower-level tasks. Byström and Hansen (2002) propose a widely adopted model that identifies three fundamental search contexts, which can be conveniently used to classify and reason about search tasks:

- Work Context (WC) tasks are directly related to higher-level context-specific user goals, such as organizing the material for a lesson.
- Information Seeking (IS) tasks are oriented towards flexible exploratory search activities, where interactively formulating precise search requirements, starting from vaguely defined objectives, is part of the task itself.
- Information Retrieval (IR) tasks, at the lower level, are related to very well specified search objectives such as those supported by a familiar Google keyword search.

This classification is very relevant in this doctoral research, which aims to consider the search tasks in the context of the wider user goals: it “defines search as the set of activities that take users from identifying a problem all the way to achieving their goals” (Wilson et al., 2010, p. 12). Clarifying the level of the tasks is critical, because design and evaluation criteria strongly depend on the level of the tasks. Ultimately, the success of search systems depends on how well they support users in carrying out their tasks. Hence, their design and evaluation should not just focus on search algorithms of classical interest to Information Retrieval, but on domain-oriented user tasks at higher IS and WC levels. For example, while minimizing task execution time could be a goal at IR level, longer execution time might be preferred at WC level, because it could induce a useful deeper understanding of the context (Järvelin and Ingwersen, 2004).

4.3.2 Information Foraging Theory: focus on QBE expansions by similarity

As discussed, the literature is a useful source of representative tasks, especially when use cases and features are also taken into account. Nonetheless this analysis requires a certain level of interpretation, for example to reverse engineer tasks from features, or to adapt tasks defined in generic search contexts to the OER search domain.

A possible solution to identify new tasks could be the experimental observation of users performing their activities using existing search engines. However, observing users working with traditional expectations and mind-sets on traditional applications or prototypes supporting only a restricted set of tasks, would not necessarily help to discover meaningful tasks. Indeed, as pointed out by Qu and Furnas (2008), users would cope with deficiencies of existing applications by altering the natural tasks they would actually need to carry out. Therefore, an experimental extraction of user tasks was not carried out with this strategy, because investigating how educators use the tools currently available might not reflect how they would ideally carry out their tasks if given appropriate support.

Instead, the tasks can be extracted and validated by means of a user behavioural model. Qu and Furnas (2008) adopted a behavioural model to suggest suitable tasks, data, and techniques for their “formative, user-centred” evaluation process. This strategy has the advantage of shifting the focus from a given prototype, to how the activities described in the model are supported – discovering missing aspects including when the users are not even aware of them. While Qu and Furnas selected a sensemaking model, Information Foraging Theory, which they also mention as a suitable model for their study, fits better the context of this research focusing on exploratory search.

Information Foraging Theory (IFT) was developed by Pirolli and Card (1999) noticing similarities between the behaviour of users (“informavores”) looking for information, and foragers hunting for food. The behaviour of hunters/gatherers searching for food was previously modelled by Optimal Foraging Theory, which considers that this behaviour is driven by their attempt to maximize their net energy intake (Pyke, 1984). Possibly because this behaviour is hard-wired in our brain, IFT has been successful in predicting human behaviour in different information search circumstances.

Here, IFT was used to suggest user tasks by transposing the behaviour of foragers to the corresponding behaviour of informavores. As an 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 checking those previously used by other users”. This can in turn suggest corresponding user tasks such as “identify OERs used by trusted colleagues”.

Inversely, IFT can help validate the soundness of previously identified user tasks, by considering the corresponding foragers behaviour. As an example, the task “identify potentially interesting OERs similar to other previously found” could be seen as corresponding to the forager task “look for prey similar to some other previously caught”. Similarity could be further refined as “sharing some common metadata”, corresponding to “sharing similar features”.

It is possible to identify easily many more similar examples: given a forager task it is possible to identify the corresponding informavore task, and vice versa. This really striking correspondence might well be due, even if not formally proved here, to an isomorphism between the set of tasks in the informavore search domain and the set of tasks in the forager domain. For this reason, the tasks that can be identified in this way with IFT are consistent with the tasks extracted from the literature.

IFT tasks, extensively based on examples and exploiting similarities, emphasize the discovery of additional resources related to resources previously identified.

4.3.3 Strategic and episodic models

In her early work into human search strategies, Bates (1979) identified 29 search “tactics”, considered elementary moves composing a search activity. Later she refined her model (Bates, 1990), structuring search activities in 4 hierarchical levels: move, tactic, stratagem, and strategy. Moves are the most elementary user actions identified in the model, such as “enter a search term”. Moves can be organized in one of 32 tactics: an example of tactic is “reduce”, which means reducing constraints in a query in order to generalize it and increase the number of hits. Tactics can be organized in stratagems, such as browsing material after locating an area of interest. Stratagems, finally, can be organized in strategies, such as locating information for a research paper. Bates’ reference context was library/information science, but the model can suggest meaningful tasks for this research.

Belkin (1995), in his Information Seeking Strategies (ISS) model, attempted to identify users’ search strategies to be supported by what he called Interactive Information Retrieval Systems. These strategies have four binary dimensions: method (scan or search), goal (learn or select), mode (recognize or specify), and resource (information or meta-information). For example, the traditional keyword search in Google corresponds to the pattern ISS15, or ISS-1111₂ in binary: search/select/specify/meta-information. Although his original model was later criticized and extended, he clearly exposed the need to identify and support search strategies also at the Information Seeking level, and to support users in moving from one to the other according to their dynamically changing requirements. While this model is not hierarchical, strategies (or more precisely, their corresponding “scripts”) are aggregated in episodes. Additionally, the various strategies are originated in the context of higher level goals, hence, once again, the importance to consider users and their tasks.

4.4 Literature related to metadata standards

A first useful source of meaningful tasks in the literature is from burgeoning activities related to the development of metadata (including paradata and social data) standards. These standards are supposed to be grounded on the analysis of the tasks that need to be supported, so that these tasks are sometimes explicitly documented.

4.4.1 LRMI / schema.org

LRMI was generally discussed in Section 2.4. Aiming to identify requirements for this standard, the Survey Report on Ease and Discoverability (LRMI, 2013b) reports useful information about how educators look for learning resources, the challenges they encounter, and some suggestions to overcome them. The aspects of this LRMI survey relevant to this research are the “most helpful search criteria” for finding educational resources indicated by educators, such as Content/Subject Area (93.3%), Grade Level (80.5%), and Alignment to Specific Standards (57.3%). LRMI (n.d.) reports a number of use cases too, implicitly identifying tasks for educators, which were used to guide the extension of schema.org to address the specific needs of the educational community. Tasks relevant to this research include searching for resources by metadata such as standard, resource type, format, age, and – at a higher level – planning a lesson accommodating multiple instructional strategies.

4.4.2 Open Discovery Space (Irma case): social aspects

ODS (Open Discovery Space) was a project aiming at fostering the sharing and use of educational resources; one of its objectives was to identify / develop a (Linked Data oriented) standard for paradata and social data, to support the discoverability of the available resources. In this context, the Irma use case reported by Drachsler et al. (2012), articulates a realistic and complex situation faced by a hypothetical teacher, related to the discovery of learning resources. The objective of this experimentally validated use case, was mainly to identify requirements for social data and paradata, and is therefore mostly limited to this domain. However, even if in the context of an over-constrained dialogue sequence typical of use cases, it identifies a set of useful relatedness metrics and information related to social aspects of OER search. Examples of these metrics are resources “used together”, “visited together”, “derivative resources”, “most highly rated resources matched to her profile”. Information displayed to users include popularity measures such as user ratings, reviews, how often viewed or downloaded. Finally, it mentions features such as tagging and searching by tag clouds.

4.5 Literature on specific search tasks

4.5.1 Search tasks used in experimental search studies: RepAST

The literature reports many experimental studies aiming at evaluating search applications or search behaviour; most of these studies involve controlled experiments where the same tasks are assigned to multiple users. Wildemuth and Freund (2009) note that the tasks reported in the literature are extremely heterogeneous and make it difficult to reuse tasks in different studies and to generalize evaluation results to different contexts. In order to make it possible to construct tasks with desirable characteristics, and to obtain generalizable findings, they collect and classify the tasks used in the most disparate experimental search studies in a very comprehensive and up to date public database (Freund and Wildemuth, 2014), the Repository of Assigned Search Tasks (RepAST). At the time of writing they have analysed 834 papers classifying 1358 different types of tasks, ranging from IR level “known-item search” and “well defined finding task”, to WC level such as “situated work tasks”.

Unfortunately, despite the notable number of tasks available, and their currentness (there are papers from 2018), there are very limited direct references to tasks related to teachers searching for OERs. Yet, there are commonalities among different contexts, at least at the lower levels, that can be exploited. Indeed, Wildemuth and Freund (2009, section 2.6) claim that the main task attributes frequently used in the literature, “can be interpreted as the research community’s common-sense understanding of what’s important about search tasks”. In particular they focussed (Wildemuth and Freund, 2012) precisely on the characteristics of search tasks intended to elicit exploratory behaviour: these tasks are usually general, open-ended, and involve uncertainty. These characteristics are useful to identify meaningful domain-oriented tasks in this research, such as, for example, finding resources related by various proximity metrics.

4.5.2 Integrating search with support for WC level tasks

As a notable exception, Kabel et al. (2004) explicitly targeted a WC level user task in the educational domain. They investigated how the task of “composing lesson material” (organize a lesson plan), involving search and retrieval of educational resources, can be improved via a specialized task oriented ontology used to annotate the educational resources. Their task ontology makes it possible to identify the instructional role of the resources at a considerable level of detail, such as introduction, motivation, description, explanation, example, reminder, testing, and feedback. They argue that such a domain and task oriented ontology decreases the effort to carry out the task and increases the quality of the output. The authors discuss also the possibility of developing a search system even more oriented to support instructional design. This is an interesting idea in principle, because one of the fundamental goals of teachers looking for OERs is to finally integrate them in their lesson plans. However, such a detailed taxonomy would need to be based on a particular instructional model, and there are many. Additionally, the challenges and efforts required to produce even basic metadata, widely recognized in the literature (Doctorow, 2001; Petrides, 2013), would be enormously exacerbated by the need to collect the detailed metadata needed to support such a task. This observation is supported by the striking contrast with the metadata currently foreseen by the pragmatic approach of LRMI. Furthermore, the approach of Kabel et al. would require a very high granularity of the learning resources, which would enable their annotation at the required very detailed level. For these reasons, this high level task of instructional design at such high level of specialization is considered outside the scope of this thesis. However, it confirms, once again, the importance of supporting the high-level task of lesson planning.

Qu and Furnas (2008) studied the way to support the high level task of making sense of the information in a domain. The items of interest, obtained from lower level search activities, would be created, labelled, dragged and dropped around at will, and annotated with embedded personal notes, with the objective to organize them in a hierarchy of topics and subtopics sketching an outline. While this task does not fit directly the needs of a teacher who presumably has considerable pre-existing knowledge about his/her domain of teaching, it is relevant considering the need of teachers to organize the information, data, knowledge and learning experiences for students who lack familiarity

with the domain. The resulting task, therefore, is similar to the previously discussed one, of organizing a lesson plan.

A similar but more domain oriented use case was identified by LRMI (n.d.). It concerns planning a lesson accommodating “multiple instructional strategies”, indicating the need for a “place to collect the selected resources”, such as a bookmarking service or a notes taking application”.

4.5.3 Search interfaces

Wilson et al. (2010) organize an excellent comprehensive monograph on search interfaces. They survey the most interesting features of existing systems and prototypes, framing them in the model they attribute to Järvelin and Ingwersen (2004), at the three levels of searching context WC, IS, and IR. Continuously emphasizing the need for an inclusive view on search, that should encompass all three levels, they discuss the following main tasks and features:

- Lesson planning.
- Learn and produce a summary of a topic, that includes sub-tasks at IS level such as comparison, synthesis, and summarization.
- Hierarchical and faceted classification for filtering.
- Classification with automatic clustering, when the previous techniques are not applicable because explicit annotations are not available. In this case, clustering can be used to automatically identify common attributes, which can be used as a dynamic classification.
- Classifications exploiting social data, in particular tagging to complement keyword search and sense-making.
- The possibility to obtain resources starting from a sample resource (“more like this”), using different similarity metrics, to allow users to follow an unanticipated path during search.

4.6 Notable learning resources discovery projects

Existing applications and research prototypes are another useful source of functionalities, which can be reverse-engineered to tasks. A few of them are briefly reviewed here, with the objective to highlight just their main characteristics of potential interest in this research.

4.6.1 MACE: domain-oriented visual interfaces

MACE (Wolpers et al., 2009) was a federated search system for learning resources in the architecture domain. While it is not operational any longer, the project was noticeable for providing exploratory oriented, innovative, visual interfaces to search for resources. As an example, the interface to browse its hierarchical topic classification, that played a key role in organizing the resources, is depicted in Figure 4.2. The figure shows the whole topic classification in a dynamic dendrogram, where it was possible to expand, shrink or pan interactively any area. By selecting one of the topics visualized, it was possible to navigate to the associated resources.

Another solution proposed was purpose-oriented tagging, where users are asked to tag resources with what they considered to be their purpose, to support a goal-oriented social search.

At the time of writing, the portal only mentions the possibility of accessing learning resources by keywords, unconstrained tags, and competencies. However, the main contributions of this project for this thesis, is the recognition that competency or educational frameworks are fundamental, but many users might not like to use them directly, and alternative techniques should be tailored to the specific needs of users and their context.

4.6.3 National Science Digital Library: exploring resource relationships

The National Science Digital Library (NSDL, 2015) provided search services based on standard aligned and curated metadata with a proprietary format. They made also use of social data and paradata (they are the ones who coined this term), to provide indicators of resource utility.

They provided a search portal including in particular an innovative search from conceptual maps. As it is possible to see in Figure 4.3, these maps let users interactively explore concepts (represented by rectangles) and their relationships (represented by arrows) in various domains. Sumner et al. (2005) argue that these active maps allow users to focus on their scientific discovery task rather than on low level aspects of the search process. The elements in the map allow one to navigate to related educational standards and learning resources. Different static and manually created maps were provided, to allow for the exploration of concepts, their relationships, and associated resources from multiple points of views.

Their search facilities have been later transferred to ISKME, so that they now share the same interface as OERCommons (Figure 4.4).

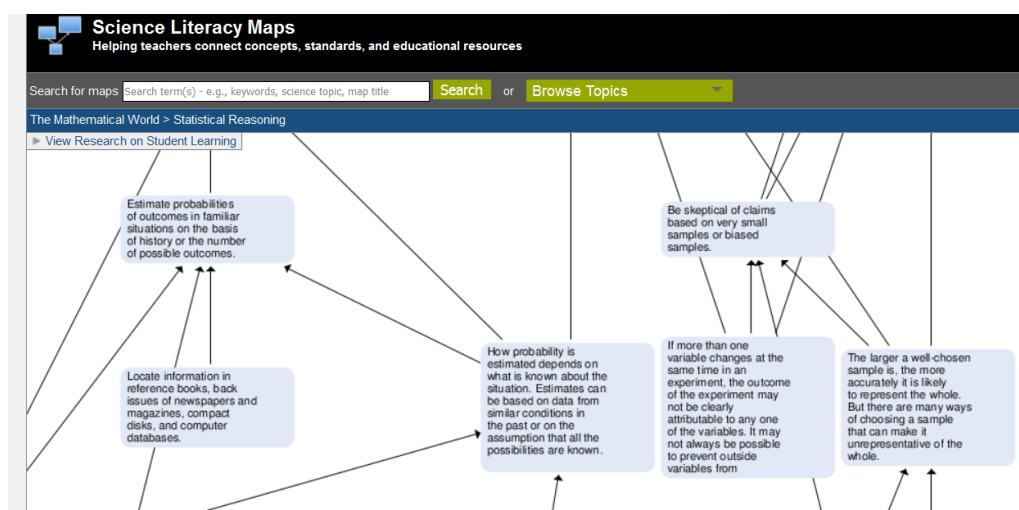


Figure 4.3 – An example of Science Literacy Map (Source: <http://strandmaps.dls.ucar.edu/>).

4.6.4 mEducator: repurposing

The project mEducator (Dietze et al., 2013), specialized in medicine, is strongly oriented to Linked Data technologies, that make it possible, in principle, to support the web-scale search of learning resources. It supports federated search, by automatically uplifting metadata from different traditional

standards to open Linked Data, and aims to automatically discover services providing information about resources (Yu et al., 2011).

mEducator has an unusually articulated strategy to support tasks related to the repurposing of existing resources and flexibly searching among these repurposed resources (Stefanut et al., 2012). Repurposing is a very common task for educators (de los Arcos et al., 2016), who constantly need to adapt and integrate existing resources (Dietze et al., 2013), for example to suit different pedagogical approaches, or different target people (Kaldoudi et al., 2011). mEducator's rich ontology supports tasks such as the identification of resources that a given resource is derived from, or which were derived from it, but it also allows one to identify the motivations for repurposing, and reconstruct the whole history.

4.6.5 OERCommons: approximate alignments to educational standards

OERCommons (2016) provides a comprehensive interface and articulated information about its resources. Metadata in OERCommons are collected from authors, but also crowdsourced from users (third party or non-authoritative metadata). However, all metadata undergo a well-organized curation process, to guarantee their quality (Campbell and Barker, 2014). The metadata schema used by OERCommons is based on LOM with some extensions. Hence, as it can be seen from Figure 4.3, it is possible to identify resources using metadata such as Subject Area, Educational Level, or Educational Use, as well as alignments to educational frameworks such as the Common Core State Standards.

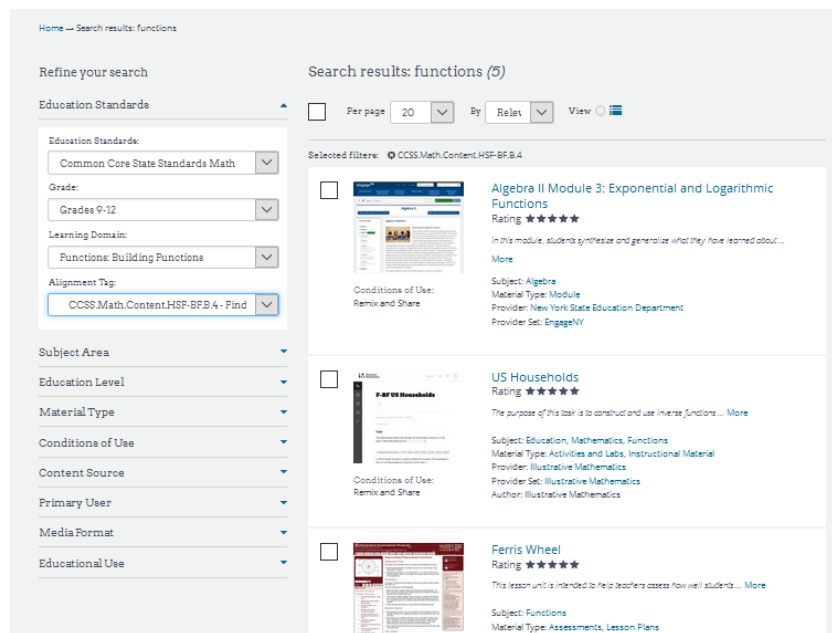


Figure 4.4 – OERCommons search interface (Source <https://www.oercommons.org/>).

Notably, OERCommons provides the possibility of specifying the “degree of alignment” of resources to educational standards. This makes it possible to search for resources which are not necessarily perfectly aligned to a standard, but just approximately aligned. This can be useful when it is not

possible to state that a resource is perfectly aligned to a standard, or to obtain additional results when there are not enough resources perfectly aligned to a standard.

A similar feature was already introduced, in the nineties, in the bibliographic cataloguing format MARC (Machine Readable Cataloguing) with the CEMARC (Curriculum-Enhanced MARC) extension, via the so called MARC 658 tag – curriculum objectives (Murphy, 1995). Yet, even the recent schema.org / LRMI only foresees, at the moment, an exact alignment of a resource to a standard.

4.6.6 SocialLearn: multiple recommender systems

Other projects in the area of OER search (and beyond) recognize the value of learning analytics, that is the process of exploiting the traces that students leave when they interact with the educational material or their educational setting. Examples of these traces are the time spent using a resource, or the performances obtained in a test. These data can be conveniently used to improve teaching and learning (Sclater et al., 2016). SocialLearn was a project focusing on *social* learning analytics (Ferguson and Buckingham Shum, 2012). This project, grounded on the firm belief that the value of OERs can only be maximized in the context of a social learning space, draws upon social learning analytics to support learners and teachers activities. It suggests, in particular, a wide range of innovative recommender systems, exploiting different classes of social learning analytics. The analysis of students' interactions in forums (social learning *discourse* analytics) could be used to understand their attitude towards the topic under discussion, and suggest OERs encouraging alternative approaches to a given subject, or challenging their point of views. Information about the students' disposition towards learning (social learning *disposition* analytics), possibly collected through profiling questionnaires, could be used to suggest educational resources to strengthen motivation.

4.6.7 OpenEd: personalized remediation activities

As previously discussed, the literature suggests to shift the attention from isolated search tasks to the context where they are originated, hence focusing on directly supporting high / WC-level tasks. This suggestion is embraced by portals that offer integrated search within Learning Management System (LMS) oriented environments. OpenEd (2016a) was a portal of this type, innovative for its commercial-oriented business model and its “formative assessments” educational model. OpenEd offered teachers a simple LMS online environment, focused on the cycle assess, analyse, and personalize teaching. Lesson plans (Figure 4.5) were basically playlists or annotated bookmarks, where it was possible to include tests retrievable from a large data bank, precisely aligned to educational standards. This made it possible to associate them with learning resources, which were recommended to students according to their test results. OpenEd supported educators to assemble a set of correlated resources (addressing complementary instructional strategies) with personal annotations in a simplified lesson plan, deliver the lesson, collect data about students' performance for formative evaluation, and advise personalized remediation activities.

Therefore, search was embedded in the WC-level tasks of planning and delivering personalized teaching, and the whole mechanism was based on the alignment of the resources to educational frameworks.

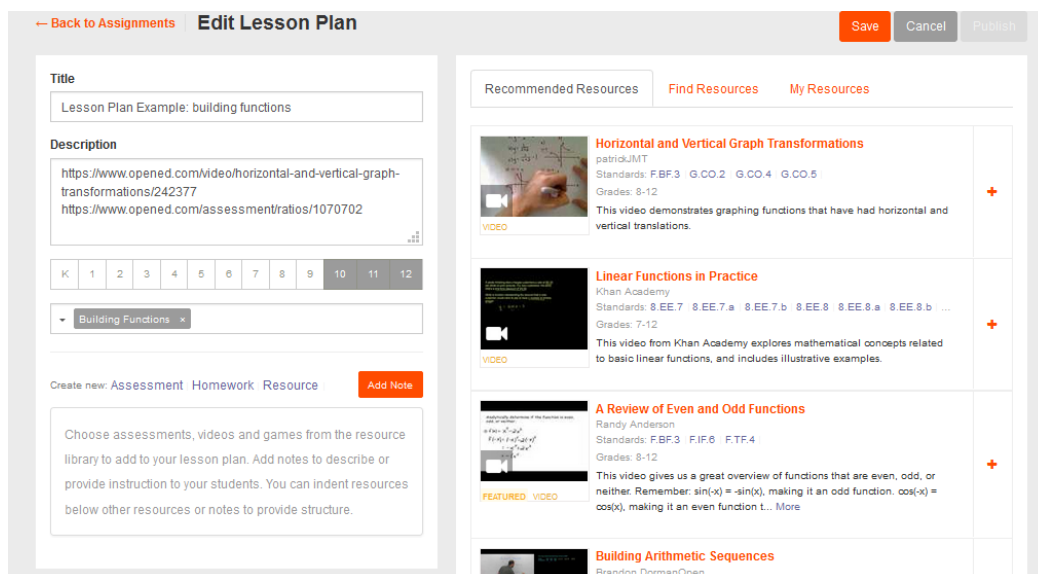


Figure 4.5 – OpenEd: editing a Lesson Plan (Source: <https://www.opened.com>).

4.6.8 Gooru: supporting different delivery strategies

Gooru (2016a) is another portal aiming to address directly high / WC-level tasks, supporting the paradigm of integrated search in a simplified LMS. Interestingly, they motivated their approach of integrating search within a wider context, as an attempt to make the whole mechanism sustainable, exploiting the synergy between searching and using. Gooru, therefore, supports teachers in planning as well as delivering their lessons. Concerning planning, educators can organize resources within collections: “playlists of multimedia resources in a variety of formats such as videos, interactives, websites, images, and more” (Gooru, 2016b). Concerning delivery, it supports the flexible organization of different delivery strategies, letting educators assign resources to students (classes), whose progress can be monitored via a Collection Progress dashboard displaying rich analytics.

Within the context of these high-level tasks, Gooru adopts (Campbell, 2014b) a rich set of descriptive metadata fields and plenty of usage data and social data, to power their search, ranking, and recommendation algorithms. Their metadata schema may be considered a variant of LOM, and includes traditional metadata such as Age Range, Educational Use, and Interactive Type. They even include metadata such as Mobile Friendliness, or 21st Century Skills, where content is categorized based on detailed skills such as Reasoning and Argumentation, Environmental Literacy, or Building of Persistence. Social and usage data include search, rate, views, clicks, assessment results, as well as student reactions such as “I can explain”, “I need help”, “I don’t understand”. Most (95%) of the resources, over 18 million, are tagged with LRMI too.

Until 2014, Gooru attempted to solve the problem of multiple existing educational standards, by having its resources aligned by curator experts, to a proprietary subject taxonomy, visualized with a

collapsing tree of 6 levels. This proprietary taxonomy would then be aligned to educational standards by users, therefore fostering, they claimed, collaboration across people using different standards. This idea looked promising in principle, yet, at the time of writing, there is no more mention of that proprietary taxonomy. Instead, as it can be seen in Figure 4.6, Gooru too has adopted existing educational standards such as Common Core and Next Generation Science.

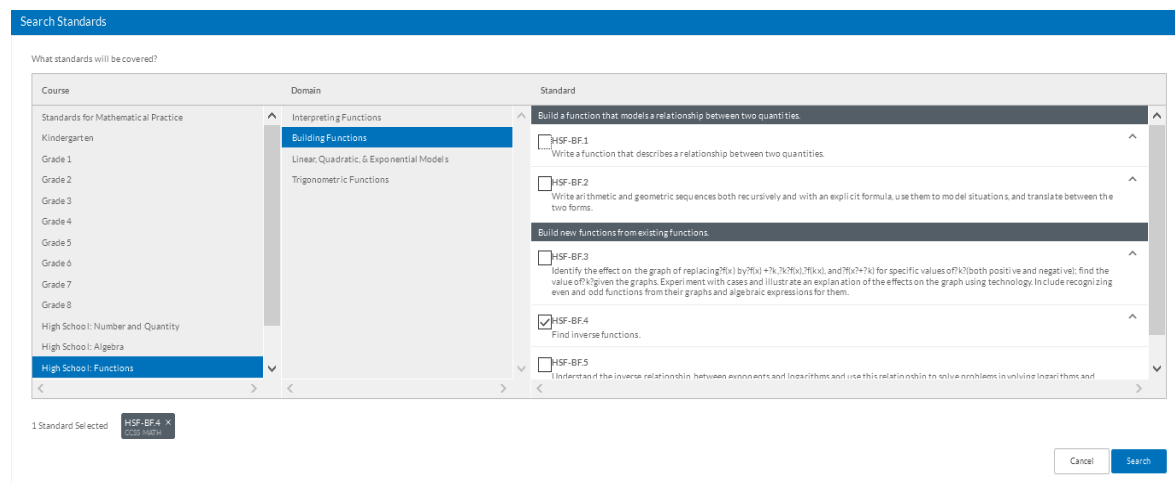


Figure 4.6 – Gooru: searching from educational standards (Source: <http://www.gooru.org>).

4.6.9 Achievement Standards Network: educational frameworks

The Achievement Standards Network (ASN, 2016), whose search website is described as a “web-scale search service for learning outcomes”, offers metadata and services related to the alignment of learning resources to achievement standards, which can be very useful in the context of OER discovery.

ASN offers access to machine-readable representations of standard educational frameworks, including globally unique Universal Resource Identifiers (URI), where educational resources can be aligned to, for example via [LRMI/schema.org](http://lrmi/schema.org). While the (K12) CCSS plays a prominent role among the standards considered, they support, or plan to support, other educational, skills, and professional standards, including from other countries in addition to the USA.

ASN offers services to search and browse the supported educational standards, to obtain different representations of the same standard, and to support standards crosswalking – that is finding corresponding alignments among different educational frameworks (Sutton, 2008). This is a mechanism that can be used to allow a user searching for OERs aligned to a given educational standard, to discover other OERs aligned to different but equivalent standards.

ASN recognized the opportunity to specify the *degree* of equivalence between standards (as well as the degree of alignment between resources and standards) (Sutton, 2008). Hence, they specified various semantic relationships to model that two standards are partially equivalent: *minor*, *major*, *narrow*, and *broad*, in addition to *exact* equivalent. A minor partial equivalence, for example, indicates a minimum overlap among two competency standards, while a broad partial equivalence indicates that a competency standard covers all the relevant concepts of another standard and

additional ones. They also model *prerequisite* alignment, which specifies that a competency is a prerequisite for another. This type of alignment between standards, could be used to identify educational resources that are prerequisite for other resources, with no need to specify this type of relationship between the resources themselves.

4.6.10 Summary of the projects reviewed

Table 4.1 summarizes the mentioned projects, indicating their main characteristics of interest in the context of this research.

Portal	Specific characteristics of interest
MACE	Strongly domain-oriented, focus on exploratory-search functionalities with innovative visual interfaces. Access to the resources via a topics taxonomy.
OpenScout	Adopted a competency-oriented framework, rather than a traditional content-oriented paradigm. Recognized the difficulty for some users to exploit these complex frameworks, and the need to provide additional alternative search techniques, tailored on users and their context.
National Science Digital Library	Exploration of resource relationships via multiple static cognitive maps. Maps aim to allow users to focus on their scientific discovery task rather than on low level aspects of the search process.
mEducator	Linked data oriented. Articulated support for repurposing, and flexibly searching among repurposed resources.
OERCommons	Particularly curated data about resource quality. LOM oriented metadata. Alignments to standards, including the possibility of specifying the degree of alignment.
SocialLearn	Articulated recommendation systems based on social learning analytics.
OpenEd	Search integrated in the WC-level tasks of planning and delivering personalized teaching. Based on the alignment of resources to standard educational frameworks.
Gooru	Search integrated in a simple Learning Management System, aiming to improve sustainability by supporting the whole life cycle of resources in education. Focus on supporting alternative delivery strategies. Recently shifted from a proprietary subject taxonomy to alignments to standard educational frameworks.
Achievement Standards Network (ASN)	Metadata and services related to the alignment of learning resources to achievement standards. Crosswalking between different standards, modelling degree of equivalence between standards.

Table 4.1 – Summary of the main characteristics of interest for the mentioned projects.

4.7 Task analysis: OER domain-oriented taxonomy and sample tasks

As previously discussed, search tasks (at IR and IS levels) should not be considered in isolation, but should rather be considered in their broader context, as they are sub-tasks of domain-oriented higher (WC) level tasks. For example, the desired level of similarity of resources identified by a discovery task at IS level has to be considered in the context of their use at WC level, such as planning a remedial rather than an in-depth teaching activity. For these reasons, the goal of this section is to produce a taxonomy as broad as possible, covering the whole OER life cycle. However, to simplify

reading, this chapter only discusses the higher level categories of the taxonomy, to provide the necessary context, and the specific Searching and Using sub-categories, more directly relevant to the specific focus of this research. (The remaining sub-categories are discussed in Appendix A.)

An additional advantage to consider the broader context, is the opportunity to obtain a comprehensive taxonomy which is potentially reusable, as suggested by Geisler (2008), in different contexts related to the design and evaluation of OER related systems.

4.7.1 Top level categories of the domain-oriented taxonomy

The generic classification at WC, IS, and IR context levels, previously introduced, is very useful to classify and think about generic search tasks. However, to structure and organize this domain-oriented task analysis, it is convenient to supplement it with a taxonomy focussed on the specific OER domain.

The starting point for the higher level categories of this taxonomy are the four simple but effective “themes” identified from the literature by Atenas and Havemann (2014) and adopted in their evaluation of OERs portals (Atenas and Havemann, 2013), covering part of the OER life cycle: Search, Share, Reuse, and Collaborate. These “themes” can be compared to the “steps” of the OER life cycle discussed in the OER Handbook (Gurell and Wiley, 2008): Find, Compose, Adapt, Use, and Share. The step Find corresponds directly to the theme Search; the aggregation of the steps Compose and Adapt corresponds to the theme Reuse. While there is no explicit step corresponding to the theme Collaborate, the Handbook emphasizes that “each step could involve some degree of collaboration”. Finally, the Handbook mentions the step Use, which has no explicit correspondence in any theme.

Search is the most directly relevant category in the context of this research, lumping together the most specific tasks related to OER search and discovery, typically at IS and IR levels. The following three categories are also related to discoverability, even if less directly, and are positioned at the higher WC level.

Share refers to publishing (making available) a resource and its description (metadata). This is the fundamental starting point of the virtuous cycle of publishing, using, repurposing, and republishing (UNESCO, 2012). It is supported, for example, by platforms such as OpenEd and Gooru.

Reuse refers to the modification of existing resources; it is related to the term “open” in the acronym “OER”, which does not simply mean that it is possible to “use” the resources as they are, but on the contrary that it is also possible to modify, adapt, and integrate them. To distinguish more easily this meaning of “reuse” from simple “use”, the term “repurpose”, widely used in other contexts, for example by UNESCO (2012), is adopted here. Repurposing is extremely important, being a very common activity for teachers, even if not frequently formally recognized as such (Dietze et al., 2013).

This was particularly curated, for example, in mEducator, but it is also supported by platforms such as OpenEd and Gooru.

Collaborate, finally, refers to another class of important tasks related to search activities, where users communicate among them, or where they provide, voluntarily or implicitly, information about the resources. This was the focus, in particular, of SocialLearn.

The previous four “themes” are appropriate as top-level categories of our taxonomy, but need to be extended with an additional top-level category, corresponding to the step Use, to accommodate the fundamental educator tasks of *using* existing resources to design and deliver lessons. “Using”, a term employed by UNESCO (2012), is adopted here as the name of this category. As previously discussed, this category is fundamental because it contains WC-level tasks that represent the ultimate goal of any educational oriented search system.

In conclusion the higher level categories of the taxonomy, reordered following the natural OER’s life cycle, are therefore:

- Publishing
- Searching
- Using
- Repurposing
- Collaborating

The categories Searching and Using, more directly relevant in this research, are further discussed and structured in subcategories in the following sections.

4.7.2 Searching: subcategories and related tasks

Goal of this section is to identify the tasks belonging to the category Searching, and their classification in sub-categories. At a sufficiently high abstraction level, these tasks are expected to be domain independent (Wildemuth and Freund, 2012). Indeed, the Information Foraging Theory assumption is that we take advantage of the same strategies our ancestors used to forage for mushrooms or rabbits, to search for *any* resource, not just OERs. However, they are mapped to specific domain-oriented tasks such as, for example, “Obtain resources aligned to a given educational standard”.

In this classification, the queries corresponding to each task are intended to be always applicable to the results of any other previous query, in an interactive navigation sequence (Stefaner et al., 2009). For example, it must be possible to apply a keyword search on the results of a previous search for resources aligned to a given educational standard, or vice versa. Of course, the first query of any sequence, is applied to the whole dataset. This is consistent with other studies, including Hearst and Degler (2013) who argue the importance of integrating different retrieval functionalities in a continuous refinement of search, fluidly moving between different operations. An LRMI (n.d.) use case too, highlights the need to filter incrementally the results by secondary search criteria.

A set of tasks of this type is discussed by Stefaner et al. (2009), who mainly focus on faceted search and “dynamic taxonomies”, building on ideas previously elaborated by Sacco (2000). Their intention was to derive minimalist orthogonal operations with a precise semantics, which have been implemented in their photo collection browser Camelis, later extended to a “logical information system”. In particular, they identify a “zoom” operation, which acts as a “zoom-in” or “zoom-out” operation according to the context. Zoom-in and zoom-out are defined as operations that enable, respectively, the specialization (restriction) or generalization of a query by adding or eliminating constraints (on facets).

Filtering

“Filtering” is generically defined in Oxford Dictionaries (filtering, 2017) as “process [...] items in order to reject those that are unwanted”. In the context of this research it means, therefore, to specialize a query by adding constraints (filters) – similarly to the zoom-in operation previously described. Of course, filters can be added as in the zoom-in operation, restricting results, as well as removed as in the zoom-out operation, generalizing results. This corresponds to the tactic “reduce” previously mentioned in the strategic model of Bates (1990). Adding, as well as removing filters, are both classified as “filtering” operations in this context (exactly as the zoom operation).

The fundamental functionality of any search application, at IR level, is the ubiquitous free keyword search (Stefaner et al., 2009). It can be used, as intended here, to filter further the results of any other query (or the whole data-set when it is used as the first operation), on the basis of the keywords specified. It is the first task to be included in this category.

Tags, as an unconstrained crowd-sourced classification, are frequently used for searching, in many modern systems. These systems make it possible to search and filter using the tags associated by users to the resources, in the same way as keywords can be used. Hence, similarly to keyword-based searching, tag-based searching can be considered an additional filtering technique, which is therefore included in this category.

Naturally following the use of unconstrained keywords and tags for filtering, comes filtering with structured data. The possibility of filtering resources with a wide range of metadata, such as interactivity type, is at the heart of modern systems and metadata standards developments.

As discussed, the possibility of filtering resources by educational or skills standards is particularly relevant in the context of this research. However, while this is applicable in the Anglophone context, where there is a strong focus on the use of educational standards, it might not necessarily apply in other contexts where formal educational standards are not widely in use.

Sutton (2008) points out the need to discover resources aligned with equivalent standards in different frameworks (crosswalking) too, to cope with the many different existing ones. Additionally, he argues for the need to support both precise alignments and approximate alignments, a feature

supported both by ASN and OERCommons. Relevant tasks are therefore: finding resources precisely / approximately aligned to a given / equivalent standard.

In addition to filtering by descriptive metadata, educators may filter by social data or paradata, for example by user ratings, most frequently used resources, or recently uploaded resources (Drachler et al., 2012).

Finally, it might be possible to filter using multilingual capabilities, as supported by MACE and OpenScout.

Navigation

“Navigation” is mainly intended in this context as moving around in the available data. With a different but related meaning, navigation also refers to moving around in the application interface. This category includes tasks such as navigating in specialized classifications of topics or educational standards.

The possibility of identifying resources via specialized topic classifications, was ranked in the LRMI (n.d.) study as the most important criteria (93.3%). It can be supported with a traditional text-based navigation in a dynamic hierarchy, or with more sophisticated graphical interfaces, such as the visually appealing dynamic dendrogram in MACE, or the graphical navigation from conceptual maps in NSDL. Variations on the theme include, as proposed in OpenScout, navigation starting from a taxonomy of domain problems.

A similar task is the navigation in educational frameworks to browse standards and explore more general, more specific, as well as equivalent or near equivalent alignments to a given standard.

Orienteering

The term “orienteering” indicates the activities aiming at gathering information about the unfamiliar structure and consistency of a dataset, which is a fundamental aim in the context of Information Seeking. The term orienteering was indicated by Stefaner et al. (2009, p. 78) as a process “to foster domain understanding by learning about characteristic metadata distributions”, and it corresponds to the pattern scan/learn/recognize/meta-information in the episodic model of Belkin (1995). Orienteering also includes ancillary functionalities intended to help users finding their way in the application interface: as an example, Wilson et al. (2009) recommend the possibility of letting users preview the effect of any selection, before committing to it.

In a number of prototypes, for example (Sacco, 2000) and (Hearst, 2009), information about the metadata available for filtering at any moment is displayed with dynamic facets, while information about the consistency of the dataset are dynamically visualized directly in the facets, with “Numerical Volume Indicators” or graphical representation with bars. A representative example task in this category is, therefore: retrieve the number of available resources related by a proximity metric to a given resource.

Tagging, as previously noted, is widely used as a crowd-sourced classification, which can be exploited for filtering. However, Wilson et al. (2010) argue that tagging is more useful for making sense of information, again at IS level, because it lets users organize personal group of documents. For this reason tagging (for sense-making) is included in this category too.

Expansion

Expansion is defined here as the process to identify new resources, related by a relatedness or proximity metric, such as likedness (liked by the same users who liked the current resource), to a sample resource previously identified. The corresponding query can be seen as a Query By Example (QBE), where the example is the initial resource this process of finding similar items starts from. This is a fundamental class of exploratory-search and discovery-oriented tasks, based on relationships among resources (Knoth, 2015). This ability to obtain resources related by a given proximity metric, can also be seen as a functionality of systems merging search and recommendation features, advocated by Chi (2015).

Wilson et al. (2010) consider that these discovery-oriented expansion functionalities have a high potential for further research, especially considering Linked Data technologies, also because they have received little attention so far. They mention Endeca, a company (later acquired by Oracle) offering search services to enterprises, that let users ask for related resources via a “More Like This” feature, allowing them to select “the dimension by which they want to see similarity” (Wilson et al., 2010, p. 58). Herrmannova and Knoth (2012) show how the various types of relationships between resources, could be visualized with a graphical notation. The metrics to be used for relatedness expansions relevant in this context can be generic, such as togetherness to indicate resources that have been used together (Drachsler et al., 2012), as well as domain specific such as similarity by common ancestor in repurposing history (Stefanut et al., 2012).

Expansion by correlation to equivalent or similar standards (Sutton, 2008), that is finding resources aligned to an equivalent (or similar) standard a sample resource is aligned to, is another significant example of relatedness metric. Drachsler et al. (2012), consider that this particular expansion could be especially useful to reduce the perceived lack of content availability by educators from small or developing countries, generally characterized by a limited availability of resources.

Table 4.2 summarizes the subcategories discussed in this section.

Category	Definition and example
Filtering	Specialize (or generalize) a query by adding (or eliminating) constraints (filters). Example: finding the resources precisely (or approximately) aligned to a given educational standard.
Navigation	Moving around in (exploring) the available data. Example: navigation in a classification of educational standards.
Orienteering	Gather information about the dataset organization (schema), its consistency, as well as ancillary functionalities to help users find their way in the application interface. Example: retrieve the number of resources related by a similarity metric to a given resource.
Expansion	Discovering items related to a QBE selection by a relatedness metric. Example: discover resources aligned to a standard similar to the standard that a selected resource is aligned to.

Table 4.2 – Category Searching.

In general, search tasks at higher levels (WC and IS) can be carried out by a combination of lower level tasks. 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.

Yet, as discovered in the next study, this is not always the strategy preferred by educators.

4.7.3 Using: subcategories and related tasks

The goal of educators searching for learning resources is, in the end, to support their teaching activities. As suggested in the literature, a search system should support, as much as possible, these high-level WC tasks.

As discussed in the previous sections, this is the strategy adopted by LMS integrated platforms such as Gooru and OpenEd, addressed – in particular – by the researches of Kabel et al. (2004), and identified in a LRMI use case. The main tasks related to this category are related to lesson development (or planning) and lesson delivery. These correspond to the two main subcategories of Using, reported in Table 4.3 with a few representative tasks.

Category: Using [WC level]	
Sub-category	Sample tasks
Lesson development	<ul style="list-style-type: none"> • Organize collection of resources in a bookmarking service or notes taking application. • Plan teaching activities with different delivery strategies (whole class instruction, small groups, individual learning, flipped classrooms, and assessment). • Collect formative evaluation analytics to improve the organization of the educational material.
Lesson delivery	<ul style="list-style-type: none"> • Organize the delivery of learning resources with different strategies (whole class instruction, small groups, individual learning, flipped classrooms, and assessment). • Track learners collecting analytics about their performances.

Table 4.3 – Category Using.

4.7.4 Summary of the domain-oriented taxonomy

The preliminary domain-oriented taxonomy is summarized in Figure 4.8. The Top-level, Searching, and Using categories are addressed in this chapter, while the discussion concerning the categories Publishing, Repurposing, and Collaborating is reported in Appendix A. A detailed list of tasks belonging to each category is made explicit in the questionnaire in Appendix B, discussed in the next chapter.

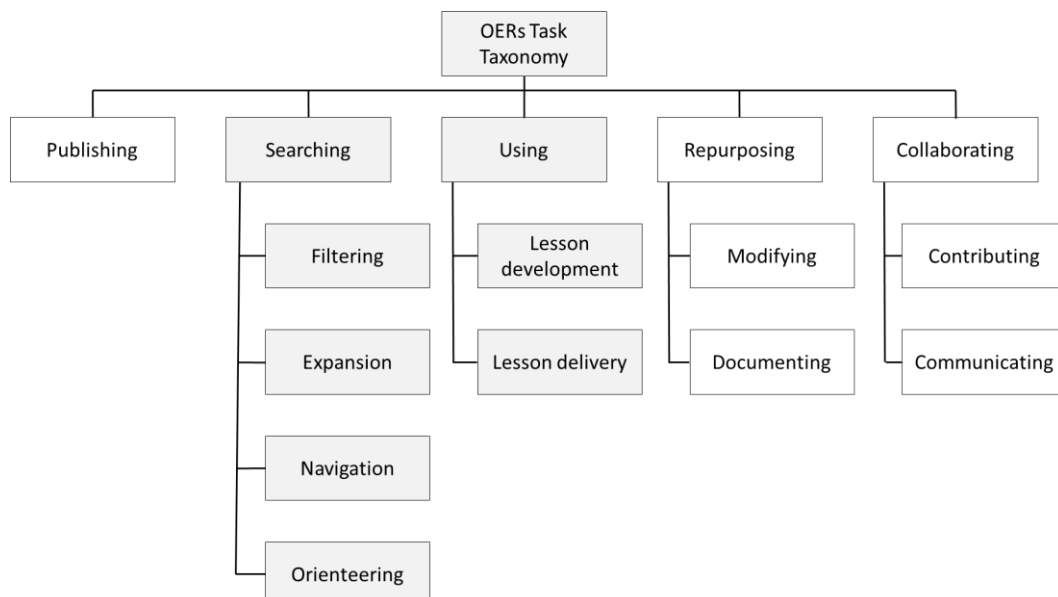


Figure 4.7 – Preliminary domain-oriented task-taxonomy.

4.8 Conclusions: an open, domain-oriented taxonomy

The analysis of the literature and numerous existing initiatives, interpreted with the help of Information Foraging Theory (Pirolli and Card, 1999), made it possible to identify and classify educators’ tasks in the domain of OER discoverability. The resulting taxonomy is based, in addition to the established but generic context model of search by Byström and Hansen (2002), on a domain specific classification purposely designed.

This doctoral research focuses on the Searching and Using categories, greyed in Figure 4.7. The subcategory Expansion, notably, is particularly promising (Wilson et al., 2010). It blends discovery oriented exploratory search, query by examples, and recommendation features under full user control – in line with latest trends in search systems (Chi, 2015).

The full taxonomy based on the complete OER life cycle is reported here, to provide a suitable context, and because it has the potential to be reusable in other situations related to the design and evaluation of OER related systems. Geisler (2008, p. 30) advocates, indeed, the development of this kind of building blocks, or reusable components, to “accelerate progress in our understanding of how to best provide information seeking system support in different domains”. While this taxonomy does not pretend to be complete, it aims to provide a well-founded initial framework open to modifications and extensions – and represents a first, even if modest, contribution of this research.

The following study empirically evaluated this taxonomy with actual educators, with the goal to identify priorities and tasks which deserved further investigation. This preliminary evaluation could then provide the initial input to kick-off the first DSR cycle of prototypes’ design, development, and evaluation.

5. OER task-taxonomy empirical evaluation

5.1 Overview: empirical evaluation of the OER task-taxonomy with actual educators

The OER task-taxonomy developed in the previous chapter was empirically evaluated in this study, with the goal to identify possible novel tasks and initial requirements deserving further research. To this end, quantitative and qualitative feedback on the OER task-taxonomy was collected from actual educators (mainly from the Italian high school context) with surveys and structured interviews.

5.2 Relationship to research questions

This study contributed further input to answer the research question

RQ1: What tasks do teachers need to carry out in relation to OER discovery, in the framework of the OER life cycle?

this time relying on actual educators, with the objective to possibly identify relevant tasks / categories that had not been considered in the OER task-taxonomy, or that would have required modifications or rearrangements.

Providing an answer to the research question

RQ2: What is the relative importance of tasks and categories identified, and which aspects deserve further investigation?

the study identified the task(s) this research would focus on, and contributed the initial input to the research question

RQ3: What are the general requirements for tools supporting the aspects of interest identified?

5.3 Methods: triangulated quantitative and qualitative analysis

The goal of this study was achieved by analysing quantitative and qualitative feedback on the proposed task-taxonomy, collected from actual educators via surveys and interviews. Feedback was initially solicited from the partners of two European research projects, from the participants in a UK OER mailing list, and from Italian high school teachers, but was mainly received from this last group. While this is the sector of main interest to the author, the follow-up studies and discussions took this limitation into due account.

The survey collected quantitative data, first with the objective to engage respondents in critical thinking, in order to elicit highly valued qualitative information. Indeed, the strategy to collect feedback about the OER task-taxonomy, attempting to maximize the quality of the data, was to ask respondents to rate first the importance of tasks and categories identified by the previous task

analysis, with single-item constant-sum questions (CSQs). That is, respondents were asked to allocate a total of 100 points to tasks or categories in a sequence of groups. Each group corresponded to a specific category in the OER task-taxonomy, traversed top-down (that is, starting from its highest abstraction level). The main disadvantage of CSQs, is the high cognitive load imposed on respondents (Sue and Ritter, 2007). Yet this was an advantage here because, removing the simplistic possibility to rate every item as “very important” as on standard rating scales, CSQs force respondents to reflect on the precise relative importance of every category and task, increasing discrimination power and engaging them in critical thinking (Timpany, 2015). An additional advantage of using CSQs is that they eliminate scale meaning bias, increasing data reliability. It is only following this activity that qualitative data were collected, by inviting respondents to answer open questions, following each CSQ, soliciting suggestions for additional tasks / categories, modification of proposed ones, or their different organization.

CSQs, in addition to being instrumental in answering question RQ1, fulfil a second general objective of collecting quantitative data indicating the importance of each task and category (RQ2). These data could be conveniently used as weights in a usability metric, as discussed by Agarwal and Venkatesh (2002).

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.

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 important tasks not covered by the OER task-taxonomy, as well as any additional comments considered relevant.

The whole questionnaire was prepended by a short introduction with goals, background information about the task analysis, instructions, privacy and data management, and contact information.

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 and easily process a potentially large number of respondents. Yet all respondents preferred to use an email version, which made it possible to contact them easily for follow-up interviews.

The questionnaire was approved by the Human Research Ethics Committee (HREC) at The Open University (UK) and is available as Appendix B (email version).

5.3.1 Quality of the data: validity and reliability

The survey attempted to maximize both the reliability and validity of the data collected. In this context, reliability is the degree to which it produces repeatable and consistent results; internal consistency reliability, in particular, is “the degree to which different test items that probe the same construct produce similar results” (Phelan and Wren, 2005). Validity indicates to what extent the test

items actually measure what they are supposed to (Rattray and Jones, 2007). There are different types of validity: content, construct, and external validity are considered here.

Content validity

Content validity “ensures that the operationalization of a construct adequately represents the domain of coverage of the construct” (Agarwal and Venkatesh, 2002, p. 173), and it is frequently assessed by expert assessments or literature reviews.

Here, questions were directly associated with precise task categories and instances, derived from the analysis of the scientific literature, hence attempting to represent the research community’s understanding. Additionally, Information Foraging Theory, a widely used (Pirolli, 2009) behavioural model, was used to help interpreting and validating the identified tasks. However, to guarantee that the whole domain (i.e. the tasks of interest to educators) was adequately covered, the survey included for every group and again in the final section, open questions asking participants if they had any additional suggestions or comments (Rattray and Jones, 2007).

Construct validity

Construct validity is the degree to which a test measures the intended construct. To provide evidence supporting construct validity, convergent and discriminant validity can be used, based on the correlation among similar or dissimilar measures (Rattray and Jones, 2007).

Here, construct validity was assured by construction: each single-item question corresponded directly to a task category or instance. It made no sense to apply convergent or discriminant validity checks, as the questions were independent by construction, and could not be expected to converge or diverge.

External validity (population and ecological validity)

External validity considers the generalizability of the results that can be obtained; it is concerned, for example, about the representativeness of the respondents. To support external validity, the survey was run among actual educators, avoiding fictitious roles such as engaging students or researchers asking them to take the role of an educator. Indeed, non-educators cannot be expected to judge reliably, for example, the importance of approximate versus precise alignment of a resource to an educational standard, or the advantage of using different proximity metrics in a QBE expansion operation.

Reliability

While participants were expected to be highly reliable, based on their experience and motivation, the reliability of the data collected could be checked by a redundant question. Additionally, outliers in the answers obtained, that is unusual weights – distant more than 3 times the interquartile range from those assigned by other participants – were double-checked in follow-up interviews, in order to exclude possible mistakes and fully understand their motivation. The cases at risk of misunderstandings because of a limited command of the survey language (English – for a potentially wider distribution), were handled by administering the survey as a “structured interview”.

5.3.2 Data collection protocol

The survey was made available in three different modalities: (1) self-administered via a document to be filled-in and sent via email, (2) self-administered via a Web form that could be filled-in also anonymously, (3) as a structured interview, including via telephone or Skype. Invitation to participate was sent to a few relevant mailing lists and to some directly known teachers in Italy and the UK.

The survey was first piloted with two respondents, and, following a discussion with them aiming to spot potential misunderstandings, modified to eliminate a few ambiguities and introduce further explanations.

Once the first round of data collection was completed, a boxplot diagram of the weights assigned enabled the identification of outliers. These were scrutinized to identify potential unwanted mistakes, and the worst cases discussed with respondents: they were invited to a follow-up interview to explain and discuss their decisions, and possibly amend the data according to the resulting improved understanding. Following this process, the number of outliers could be decreased, contributing to improve the quality of the data and the applicability of parametric analysis techniques.

Finally, every comment collected via open questions was followed-up, to fully understand the underlying motivations and to elicit additional information. These open discussions enabled the collection of additional qualitative information, which helped in building a broader perspective on educators' way of thinking, which proved to be the most valuable information.

5.4 Data analysis methodology

5.4.1 Quantitative data analysis

Applicability of parametric statistics

The data collected through CSQ are numerical data, which capture subjective estimations of magnitude. Their level of measurement is clearly ordinal as a minimum (a higher weight assigned to Task A compared to Task B, implies that the participant considers Task A more important than Task B). It could be argued that it is not interval, because it is difficult for people to score the importance of task on an interval rather than an ordinal scale. Yet respondents have considerable experience in the domain, so that it could be claimed that approximate equal intervals between points can be reasonably assumed. If these data approximate interval data, then they also approximate ratio data: zero is meaningful (corresponding to a useless task), hence it could be reasonably assumed that, for example, 40 is double than 20. However, because of the limited sample size, their quantitative analysis required caution. In particular, the weights collected were tested for Normality before applying parametric statistics, with the Shapiro-Wilk Test. The null hypothesis of this test is that the sample comes from a normally distributed population; this hypothesis can be rejected with p-values lower than the chosen alpha level. In addition to the test, as with such a sample size there are significant chances of type II errors (false positives), a visual check was performed on the histogram and Q-Q plot of each set of weights.

Whenever any doubt persisted about the applicability of parametric analysis, the results of parametric tests were cross-checked with the results of corresponding non-parametric tests – which are suitable for ordinal data too.

Weights analysis: central tendency with confidence intervals

The weights assigned to tasks and categories (RQ2) were collected via CSQs primarily to foster critical thinking among respondents. Their mean was plotted on a bar chart, to give an immediate appreciation of the relative importance of the task. When necessary to compare items from different CSQs, they were scaled on the same interval between 0 and 10. As a measure of the mean accuracy, that is to understand how well the sample means represent the population means, an estimation of 95% confidence intervals was computed via One-Sample T-Test analysis. In addition, to improve estimate accuracy, bootstrap analysis (Efron, 1979) was employed.

Planned comparative analysis - relative importance of specific tasks

The research question investigating the relative importance of different tasks (RQ2), was investigated by charting basic descriptive statistics, and by using parametric Paired-Samples T-Test. This test is used to determine whether the mean of the populations corresponding to two paired samples is the same (the null hypothesis), that is, whether the differences observed in the mean of the samples could be justified by random noise.

Because some doubts about the applicability of parametric statistics persisted, the non-parametric Wilcoxon Signed Rank Test was used as a cross-check. This test has the same objective as the previous Paired-Samples T-Test, but does not assume that the populations are normally distributed.

Explorative comparative analysis - educators with different profiles

While the sample size did not allow an extended multiple (post-hoc) comparison with the desired reliability, via ANOVA for example, to identify significant differences depending on the different profile of surveyed educators, it was still possible to test a few specific interesting cases. In particular, the explorative analysis of possible different priorities of educators having different ages, in relation to aspects that might be influenced by the amount of experience, was attempted. This analysis used parametric Multiple Independent-Samples T-Tests, with a confidence interval of 95%, to test the null hypothesis that the two populations means were equal at $\alpha = 0.05$, following Levene's Test for Equality of Variances. The effect-size, standardized group mean difference (Cohen's d) and r , were also evaluated.

Again, because of persisting doubts about the applicability of parametric statistics, the non-parametric Mann-Whitney U-Test was used as an additional cross-check. This test can be considered as a non-parametric version of the previous Independent-Samples T-Test, having the same objectives but without assuming that the populations are normally distributed.

Explorative correlation analysis

An explorative correlation analysis among weights was carried out, to identify possible correlations to be further investigated. Here too, the non-parametric Spearman Correlation Test was used in addition to the parametric Pearson Correlation Test.

5.4.2 Qualitative data analysis

A limited but precious amount of qualitative data were collected via open questions in the survey, as well as in structured follow-up interviews. The data were analysed according to the method of content analysis (Cho and Lee, 2014). The text was first split in parts (set of sentences, whole sentences, or part of sentences) referring to single concepts. Then, the following simple schema with the most relevant or frequently expressed concepts was extracted:

1. Comment on Google limitations
2. Suggestions for additional tasks in the category Expansion
3. Educational alignments interesting, but unwilling to use explicitly learning objectives
4. More familiar with filtering type of tasks
5. First topics, then learning objectives
6. Comment on overall completeness
7. Suggestion for additional metadata
8. Interested on Query By Examples
9. Increased score for non-authoritative metadata
10. Publishing resources needs economic incentives
11. No interest in using OERs

Finally, the schema was used to colour-code the original text – to make it possible to scan quickly the main concepts expressed. These qualitative data were used to complement quantitative data by cross-checking them, and made it possible to understand motivations. In particular, they made it possible to explain a totally unanticipated way of thinking that proved to be the most useful output of the study.

5.5 Results

5.5.1 Participants demography

Nine educators participated in the survey (Table 5.1). This sample is smaller than hoped, yet the data analysis methodology (Section 5.4.1) was suitable for samples of this size – in particular the use of non-parametric statistics and the triangulation with qualitative analysis. Five participants opted for the email solution, 2 participants opted for a face to face supported structured interview, 2 participants opted for a telephone supported structured interview; no one selected the Web-based solution.

Participant ID	Gender	Age	Working Country	Teaching Experience [years]	Teaching Educational Level	Teaching Subject	OERs searching frequency
A	M	41-50	IT	>20	Sec. Educ.	Technology	21-50
B	M	>50	UK	6-10	Sec. Ed., Uni	Technology	6-20
C	M	31-40	IT	11-20	Sec. Educ.	Technology	21-50
D	F	41-50	IT	>20	Sec. Educ.	Technology	21-50
E	M	>50	IT	>20	Sec. Educ.	Technology	6-20
F	M	>50	IT	>20	Sec. Educ.	Math	<6
G	M	>50	INT	11-20	University	Technology	21-50
H	M	41-50	IT	>20	Sec. Educ.	Technology	21-50
I	M	>50	IT	>21	Sec. Educ.	Technology	>50

Table 5.1 – OER task-taxonomy Empirical Evaluation: Participants profile.

Their age was distributed as in the bar chart in Figure 5.1.

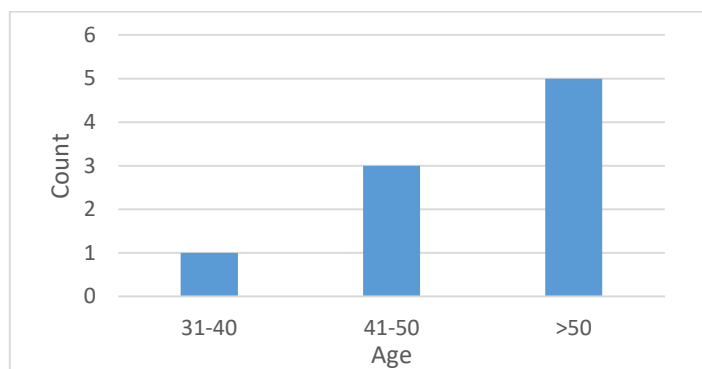


Figure 5.1 – Age distribution in the sample.

5.5.2 Internal consistency reliability check

Two different variables measured the same construct, the importance of adding non authoritative metadata, in different contexts: in the category Publishing (variable named tPaddmeta), and in the category Collaborating/Contributing (variable named tCRaddmeta). The Spearman's correlation coefficient indicates a strong positive monotonic relationships ($r_s = 0.674$) with a significance level of $p = 0.047$. Figure 5.2 shows the ranks correlation of the two variables standardized in the [0-10] interval, and their mean and confidence intervals. The correlation cannot be expected to be perfect, because while it is the same task, it is in different contexts, hence it is compared to different tasks. However, the resulting means and confidence intervals are remarkably similar.

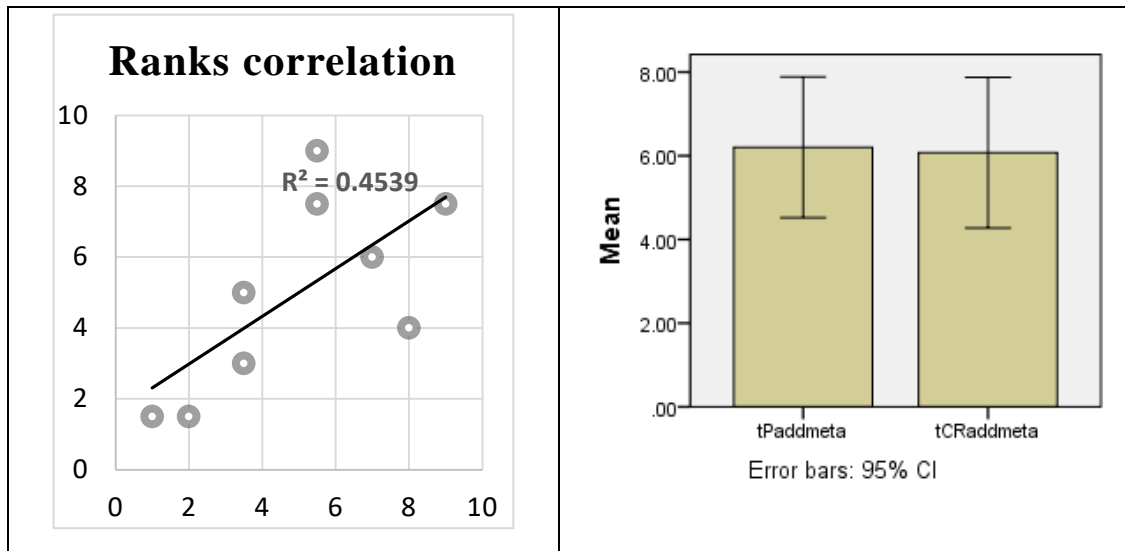


Figure 5.2 – Comparing weights of “Adding non authoritative metadata”.

5.5.3 OER task-taxonomy completeness

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 (Figure 5.3). The three scores that were less than the maximum possible were followed-up: the respondents motivated their score with their own lack of confidence due to limited knowledge, but could not pinpoint any shortcoming in the analysis:

“Because I cannot trust my own judgment, but I cannot indicate any shortcoming”

[User C]

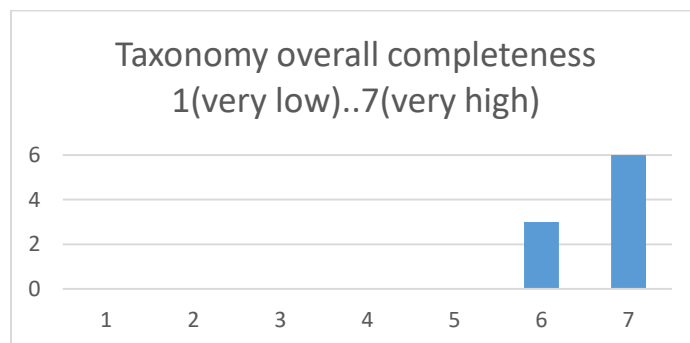


Figure 5.3 – Completeness of OER task-taxonomy.

The high rating obtained is consistently supported by lack of suggestions for the modification or reorganization of proposed categories/tasks, and lack of indication of missing categories/tasks. Three participants left explicit positive comments about the completeness of the OER task-taxonomy:

“Extremely detailed and well structured” [User E],

“A useful taxonomy of the various dimensions and attributes” [User B],

“I wish there was a system supporting all these tasks” [User A].

In one case a respondent left a comment, suggesting the use of additional metadata for filtering not explicitly mentioned in the survey: in the follow-up discussion it was fully acknowledged that the suggested metadata (and many more) were already implicitly included.

However, five respondents, while asking for clarifications and during post interviews, suggested the idea of including additional “expansions”, such as on same topics, same educational standards, and even same authors; for example:

*“Once you have found something, you may expand your search to similar resources:
resources with same authors, same standards, same tag” [User G],*

*“The possibility of expanding to similar resources, for example resources with the same
topics, objectives, or authors” [User F],*

“Expansion could include same topic, same educational standards” [User E].

These suggested tasks could be achieved with a suitable combination of filtering, generalization and expansion operations, already foreseen in the taxonomy proposed, so that it could be argued that there is no need to add them. In reality, these totally unforeseen but consistent suggestions, reveal a task-oriented thinking strategy of educators, which is not taken into due account by the proposed taxonomy. These unforeseen tasks were the focus of the following Design Science Research activities, which designed and evaluated a number of prototypes to research the best way to support them.

5.5.4 Normal distribution of task weights

Concerning the applicability of parametric statistics, Shapiro-Wilk failed to reject the null hypothesis that sample distribution is not statistically different from normal distribution, at a significance level of $\alpha = 0.05$ for every weight measured. While this cannot guarantee that the sample distributions are indeed normal, a visual check on histograms and Q-Q plots further supported the hypothesis that most variables have a normal distribution, even if doubts persisted in a few cases (such as Search Expansion by togetherness). Figure 5.4 reports, as an example, boxplot, histogram, and Q-Q plot for the weights assigned to Publishing, that look approximately normally distributed.

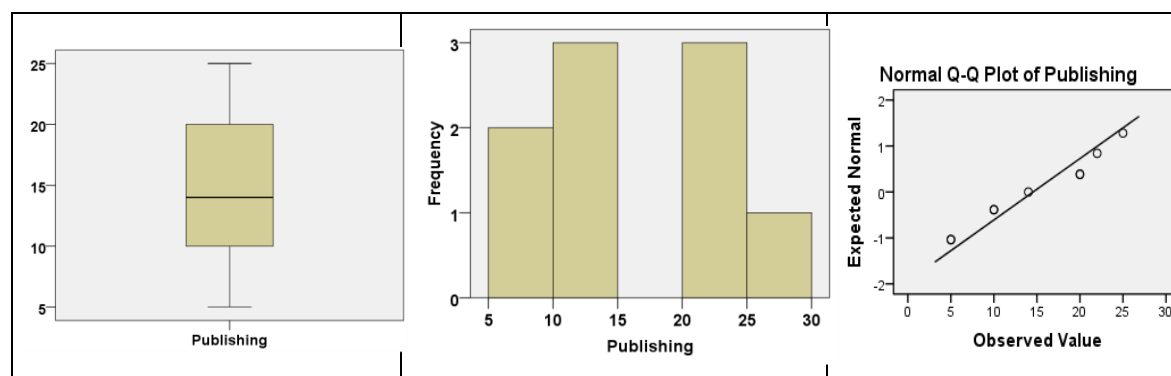


Figure 5.4 – Boxplot, histogram and Q-Q plot for the weights of the task “Publishing”.

5.5.5 Task weights

Figure 5.5 displays the mean with 95% confidence intervals of the weights obtained for the Top-level categories, while Figure 5.6 displays the same information for the sub-categories of Searching. The confidence intervals were obtained using Student's t-test.

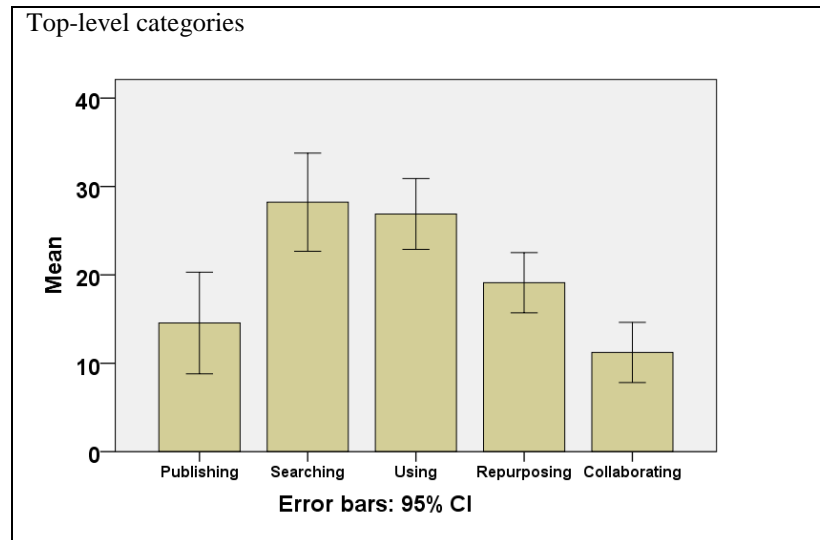


Figure 5.5 – Weights with 95% confidence intervals for top level categories.

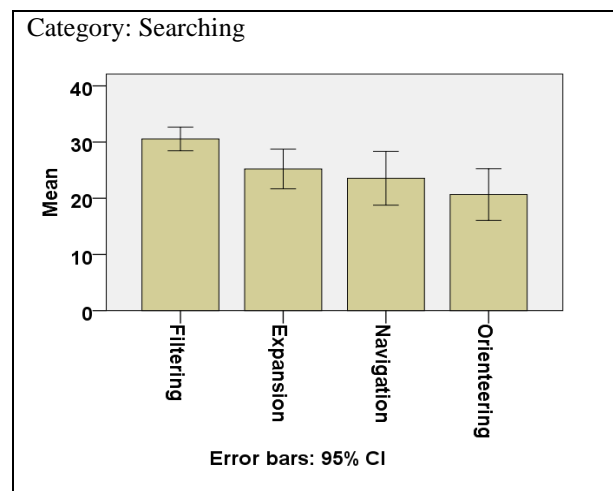


Figure 5.6 – Weights with 95% confidence intervals for the Searching subcategory.

Similarly, Figure 5.7 displays the mean with 95% confidence intervals (obtained with Student's t-test) of the weights obtained for the tasks of category Searching/Expansion.

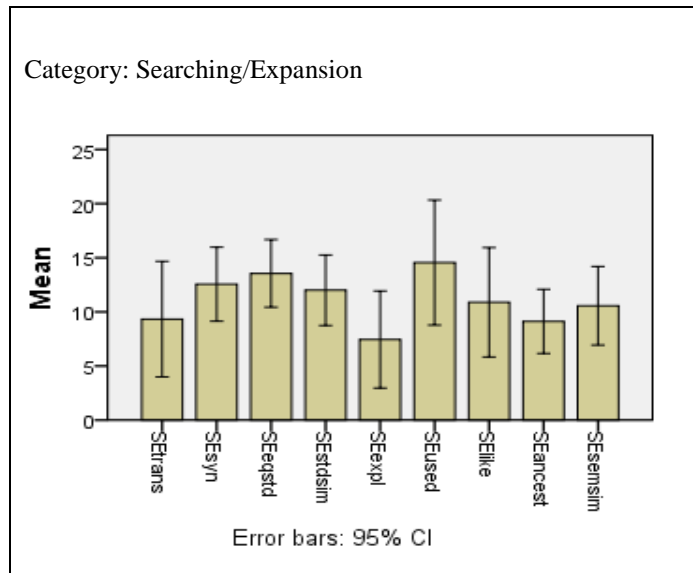


Figure 5.7 – Weights with 95% confidence intervals in the category Searching/Expansion.

Here is a description of the tasks (proposed in the questionnaire), corresponding to the variables in the figure:

- SEtrans: expand the current search by adding the translation of its query terms in different languages.
- SEsyn: expand the current search by adding synonyms of its query terms.
- SEeqstd: discover resources aligned to educational standards equivalent to those of selected resources (QBE). [relatedness metrics: equivalent standards]
- SEstdsim: discover resources aligned to educational standards similar to those of selected resources (QBE). [relatedness metrics: standards similarity]
- SEexpl: discover resources frequently explored with selected resources (QBE). [relatedness metric: explored together]
- SEused: discover resources frequently used with selected resources (QBE). [relatedness metrics: togetherness (used together)]
- SELike: discover resources liked by the same users who liked selected resources (QBE). [relatedness metrics: likedness (liked by same users)]
- SEancest: discover resources derived from the same ancestor of selected resources (QBE). [relatedness metrics: common ancestor in repurposing history]
- SEsemsim: discover resources whose content is semantically similar to selected resources (QBE). [relatedness metrics: content semantic similarity]

Bootstrap analysis could improve the quality of previous results by shrinking the confidence intervals range for the same level of confidence. For example, the confidence interval of the mean of weights assigned to the top-level category “Filtering” obtained by a simple t-test, could be shrunk by about 30% with bootstrapping analysis – at the same level of confidence (95%).

Weights and basic descriptive statistics for each task and category are available in Appendix C.

5.5.6 Planned comparative analysis: relative importance of specific tasks

To find out whether there was an interest on expansion as well as on filtering, Figure 5.8 compares the weights assigned to filtering and expansion: a Paired Samples T-Test with bootstrap gives a means difference of 5.3, with a very significant $p = 0.013$.

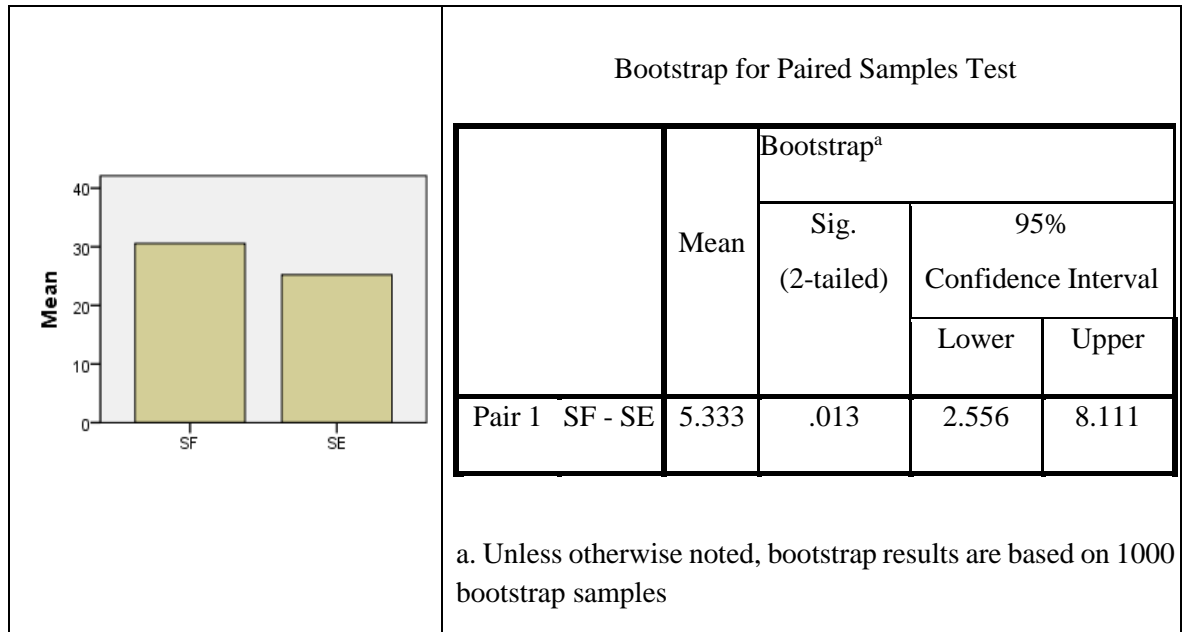


Figure 5.8 – Comparative analysis for categories Filtering (SF) and Expansion (SE).

A significant difference is confirmed also by a non-parametric Wilcoxon Signed Rank Test as in Figure 5.9.

Hypothesis Test Summary				
	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between SF and SE equals 0.	Related-Samples Wilcoxon Signed Rank Test	.026	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 5.9 – Comparing Filtering (SF) and Expansion (SE) with non-parametric test.

There are no doubts that filtering is considered more important than expansion; the difference, however, just amounts to about 17%. This was explicitly motivated by a participant:

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

This was also justified by the higher familiarity of participants with filtering, compared to expansion:

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

Figure 5.10 compares filtering via subject taxonomies (SFsubj) and filtering via alignments to educational standards (SFedst). In this case, the Paired Samples T-Test with bootstrap fails to indicate a statistically significant difference, even if the histogram seems to indicate that filtering via subject taxonomy is more important.

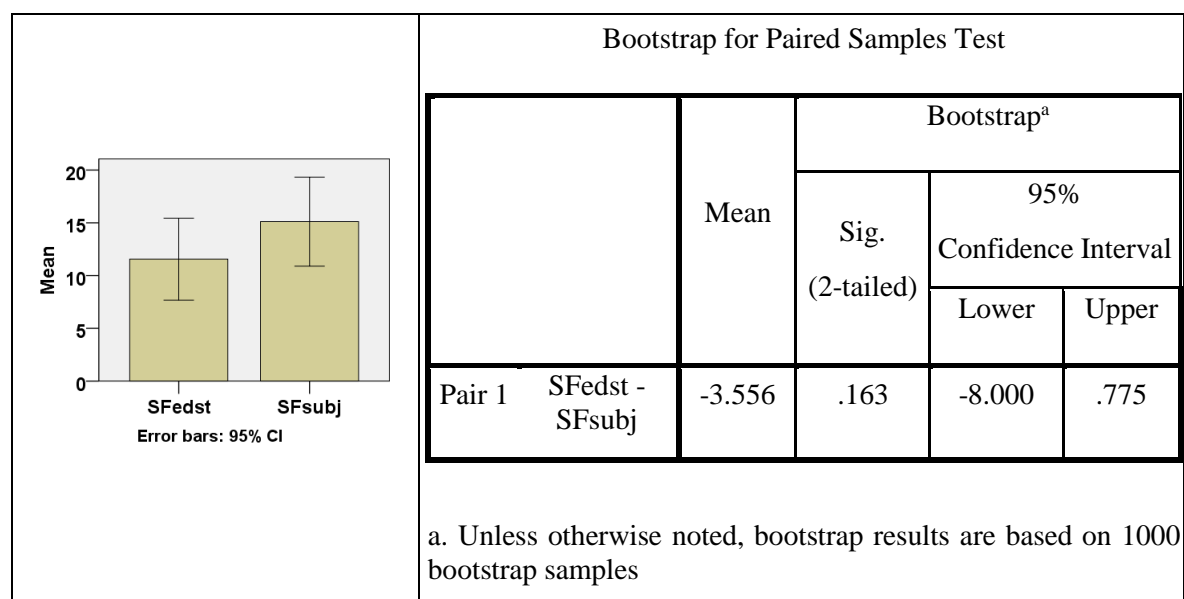


Figure 5.10 – Comparing filtering via subject taxonomies (SFsubj) versus filtering via alignment to educational standards (SFedst).

Qualitative data collected in post-interviews, also suggest a preference to handle explicitly subject taxonomies compared to educational standards. Three participants explicitly expressed this idea, exemplified by the following comments:

“My fundamental reference points are the topics, then I will evaluate myself how the associated learning resources fit my educational objectives in my specific context”

[User E],

“I feel I am more in control by using topic taxonomies” [User G].

This was likely influenced by the strong familiarity of the sampled educators with topics, contrasted to their lack of familiarity with learning objectives:

“The possibility to target a specific educational standard is quite interesting in principle.

But we don’t use formal educational standards!” [User D].

Yet, participants did consider 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” [User G]

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

5.5.7 Explorative comparative analysis among educators with different profile

There was a remarkable difference of perceptions concerning the importance of using non-authoritative metadata, between mature versus young educators, in the preliminary answers that younger educators provided *before* they asked to have their scores modified. Indeed, the initial scores showed that mature educators had a stronger interest in non-authoritative metadata. However, this difference disappeared, after the scores were modified. In order to check if educators with different age – older or younger than 50 years – presented markedly different perceptions on other specific aspects, a number of attempts were carried out following an explorative research approach. These consisted in looking for significant relationships among data potentially worth of further investigation, post-hoc, that is without making hypotheses in advance.

In particular, it was discovered that younger educators gave more importance (mean 31.25) to the need to support the *use* of resources (category “Using”), compared to more mature educators (mean 23.40). The parametric Independent-Sample T-Test at $\alpha = 0.05$, following Levene’s Test for Equality of Variances, shows that the difference of the means of these two groups (mean difference 7.85) is indeed statistically significant ($p = 0.011$). The non-parametric Mann-Whitney U-Test on the same data, confirms a meaningful difference among the two groups, again with a significant $p = 0.032$. The effect-size is also significant: standardized group mean difference Cohen’s $d = 2.39$, effect size $r = 0.77$.

Younger educators considered more important, compared to older educators, also the possibility to navigate to equivalent educational standards. Indeed, the parametric Independent-Sample T-Test at $\alpha = 0.05$, following Levene’s Test for Equality of Variances, shows that the two groups (having means, respectively, 21.3 and 14.2) have indeed mean difference in the 95% confidence interval [3.19, 10.9]. This confidence interval does not contain 0, which indicates that this difference (mean 7.05) is indeed statistically significant ($p = 0.004$). The non-parametric Mann-Whitney U-Test on the same data, confirms a meaningful difference among the two groups, with $p = 0.016$. The effect-size, again, is very significant: standardized group mean difference Cohen’s $d = 2.88$, effect size $r = 0.82$.

It was not possible to find similar results, taking into account the respondents’ amount of experience as teachers, likely because a number of educators had additional working experience in other sectors, and this experience too contributed to the development of their overall way of thinking.

5.5.8 Explorative correlation analysis

An explorative correlation analysis among the tasks/categories weights, identified the following potentially interesting aspects:

- Search/Expansion on likedness (resources liked by the same users) is positively correlated to Search/Expansion on togetherness (resources used together): Spearman $r_s = 0.7$, $p = 0.037$. This indicates, as supported by qualitative information collected in the interviews, that either educators were “socially oriented” and valued both their peers’ opinions and behaviours, or were not interested at all.

- Search/Filtering on educational standards is negatively correlated to Search/Expansion on equivalent or similar standard: Spearman $r_s = -0.716$, $p = 0.03$. This indicates that most educators either are interested to use educational standards for filtering, or for expansion – but not for both.

5.5.9 Additional qualitative data: use of Google

Most qualitative data collected have already been reported in the previous sections to triangulate, that is cross-check and integrate, the results of quantitative analysis. Additionally, consistently with the literature (LRMI, 2013b; Abeywardena et al., 2013), Google was the search engine used by every educator in the sample to look for OERs, even if everybody lamented its limitations for this particular task. For example:

“Too many irrelevant results” [User E],

*“Plain keyword search should be complemented with more advanced functions”
[User B],*

“Too time consuming” [User F].

Despite this awareness, however, just one educator out of nine complemented the use of Google with other specialized OER search engines (in this case Merlot, Curriki, and OER Commons).

5.6 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 indications for modifications. These results are in line with the research literature, as the taxonomy proposed represents by construction a synthesis of the research community’s understanding.

However, the exploratory study indicated that respondents included in the expansion category, tasks that could be alternatively carried out with a combination of less-intuitive lower-level filtering, generalization, and navigation operations on faceted interfaces. Examples of these tasks are finding additional resources with the same educational alignments, classified under the same topics, or even produced by the same authors, as a previously identified resource. While these expansion operations could be considered technically redundant, they represent very useful short-cut operations that are close to the natural task-oriented thinking strategy of educators. Forcing educators to decompose these “natural” tasks in a number of sub-tasks, obliges them to think in procedural terms and to take into account complex underlying data structures, imposing an unnecessary cognitive overload. This is the main finding of this preliminary study, in line with the need to identify domain and user oriented tasks, widely advocated in the scientific literature by authoritative authors such as Marchionini (2006) and Wilson et al. (2010), to advance the research in the field.

Estimating the importance of the tasks and categories identified in the taxonomy, was primarily used as a strategy to promote critical thinking among respondents, in order to elicit valued qualitative

information. However, the internal consistency reliability, the relatively limited width of the 95% confidence intervals denoting a shared view among participants, and the external validity of this study, suggest that the weights collected are of reasonable quality and could be exploited for additional uses. Indeed, Agarwal and Venkatesh (2002) developed an evaluation instrument for the usability of Web sites by following a process similar to this study. They developed first a taxonomy of categories and subcategories of relevant aspects, and then identified weights for each category. Actual systems were then evaluated by collecting ratings for each category from evaluators, and by computing an overall index of usability, summing up all the weighted ratings. Similarly, the OER-search task taxonomy previously identified, and the related weights obtained in this study, could be used as a metric in a framework for the evaluation of educational resource search/discovery applications.

The analysis shows that the scattering of the task importance indicated by different educators, as represented by the range of confidence interval for some of the variables in Figures 5.5, 5.6, and 5.7, can be partially motivated by structural reasons. Indeed, the comparative analysis suggests that educators sharing a similar profile do present more coherent views. For example, the explorative comparative analysis among educators with different profile shows that less experienced educators (considering the overall working experience) consider more important to support the use of resources, compared to their more experienced colleagues. Also, the explorative correlation analysis shows that educators split between those socially oriented, who value both other educator's opinions and behaviour (social data and paradata), and those who do not.

While filtering emerged as the most important operation, expansion is considered highly important too. The statistically significant difference identified, corresponding to less than 20%, is even less important, considering that educators' familiarity with filtering and their lack of familiarity with expansion may have biased their answers in favour of filtering. This is fully confirmed by the qualitative feedback obtained from respondents. This result is in line with the scientific literature. Marchionini (2006) in particular argues that Information Retrieval and information 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.

Filtering via subjects' taxonomies looks, in the bar chart of Figure 5.10, more important than filtering via alignments to educational standards. Yet the Paired-Samples T-Test with bootstrap fails to find any statistically significant difference, even if the hypothesis that subject's taxonomies are privileged is supported by explicit comments expressed by participants. This is certainly influenced by the majority of educators in the sample operating in a context in which formal educational standards are not adopted. It is also consistent with the large efforts devoted by the bibliographic community to topic classification (Frické, 2012), and even with the attempt by the (recently restructured) Open Linked Education Data Community Group (W3C, 2015) to develop a universal topic taxonomy. At the same time, this is not really in contrast with the emphasis on alignment to educational standards for example in the context of LRMI/schema.org (LRMI, 2013c; Barker, 2014), which is originated

in a context in which there is a strong orientation to make formal use of educational standards (Porter et al., 2011). Indeed, while respondents showed higher familiarity with the use of topic taxonomies, qualitative data clearly indicated that they fully realized the advantage to target precisely resources with standard alignments (Section 5.5.6).

While the research by Drachsler et al. (2012) might give the impression that all educators are extremely keen on “social” functionalities, 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, considering the limited representativeness of its sample, demonstrates anyway the importance of obtaining a global picture of the educators’ needs.

Coherently with many other studies (for example LRMI (2013b); Abeywardena et al. (2013)) Google is by far the engine most frequently used by educators searching for OERs, even if they are fully aware of the limitations of its keyword-based mechanism for these particular tasks.

5.6.1 Limitations: sample size and representativeness

The major limitations of this study are the relatively limited sample size, and the limited representativeness of the sample: most respondents teaching technology-related subjects, from Italian high schools. While the Italian context is the main target of this research, the generalization of its results would certainly benefit from a larger size, more representative of a wider population such as educators teaching different subjects.

The study is based on respondents’ current views, which can rapidly change, as demonstrated by the different importance assigned to non-authoritative metadata by young educators, before and after discussion. Kules and Shneiderman (2008) show that perceptions may also change following a direct experience with applications supporting new functionalities. Thus, a longitudinal research study would provide useful insights concerning the evolution in time of educators’ priorities.

Finally, the empirical study is based on the opinions of educators, even if well motivated and experienced, not on their actual behaviour: while opinion and behaviour can be expected to be strongly correlated, it may be wise to verify it in a more realistic setting.

However, while the sample was relatively limited in size for the quantitative analysis, and the study would certainly benefit from additional longitudinal studies or research based on actual behaviour rather than opinions, it was still possible to obtain significant results. In some cases p-values lower than 0.01 and effect size higher than 0.8 were obtained, with results confirmed by both parametric and non-parametric analysis. More important, the qualitative analysis helped to triangulate quantitative data and better understand educator concerns – which did not always correspond to what

is commonly assumed. These preliminary results, which were used to initiate subsequent studies, were anyway further tested and confirmed.

5.6.2 Findings: new domain-oriented tasks

The study confirmed the overall validity of the OER task-taxonomy, produced by the previous task analysis (RQ1), and contributed weights indicating the relative importance of each task and category (RQ2).

The main finding of this study was the identification of new expansion functionalities to support the natural task and domain-oriented thinking strategy of educators (RQ2).

Another important finding was the understanding that educators, in a context in which educational alignments are not formally used, are fully aware of their potential for precisely targeting learning resources, but are quite reluctant to use them explicitly (RQ3).

A minor finding, fully consistent with the literature (LRMI, 2013b; Abeywardena et al., 2013), was the confirmation that the great majority of educators makes use of Google to look for educational resources, and despite being fully aware of its limitations, do not make use of more specialized portals.

5.7 Conclusions: new expansion task transparently based on educational alignments

This empirical evaluation could support the validity of the OER task-taxonomy, previously synthesized from a review of the literature and existing platforms, and interpreted in the light of the Information Foraging Theory (RQ1). The study also produced a set of weights indicating the importance for educators, of each task and category in the OER task-taxonomy. These weights could be used as a usability metric in a framework for the evaluation of applications for search/discovery of educational resources (Agarwal and Venkatesh, 2002).

The study revealed that educators tend to conceptualize certain high-level tasks as additional ones in the Expansion category, even when they could be carried out as a combination of lower-level operations. An example of these expansion tasks is finding additional resources sharing the same educational alignments (or the same authors, tags, topics) as a sample resource. Carrying out these conceptually simple domain-oriented tasks on commonly available interfaces, requires users to decompose their original task in a non-intuitive sequence of lower-level specialization, generalization, and navigation operations. This requires users to think in terms of a complex procedural approach shaped by the underlying data structure. Providing an answer to research question RQ2, this suggested the opportunity to research further the possibility of assisting users with tools that can directly support these domain-oriented tasks (Marchionini, 2006; Wilson et al., 2010), at a declarative level, without requiring their procedural decomposition in sub-tasks.

Educators in the sample were definitely aware of the suitability of alignments to educational standards to target precisely educational resources. Yet, there was resistance to handle explicitly

educational standards, at least in a context where they are not formally used: educators felt more comfortable dealing with more familiar topic taxonomies. A similar situation was identified by OpenScout designers (Kalz et al., 2010), who recognized the importance of aligning the resources to competency frameworks but, at the same time, realized the need to provide alternative mechanisms. This suggested the opportunity to exploit the recognized power of educational alignments in an expansion by similarity operation, without the need for users to be aware of them, contributing to RQ3.

These suggestions for further exploration in this research are also supported by the literature, as discussed in Chapter 4. For example, Wilson et al. (2010, p. 78) claim that “*some techniques have received little published study and evaluation, like allowing users to view similar results to any one result returned, and so our understanding of their benefits is [only] through experience and intuition*”. Additionally, the similarity metric proposed is very much domain-oriented, and has therefore potential to support tasks at Work Context. This makes it a good candidate for further exploration, because there has been “*comparatively little work that has focused on work-contexts*” (Wilson et al., 2010, p. 79).

As widely reported in the scientific literature, Google was indicated as the search engine most frequently used by educators looking for learning resources; despite being fully aware about the limitations of its keyword-based search mechanism, the great majority of educators in the sample did not make any use of existing dedicated portals. This suggested the opportunity to provide the proposed functionalities on top of this widely used search engine.

As summarized in Figure 5.11, the results of this study provided the initial input to the following DSR sequence of prototypes design and evaluation activities, which focused on providing expansion by similarity functionalities directly from Google result pages. These functionalities make use of a similarity metric based on alignments to educational standards, without requiring educators to be aware of them.

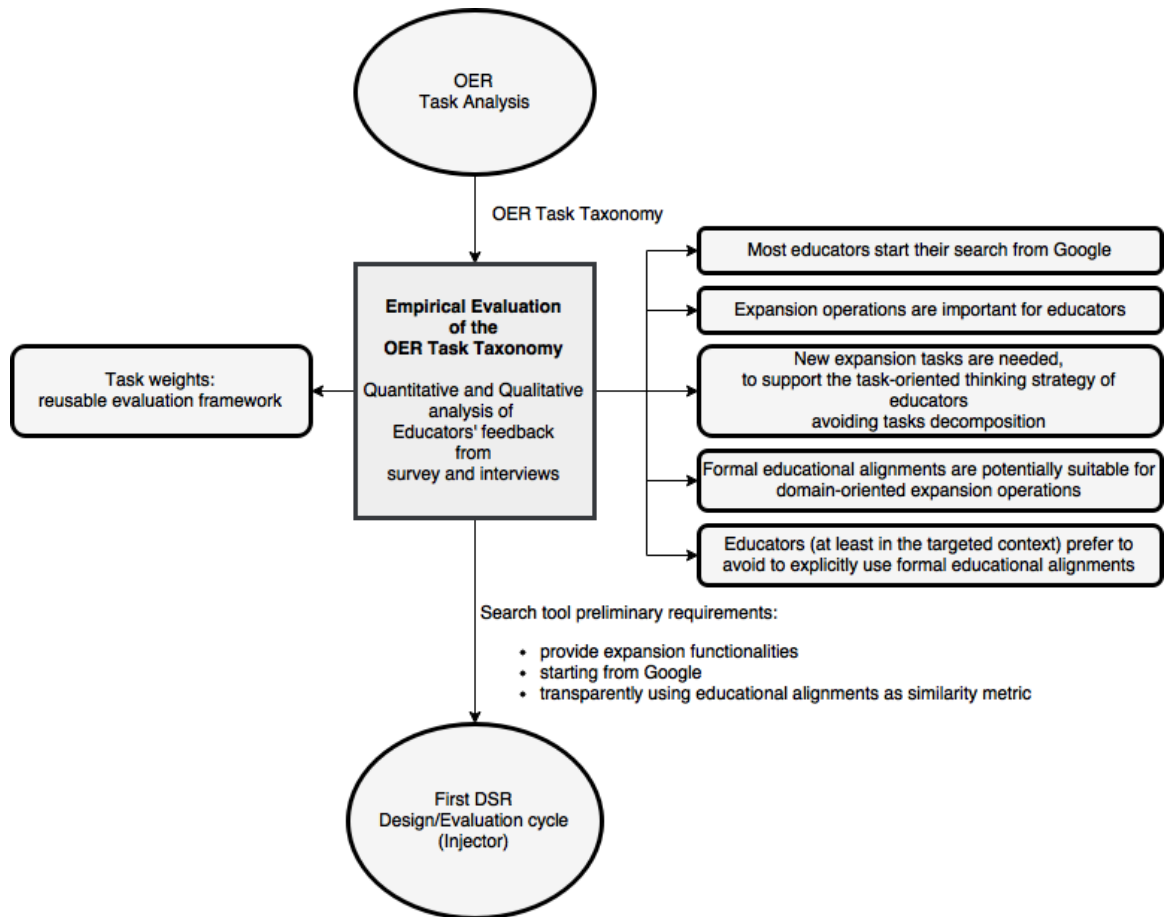


Figure 5.11 – OER Task-taxonomy Empirical Evaluation: an overall map.

6. Injector: first design and evaluation DSR cycle

The previous taxonomy empirical evaluation confirmed the importance of the expansion category of tasks indicated by the literature. More importantly, it showed that educators tend to conceptualize certain high-level tasks in this expansion category, as shortcut combinations of lower-level filtering and navigation operations on modern faceted interfaces.

The findings of the previous study suggested the opportunity to further explore the possibility to exploit the acknowledged power of educational alignments, without users being aware of them, to identify similar resources in an expansion operation, starting from Google Search Engine Result Pages (SERPs). According to the core methodology adopted in this research (Chapter 3), a sequence of prototypes was developed in three design/evaluation DSR cycles, to clarify user requirements preliminarily identified in the previous study, establish their relevance, provide design knowledge on how to build tools to support them, and establish to what extent the proposed solution can support educators in their high-level domain-oriented tasks.

This chapter describes Injector, the prototype designed and evaluated in this first DSR cycle. Injector was evaluated, following its design and development, in a discounted heuristic evaluation with actual educators, mainly to identify possible shortcomings and new research questions to be addressed in the following cycle.

6.1 Injector research questions: design and knowledge oriented

The design and evaluation of Injector, the first of a sequence of three prototypes, mainly aimed to contribute to the investigation of the following research questions:

RQ3: What are the general requirements for tools supporting the aspects of interest identified?

RQ4: Which tools could support the identified requirements, and how do teachers respond to them?

The second question (RQ4) reflects a design oriented goal and has the implicit typical structure of “know how” oriented research questions in Design Science Research. However, this research aims to answer “know what” oriented research questions too, reflecting more knowledge oriented goals (Wieringa, 2014). In particular, while the requirements for the tool had been already preliminarily identified from the previous study, the evaluation of this prototype, as well as the following two prototypes, also aimed to contribute incrementally to the first research question.

Relevant lower level questions, which expand on the original RQ4 research question, were:

RQ4.1: How can an OER discovery tool be embedded into web search?

RQ4.2: How can Query By Example be added to an OER discovery tool?

RQ4.3: How can similarity based on educational alignments be added to an OER discovery tool?

Coherently with the methodology adopted in this research, the experimentation with this first prototype was not supposed to provide definitive answers to the above research questions, but preliminary contributions which would be complemented by further contributions in the following DSR cycles. The development and evaluation of this first prototype generated, as intended, additional more specific design research questions, addressed in subsequent cycles.

6.2 Preliminary requirements: resources with similar educational alignments from Google SERPs

Following the findings from the task-taxonomy empirical evaluation, recalled in Section 6.1, the entry requirements for this first prototype were:

- the prototype should make available its discovery functionalities starting from the familiar context of the search platform educators most frequently use (that is, Google);
- the prototype should implement the “expand to similar resources” expansion operation, avoiding its decomposition in unnatural sequences of specialization / generalization / navigation operations, aiming to support the natural task-oriented thinking strategy of educators;
- the prototype could take advantage of the appreciated domain-oriented educational alignments. However, according to the preferences expressed by educators, educational alignments should be exploited in a transparent way. In other words, educators using the system should not be required to deal with educational alignments, which should be handled automatically, without the need for educators to be aware of them.

It is worthwhile reiterating that while these preliminary requirements, derived from the empirical evaluation of the task-taxonomy, were used as input to this first prototype (Injector), 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 all the following DSR cycles.

6.3 Injector Design and Development

6.3.1 Overview: custom rich snippets injected in Google SERPs

In order to satisfy the previously defined requirement of providing discovery functionalities directly from the familiar context of the search platform educators most frequently use, that is Google, the prototype makes use of Chrome (2016) extension technologies. They are particularly suitable to intercept and analyse HTML pages received from the network, and to modify them by injecting additional information. These basic techniques were previously demonstrated by the “Browser Plugin” (Lockley, 2011) and “AMPS” (Klo, 2011) prototypes, even if they were just limited to the injection of static metadata and paradata in Google SERPs, and they are no longer working because

of the evolution of Chrome APIs, and the new orientation to schema.org of the metadata schema of the services they were using.

Injector, the browser extension prototyped in this study, works as follows:

- it identifies (“scrapes”) the URLs of results from Google SERPs, and checks if they correspond to educational resources for which there are metadata available;
- if this is the case, it also attempts to obtain similar resources, which are identified and ranked according to a similarity metric designed in this research, based on educational alignments;
- it builds, for each resource, a tiny HTML document segment, or custom “rich-snippet”, containing the available educational metadata and an active button, to allow users to expand it to (or request suggestions for) similar resources;
- these custom rich-snippets are finally injected in the original Google SERP, which is therefore extended with specific educational metadata and “transparent” expansion functionalities to (or suggestions for) “similar” resources.

The user-oriented characteristics of the prototype are best illustrated by the following sample session.

6.3.2 Injector sample session

This section illustrates a search/discovery session actually carried out with the prototype, with a short but meaningful sequence of screenshots.

Figure 6.1 displays the SERP page returned by Google, following a search with the keywords “tree diagrams and figure probabilities of events”.

Following a mouse-click on the Injector extension button, visible in the top right position in Figure 6.1, the system was able to identify a learning resource in the previous Google SERP, where a corresponding custom rich-snippet is injected, as displayed in Figure 6.2. This custom snippet contains the available LRMI/schema.org metadata, plus an expansion button, to allow users to request additional similar resources. The expansion button contains a “volume indicator”, that is the number of similar resources available, to provide users with orienteering information (Sacco, 2000; Hearst, 2009).

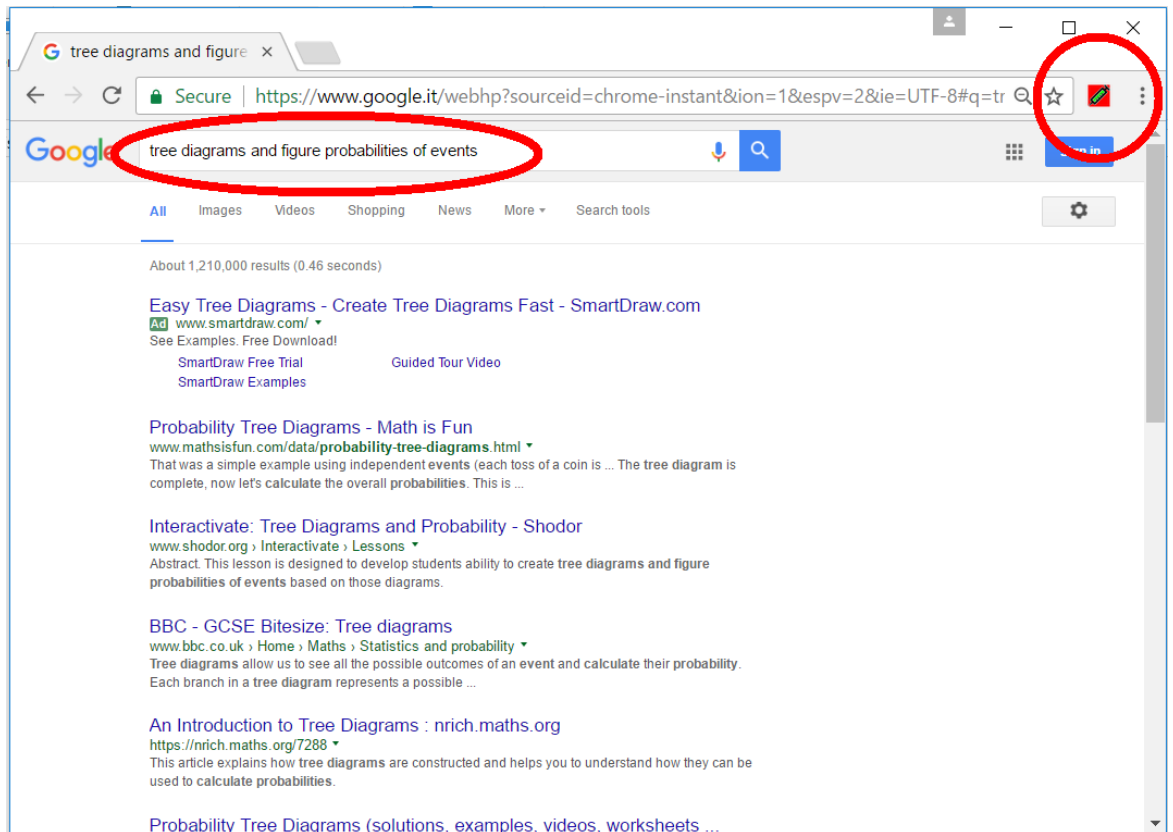


Figure 6.1 – Traditional Google SERP page.

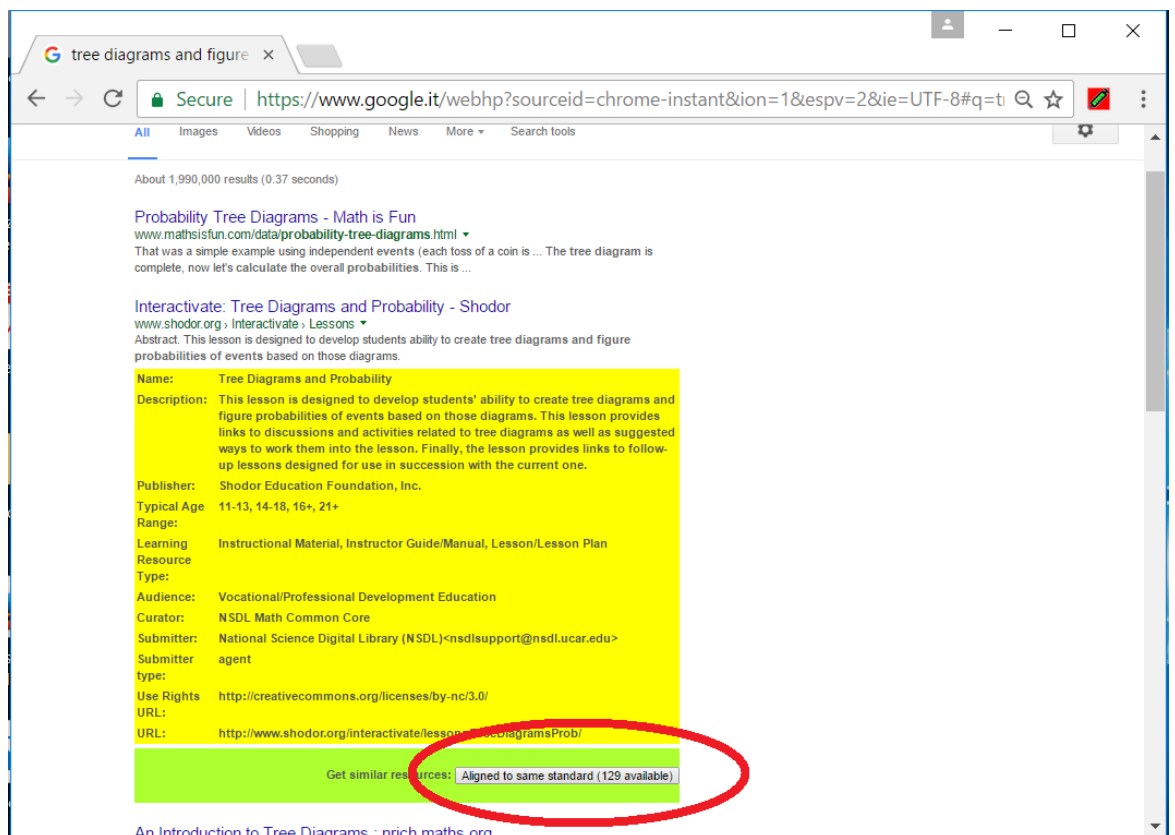


Figure 6.2 – Previous Google SERP with a custom rich snippet inserted, containing the available metadata and an expansion button.

Clicking on the available expansion button, displays a new “expansion” window, shown in Figure 6.3, containing similar resources, again with their relevant educational (LRMI) metadata. The resources listed in the expansion window are identified and ranked with the similarity metric implemented in the prototype, and displayed in order of similarity. Of course, the expansion window also contains links to navigate directly to the educational resources of interest.

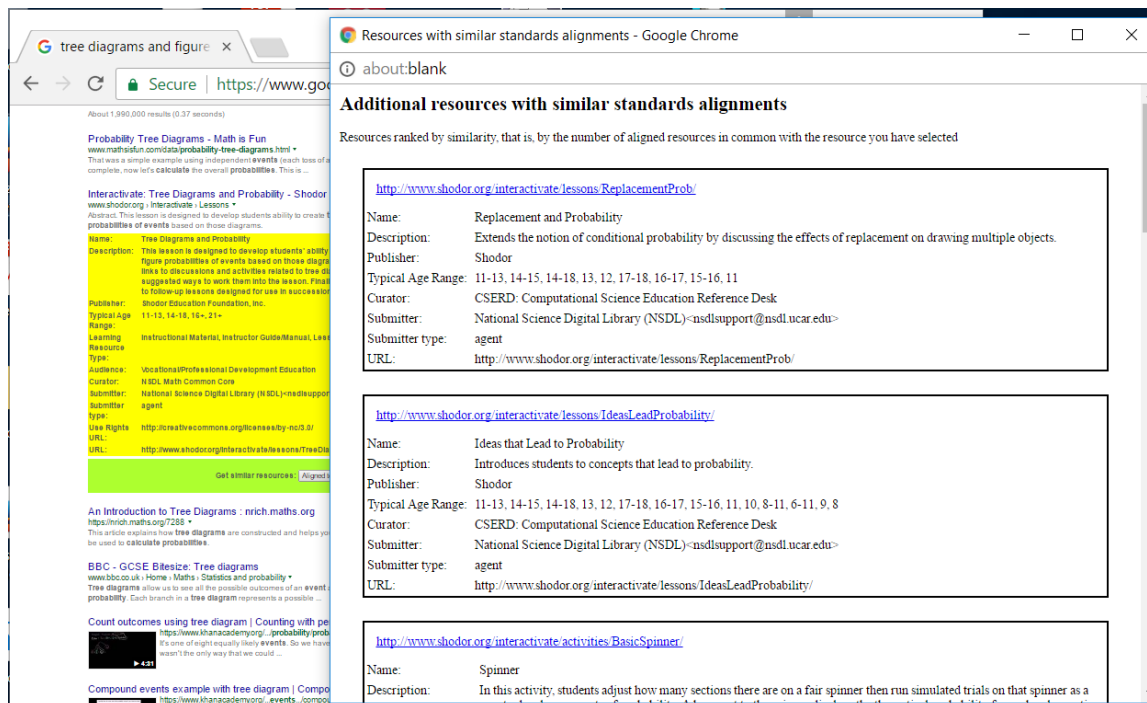


Figure 6.3 – Expansion window: related resources ranked by similarity, with relevant educational metadata.

6.3.3 Similarity metric based on shared educational alignments

Similar resources are identified through a similarity metric, defined in this research in terms of educational alignments as follows:

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

This similarity metric makes it possible to reuse the outstanding efforts in aligning resources to educational standards, without the need for educators to be aware of them, conforming to the requirement to avoid the need for educators to handle explicitly educational standards.

The identification of similar resources with this metric could be carried out, in principle, on existing faceted interfaces, but that would force educators to split their original simple task oriented goal (find resources with similar alignments) in a number of complex subtasks. Yet the prototype, making use of this metric in a simple expansion operation, conforms to the requirement to support the task oriented thinking strategy of educators, completely hiding the complexity of the procedural aspects involved, such as metadata extraction and processing involving multiple requests to different services, behind a single button click.

This novel similarity metric, novel in the sense that I could not find any mention to it in the literature, makes use of precisely-targeted and professionally-curated crowdsourced data. It is a strongly domain oriented metric that has the potential to generate more relevant results than generic metrics based on keywords.

6.3.4 Selection of datasets and services providing alignments metadata

The prototype, specifically the similarity metric defined in this study, needs educational alignments metadata for a volume of educational resources sufficiently large to carry out realistic experiments. This called for an analysis of schema and content of available datasets and services. The Learning Registry (2011), an aggregator of metadata, paradata and social-data undertaken by a number of federal agencies in the USA, was finally selected as the preferred option. Its value in the context of this research lays in the rich metadata that it collects and makes openly available, and its (RESTful) services that make it possible to experiment and develop innovative user oriented solutions.

Other potential solutions were examined but discarded for a number of reasons. In most cases, such as the ARIADNE (2014) SPARQL (a semantic query language to retrieve and manipulate data expressed in RDF) endpoint, or the Open University (2016) Linked Data datasets, there were no alignments to comprehensive educational frameworks available. An alternative potential solution was to exploit schema.org educational metadata embedded in web pages via Google Custom Search (Google, 2017), but at that moment there was a limited amount of educational metadata available (Cortinovis, 2016), especially related to educational alignments (Dietze et al., 2017). Gooru (2016a) announced the availability of API services to access their data, which also include educational alignments; however, at that moment, these services were still under development. Finally, OpenEd was another interesting potential solution, as they enrich with LRMI metadata, including alignments to educational standards, resources harvested from the Learning Registry (OpenEd, 2016b) and other repositories. Indeed, they even suggested to use their APIs to access information about Learning Registry resources, because they claimed that their search API enabled more fine-grained queries. Unfortunately, they had very restrictive access policies: in particular, they required the copyright on any tool developed using their data. Therefore, they refused to grant access to the data as the intention to make the prototype available in the public domain was expressed (Drabkin, 2016b). However, they also announced (OpenEd, 2016c) that they “will be publishing alignments back to Learning Registry”, in addition to other educational metadata (Drabkin, 2016a), which further supported the Learning Registry as a suitable solution.

6.3.5 Injector architecture

Figure 6.4 illustrates the general architecture of Injector. It makes use of the Chrome extension API (Chrome, 2016) which, as previously observed, is an effective technology to parse programmatically and modify the content of Web pages and data obtained from the network (SERPs from Google and JSON_LD data from the Learning Registry in this case), as well as to perform cross-origin XMLHttpRequests. These requests provide the basic mechanism for a client browser to obtain data from a server. The term “cross-origin” refers to the possibility to obtain data from a server which is

different from the server a web page was loaded from. Despite the acronym “XML” in the name of the request, it is also possible to retrieve data in different formats, such as JSON_LD in this case. These network requests are used to access the Learning Registry services; the Obtain Services to get general educational resources metadata; and the Data Services to obtain specific information concerning the alignment of resources to educational standards (Learning Registry, 2016b).

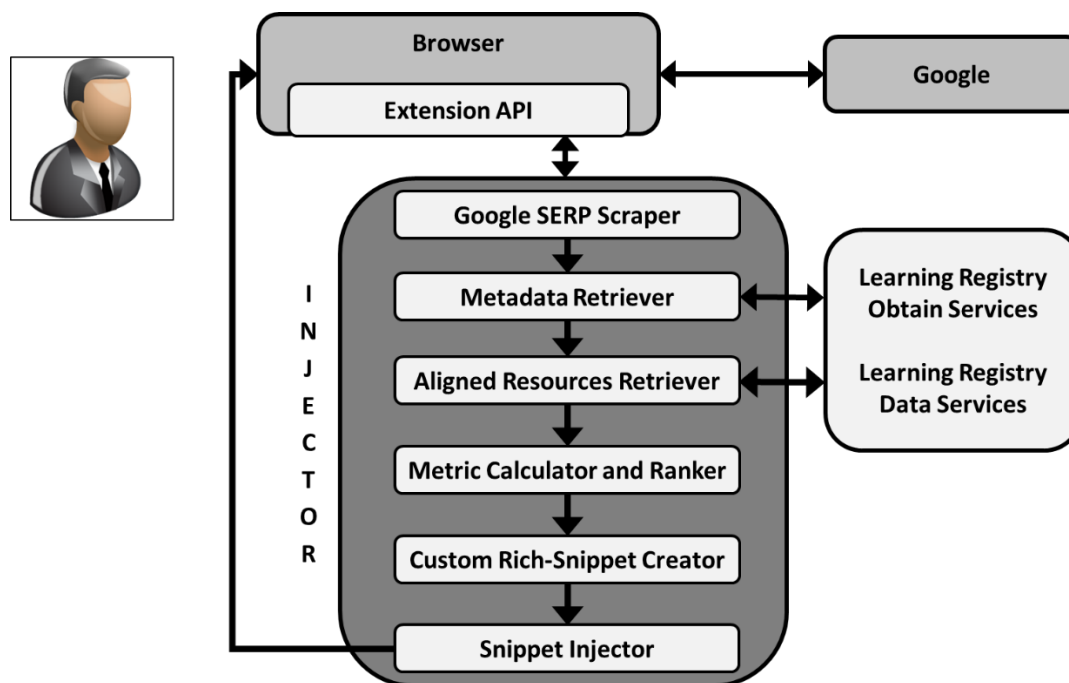


Figure 6.4 – Injector architecture.

The SERP Scraper extracts the URLs of Google results from its SERP. The Metadata Retriever attempts to find educational metadata for the extracted URLs, which potentially identify educational resources, by querying the Obtain Services of the Learning Registry and parsing its results. For each URL for which educational metadata are available, the Aligned Resources Retriever attempts to obtain the educational standards the corresponding resource is aligned to, by querying the Learning Registry Data Services. If the operation is successful, the Metric Calculator and Ranker identifies similar resources by requesting the resources aligned to each standard previously identified, again to the Data Services, computes their distance according to the metric identified in this research, and ranks them in order of similarity. At this point, custom rich-snippets are created with relevant educational metadata and, when similar resources could be identified, expansion buttons. These custom rich-snippets are finally injected in the original Google SERP. Whenever the user clicks an expansion button, the associated similar resources are displayed in a separate window with relevant educational metadata.

The algorithms

The conceptual simplified algorithms, abstracting from concurrent processing and optimizations, are reported in the listings in Figure 6.5 and Figure 6.6. The algorithm in the first listing is activated when a user selects the browser Injector extension button from a Google results page. The algorithm

in the second listing is activated when a user selects an expansion button in one of the custom snippets injected by the previous action. The calls to the Learning Registry Obtain and Data Services to get the necessary data, are indicated by the label “REST”.

```

When user clicks on the Injector browser extension button do:
  HitsList <- Scraped URLs of hits from Google SERP
  For-each potential resource ResID in HitsList do:

    [M – inject available educational metadata:]
    ResIdMetadata <- REST (Attempt to get ResId metadata)
    If ResIdMetadata is not null, create snippet and inject it in original Google page

    [M – inject expansion functionalities:]
    ResIDEdStdList <- REST (educational standards the resource ResID is aligned
to)
    AlignedResList <- []
    For-each educational standard EdStd in list ResIDEdStdList do:
      AlignedResList <- concatenate ( AlignedResList,
        REST (resources aligned to standard EdStd) – {ResId})
    End for-each
    NumRes <- number of unique resources in AlignedResList
    Inject expansion button in Google SERP with Volume Indicator NumRes
  End for-each
End when

```

Figure 6.5 – Clicking the browser Injector extension button: conceptual simplified algorithm.

```

When user clicks on an injected expansion button do:
  [M - compute number of standards in common between ResID and AlignedRes:]
  For-each unique resource AlignedRes in list AlignedResList do:
    NumStdInCom[AlignedRes] <- number of occurrences of AlignedRes in
    AlignedResList
  End for-each

  [M - rank similar resources:]
  AlignedResRankedList <- unique resources in AlignedResList, ranked by NumStdInCom

  [M - create snippets for the identified similar resources create custom SERP:]
  For-each resource SimRes in AlignedResRankedList do:
    SimResMetadata <- REST (Attempt to get SimRes metadata)
    Create snippet with SimRes and SimResMetadata
  End for-each
  Display in a separate expansion window snippets of ranked resources in
  AlignedResRankedList
End when

```

Figure 6.6 – Clicking the resource expansion button: conceptual simplified algorithm.

Algorithms optimization

In the previous sequential algorithms formulation, the amount of calls to Learning Registry services, that runs in the dozens calls, would result in queries being returned on an unacceptable timeframe. Therefore, the above algorithms have been implemented using concurrent programming techniques to achieve an acceptable response time, at least when the Learning Registry infrastructure is not too taxed by other requests. The asynchronous updating of the visualized page as soon as any data becomes available, generally starts in less than half a second and completes in less than two seconds. According to the literature, these figures are suitable to guarantee a comfortable user experience. Indeed, Miller (1968, p. 270) indicated more than fifty years ago that a couple of seconds is an appropriate response time for this type of interaction (“response category”), where it is important “to maintain the continuity of human thought processes” with a natural “conversational speed”. Nielsen (1993) writes that 1 second is approximately the limit to avoid the interruption of user’s thinking flow. More recently, the Google RAIL model (Kearney, 2017) specifies more precisely that delivering the essential parts of interactive content within a window of 300 – 1000 ms, and loading remaining data progressively, is appropriate to keep users engaged.

The algorithms have been optimized also in other aspects. In particular, in order to reduce the load on the Learning Registry infrastructure, the prototype caches all the Learning Registry responses, in order to avoid multiple identical requests in the same session.

6.4 Injector Discounted Heuristic evaluation

6.4.1 Goals: identify Injector major drawbacks and get preliminary feedback

The adopted Design Science Research (DSR) methodology envisages the iterative development and evaluation of a prototype in each cycle of a series, to drive further developments and evaluations in the following cycle. The main objective of this evaluation was therefore to collect feedback to efficiently identify potential major drawbacks of the first prototype, to be addressed in the following cycle, as well as to collect preliminary evidence about the relevance of the problem identified and the suitability of the proposed conceptual solution. This preliminary feedback was intended to be supported by further evidence collected in subsequent cycles.

6.4.2 Methods: discounted heuristic evaluation

Injector was evaluated in this first cycle with a discounted heuristic evaluation, by a small but representative sample of users from different contexts. A discounted evaluation is appropriate for early prototypes, because it makes it possible to quickly identify major problems at an early stage (Nielsen, 1995). As recommended by Nielsen (1994), it is more effective to run multiple inexpensive agile evaluations distributed along the design cycle, rather than a single expensive summative evaluation. This suggestion is consistent with the adopted methodology, where this first evaluation was complemented by additional evaluations conducted on more advanced prototypes in the following cycles (involving an incrementally larger number of users from different contexts).

An evaluation questionnaire was designed (available as Appendix D) that, following a preliminary section to collect demographic data: proposes a scenario with predefined tasks to be carried out with the prototype; invites test-users to reflect about how they could carry out the proposed tasks with the tools they normally use; and provides a selection of questions to act as a heuristic checklist.

Some of the questions were obtained by selecting and adapting, from evaluation heuristics available in the literature, the items applicable to this case. In particular, question 5 in the evaluation form is derived from Molich and Nielsen (1990), later also adopted by Muller et al. (1998), question 8 from Gerhardt-Powals (1996), question 7 from the widely used and reliable (Bangor et al., 2008) System Usability Scale (Brooke, 1986). Here it made little sense, for example, to consider other aspects such as learnability, memorability, or help and documentation. Other questions were developed considering the specific objectives of this activity. The heuristics covers elements related to three fundamental aspects: functionalities (major one), usability, and user experience. While strongly interrelated, functionalities are more concerned about the application, usability about the interaction, and user experience about more general or holistic aspects (McNamara and Kirakowski, 2006). The questionnaire also includes open questions to collect some qualitative feedback, including about possibly unanticipated aspects.

The evaluation questionnaire was intended to be administered as a structured interview when feasible, or as a self-administered survey supported by synchronous and asynchronous communication. It was written in English, because the targeted participants were known to be conversant in this language.

The evaluation of Injector in this first DSR cycle was also intended as a pilot of the evaluation questionnaire. Indeed, the questionnaire would be substantially reused, even if with the necessary modifications reflecting the evolving goals, in subsequent evaluations.

Because of the limited sample size, there was no intention to carry out a quantitative analysis at this stage, which was left for the summative evaluation in the last cycle. However, basic descriptive statistics were considered useful to give a first appreciation of the position of sampled educators. The qualitative data were expected to provide the sought-after indications of current prototype drawbacks, to be addressed in the subsequent cycle.

6.4.3 Injector evaluation: results and discussion

A number of experienced educators from different contexts, were invited to participate in the evaluation (Table 6.1). Four of them accepted to participate in this first heuristic discounted evaluation, one female and three males, three from Italy and one from the UK, all involved in teaching technical subjects at secondary education (one including experience at University level), with 6 to 20 years of experience.

User ID	Gender	Working Country	Teaching Experience [years]	Teaching Educational Level	Teaching Subject
User1	M	UK	6-10	University	Technology
User2	M	Italy	11-20	Sec. Educ.	Technology
User3	F	Italy	11-20	Sec. Educ.	Technology
User4	M	Italy	11-20	Sec. Educ.	Technology

Table 6.1 – Injector evaluation Test-users profile.

There was no overlap of participants in this study with participants from the previous study, except for User1 (corresponding to UserB). The evaluation questionnaire previously discussed was administered as a structured interview in two cases, and as a self-administered survey in two other cases. In the first two cases, field notes were annotated directly on the survey, and test-users were encouraged to modify or comment on them. In two cases the system was utilized directly, while in the two cases managed remotely, a representative demo screencast was provided in addition to a remote demonstration, with the possibility of requesting any clarification.

The demographic data requested were very basic and optional, and no other sensitive information was collected. However the data were anonymized and all results reported anonymously or aggregated.

The relevance of the tasks in the proposed scenario, measured on a Likert Scale anchored from 1 (totally irrelevant) to 7 (very relevant), obtained a mean of 5.8. This value is definitely higher than the neutral value of 4, even if not the highest possible, likely because the prototype addresses, by design, only a subset of all possible requirements. This is supported by some educators' suggestions to include potential extensions unrelated to the current focus of this research, such as the possibility for users to contribute to the classification of the resources, the inclusion of additional metadata, or the inclusion of resources in different languages.

Educators indicated that it would have been very difficult to carry out these tasks by means of alternative tools they normally employ, with a mean of 6 on a scale from 1 (very easy) to 7 (very difficult). In this regard, they consented – mean of 6.8 on a scale from 1 (very easy) to 7 (very difficult) – that the prototype automatizes unwanted workload.

Consistent with the above data, educators unanimously indicated that they would like to use the system frequently: mean of 7 on a scale from 1 (fully disagree) to 7 (fully agree).

This is consistent with a number of very positive comments. Some of them were quite general:

“I think this is the potentially perfect ‘all-in-one’ instrument for us educators” [User2],

“Wonderful idea” [User4],

“Opens up, literally, a whole new dimension in knowledge and content searching”

[User1],

Other comments were more specific:

“Educational metadata are much more useful than the traditional snippets provided by

Google” [User3].

Coherently with the results of the empirical evaluation of the task-taxonomy, the discovery oriented functionality based on educational alignments was, again unanimously, considered very useful, with mean 7 on a scale from 1 (totally useless) to 7 (very useful). In particular, 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). One of the participants explicitly reported:

“The transparent use of educational alignments is very much appreciated, as I am not

familiar with existing standards” [User3].

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).

Shortcomings and areas for improvements

However, despite the fact that one of the educators even managed to actually identify a useful resource

“I have even found a very useful resource for the lessons I am currently planning”

[User3],

they clearly indicated that the tool was of limited practical utility, because of the modest number of educational resources it could identify in Google pages. Some comments were definitely explicit:

“In many cases it does not produce any result” [User3],

“It obviously identifies too few resources” [User4].

The following comments were less direct, but sufficiently clear beyond their face value:

“As a prototype this is fine” [User1],

“Potentially very useful” [User3].

Indeed, while very positive on the potential of the idea, the expressions “as a prototype” and “potentially” purport that the system was not considered to be really practically useful.

Two participants in particular, expressed the wish to see exclusively the educational resources:

“I would prefer to see just the educational resources in [Google] results pages” [User2],

“It would be better if it could offer only educational resources” [User3].

This is understandable, because Injector results look sparse among many other hits when Google returns results.

6.4.4 Evaluation conclusions: the challenge of sparsity

The analysis of the qualitative data obtained from the educators involved in this first discounted evaluation, achieved its overall goal by clearly and efficiently spotting the major drawbacks of the prototype. The evaluation provided also some preliminary data about the relevance of the problem addressed and the suitability of the solutions proposed. These data, considering the limited size of the sample, were not sufficient to carry out a quantitative statistically significant analysis, yet could provide useful preliminary indications. A numerically larger sample of test-users, better suited for quantitative analysis, was going to be involved in the evaluation of more advanced prototypes.

The feedback obtained from this initial sample of educators, both quantitative and qualitative, indicates a consistent enthusiastic appreciation for the proposed approach to “augment” Google’ search functionalities by injecting relevant educational metadata in its traditional snippets, and providing expansion by similarity functionalities transparently based on educational alignments. This is consistent with additional positive feedback received from the Learning Registry management, who expressed interest in discussing the integration of the prototype on their platform (Midgley, 2016).

However, while the tool was highly appreciated conceptually, qualitative feedback from test-users, interpreted beyond its immediate face value, provided more or less direct, yet very consistent indications (triangulation among different participants), that the prototype, while much appreciated in principle, was considered to be of limited practical utility, because of the modest number of resources that could be identified within Google result pages. More precisely, the Injector prototype presents the following main problems that needed to be addressed in the following DSR cycle:

- **intrinsic sparsity:** there are a limited number of educational resources identified as such, within unconstrained Google SERPs. This is a structural limitation, because Google results contain heterogeneous resources that are related just because they share some search keywords. Hence they contain many items which are not “educational” resources. Of course, it might also happen that genuine educational resources do not get identified, because no related educational metadata could be found in the Learning Registry.
- **alignments sparsity:** among the few educational resources identified, there is an even more modest number of resources for which alignments metadata are available. In this case the expansion is not possible, because no similar educational resources can be found with the adopted metric.

The explicit suggestions to include exclusively educational resources in the result pages are particularly intriguing, because apparently in contradiction with the satisfaction expressed towards the proposed strategy to identify, highlight, and enrich resource snippets **within** standard Google pages, which could necessarily contain non-educational resources too. However, the idea was further developed and evaluated in the following prototype.

Limitations

A first limitation of this evaluation was the limited sample size: a larger sample, especially from different contexts, would increase the generalizability of results. Yet, it is generally accepted that discounted formative evaluations (that is with a few participants) are effective in providing adequate evidence to identify deficiencies in a prototype, especially in a context of frequent prototype development and evaluation iterations, and when established heuristics are used (Nielsen, 2013). Indeed, this limitation did not prevent the achievement of the main goal of finding shortcomings in the current design (that is sparsity). Nielsen (1994) also argues that it is generally more efficient to run frequent studies with fewer participants, rather than a single study with a large number of participants. More specifically, he suggests that it is usually more efficient to involve a limited number of test-users in the early phases of a design, and to progressively increase it in later stages. As this was just the first evaluation carried out in the context of the adopted DSR methodology, it was considered more efficient to eliminate first the drawbacks already identified in the current prototype, and concentrate additional resources on the evaluation of more mature prototype versions – where the identification of additional areas for improvements was likely going to be more difficult. Using a larger sample in this cycle would have been a waste of resources, more efficiently used in the following cycles.

A second limitation of the evaluation was the excessively constrained tasks that were proposed to test-users during evaluation. Again, this is common for an early prototype, and the following prototypes could be evaluated with more realistic tasks, strengthening the preliminary evidence collected here.

6.5 Injector design and evaluation: conclusions

In this study a first prototype, Injector, was developed and evaluated in a first of a series of studies following the DSR research paradigm.

Coherently with the findings and requirements arising from the previous study, the task-analysis empirical evaluation, Injector provides educational metadata and expansion functionalities to similar resources, directly within Google SERPs. To this end, it relies upon a novel similarity metric developed in this research, defining the degree of similarity between two resources as the number of shared educational alignments.

The discounted formative heuristic evaluation of the prototype, utilized with a restricted set of search keywords (constrained tasks) by a small number of representative users, was carried out with the

main objective to identify possible design deficiencies and areas for improvements, to be addressed in subsequent studies.

The map in Figure 6.7 provides an overview of the main findings involved in the design and evaluation of Injector.

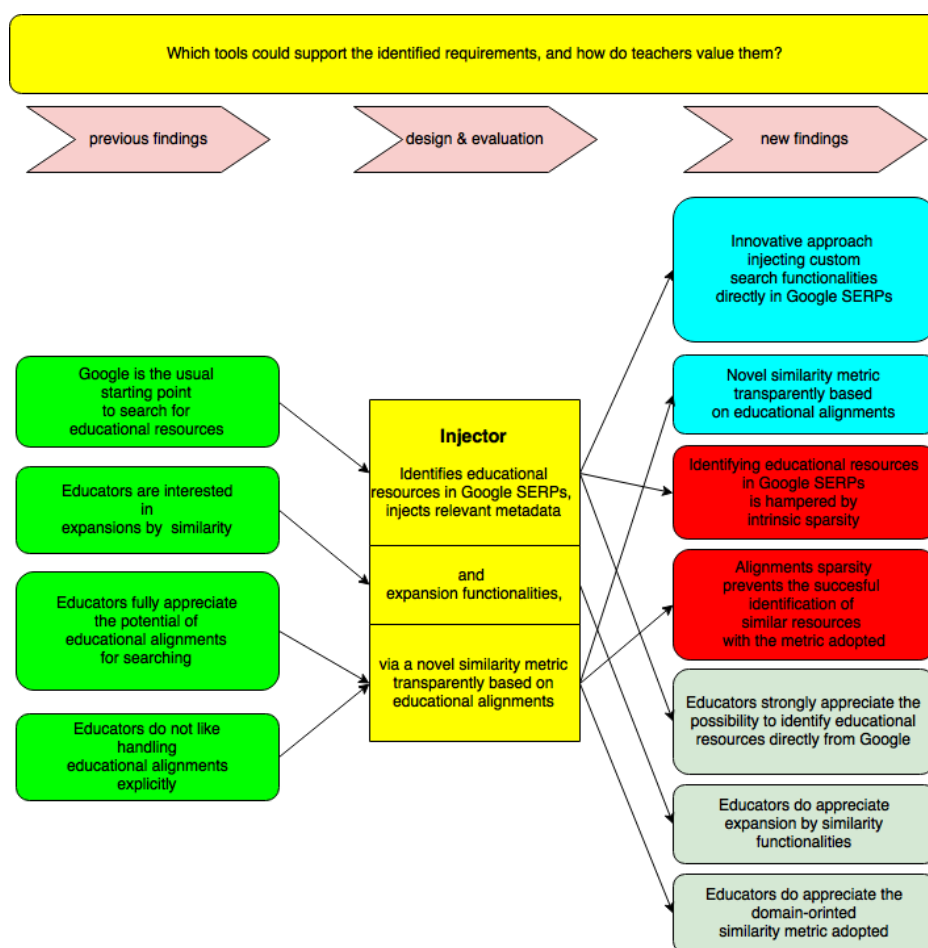


Figure 6.7 – Overview of the findings related to the design and evaluation of Injector.

The top of the figure highlights the main research question guiding the activities in this first DSR iteration. The figure maps the relevant findings (knowledge) available from the previous study, indicated in the first column, to specific prototype features, reported in the central column; these features are then mapped to the findings derived in the current iteration, summarized on the right. These resulting findings are colour coded to indicate their type: blue for design knowledge, red for constraint knowledge, while general claims just supported by preliminary evidence so far, are indicated in grey.

As was found in the formative evaluation, test-users strongly appreciated the availability of specialized discovery functionalities “directly” in Google results pages, largely their most popular search context, avoiding the time-consuming need to move between specialized portals. They also considered very positively the possibility of discovering “similar” (or “more like this”) resources, according to the domain-oriented metric defined in the study (grey boxes on the right). This contributed directly to RQ3 and RQ4, in particular RQ4.1 and RQ4.2. However, test-users

consistently identified the problems of intrinsic and alignments “sparsity” (red boxes on the right), that make the experimented design unsuitable for practical use, and needed to be addressed in an enhanced prototype in the following study.

7. RepExp: Replicator-Expander prototype

In this second DSR cycle, a new enhanced prototype was designed and developed to address the challenges identified in the previous cycle: intrinsic sparsity and alignments sparsity. The prototype was then evaluated to identify possible additional shortcomings and new specific research questions to be addressed in the following cycle.

7.1 RepExp goal and research questions: mitigate intrinsic and alignments sparsity

The main specific research question for this DSR iteration, which arose from the previous cycle, and expands on the original research question RQ4, can be explicitly expressed as follows:

RQ4.4: How can the intrinsic and alignments sparsity of web resources be addressed by an OER discovery tool?

This activity also contributed further input to help answering other more general questions:

RQ4.1: How can an OER discovery tool be embedded into web search?

RQ4.2: How can Query By Example be added to an OER discovery tool?

RQ4.3: How can similarity based on educational alignments be added to an OER discovery tool?

7.2 RepExp Design

7.2.1 Sparsity: problems and possible solutions

The term sparsity indicates, in general, a low density of relevant data. It is a typical challenge of recommendation systems adopting user-based collaborative filtering, where there is generally a limited amount of recommendations compared to the number of users and items. In such cases the problem can be mitigated, for example, by the use of item-item algorithms: these first identify relationships between different items, which are then used to compute recommendations (Sarwar et al., 2001).

Injector too can be considered a recommendation system under user control (Chi, 2015), which makes use of an item-item metric to derive recommendations. Yet, what has been called “intrinsic sparsity” here, is a challenge of a different origin. It is mainly due to the inevitable limited number of actual educational resources in Google SERPs, which usually contain resources of heterogeneous types, such as commercial companies, products, or news, in addition to learning resources. The problem is aggravated by other factors, such as the lack of a unique global identifier capable of identifying multiple instances of the same resource, which makes it difficult to properly identify a resource, and collate metadata referring to the same resource. The use of the URL as a resource identifier is not an ideal solution, because the same resource can be hosted in different repositories,

hence can be designated by different URLs in the SERPs. A number of solutions to this problem were attempted, such as the use of entity reconciliation techniques, or sameAs relationships to state the equivalence of different instances (Bienkowski, 2012), but so far there are few data available (Guha et al., 2016). The very limited data about repurposing history further exacerbates the problem, making it difficult to collate efficiently data referring to repurposed resources. Finally, even if the Learning Registry aimed to be a hub centralizing as many metadata and paradata about learning resources as possible, the reality is that Google SERPs could contain educational resources for which there were no metadata available in the Learning Registry.

Intrinsic sparsity: intercepting and replicating elsewhere a Google keywords search

A possible solution to avoid intrinsic sparsity, corresponding to the suggestions of test-users B and C in the evaluation of Injector, would be to filter out irrelevant hits from Google SERPs, requesting multiple pages, parsing and filtering them until the required number of educational resources are identified and can be displayed to the user. For this, Google could be queried programmatically via its XML API Google Custom Search interface, but this would legally require the payment of a fee or the display of the SERPs obtained from the non-XML interface version without modifications to the users. None of these options was considered appropriate: the first one for economic reasons, the second one because it would not permit to retain full control on the information presented to the users.

Another possible solution to mitigate intrinsic sparsity, could be to start the search by similarity directly from more relevant SERPs, that is search results pages containing exclusively, or at least predominantly, educational resources. These more relevant SERPs could be obtained by simply inviting users to leave Google and repeat their initial queries on a specialized portal. This option, however, was not considered appropriate, because educators in the evaluation of Injector definitely appreciated the availability of discovery functionalities directly from their familiar Google SERPs.

Therefore, it was decided to *automatically* intercept the initial user Google search and *transparently* replicate it in a specialized educational resources metadata repository (Learning Registry), to display resulting hits with customized rich snippets in a separate “Google-like” window. This way, users would not need to care – and most of them would likely not even realize – that such data did *not* come from their habitual Google search engine.

Alignments sparsity: restricting hits to resources having alignments metadata

Alignments sparsity is caused by the limited alignments metadata available for educational resources. When this study was designed, there were about 60,000 educational resources aligned to educational frameworks, within a total of little more than 600,000 resources, available through the Learning Registry Search API used in this prototype. Under these conditions there would still be considerable alignments sparsity, as just an average of 10% resources could be expanded for similarity.

Once again, a possible solution to the problem, sufficient at least for a realistic evaluation of the prototype, was to get more relevant SERPs by dynamically restricting the resulting hits to those

resources actually having alignments metadata. Because of this, the large majority of the resources identified by the prototype would have similar resources, according to the metric used. This would enable their expansion, hence eliminating the problem of alignments sparsity. Whenever alignments metadata are added to the existing resources, or new resources with alignment metadata are added in the repository, these resources would not be filtered out any longer and would therefore be automatically considered by the prototype.

7.2.2 Separating Replicator and Expander: RepExp

One of the Injector test-users (User3), remarked that it may be better to expand a resource without the need to go back to the SERP page. Indeed, it is reasonable to imagine that users are more likely willing to explore a given resource, before deciding to get similar resources by expanding it. This observation suggested an additional enhancement to the prototype, namely to allow the expansion to similar resources starting directly from the resource being explored (the “current” URL), eliminating the need to navigate back to the SERP context where the previous prototype, Injector, used to inject expansion functionalities immediately following the first query. This solution has the additional advantage to enable the use of the expansion functionality even starting from resources identified via any other unplanned search portal, not just Google or this prototype. It is indeed a solution that fully supports QBE: additional similar resources are just identified starting from any sample resource. Another advantage is that this solution is computationally more efficient, as it computes potential expansions exclusively for resources the user is interested in. A minor drawback of this solution, though, is that it cannot display in advance volume indicators (Sacco, 2000; Hearst, 2009) with the total number of possible similar resources, so that it provides less orienteering information, or, in terms of Information Foraging Theory, less scent for users to help them decide whether it is worthwhile moving to a different patch. However, weighting this minor inconvenience with the previously described advantages, it was decided to separate in this second prototype the keywords search replication functionalities from the expansion functionalities, with two physically independent components called, respectively, Replicator and Expander. This accounts for the origin of the name RepExp for the prototype.

7.2.3 RepExp sample session

This section illustrates a search/discovery session actually carried out with the prototype, with a commented sequence of screenshots.

Figure 7.1 reports the SERP page returned by Google following a traditional search with the keyword “biology”, as indicated by the arrow on the left.

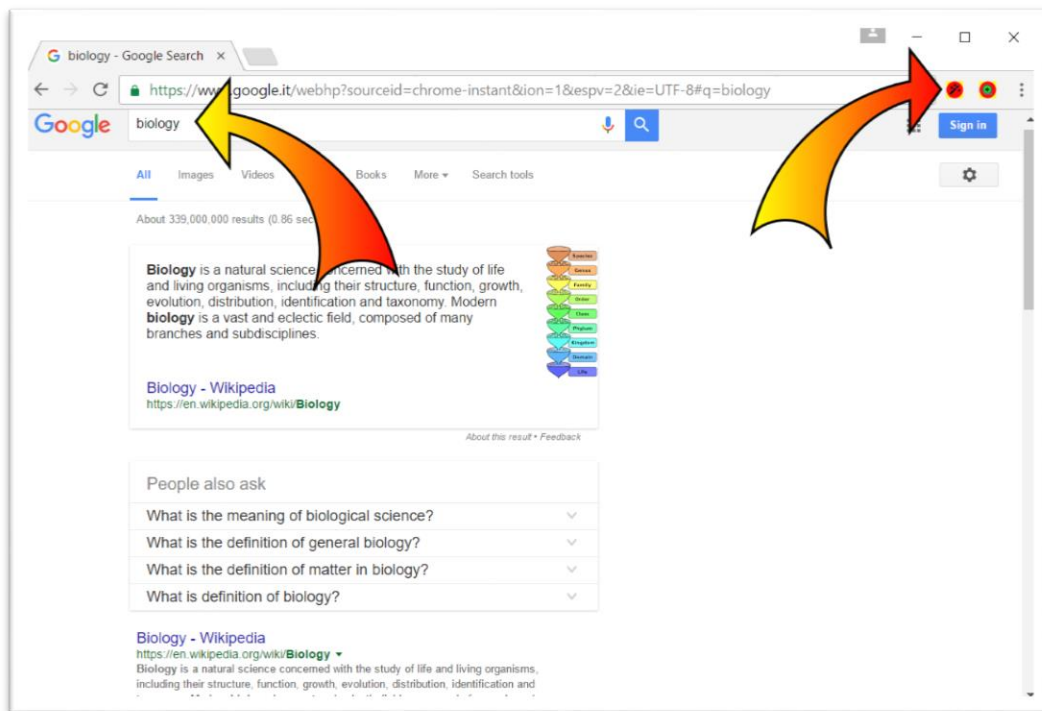


Figure 7.1 – Google SERP resulting from search with keyword "biology".

The right arrow points to the extension icon corresponding to Replicator, which is active because the current page is automatically recognized by RepExp as a Google SERP. Following a mouse click on this icon, Replicator automatically intercepts the keyword “biology”, replicates the search in the Learning Registry, and formats and displays the results in its own SERP window, reported in Figure 7.2. This window signals that 399 resources have been identified by the keyword “biology”, and displays the available educational metadata for the first few most relevant resources.

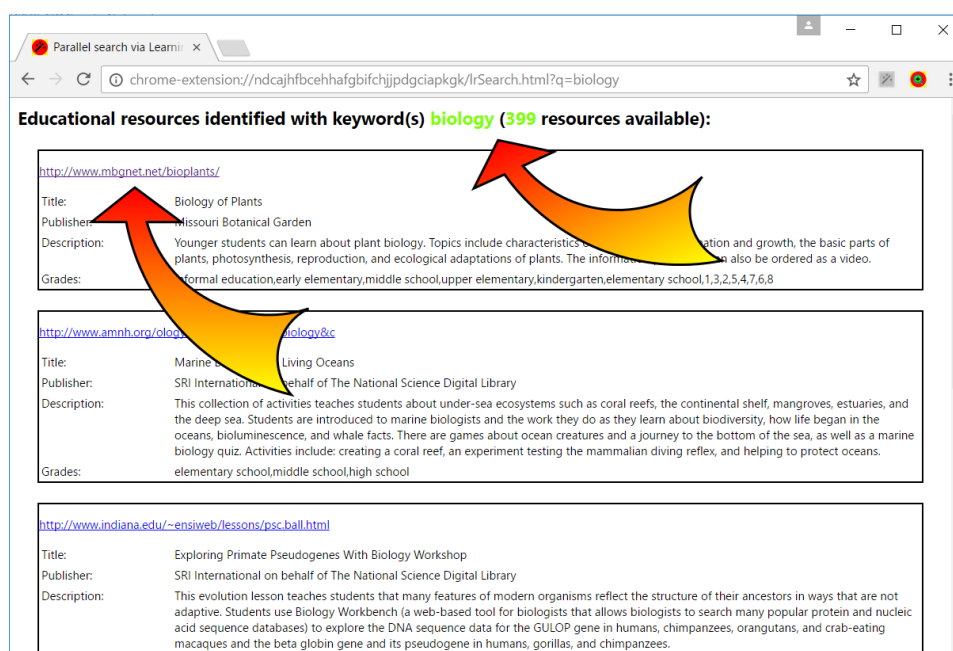


Figure 7.2 – The new SERP automatically obtained by Replicator.

The user is now free to browse and navigate to any resource displayed, and selects the first one, obtaining the window in Figure 7.3.

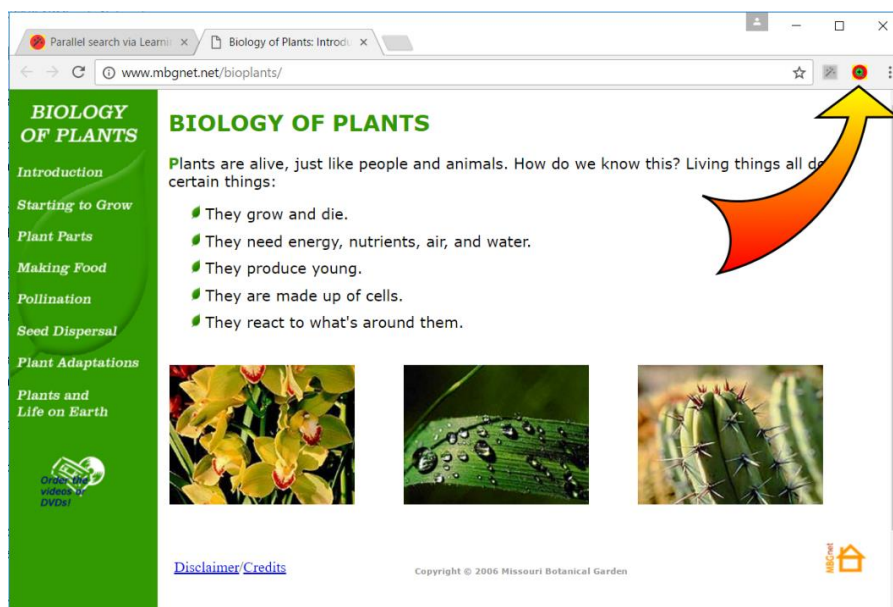


Figure 7.3 – Expanding the current educational resource.

In this new frame, the icon corresponding to Replicator has been automatically deactivated, because the application has detected that this page is not a Google SERP, and it would make no sense to use Replicator. Here the user, in case he/she is interested in obtaining other resources similar to this one, similar not just because they share the same keywords, but much more precisely because they share the maximum number of learning objectives, can select the extension icon corresponding to Expander, indicated by the arrow, so that Expander replaces the current page with the one available in Figure 7.4.

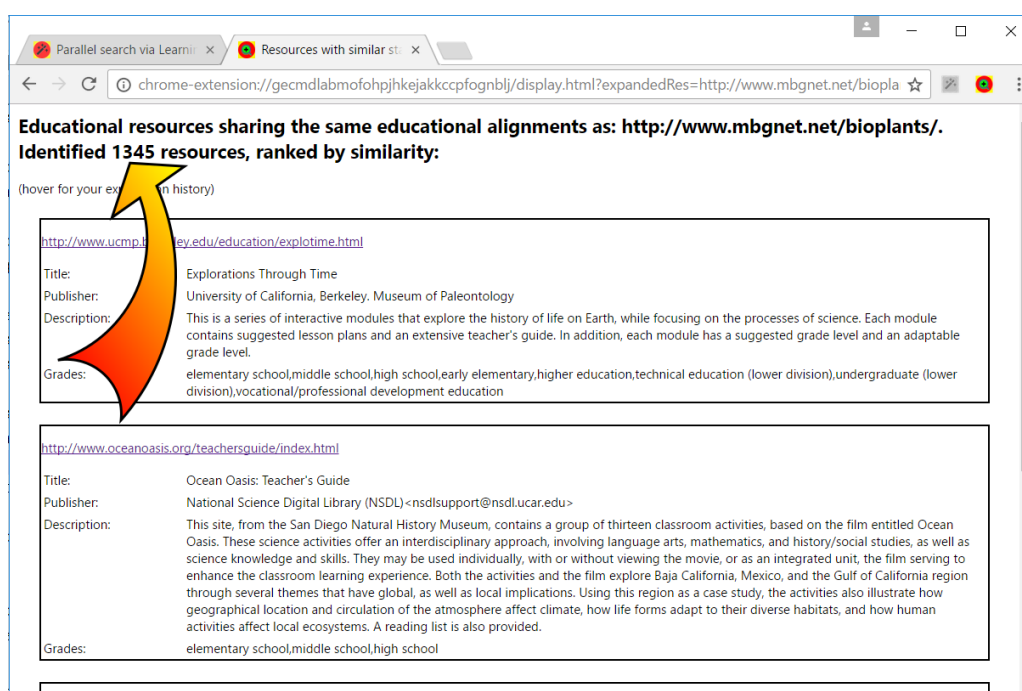


Figure 7.4 – Expander SERP page with the similar resources identified.

In this case the frame reports that 1345 similar resources have been identified, and displays the snippets corresponding to the first few resources with higher similarity, containing the available educational metadata. Again, from here, the user can navigate to the resources he/she is interested in, and at any time he/she can iteratively request further resources similar to the resource currently explored.

It is important to note that Expander can be used starting from any resource being explored, not necessarily a resource identified via Replicator. This makes it possible to expand an educational resource identified via any other search engine or portal. Or, for example, an educator interested in finding resources for remediation activities, could also navigate directly to a resource he/she is already using, and use Expander to search quickly for additional resources sharing the maximum number of learning objectives. Obviously, this is only possible if metadata for the resource are available in the Learning Registry.

Finally, users can continue iteratively exploring and expanding resources, and can consult at any time their exploration history, including information about visited and/or expanded resources, as shown in Figure 7.5, following the further exploration (visit) of two additional resources and the expansion of the last explored one.

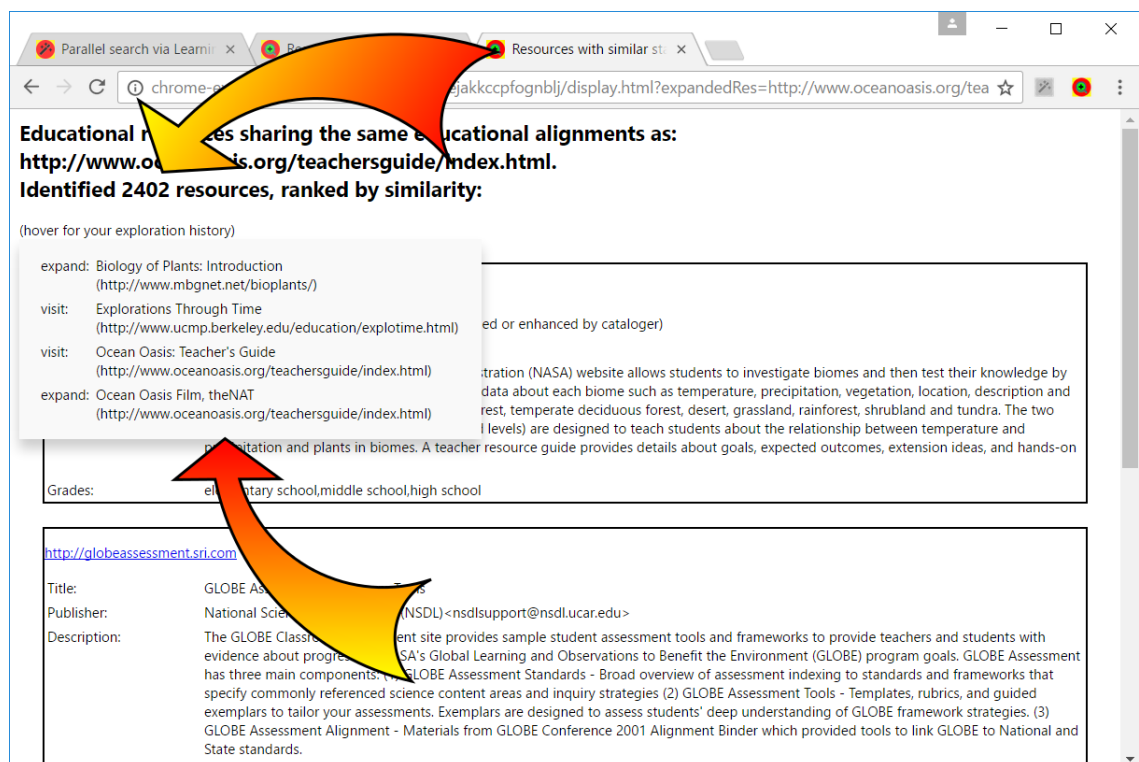


Figure 7.5 – Exploration history.

7.2.4 RepExp architecture

The prototype, as previously discussed, is composed of two physically independent but logically related components, Replicator and Expander, which can be used together or independently, for maximum flexibility. As for the previous prototype, they were built as Chrome Extensions, and make

use of the Registry Extract Data Services; in this case, however, they make use of the Learning Registry Search API (Learning Registry, 2016b) too.

Replicator architecture

The Replicator architecture is illustrated in Figure 7.6.

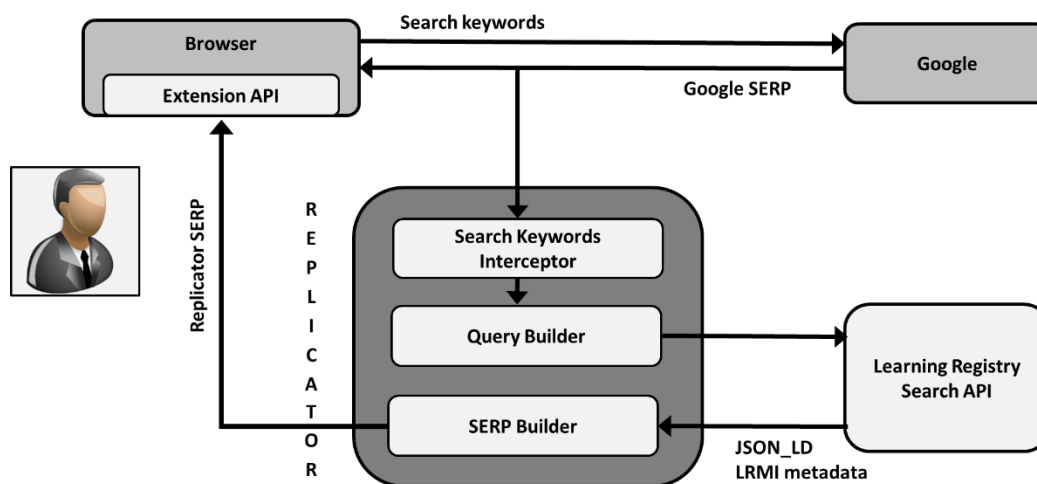


Figure 7.6 – Replicator architecture.

Replicator is normally inactive, but it becomes automatically active whenever a Google SERP page is visualized in the browser. This way, it signals to the user the possibility of getting additional more targeted educational resources. This context-sensitive clue is consistent, as usual, with the suggestion from the Information Foraging Theory to provide users with as much scent as possible.

Search Keyword Interceptor captures the search keywords used in Google by the user, and Query Builder creates and sends a query to the Learning Registry Search API requesting resource metadata associated with the intercepted keywords, satisfying the additional condition that they have some alignments metadata specified. The way to specify this last condition was not documented, nor the Learning Registry developers could provide any help, but could be identified by tracking the network calls of their tools. This restriction is necessary to make sure that most resources identified can be later expanded by Replicator. The volume of more than 60,000 educational resources with alignments metadata in the Learning Registry dataset, was sufficient to test the prototype in realistic conditions.

Finally, SERP Builder receives the answer from the Learning Registry containing the LRMI metadata for the resources identified in JSON_LD format, parses them and creates the customized SERP HTML page with the resource snippets to be displayed in the browser.

Expander architecture

The Expander architecture is reported in Figure 7.7. When Expander is called, by clicking on its extension icon, it captures the URL of the page being visited, which is used as the identifier of a possible educational resource. This is the “example” resource in a QBE: it might be a resource previously identified by Replicator, by any other search engine, or just known to the user in any other way. The Alignment Standards Retriever attempts to get all the standards that particular educational

resource is aligned to, by interacting with the Learning Registry Data Services. Once the standards have been obtained, the Parallel Aligned Resources Retriever attempts to obtain from the Learning Registry Data Services the metadata for the resources aligned to each standard previously identified. Given the considerable volume of data involved, these requests are performed asynchronously, in order to maximize performances. Once all the requests have been completed, the Metric Calculator counts the number of occurrences of each resource in the responses received, and sorts the resulting similarity vector. Finally, the SERP Builder creates the HTML Expander SERP page with the rich snippets containing the available metadata for the highest ranking (most similar) educational resources, which is displayed to the user in a separate browser window.

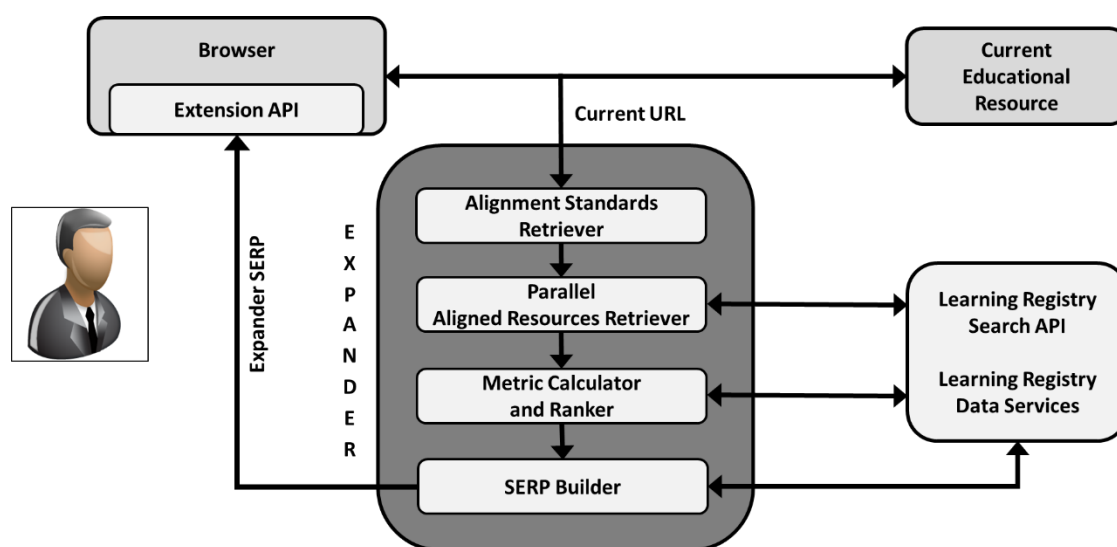


Figure 7.7 – Expander architecture.

7.3 RepExp Heuristic evaluation

7.3.1 Goals: identify RepExp further drawbacks and possible areas for improvements

According to the adopted DSR methodology, the main objective of this evaluation was to understand whether the previously identified challenge of sparsity was adequately addressed, and to identify possible additional shortcomings and areas for improvements that could not be identified before, because likely masked by the larger problem of sparsity. The evaluation also aimed to collect additional incremental feedback concerning the usefulness and relevance of the proposed solutions.

7.3.2 Methods: a second formative heuristic evaluation with content analysis

A second formative heuristic evaluation was carried out in this new DSR cycle. Because of the reasons explained in the previous chapter concerning Injector, a discounted evaluation was considered suitable for the scope of this formative evaluation of a prototype, which was carried out, however, with a slightly larger sample of experienced educators from Italy, UK, and Brazil.

A new evaluation questionnaire was designed (Appendix E), by improving and customizing the one previously used in the evaluation of Injector, which also presented the opportunity to pilot the initial

evaluation form. With the intention to foster the elicitation of some more in-depth qualitative feedback, a first modification consisted in the introduction of additional open questions, soliciting test-users to provide reasons justifying their answers to existing closed questions, as well as any additional comments. A second modification concerned questions in the “functionalities” section, which proved to require further explanations and clarifications; these questions were reformulated to focus more precisely on some of the objectives of the evaluation: getting user feedback on the strategy of offering educational resources discovery functionalities directly from Google pages, on the relevance of the generic task of expansion by similarity, and on the appropriateness of the specific similarity metric proposed. Finally, instructions and task proposed to test-users were customized for the new prototype. In particular, in the case of Injector it was necessary to propose a discovery task starting from very constrained keywords to make sure that users could get some results. In this case, users could use any keywords of their own choice, reinforcing the ecological validity (generalizability to real-life situations) of the evaluation.

The questionnaire was written in English as in the previous case, but participants were invited to answer in English, Italian, or Spanish, to improve the likelihood of contributing qualitative data.

As in the previous case, the quantitative data collected were planned to be analysed with simple descriptive statistics in order to get a first appreciation of the position of sampled educators. In particular, the mean was used as a measure of central tendency, in case there were no outliers among the data collected (outliers could negatively influence the validity of the mean as indicator of central tendency). Preliminary findings obtained from these data, however, would have to be investigated further in the summative evaluation in the last DSR cycle, to obtain more statistically meaningful results.

This second evaluation planned to collect a larger volume of qualitative data, compared to the previous evaluation. Hence, a more formal approach was required for its analysis, following the method of qualitative content analysis (Cho and Lee, 2014):

- Subdivision of the text in chunks expressing single concepts;
- Iterative extraction of a set of categories or themes from the key concepts recurring in the textual data themselves, organized hierarchically as convenient. The objective was to identify representative categories, mutually exclusive and covering all the chunks, hence dividing the concept space of interest in partitions;
- Coding of the original chunks with the categories extracted.

The analysis of the qualitative data was expected to produce the main outcomes demanded from this activity, that is the identification of possible shortcomings and areas for improvements in the current RepExp prototype.

7.3.3 RepExp evaluation results and discussion

Six educators participated as test-users in this second heuristic discounted evaluation (Table 7.1): four males and two females, five teaching at level of Secondary Education, one at University level, four from Italy, one from the UK, and one from Brazil, all involved in teaching technical or scientific subjects, with 6 to more than 20 years of experience. Four of them were selected opportunistically among educators personally known. Two of them were identified with a snowball strategy, that is asking to the initially identified educators to encourage other educators to participate. One of the test-user (User1) overlapped with a test-user (User1) from the previous study.

User ID	Gender	Working Country	Teaching Experience [years]	Teaching Educational Level	Teaching Subject
User1	M	UK	6-10	Sec. Educ.	Technology
User2	M	Italy	>21	Sec. Educ.	Math
User3	M	Italy	11-20	Sec. Educ.	Technology
User4	M	Italy	>21	Sec. Educ.	Technology
User5	F	Italy	11-20	Sec. Educ.	Math
User6	F	Brazil	11-20	University	Technology

Table 7.1 – RepExp evaluation Test-users profile.

Six participants were sufficient to reach the goal of this second formative evaluation, which could indeed identify new challenges to be addressed in the following DSR cycle.

The evaluation questionnaire (Appendix E) 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.

As in the previous case, no sensitive data were collected from test-users, and the basic demographic data were optional. However, data were anonymized after collection, and results only reported anonymously or in aggregated form.

The qualitative data were analysed with a traditional content analysis approach, as indicated in the methods section. Following the sub-division of the qualitative data in chunks expressing a single concept, and the iterative extraction of the key concepts, the following schema was obtained:

- Generic positive comment
 - Specific positive comment on domain orientation
 - Specific positive comment on speed, simplification, automation
- Comment about the interface
- Suggestion for additional uses of the tool
- Generic suggestion for additional functionalities
 - Specific suggestion for additional functionalities

- Comment about large number of results
- Comment about sense making
- Comment about degree of similarity

Finally, qualitative data chunks were colour coded according to the previous schema, to provide an immediate overall view of the concepts expressed. Major findings are discussed in the following sections.

General feedback on usefulness and relevance of proposed functionalities

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). Justifications provided were sometimes generic, such as:

“I found fantastic the option” [User6],

or more specific:

“I always use Google, I ignore other portals” [User2],

“Very convenient for educators” [User1],

“Most of us educators start, and stop, in Google” [User3].

Test-users also definitely 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). Reasons abducted explicitly mentioned, in five over six cases, its strongly domain oriented characteristics, for example:

“Precisely focused on the educational domain” [User2],

“Very appropriate in education” [User3].

An educator remarked that such a metric could be useful also to other types of users, in addition to targeted educators:

“It would be useful for students too” [User2].

Yet, some test-users expressed some concerns too, related to the appropriateness of ranking by degree of similarity, and to the large number of results: these aspects deserve special attention and will be addressed in detail later in this section.

The usefulness of expansion functionalities in general, that is including the use of other similarity metrics, was considered pretty high: mean 1.8 on a scale between 1 (very useful) and 7 (totally useless). This confirmed the results of the task-taxonomy empirical evaluation, and was reinforced by explicit comments such as:

“Very useful” [User1],

“I liked that idea too” [User6].

Three out of six educators noted, however, that the functionalities supported by the prototype should be integrated with additional ones, again in line with the task-taxonomy empirical evaluation:

“Needs to be complemented by other functionalities” [User1],

“Not the only task that needs to be supported” [User4].

Finding with alternative techniques the same type of similar resources that could be found with the prototype, was considered – as in the previous evaluation – quite difficult: mean 6.2 on a scale from 1 (very easy) to 7 (very difficult).

As observed in the previous study, educators consented that the prototype used familiar user oriented rather than system oriented terms: mean 6.7 on a scale from 1 (fully disagree) to 7 (fully agree). This was reinforced by the following comment:

“The user interface is delightfully simple” [User1].

Again, in this evaluation too, the transparency of the tool using educational alignments with no need for users to explicitly handle them, was considered very positively: average 6.3 on a scale from 1 (totally useless) to 7 (very useful). One educator, however, suggested the possibility of making explicit use of educational alignments in an advanced modality:

“It may be useful, though, to have also a sort of more advanced user mode, offering the possibility of examining and manually selecting the learning objectives.” [User4].

Concerning User Experience, test-users 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” [User4].

Test-users also agreed that the prototype automates unwanted workload: mean 6.3 on a scale from 1 (fully disagree) to 7 (fully agree). This was supported by comments such as:

“It saves a lot of work, automatizes a complex task” [User4],

“Speeds-up searches” [User5].

The reference task proposed in the evaluation was considered relevant, with a mean of 6 on a scale from 1 (totally irrelevant) to 7 (very relevant). Yet, while this score indicates quite a high relevance, it was the lower score obtained among all the questions. The likely reason was that test-users did consider the task to be very relevant, but they deemed that it covered only partially their overall needs, as suggested by the following comment:

“It is very relevant, but of course it does not cover all needs” [User4].

Test-users suggested a number of functionalities to be added to the current prototype, that are certainly important but not in the focus of this research. For example, the following remark,

“The possibility of selecting the type of metric” [User4],

is definitely in line with the strategy of this research, which however focuses on a specific metric;

“More functionality in terms of actually making use of the information that has been found” [User1]

is another excellent suggestion, showing the way for a very relevant domain-oriented further step, which however is considered outside the current scope of the research. Another useful suggestion previously mentioned, concerned the possibility of offering an advanced modality where users could indicate explicitly the educational alignments of interest.

Shortcomings and areas for improvements

Fundamental to the objectives of this evaluation, test-users produced some critical observations about core features of the prototype, which pinpointed areas of concerns and potential improvements, to be addressed in the next prototype version.

Degree of similarity and diversity

The first important concern was related to the degree of similarity of the presented resources. The prototype ranks and presents the identified resources in order of similarity, starting with the most similar. Consequently, when there are many similar resources available, the first few resources presented are characterized by the highest degree of similarity. These first few resources are usually the only ones examined by users. The resulting *unintentional* effect, is that the prototype frequently ends-up by presenting to users *exclusively* the most similar resources. Therefore, while all test-users,

as previously noted, did appreciate the strong domain orientation of the adopted similarity metric, some of them remarked that presenting exclusively resources with the highest similarity is not the best option:

“Resources should not be too similar; that is, they should be somewhat similar, but not equals” [User2].

Indeed, Smyth and McClave (2001, p. 348) claim that the “standard pure similarity-based retrieval strategy is flawed in some application domains”. They also point out more precisely that “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 needs to take into account diversity too: as argued by Bradley and Smyth (2001, p. 90), “in many recommender system application scenarios, similarity and diversity both have roles to play in the recommendation process”.

Yet, the problem of excessively similar resources is not the key challenge in this particular case. Indeed, while in general there is a risk of offering excessively similar resources, because they might be so similar to the original resource to be useless, this is not necessarily the case in this context – because even when two resources have exactly the same learning objectives, that is they are “equivalent” according to the similarity metric adopted, they can still be vastly different in their contents. For example, they might use different metaphors or different educational strategies to attain the same learning objectives.

However, as pointed out by another test-user,

“Maximum similarity is not necessarily what one looks for in every opportunity” [User4].

Indeed, maximum similarity is not necessarily the best solution to maximize utility. A resource with the maximum possible similarity would be an identical resource, which would be obviously inappropriate. 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. Indeed, the goal in this case is to provide students with the opportunity to face new challenges and achieve additional learning objectives.

This opportunity to support educators in their WC level tasks, by offering them some control on the degree of similarity, was experimented in a new prototype in the next DSR cycle.

Sense-making and information overload

A second 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 test-users expressed concerns for the opposite reason, that is for the *large* number of resources identified:

“Sometimes there are too many [resources]” [User5],

“The large volume of hits might be a problem” [User2],

“How can I handle all those results?” [User2].

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

*“It would be useful to get a quicker global picture of the available similar resources”
[User5].*

Lock-in / inability to re-patch

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” [User2].

This effect was noticed by one participant because, in attempting to make sense of the whole set of similar resources available, he insisted in iteratively expanding resources from each resulting set. Indeed, this inconvenience becomes rather more evident when there is a limited set of similar resources with the highest degree of similarity, which are repeatedly expanded. In this case, the few most similar resources get prioritized by the prototype and consistently end up (repeated) in the first resources displayed to users. This problem went unnoticed most of the time, because when there are many resources with the same high degree of similarity in the result-set, the probability that one of these resources gets repeated in the first few resources displayed to the user in a sequence of expansions, is sufficiently low.

Of course, repeated resources could be easily eliminated by keeping track of the resources already shown to the user, and eliminating them from subsequent expansion results. Yet, this solution would not really solve the problem, because there might be many highly similar resources, even if not identical.

The repetition of resources following repeated expansions, is just the visible symptom of a larger problem, hereafter 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.

The problem of lock-in can be explained in terms of the patch model in the Information Foraging Theory (Pirolli and Card, 1999). In this model, users (informavores) carry out their discovery activities by moving within a patch, and by moving between patches (re-patching). Moves within a patch as well as re-patching have an associated operational and cognitive cost. The Information Foraging Theory postulates that user's decisions are driven by their attempt to maximize their results while minimizing the overall cost of the activities. In the case of RepExp, a group of very similar resources can be considered a "patch". Repeated expansions do not allow users to re-patch. Eliminating previously seen resources from newly expanded results, would allow re-patching, but at a very high cost (that is, by sequentially expanding all the resources). Another problem with this approach is that resources that have already been seen would disappear and could not be re-found. The problem of "lock-in" can therefore be considered, in term of the Information Foraging Theory, as the inability or high cost of re-patching.

The problem of "lock-in" can also be explained in terms of the characteristics of the similarity relationship adopted, coupled with the strategy of ranking resources by similarity. This relation of *highest similarity* is an approximate equivalence relation, hence it partitions the set of resources in approximate equivalence classes. A relation is an equivalence relation if and only if it is reflexive, symmetrical, and transitive. Indeed, the relationship is obviously reflexive: a resource A is always very similar to itself, because it shares all its alignments with itself. Additionally, the relationship is usually symmetrical: if resource A is highly similar to resource B, because they share many educational alignments, resource B is usually highly similar to resource A. Finally, the relationship is also approximately transitive, at least for a very limited number of steps: if resource A is highly similar to B, and resource B is highly similar to C because they all share many educational alignments, resource A is usually highly similar to resource C. Therefore, the most similar resources tend to be organized in "approximate" equivalence classes, whose elements are the only ones that keep showing up in repeated expansions.

It is interesting to note, however, that there can be resources for which the symmetrical and transitive properties do not hold. For example, in Figure 7.8, resource A shares as many educational alignments with resource B as with its other most similar resources (resource AA in the example), but resource B shares a *higher* number of *different* educational alignments with another resource C.

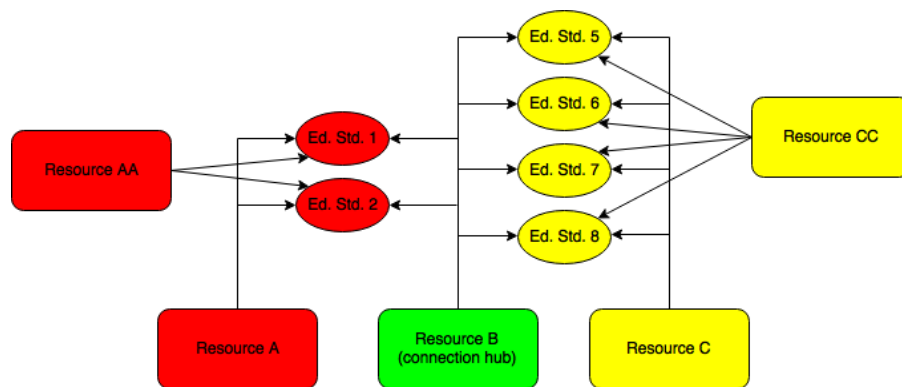


Figure 7.8 – Symmetrical and transitive properties do not always hold.

In this case, A would be considered highly similar to B, but B would not be considered highly similar to A, because it would be more similar to C, hence the symmetrical property would not be satisfied. Moreover, A would be highly similar to B, and B would be highly similar to C, yet A would not be similar to C at all: hence the transitive property would not be satisfied. As a consequence, while proper equivalence classes are either equal or disjoint, in this case, where the symmetrical and transitive properties are not always satisfied, there are resources (resource B in the example) belonging to more than one class (approximate equivalence class), which are classified, in network theory, as connector hubs.

In the educational context, a connector hub resource would play an interesting pivotal or bridging role among different sets of educational alignments, therefore representing a likely interdisciplinary educational resource.

It can be observed, finally, that the likelihood of having resources bridging from one patch to another increases, when we consider also resources with a lower degree of similarity, rather than just those with higher similarity. Indeed, in this case, the resources sharing few of their alignments with the expanded resource, therefore having low similarity, can more likely share more alignments with other resources, having higher similarity. As it will be discussed in the design of the next prototype, this is the key to solve the lock-in problem.

Interface

Another class of remarks concerned the visual presentation of the information, which was the object of a consistent number of negative generic comments, such as:

“Improve the markings/colors in the preview” [User6],

“The results page is visually pretty poor” [User5],

“The appearance is definitely not appealing” [User4].

There were also a few more specific comments. In particular the use of two extension buttons was considered confusing by two educators, which likely saw them as having the same basic purpose –

just “find other resources” – whether exploiting keywords as in Replicator, or educational alignments as in Expander:

“The use of two different extensions for such related functionalities is not intuitive”
[User5],

“The two different buttons may create some confusion” [User4].

Another specific indication, even if it may possibly sound too drastic,

“Just show Open ER ...” [User6],

suggests that it may be worthwhile, at least, to visually discriminate between open and non-open educational resources.

7.3.4 RepExp evaluation conclusions: degree of similarity, sense-making, lock-in, interface

The main goal of this second formative evaluation was to identify shortcomings and areas for improvements to be addressed in a new prototype in the subsequent DSR cycle. Some more feedback on how educators considered the strategies adopted in the prototype was collected too.

The questionnaire available in Appendix E was administered as survey and structured interview to six educators from Italy, UK, and Brazil. The quantitative data collected were analysed with basic descriptive statistics, and triangulated with results arising from the content analysis of the qualitative data collected.

The evaluation indicated, coherently with the previous evaluation, that the functionalities supported by the prototype are very much appreciated by educators, and considered very relevant to their specific domain. Despite the limited number of test-users, the evaluation could efficiently identify a number of challenges to be addressed in the next prototype version:

- degree of similarity: an unintentional effect of the prototype is to present to users exclusively the resources with the highest similarity. There seems to be an opportunity to support users in their WC-level tasks, by offering them some control on the degree of similarity of the resources identified, according to their specific educational goals;
- sense-making / information overloading: the large volume of hits made it difficult for participants to get an overall picture of the available resources;
- lock-in: because of the way the similarity metric is used in the prototype, resources tend to be partitioned in approximate equivalence classes, and the repeated expansion of educational resources tend to produce the same group of resources over and over again, so that users cannot navigate out from that group;
- interface: the visual presentation of the information was considered pretty poor; the use of two expansion buttons was considered confusing, and it was suggested to differentiate between open and non-open resources.

Test-users also suggested interesting additional functionalities, which while not within the current scope of this research, will be taken into account for future activities.

Limitations

The main limitation of this evaluation was, as in the case of Injector, the modest number of educators involved, even if in this case the sample was slightly larger (six users) and included test-users from three different countries. As argued in the previous chapter, this is commonly accepted for a formative evaluation in a context in which a series of prototypes are developed and repeatedly evaluated, especially when established heuristics are used (Nielsen, 2013). 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. Indeed, the evaluation reached its main objective to identify further challenges to be addressed in the following prototype, and once again it made it possible to collect additional incremental evidence on the positive attitude of users towards the solutions proposed in this research.

7.4 RepExp design and evaluation: reflection and further work

In this second DSR iteration, the prototype RepExp was designed and evaluated, with the main objective to address the challenge of intrinsic and alignments sparsity, and to identify possible further design deficiencies and areas for improvements.

The challenge of sparsity (RQ4.4) was addressed by increasing the relevance of the starting SERP restricting results to educational resources (hence reducing intrinsic sparsity), having educational alignments metadata available (hence reducing alignments sparsity). For maximum flexibility, similar educational resources can be identified starting from any resource being explored, rather than from the custom SERP: this QBE oriented strategy has the advantage to let users expand resources (that is to find similar resources) identified via the custom SERP generated by this prototype, as well as directly by Google or in any other way.

The prototype was evaluated with a discounted heuristic evaluation as the previous prototype, but with a slightly larger sample of educators from three different countries, and with a more realistic (less constrained) task.

Figure 7.9 provides an overview of the findings related to the prototype RepExp. At the top, the main research question highlights the main problem that needs to be solved by the prototype (RQ4.4). The first column reports the most relevant findings resulting from the previous DSR iteration, which are mapped to specific characteristics of RepExp reported in the central column. These characteristics are then mapped to the knowledge (findings) that could be identified with the design and evaluation activities of RepExp. The types of knowledge are colour coded: red for constraint knowledge, blue for design knowledge, and grey for general knowledge that the additional evidence collected in this iteration contributes to support, but for which additional evidence is needed.

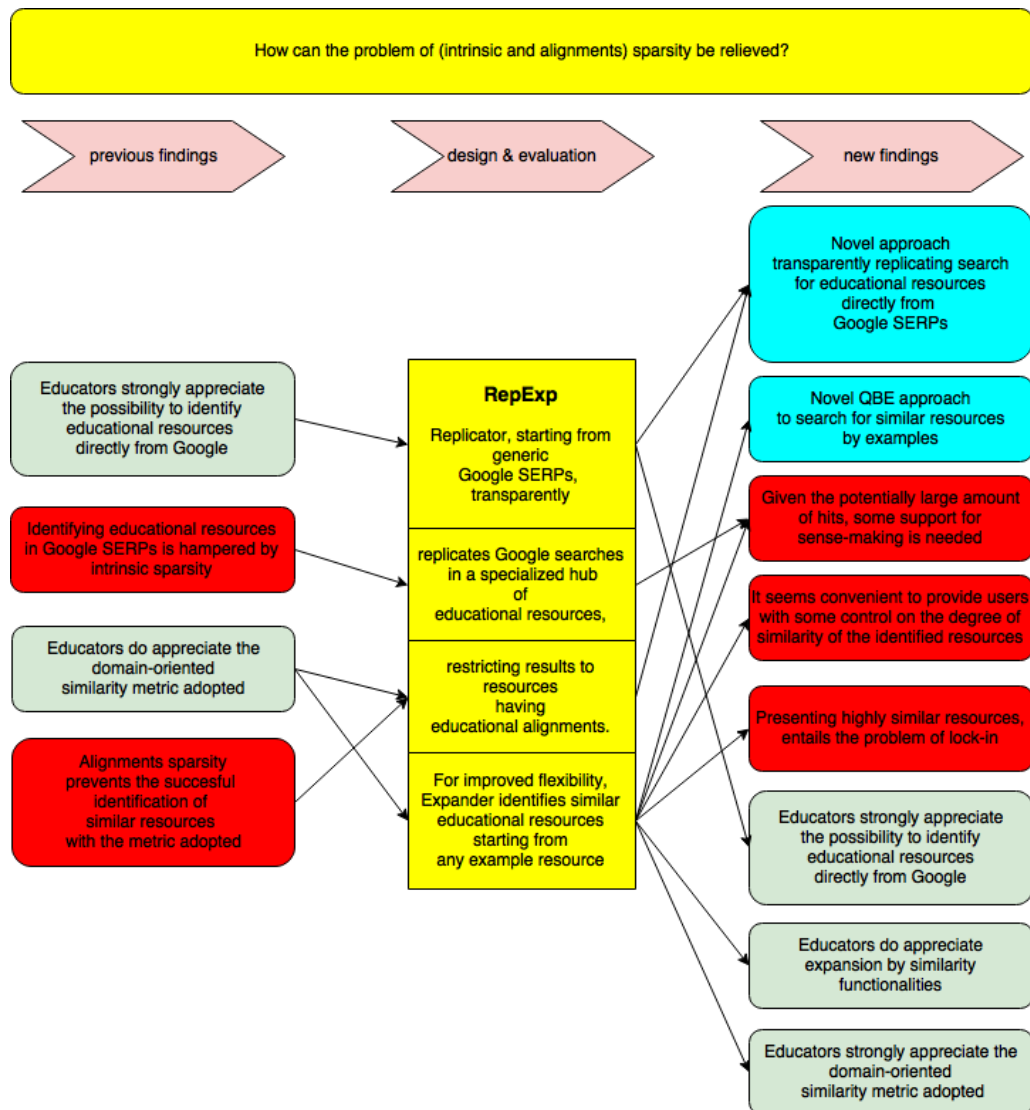


Figure 7.9 – Overview of the findings related to the design and evaluation of RepExp.

Results indicated that the problem of sparsity (red boxes on the left) had been effectively addressed (RQ4.4). Consistently with the evaluation of the previous prototype Injector – test-users expressed a high appreciation for the possibility of using the tool directly from Google pages (RQ4.1), its general approach (RQ4.2), and the specific similarity metric adopted (grey boxes) (RQ4.3). The positive characteristic of the prototype most frequently and consistently indicated, was its relevance to the educational domain.

According to the goal, the evaluation could pinpoint precise sought after challenges and areas for further improvements of the current prototype, as indicated in the red boxes on the right. First, now that the problem of sparsity was solved and plenty of resources could be returned by the prototype, test-users reported difficulties in making sense of sometimes large result sets. Second, test-users wished to be offered resources with different degrees of similarity, not just resources with the highest similarity. Third, the system should avoid to lock users in clusters of resources following repeated expansions. Finally, the need emerged to simplify the interface and improve / enrich the presentation of the data.

8. The prototype Discoverer

In this third DSR cycle, a final prototype was developed to address the challenges identified in the previous cycle. The prototype was evaluated again, this time with a larger sample of twenty-nine educators from three different countries. The goals were: to identify to what extent the new prototype overcomes the challenges previously identified; to spot possible additional shortcomings; and to collect further supporting evidence about findings identified in the previous formative evaluations.

8.1 Discoverer research questions: sense-making, lock-in, domain-orientation

The evaluation of the prototype in the previous DSR cycle, RepExp, provided some additional supporting evidence that test-users had a very positive attitude towards the general strategies adopted: offering discovery functionalities directly from Google pages, expansions by similarity starting from any sample resource, and the specific similarity metric transparently based on educational alignments. However, quite a number of new challenges were identified:

- a) Allow users to select resources with different degrees of similarity, to better support their high-level tasks;
- b) Support users in making sense of the potentially large number of hits, by offering a compressed view of the results;
- c) Avoid the “lock-in” problem, offering the possibility of moving to different patches;
- d) Reduce confusion in the use of the two separate extensions;
- e) Discriminate open from non-open educational resources;
- f) Improve the presentation of the information.

The main research question for this DSR iteration, where a new enhanced prototype was designed and evaluated, mainly concerned the first three challenges. It expanded on the original research question RQ4, and could be expressed as follows:

RQ4.5: How can a large number of similar educational resources be presented by an OER discovery tool to support work context level tasks?

As in the previous DSR cycles, this activity contributed again further evidence, this time from a larger and more representative sample, to answer the more general knowledge oriented questions:

RQ4.1: How can an OER discovery tool be embedded into web search?

RQ4.2: How can Query By Example be added to an OER discovery tool?

RQ4.3: How can similarity based on educational alignments be added to an OER discovery tool?

8.2 Discoverer Design

The first three challenges identified, apparently very distinct, are in reality strongly interrelated: (a) substantially subsumes (b) and (c), as discussed below.

Concerning challenge (a), as discussed in the previous chapter, it seems convenient to offer educators some control on the degree of similarity of the returned resources. The hypothesis is that offering educators the possibility of balancing similarity as well as diversity (Bradley and Smyth, 2001), can support them in their high-level tasks, such as finding resources for remediation versus in-depth activities. In order to make it possible to test this hypothesis, the new prototype, Discoverer, clusters results in groups of different degrees of similarity, and lets educators explore and select resources from each group.

Concerning challenge (b), test-users of the previous prototype reported that it was difficult to make sense of large result-sets, just by looking to a small number of resources having maximum similarity with the expanded one. This is a common problem for search engines, which frequently display a long list of ranked resources, targeting maximum relevance. Many researches attempted to solve this problem, by clustering (Jain, 2010) search engine results, with the objective to summarize them (Zamir and Etzioni, 1999). Alam and Sadaf (2013, p. 153), in particular, argue that “The main use for web search result clustering is [...] to give the user a quick overview of the results”.

Indeed, a possible solution to support users in making better sense of the potentially large number of hits, is to make the few resources presented more representative of the whole result set. This can be achieved by “compressing” them, that is by grouping them in a few clusters of similar degree of similarity, and presenting to users only a few “prototype” resources, representative of each cluster. In this way, it is possible to present a smaller number of results, at the same time more representative of the whole result set: they can be easily examined and can offer users a quicker and representative picture of the whole result-set. Hence, presenting users resources clustered in different levels of similarity, as required by (a), can contribute to solve the challenge (b).

Finally, concerning challenge (c), as discussed in the previous chapter, resources tend to be partitioned into equivalence classes by the relation of maximum similarity, so that users get locked-in a class and cannot get out of it by repeated expansion operations. However, resources with lower similarity are likely to act as connector hub resources, thus can naturally bridge between different classes. This means again that offering users the possibility of selecting resources with lower degree of similarity, as required by (a), is also a strategy providing a solution for the challenge (c).

Lee (2017) discusses an additional reason to avoid offering users just resources with the top-ranked results of a query, that is, exclusively with the highest similarity. His goal was to explore the possibility to support users in their iterative search activities, helping them to iteratively modify previous queries on the basis of the results obtained. He argues that showing exclusively the top-

ranked results, does not offer users any opportunity to create an additional query. On the contrary, providing results with different degrees of similarity offers users more opportunities to issue new modified queries by expanding resources with different degrees of similarity (in an iterative, exploratory oriented sequence of queries).

8.2.1 Clustering results in classes of comparable similarity

The new key feature of this prototype, is therefore to present users with resources grouped in a number of clusters of comparable degree of similarity. It was initially decided to group the resources in a maximum of three different clusters, respectively including resources with high, average and minimum degree of similarity. It was considered that three clusters could be sufficient to provide users with a good degree of control, without over-loading them with a likely ineffective higher degree of control and an unnecessarily complex interface. The truth of this assumption, however, was going to be tested in the evaluation.

A method was needed to assign each resource to one of the three desired clusters, groups, or classes, based on the numerical value of their degree of similarity. This task is commonly referred to as “classification” or “clustering”.

From a quick analysis of the dataset available in the Learning Registry, the degree of similarity among related resources was a number generally comprised between 1 and 10. Hence, a strategy was required to map up to a maximum of about 10 degrees of similarity down to three, grouping the corresponding resources in one of the three targeted clusters. This classification can be seen as the identification of disjoint intervals of numbers, whose endpoints are called “breaks”, which split the range of degrees of similarity data in partitions corresponding to the clusters:

$$\text{Breaks: } X_0, X_1, X_2, X_3 \quad \text{Intervals: } [X_0, X_1], (X_1, X_2], (X_2, X_3] \quad X_i < x < X_{i+1} \leftrightarrow x \text{ in Cluster}_i$$

In this section, a number of possible solutions are discussed, and illustrated with the same reference similarity vector reported in Figure 8.1. This reference vector considers 13 degrees of similarity rather than 10 as previously discussed, to make it easier, with the same vector, to compare all the different solutions; the alternative of using different ad-hoc vectors to separately illustrate each alternative solution would be less compelling. In this figure, it should be possible to identify visually three “natural” clusters, indicated by red circles, which are mainly characterized by the three groups of resources with the highest number of occurrences. The resources with similarity 10, for example, are grouped with resources having similarity 8 rather than with those having similarity 12, because while they are equally apart, they have a closer cardinality. Of course, there is a certain amount of subjectivity in this “natural” identification of the clusters and corresponding breaks, which calls for the adoption of a formal, objective technique. As it will be explained, these three plausible “natural” clusters can be formally identified by minimizing a measure of dispersion within clusters, while maximizing it between clusters (Slocum et al., 2009).

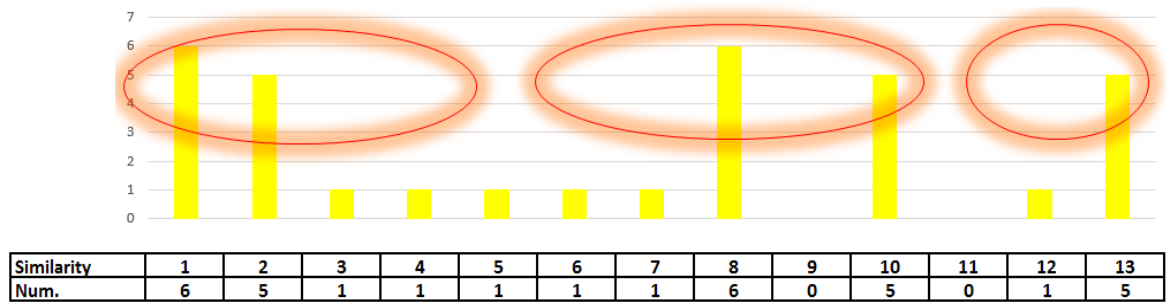


Figure 8.1 – Reference similarity vector.

A first possible solution, called Equal Intervals, consists in dividing the total range of similarity degrees in three sub-intervals of equal range. This technique just considers the degrees of similarity disregarding the number of data (hence resources), thus it has the main disadvantage of being unrelated to the distribution of data. This method was not considered convenient in this case, because similar resources may easily end up in separate classes, and dissimilar ones may end up in the same class. For example, applying this method to the sample similarity vector in Figure 8.1, the range of similarity degrees ($13 - 1 = 12$) is divided by the number of desired classes (3), obtaining the value of 4. The resulting breaks indicating the class boundaries are therefore 1, 5, 9, and 13. The resulting classes, coded with different colours, are shown in Figure 8.2, where it is possible to note that the resources with close degree of similarity 8 and 10 are incorrectly assigned to separate classes.

Equal Intervals classification													
Similarity degree	1	2	3	4	5	6	7	8	9	10	11	12	13
Num. resources	6	5	1	1	1	1	1	6	0	5	0	1	5

Figure 8.2 – Clusters resulting from the application of the Equal Intervals classification.

A second possible solution, called Quantile classification, organizes the data in groups, each one having the same quantity of data (as far as practicable). This technique just considers the number of ranked data (resources), disregarding their actual degree of similarity. Therefore, as in the previous case, it does not consider the distribution of the data. While this may be appropriate when the data are uniformly distributed, it would make little sense in general, because nothing guarantees a uniform distribution. Applying this method to the same reference similarity vector again, the total number of 33 resources would be split in three classes, hence 11 resources per class. The resulting breaks indicating class boundaries are in this case 1, 2, 8, and 13, as shown in Figure 8.3. Here, resources with similarity 8 and 10 are again assigned to different classes, and the only resource with degree of similarity 3 is not in the same class as resources with similarity 2.

Quantiles classification													
Similarity degree	1	2	3	4	5	6	7	8	9	10	11	12	13
Num. resources	6	5	1	1	1	1	1	6	0	5	0	1	5

Figure 8.3 – Clusters resulting from the application of the Quantiles classification.

The Standard Deviation classification method is similar to the Equal Intervals classification, but it does take the distribution of the data into account. Indeed, it creates breaks with ranges that are a proportion of the standard deviation, for example at intervals of 1 or ½ standard deviations below and above the mean. However, this classification method is useful when it is important to discriminate how normally distributed data differ from the mean, which is irrelevant in this case.

The Maximum Breaks method, breaks the data where the largest difference among adjacent values, previously sorted, occur. This simple method, focusing solely on the gaps between adjacent data, would fail to recognize clusters of data if separated by gaps, as in the case of the resources with degree of similarity 8 and 10 in Figure 8.4. This method also fails to separate data in different classes when they are not separated by gaps: for example the degree of similarity 8 is classified in the same group with very different degree of similarity 1.

Maximum Breaks classification													
Similarity degree	1	2	3	4	5	6	7	8	9	10	11	12	13
Num. resources	6	5	1	1	1	1	1	6	0	5	0	1	5

Figure 8.4 – Clusters resulting from the application of the Maximum Breaks classification.

A more suitable method to subdivide the resources in the result-set in clusters of similar characteristics, is to use the Jenks (sometimes referred to as Fisher-Jenks) Natural Breaks Optimization technique (Slocum et al., 2009), which attempts to partition data into classes based on natural groupings inherent in the data. This technique is widely used by cartographers to identify “natural” but objective groupings of data to create choropleth maps (that is, maps that are shaded in proportion to a numerical variable, such as the population density). The method identifies breaks in such a way that data in the same group are as similar as possible among them, and as dissimilar as possible with those in other groups. This is accomplished by minimizing a measure of dispersion (for example the sum of absolute deviations from the median, or the standard deviation) within classes, and maximizing it between classes. A number of slightly different implementations are available in common cartographic and statistical packages, derived from the Fischer-Jenks algorithm published in 1977, that guarantees an optimal solution to the oldest empirical solution of Jenks-Caspall published in 1971 (Slocum et al., 2009). Jenks Natural Breaks Optimization can be considered a one-dimensional version of the more general and popular K-Means Clustering technique, frequently used in machine learning (Jain, 2010), simplified by the possibility of sorting the data, which is not possible in case of multiple dimensions as in K-means.

Figure 8.5 shows the clusters resulting from the application of this technique to the usual reference similarity vector, that generates the breaks 1, 5, 10, and 13. Contrary to the previously examined techniques, it shows that resources having non-contiguous degree of similarity, such as 8 and 10, can be properly grouped in the same class, while resources with contiguous values, such as 5 and 6, can be grouped in different classes. The clusters identified correspond, indeed, to the clusters “visually” identified in Figure 8.1: this is the reason why this technique is called “natural”.

Jenks Natural Breaks classification													
Similarity degree	1	2	3	4	5	6	7	8	9	10	11	12	13
Num. resources	6	5	1	1	1	1	1	6	0	5	0	1	5

Figure 8.5 – Clusters resulting from the application of the Jenks Natural Breaks classification.

Jenks Natural Breaks Optimization is the method selected to cluster resources (according to their degree of similarity) in the prototype Discoverer.

8.2.2 Additional metadata including licensing information to highlight open resources

Test-users expressed high appreciation for the educational oriented data (metadata) visualized by the previous prototypes. They even expressed the wish to obtain additional metadata, including – at least implicitly – about licensing information, in order to highlight open resources. For these reasons, the Learning Registry Search API used in the previous prototype RepExp was used again to identify resources by keywords, but had to be integrated with additional services. To this end, the Learning Registry Obtain Services were used, as in the first prototype Injector, to retrieve the maximum possible amount of available metadata.

While this solution proved effective for some metadata, such as Age Range, Learning Resource Type, Audience, and Author, it was found out that licensing information are specified for a very limited number of educational resources available from the Obtain Services. Hence, the prototype was extended to cope with the slightly different format of GoOpen (Hobson, 2016), a Learning Registry Node specifically dedicated to Open Educational Resources. Licensing information must be mandatorily specified for the metadata uploaded to this node. Unfortunately, it turned out that, while licensing metadata are indeed specified for all its resources, there is an extremely limited number of educational alignments specified, so that at the time of writing, most of its resources cannot be exploited by Discoverer. Yet, the prototype is ready to handle them, as soon as alignment data will become available. Additionally, the large majority of resources returned by the prototype are indeed open educational resources, even if this is not yet formally documented in the available metadata.

8.2.3 Improving the presentation: resource thumbnail images

To contribute to the improvement of the prototype interface, that was purely textual in the previous prototypes, it was decided to include a thumbnail image for each resource. The intention was to satisfy the requests from test-users to provide a more visually attractive presentation, but also to improve its effectiveness. Indeed, mixing textual information and thumbnail previews, should improve the recall of previously seen resource snippets or visited resources (Dziadosz and Chandrasekar, 2002), (Woodruff et al., 2001).

Unfortunately, the thumbnail, that is supposed to be specified by the “thumbnailUrl” attribute in the metadata (schema.org, 2013), is available only for a very limited amount of educational resources. Hence, whenever it is not available, a secondary identifier of the resource is retrieved via the API Search Service, which is then used to retrieve a thumbnail previously obtained by automatically

scraping the original resource. This procedure requires considerable time to be completed: usually between two and four seconds. The potentially negative impact of this delay on the user experience was minimized by asynchronously updating the resulting SERP. This returns most of the information within the first half second, and completes the missing details incrementally, as soon as they become available. This fully satisfies the Google RAIL performance model (Kearney, 2017), that considers appropriate – for an engaging user experience – to load data incrementally, provided that the essential information is delivered within a window of 300 - 1000 ms.

8.2.4 Discoverer sample session

This section illustrates a sample search/discovery session actually carried out with the prototype, with a commented sequence of screenshots.

Figure 8.6 shows the SERP returned by Google following a search with the keyword “add”. It includes entries concerning Attention Deficit Disorder, and the brand “add” selling down jackets, as well as entries related to the mathematical operation addition.

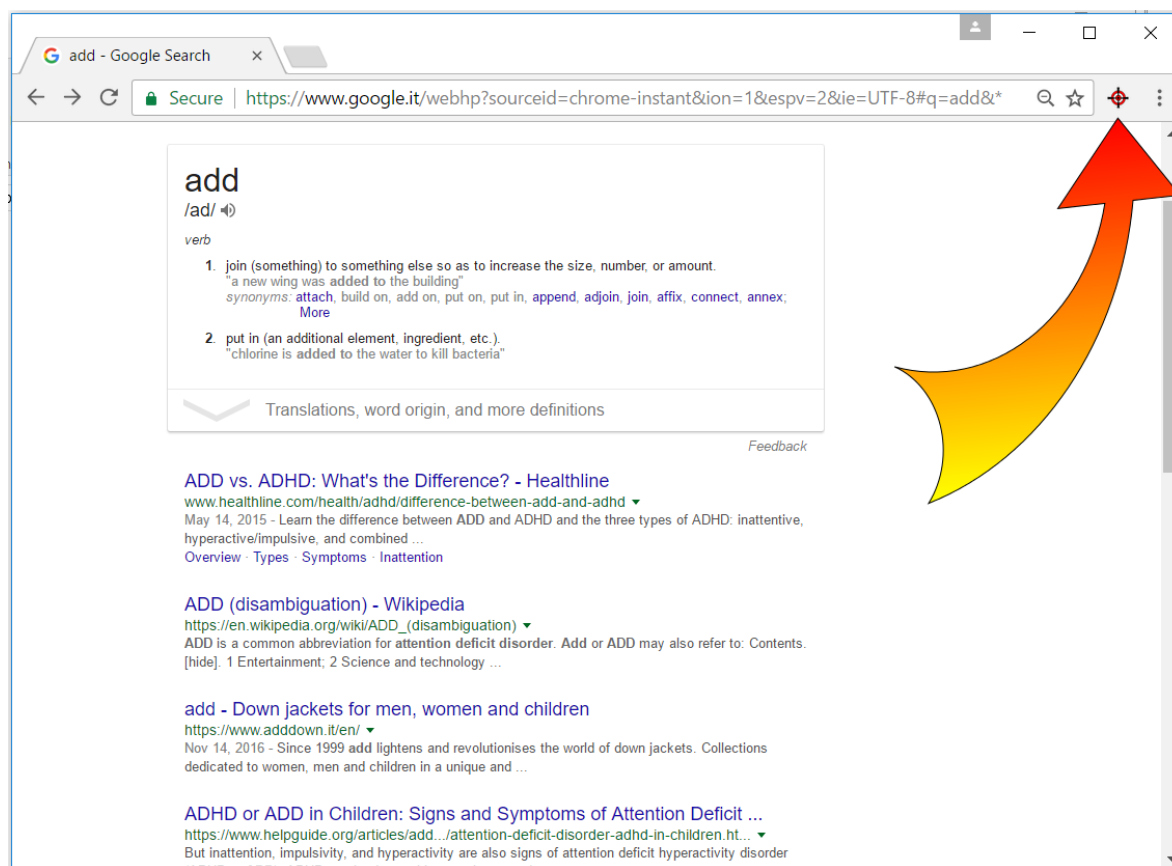


Figure 8.6 – Activating Discoverer from a Google SERP.

The user is not satisfied by these results, and decides to activate Discoverer to get more relevant resources and metadata, by clicking on its icon in the top right corner of the screen. Discoverer detects that it has been called from the context of a Google SERP, hence it intercepts the keyword(s) used in the previous search (just “add” in this case), repeats the query in a specialized hub of educational resources (the Learning Registry), and displays the new resulting custom SERP reported in Figure

8.7. This search has identified 30 educational resources, displayed in the custom SERP with related educational metadata.

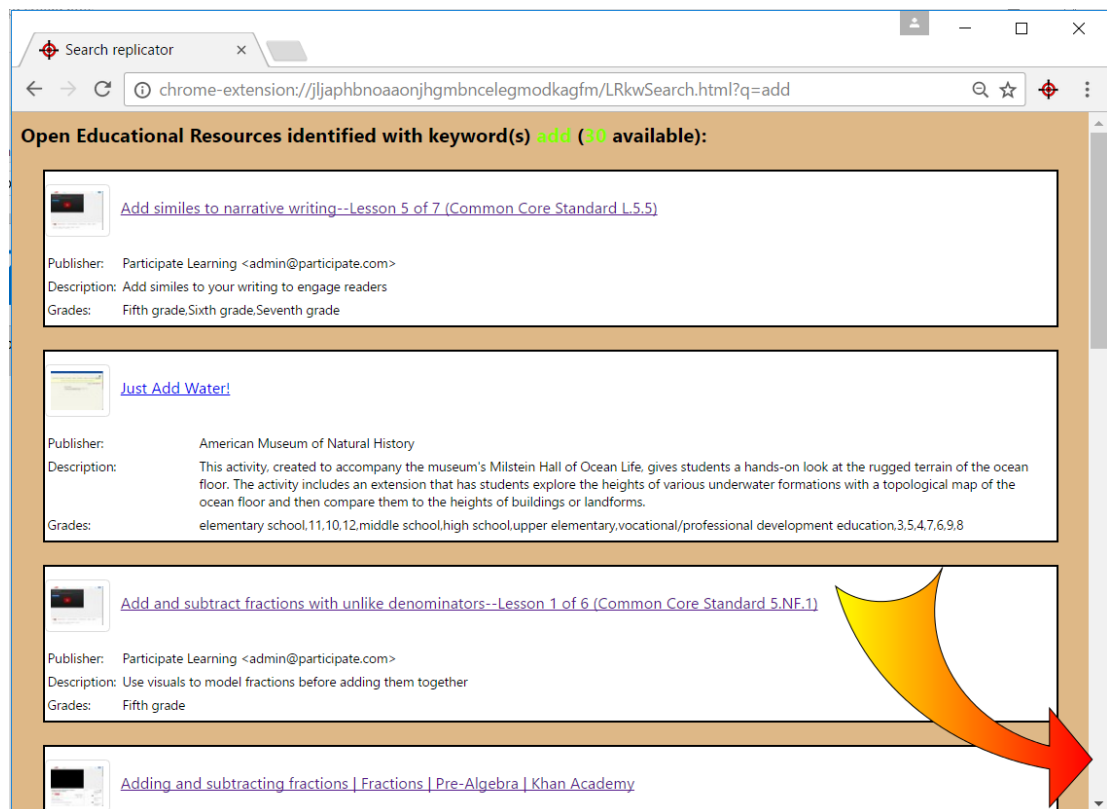


Figure 8.7 – SERP returned by Discoverer, following a query with the intercepted keyword "add".

The user scrolls down the list obtaining Figure 8.8, and being interested about adding fractions, he/she clicks on the potentially useful resource “Fraction Tutorial” to explore it.

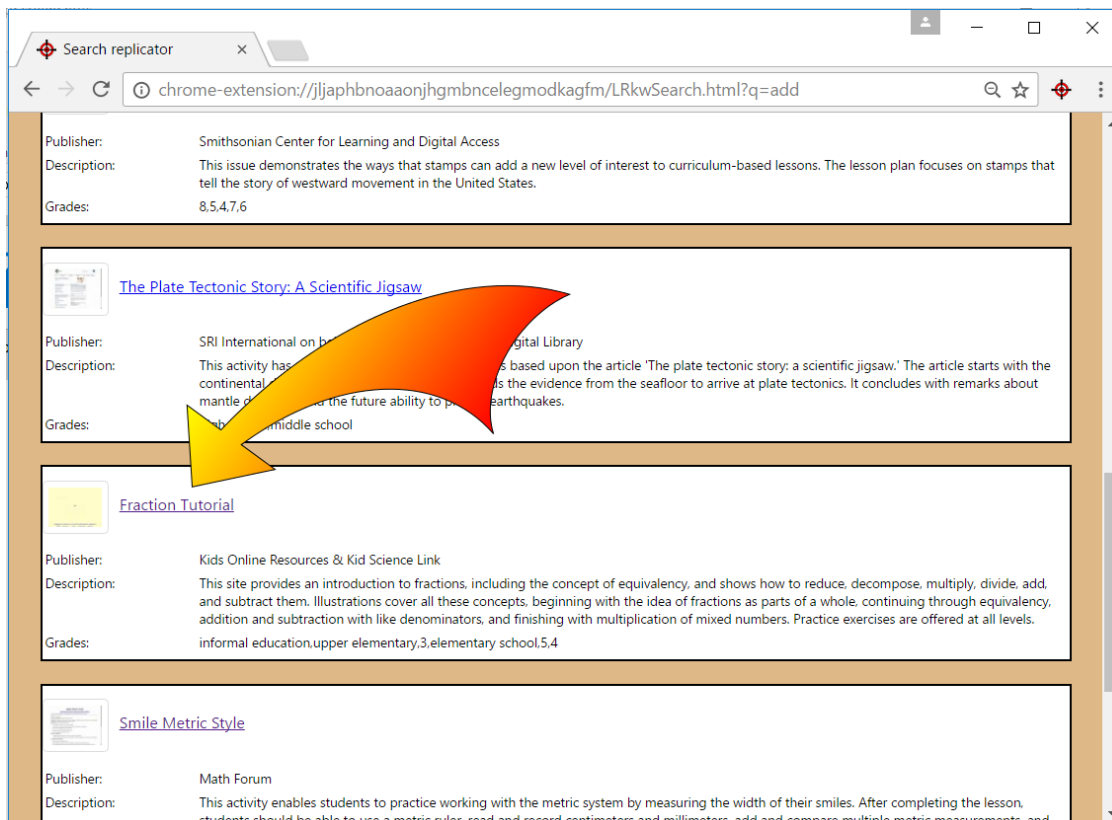


Figure 8.8 – Scrolling down the SERP and visiting a resource of interest.

The educational resource is displayed in a new tab, as shown in Figure 8.9, so that the user can freely explore it.

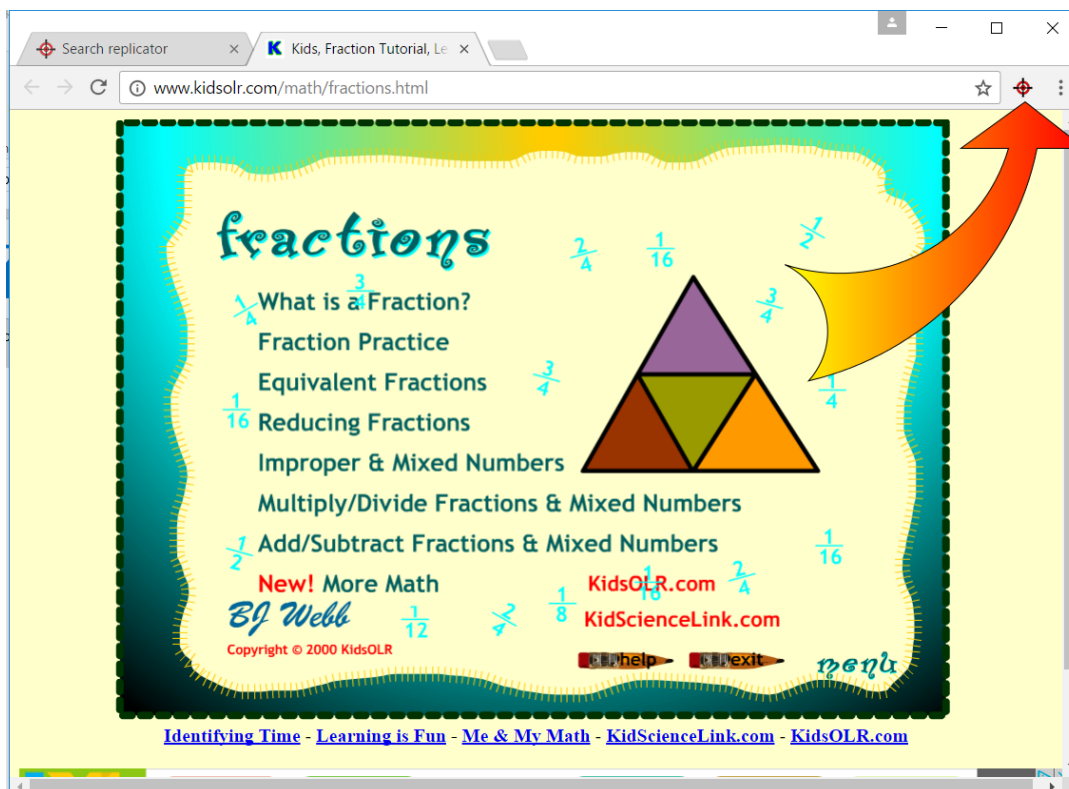


Figure 8.9 – Exploring the educational resource.

The user is now interested in getting additional resources similar to this one, so he/she clicks on the Discoverer's icon again. This time, Discoverer detects that it has not been called from a Google SERP, therefore it reads the current URL, assuming that it is the identifier of an educational resource. Discoverer attempts then to identify the educational alignments of the resource, searches for additional resources sharing some of those educational alignments, ranks and clusters resources by degree of similarity, and creates and displays the SERP page in Figure 8.10. This page indicates that 617 similar resources have been identified, and shows the first few having maximum similarity. The snippets displayed for each resource contain a rich set of educational metadata such as, in this case, Publisher, Author, Description, Audience, Learning Resource Type, and Age Range.

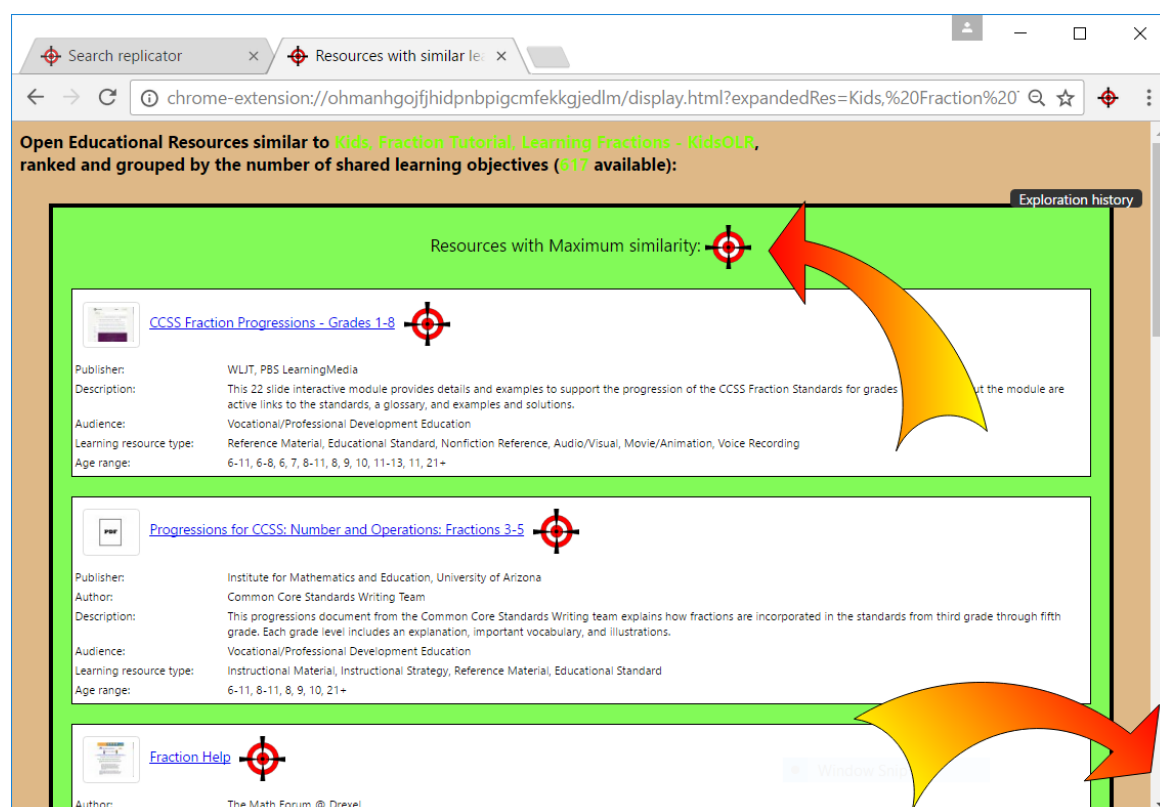


Figure 8.10 – SERP page with similar educational resources, ranked and clustered by degree of similarity.

By scrolling down the window, it is possible to see another cluster of resources, this time with medium degree of similarity, shown in Figure 8.11. Scrolling further down would show a third cluster containing resources with minimum similarity.

Discoverer only shows a few sample resources (the three with the highest similarity with the original resource) for each cluster, but the user can get more by clicking on the corresponding button. For example, the user requests more resources with high degree of similarity in Figure 8.11, obtaining Figure 8.12. Here, the user decides to hide again the additional resources with maximum similarity, by clicking on the corresponding button.

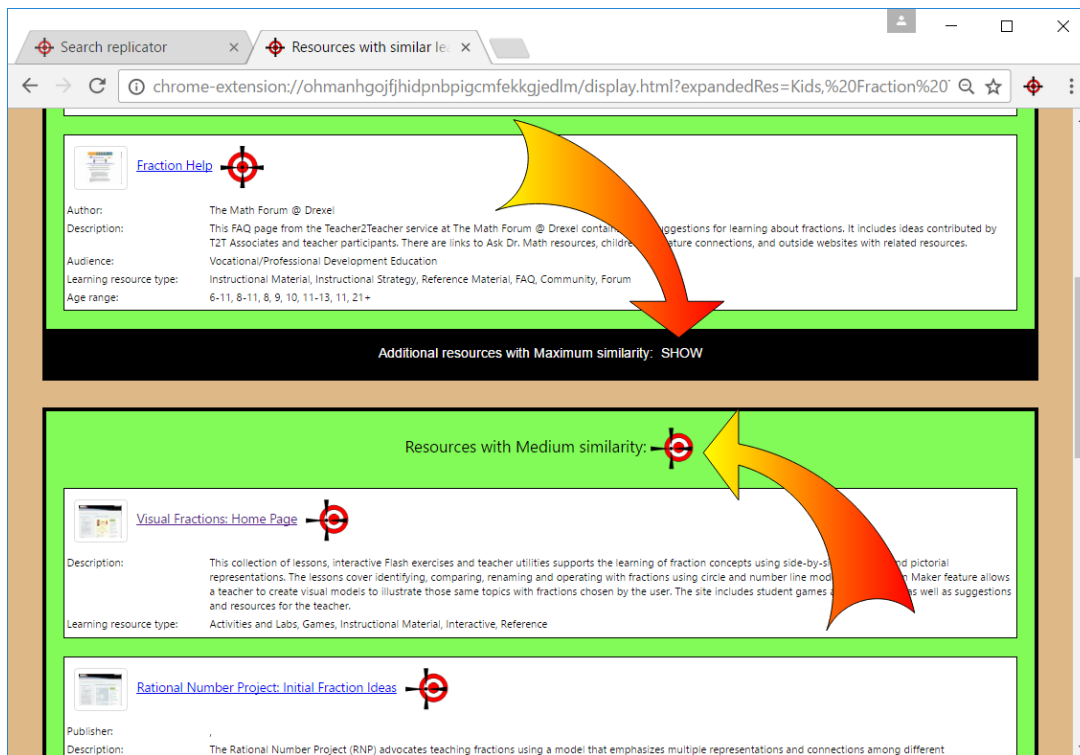


Figure 8.11 – Educational resources with medium degree of similarity; requesting additional resources with maximum similarity.

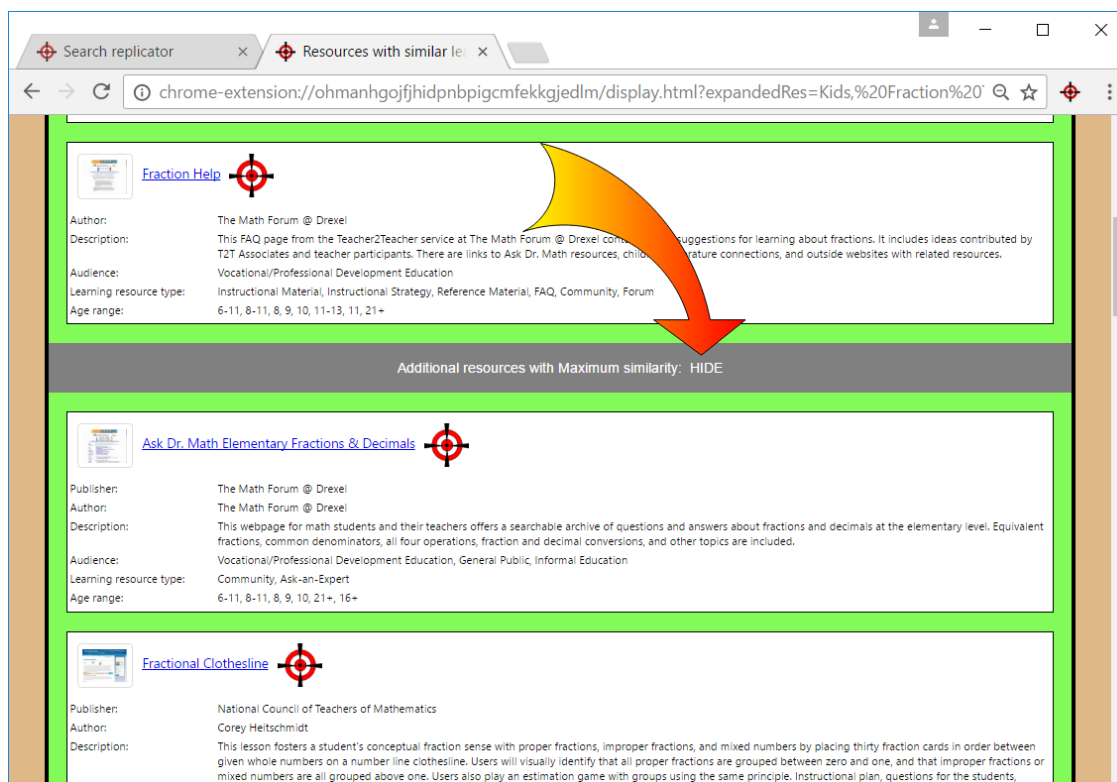


Figure 8.12 – Hiding the additional resources with maximum similarity.

The user is now interested in obtaining additional educational resources, to be used for in-depth activities, therefore not too similar to the previous one. Hence he/she decides, always from the same SERP in Figure 8.13, to explore one of the listed resources with medium similarity. Clicking on its title, the user gets access to the resource as shown in Figure 8.14. It is now possible to request further

similar resources, in this case 2854 as shown in Figure 8.15, again ranked and clustered according to their degree of similarity.

It is important to note that such an expansion could be triggered from any resource, found in any possible way, not just via Discoverer. For example, an educator might start directly from a resource he/she is already using in the classroom, and easily find other very similar resources for remediation activities, or moderately similar resources for in depth activities.

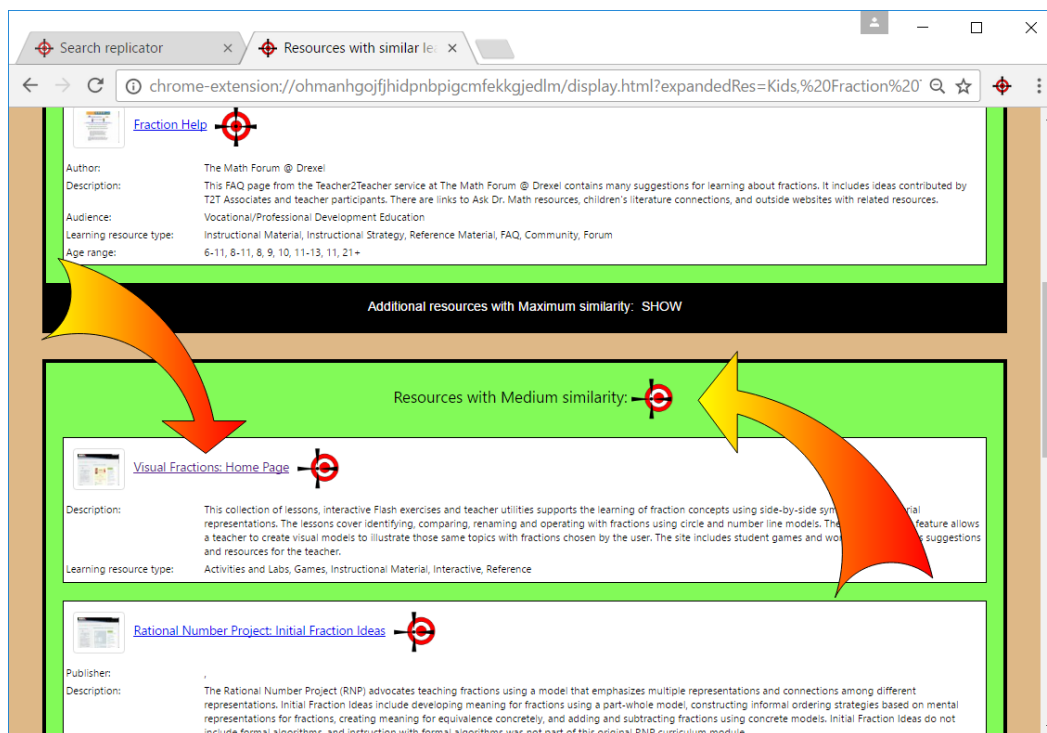


Figure 8.13 – Exploring an educational resource with medium degree of similarity.

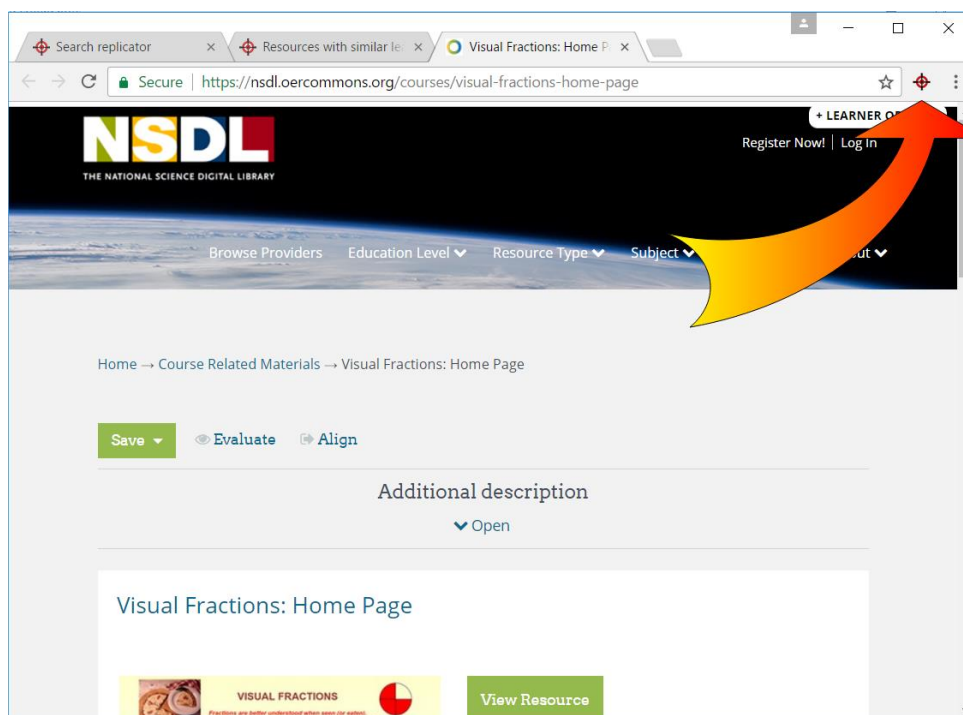


Figure 8.14 – Exploring the new resource and getting further similar resources.

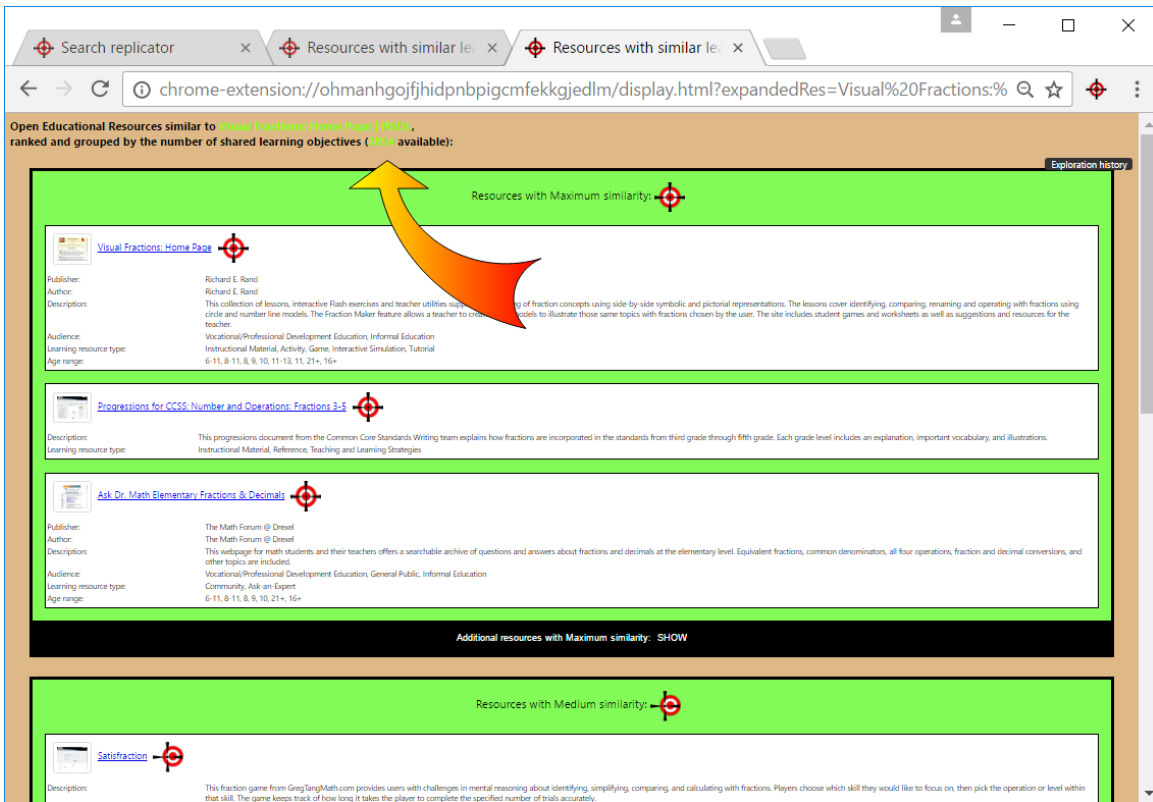


Figure 8.15 – New SERP with additional (2854) similar resources, again ranked and clustered by degree of similarity.

At any time, the user can display his/her navigation history, as shown in Figure 8.16.

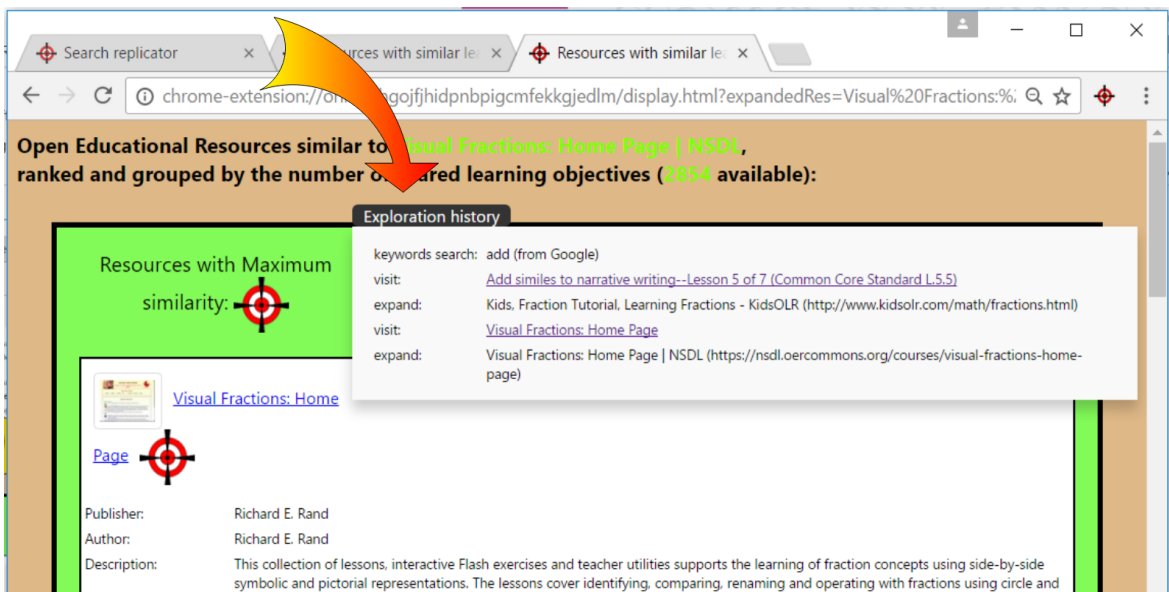


Figure 8.16 – Displaying the user navigation history.

8.2.5 Discoverer architecture

Discoverer is an evolution of the two previous prototypes Injector and RepExp, and as such it replicates many of their components. The most meaningful differences concern the component for

results clustering, and the enrichment of the visual interface. The overall architecture is schematized in Figure 8.17.

In order to reduce the confusion arising from the use of two separate extensions (Replicator and Expander) in the previous RepExp prototype, Discoverer was designed with just one extension. Whenever the prototype is activated, by clicking on its corresponding (unique) extension icon in the browser, it detects first the context from where it was called. If it was called from a Google SERP page, it activates its functionalities as a keyword-based search replicator. Otherwise it assumes that the user is interested to expand the educational resource that he/she is currently visiting. In other words, the prototype tries to find more resources with whatever data are available from its context: the search keywords used in a query, or the characteristics of the educational resource being explored (QBE).

As a keyword search replicator, Discoverer works like the Replicator subcomponent of its predecessor RepExp: the Search Keywords Interceptor transparently intercepts the keywords used in the current Google search, builds a corresponding query that is sent to the Learning Registry Search API, obtains the JSON-LD coded results and formats them as an HTML document. This document is indicated as Keywords Search SERP in Figure 8.17, and is displayed in a separate tab in the browser. The user is free to explore this results page and navigate to any of the listed resources.

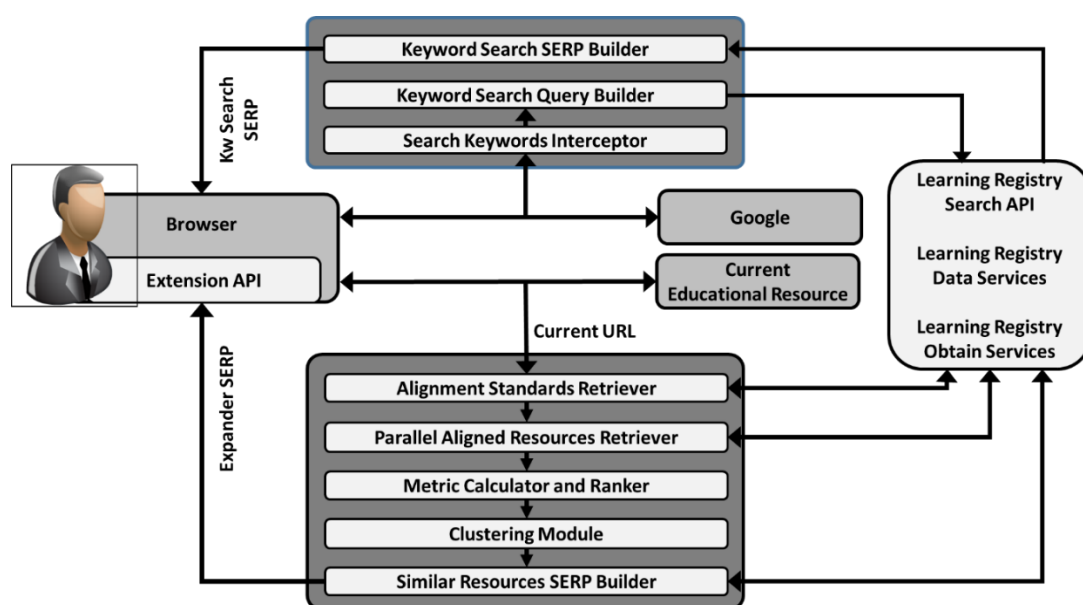


Figure 8.17 – Discoverer architecture.

When acting as an educational resource expander (that is, when activated while the user is exploring an educational resource) Discoverer reads the URL of the resource. The Alignment Standards Retriever attempts to get the educational alignments of the resource via the Data Services or, when no results are returned, via the Search API of the Learning Registry. The Parallel Aligned Resources Retriever then attempts to asynchronously identify the resources aligned to each previously identified standard.

When all the data have been successfully retrieved, the Metric Calculator and Ranker computes the similarity vector (NumStdInCom) containing the degree of similarity between the expanded resource and all other resources identified.

In this similarity vector, NumStdInCom [Res_i] is an integer value that indicates the number of alignments that the expanded resource has in common with the related resource Res_i. This is their degree of similarity, according to the metric defined in this thesis. This vector – here is the new element of this prototype – is now exploited by the Clustering Module, making use of the Jenks Natural Breaks optimization method, to classify the available resources in three (or less, in case there are less than three different degrees of similarity) clusters of comparable degree of similarity: high, medium and low. The Similar Resources SERP Builder creates an HTML page, named Expander SERP in Figure 8.17, which lists the resulting resources with related educational metadata and icon, in the user browser. The system displays a few representative resources for each one of the available clusters directly in the SERP page, and offers the possibility of obtaining additional resources from any cluster, by expanding it with an accordion user interface technique. Differently from the previous prototype, the system makes also use of the Learning Registry Obtain Services to retrieve additional educational metadata not (yet) available from the Learning Registry Search API, to enrich the resulting SERP. From here, of course, the user may navigate to any resource and further expand those of interest.

8.3 Discoverer Summative Evaluation

Discoverer was evaluated with a summative evaluation in this last iteration of design/evaluation DSR cycles. The aim of this evaluation was not so much to identify weaknesses and areas for improvements as in previous iterations: it was more oriented to establish a credible evaluation of the general effectiveness of the artefact developed, in the target context.

Venable et al. (2012) classify this kind of prototype as a “socio-technical” artefact, because “humans must interact” with it to provide its utility. According to their framework for evaluation in DSR, a suitable strategy in this case is a naturalistic, ex post evaluation; a suitable research method is its observational or naturalistic evaluation. This was realized by collecting and analysing feedback from real users carrying out realistic tasks with the prototype.

The evaluation criteria focused on utility as in most DSR researches. The objective was to evaluate the abstract model of the specific artefact instantiation. Obviously, the specific instantiation is to be intended just as an instrument to support the evaluation of the abstract artefact (Prat et al., 2014).

8.3.1 Updating the evaluation questionnaire

The questionnaire used in the previous evaluation, to be administered again as structured interview or survey to representative users, was adapted to the new goals, in particular to establish the relevance of the proposed scenarios / tasks, and the overall suitability of the proposed solution to support users in their WC-level tasks, mainly:

- Transparent and direct access from Google
- QBE search by similarity
- Similarity metric transparently based on educational alignments (learning objectives)
- Clustering in three expandable classes of comparable degree of similarity

Taking advantage from the experience with the previous evaluation and from consultations with experienced academics, the questionnaire was further improved. It was decided to collect additional information concerning familiarity of test-users with OERs, to identify possible correlations with their answers. It was decided to solicit further qualitative input with open questions to motivate the numerical answers. A question was added, by adapting Reichheld's (2003) original question "How likely is it that you'll recommend this product to a friend or colleague?" The wording of some questions and possible answers was improved. A question (original number 11) was rephrased because it was ambiguous. In some cases these modifications converted the corresponding original bipolar questions to unipolar questions, so that the previous 7 points scales were reduced to 5 points scales, with a view to increase validity and reliability (Krosnick et al., 2014).

Finally, the opportunity to label all scale points with words was analysed, to "ensure consistent interpretation of rating scales" (Krosnick et al., 2015, p. 7), hence improving validity and reliability. In this case, however, the evaluation questionnaire was going to be used also with test-users not having English as their first language. Hence, on the contrary, it was considered less ambiguous and more reliable, to label options with numbers. The use of numbers was also expected to increase the legitimacy of data processing with statistical parametric techniques, making it more legitimate to interpret the scoring assigned by test-users as interval data rather than ordinal data (Norman, 2010). Yet as an exception, for increased clarity and to help quickly differentiate between unipolar and bipolar questions, the central scale point in bipolar questions was labelled with the word "neutral".

The updated questionnaire, available as Appendix F, obtained institutional ethical clearance. This time an Italian version was prepared too, considering its wider distribution to educators whose command of the English language was unknown. Participants could reply in English, Italian, French, or Spanish.

Considering the high realism reached by the current version of the prototype, it was also planned to collect additional "field notes" resulting from the observation of test-users using the prototype, when the questionnaire was administered as structured interview.

8.3.2 Discoverer Evaluation – results and discussion

The invitation to participate in the prototype evaluation was sent to about 50 educators in three different countries, identified via a snowball strategy. While no sensitive data were collected from participants, they were anonymized once the necessary clarifications could be obtained.

Participants profile

Twenty-nine educators, whose profile is summarized in Table 8.1, accepted the invitation to participate in the evaluation. Two of them, User17 and User24, overlapped respectively with User1 and User6 in the previous study.

Age		Sex		Teaching Nation		Teaching Experience		Teaching Level	
<=30	6.9%	M	55.2%	IT	79.3%	<=5	6.9%	High School	79.3%
31-40	10.3%	F	44.8%	BR	6.9%	6-10	3.4%	Primary	6.9%
41-50	17.2%			UK	13.8%	11-20	24.1%	University	13.8%
>=51	65.5%					>=21	65.5%		

Subject		Knowledge about OERs		Frequency of use of OERs	
Humanities	31.0%	None	3.4%	Never	10.3%
Math	13.8%	Some	55.2%	Occasionally	55.2%
Science	17.2%	Good	41.4%	Very often	34.5%
Technology	37.9%				

Table 8.1 – Test-users' profile.

The teaching experience of educators in the sample is highly correlated with their age: Kendall's tau test indicates a correlation coefficient of about 0.8, significant at the 0.01 level of confidence.

The frequency of use of OERs by test-users is only moderately correlated with their reported level of knowledge about OERs: Kendall's tau test indicates a correlation coefficient of about 0.38 with a p-value of 0.036. The frequency of use of OERs is not correlated with working experience nor with age of educators.

Likert-type scales: central tendency and dispersion

Quantitative data were collected in the questionnaire via Likert-type scales, so that it would be appropriate to treat them as ordinal data. While the scale design discussed in the previous section could justify their treatment as interval data, they are not normally distributed. This can be appreciated from the visual examination of their distribution, because the data are skewed towards the maximum possible value. Additionally, this is confirmed analytically by the p-values obtained from the Shapiro-Wilk test, which is lower than 0.01 for all variables: the null hypothesis that the samples are drawn from a population with a normal distribution needs to be rejected. Hence, it is convenient to consider multiple measures of central tendency. Median and mode are appropriate for ordinal data, but it is also considered legitimate to use the mean, because the data do not contain outliers that could result in bias. A safe measure of dispersion is the range across quartiles (IQR), which is suitable for skewed distributions and ordinal data too (Leppink, n.d.).

These statistics are reported in Table 8.2, for all variables corresponding to the Likert-type scales used in the evaluation survey, together with their description and short name used in this section. 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.

Variable name	Description and possible range	Mode	Median	Mean	IQR
TranspFrmGoogle	Usefulness of transparently starting a specialized search directly from Google, compared to dedicated portals. [1..7]	7	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	7	6.7	0.5
SimilarityGeneral	Usefulness of the expansion by similarity in general, considering also other metrics (such as togetherness, likedness). [1..5]	4.5	4	4.0	1.5
DiffcltyAltTechn	Difficulty to find resources that share the same learning objectives with alternative tools currently used. [1..7]	7	6	6.4	1
ClustrForOverview	Usefulness of clustering to help educators making sense of large volumes of hits. [1..5]	5	5	4.8	0
ClustrForEducStrat	Usefulness of clustering to support search of resources targeting specific educational strategies. [1..5]	5	5	4.8	0
UserOrntTermin	Familiarity / user orientation of terminology and concepts. [1..7]	7	7	6.6	1
TranspUseLO	Usefulness of the tool transparency, which avoids explicit handling of formal learning objectives. [1..5]	5	5	4.8	0
WdLikeUsing	Willingness to use the tool. [1..7]	7	7	6.8	0
WorkloadReduct	Effectiveness in reducing workload.	7	7	6.6	0
WouldRecomm	Willingness to recommend the tool to a colleague. [1..7]	7	7	6.9	0
ScenRelevance	Relevance of the scenarios proposed in the evaluation.	5	5	4.8	0

Table 8.2 – Basic descriptive statistics of Likert-type scales.

As it can be seen, the feedback collected was very positive: the mode corresponds, in all but one case, to the maximum possible positive value. The only case where test-users did not give the highest favourable feedback, was about the question asking about the usefulness of the expansion by similarity in general, considering also other metrics such as togetherness or likedness. These additional metrics were not implemented in the prototype, because they were considered of limited priority by some educators in the OERs task taxonomy empirical evaluation. However, they were mentioned in the questionnaire because frequently considered in the literature. While IQR is anyway robust against potential outliers, it is worthwhile remarking that not a single negative or neutral opinion was expressed.

Figure 8.18 shows, as an example, the frequency distribution for the variables SimilarityByLo and WouldRecomm (max range 1..7), and TranspUseLO (max range 1..5).

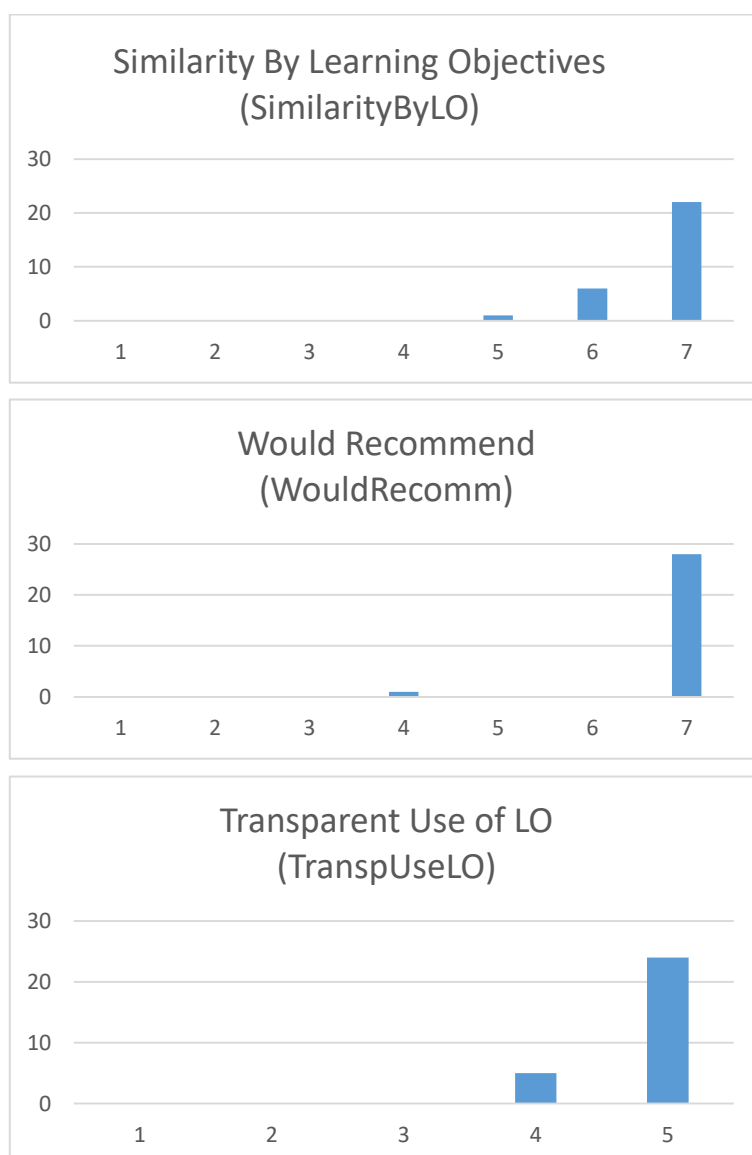


Figure 8.18 – Frequency distribution for the variables SimilarityByLo, WouldRecomm, and TranspUseLO.

In the case of WouldRecomm, the only value lower than the maximum was motivated by the participant saying:

“Because I’d like to use it for a while first” [User10]

Table 8.2 and the boxplots for all Likert-type scales in Figure 8.19, show that the variable with the largest dispersion is SimilarityGeneral: educators consistently consider that learning objectives are much better than keywords for a similarity metric, but their interest in other similarity metrics is less firm. The reason, as further discussed in the qualitative analysis section, is that some educators were quite negative about the idea of exploiting social data. This result is consistent with the outcomes from the TA empirical evaluation.

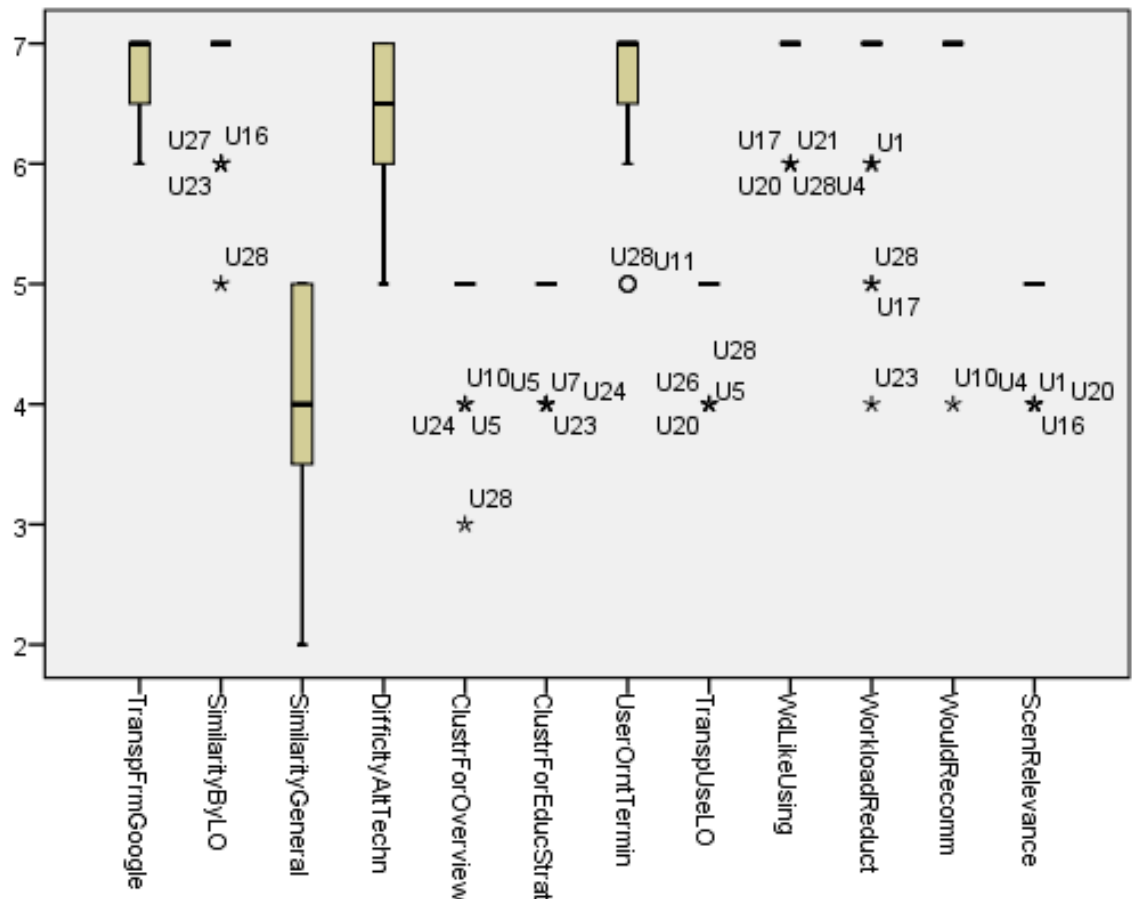


Figure 8.19 – Boxplots for variables corresponding to Likert-type scales.

Analysis of distribution across different profiles

Data do not show significant differences among educators with different profiles, such as sex, age, or teaching subject. The clustered bar charts in Figure 8.20, for example, shows the mode for some meaningful variables 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 would be expected, given that most data consist anyway in the highest possible value.

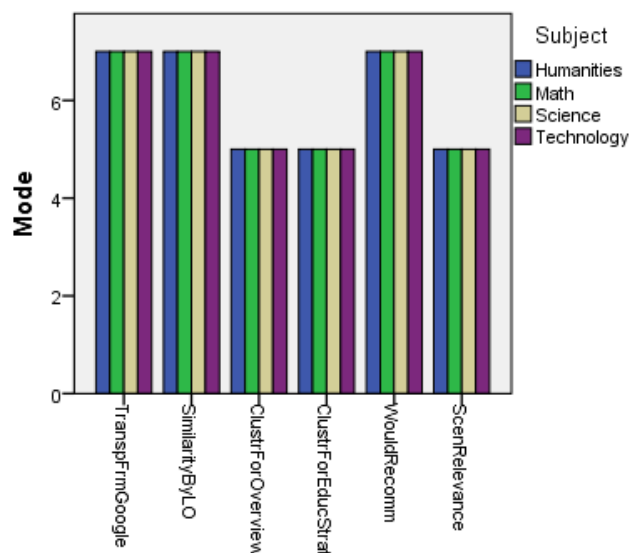


Figure 8.20 – Distribution across teaching subjects.

Experience in using OERs: the main factor that differentiates participants responses

Yet, the characteristic that most differentiate educators' responses, is their experience in using OERs (never, occasional, very frequent), and, to a more limited extent, their reported knowledge about them. It is evident from the boxplots in Figure 8.21, that the more participants are experienced with OERs, the more they are 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. Similar results can be obtained by considering the subgroup of participants from Italy.

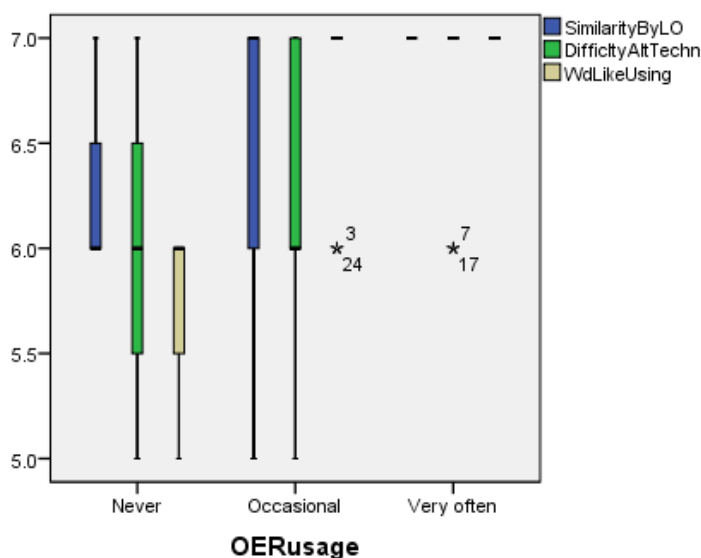


Figure 8.21 – Experience with using OERs is correlated with positive opinion.

The non-parametric independent-samples Kruskal-Wallis test (that can be used to identify statistically significant differences between two or more groups of an independent variable) allows us to reject the null hypothesis that the distributions of SimilarityByLO ($H(2) = 6.055$, $p = 0.048$), DifficltlyAltTechn ($H(2) = 6.212$, $p = 0.045$), and WdLikeUsing ($H(2) = 16.891$, $p < 0.01$), are the same across categories of educators with different experience in the use of OERs. These results can

be confirmed by a Jonckheere-Terpstra test, that has a similar purpose as Kruskal-Wallis but has more statistical power in this case, because it can exploit the a-priori ordering of the populations. The conclusion indeed is the same, but in this case with an even better p-value (Figure 8.22).

Hypothesis Test Summary				
	Null Hypothesis	Test	Sig.	Decision
1	The distribution of SimilarityByLO is the same across categories of OERusage.	Independent-Samples Kruskal-Wallis Test	.048	Reject the null hypothesis.
2	The distribution of SimilarityByLO is the same across categories of OERusage.	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	.015	Reject the null hypothesis.
3	The distribution of DifficltyAltTechn is the same across categories of OERusage.	Independent-Samples Kruskal-Wallis Test	.045	Reject the null hypothesis.
4	The distribution of DifficltyAltTechn is the same across categories of OERusage.	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	.013	Reject the null hypothesis.
5	The distribution of WdLikeUsing is the same across categories of OERusage.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.
6	The distribution of WdLikeUsing is the same across categories of OERusage.	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	.002	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 8.22 – Differences in distributions across experience in using OERs.

In order to obtain an estimation of the effect size, it is possible to use correlation analysis with Kendall's tau. In this case the effect size corresponds to the correlation coefficient, while the p-value obtained is necessarily the same as with Jonckheere-Terpstra. The sign of the correlation coefficient indicates the direction of the correlation. For example, the correlation between OERUsageCode and SimilarityByLO computed with Kendall's tau is $r_t = 0.436$: a positive correlation coefficient confirms that the more educators make use of OERs, the more they appreciate the proposed similarity metric. Correlation coefficients with DifficltyAltTechn and WdLikeUsing are, respectively, $r_t = 0.439$ and $r_t = 0.547$.

Prototype appreciated for different reasons in different nations

Distribution of the various variables are pretty uniform across nations, except for those in Figure 8.23, that show a difference among Italy and UK, amounting to about one unit in each case.

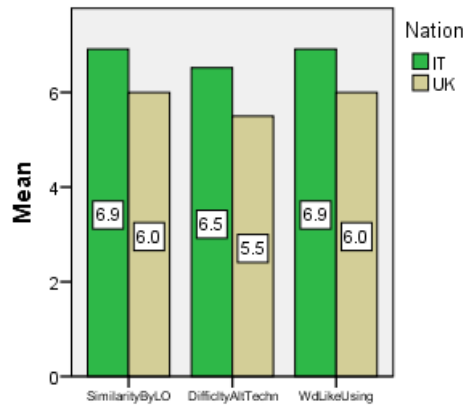


Figure 8.23 – Distribution across Nations: Italy versus UK.

Indeed, the independent samples Kruskal-Wallis test identifies statistically significant differences among the distributions of SimilarityByLO ($H(2) = 14.62$, $p = 0.01$), DifficultyAltTechn ($H(2) = 8.95$, $p = 0.011$), and WdLikeUsing ($H(2) = 11.13$, $p = 0.004$) across Nations. The Mann-Whitney U-test can be used to identify which independent groups are responsible for these differences. This test shows that the statistically significant difference is between Italy and the UK. However, while there are too few test-users from Brazil, to obtain statistically significant results, their data are very much similar to those of Italian participants.

These types of differences were not unexpected, because the prototype addresses the Work-Context user domain, which is inevitably affected by organizational and cultural aspects (Wilson et al., 2010). This difference, as anticipated in the preliminary literature review, might be linked to the higher familiarity of British educators with formal learning objectives, compared to Italian (and Brazilian) educators. This higher familiarity is revealed by the feedback from British educators, which make direct references to learning objectives, for example:

“Often when researching I might get interested in another area and not focus enough on my Learning Objective. This would enable me to carry out more research, have more choice, but also to stay within my planning brief”. [User29, UK]

“Declared learning objectives provide a more reliable guide for the content”.
[User17, UK]

This is further supported by a comment from a British test-user who even expressed the desire to handle learning objectives directly:

“I would like learning objectives to be an explicit part of the process”. [User25, UK]

This might explain, in particular, the slightly lower score assigned by UK educators to DifficultyAltTechn. Indeed, given their higher familiarity with formal learning objectives, they may feel better equipped to handle the task of finding resources sharing similar learning objectives with alternative techniques. In other words, they seem a bit more at ease to deconstruct the high level task in a sequence of subtasks.

Despite these minor differences, however, all British educators assigned the maximum possible score (7) to their willingness to recommend the prototype to their colleagues. The previous participant, for example, commented indeed:

“I’m impressed and will recommend this to friends. I hope you will let me know when it is available”. [User25, UK]

The prototype seems highly appreciated by educators from both Countries. However, there is some evidence that this could be due to partially different reasons. Italians like it mainly because it does not force them to use directly formal learning objectives they are not familiar with, coherently with the outcomes of the taxonomy empirical evaluation, and the explicit goal of the prototype. Yet British participants like it too, but perhaps because they are even more conscious about the discriminating power of learning objectives.

Analysis of correlation among Likert-type scale variables

The methodology that lead to the final prototype design, makes it reasonable to expect that educators who appreciate the similarity metric based on learning objectives, would also appreciate clustering of resources by degree of similarity, to provide a quick overview and to support different educational strategies. Indeed a Spearman's rank-order test identifies a statistically significant positive correlation between SimilarityByLo and ClustrForOverview ($r_s = 0.430$, $p = 0.020$), ClustrForEducStrat ($r_s = 0.535$, $p = 0.003$), WorkloadReduct ($r_s = 0.450$, $p = 0.016$), WdLikeUsing ($r_s = 0.421$, $p = 0.023$).

The feature of clustering in Discoverer was included for multiple reasons, including to help making sense of sometimes large volumes of results, and to support educators at WC-level in their lesson planning activities. Indeed, there is a positive correlation between ClustrForOverview and ClustrForEducStrat, which shows that whenever clustering is appreciated, it is appreciated for both reasons.

A strong correlation could also be expected between WdLikeUsing and WouldRecommend. Yet, this is not identified by Spearman (nor by Kendall's tau). Indeed, curiously, while educators assigned 28 scores of value 7 to WouldRecommend, they assigned “only” 24 scores of value 7 to WdLikeUsing. This could be due to random fluctuations.

There are a few other statistically significant correlations. However, the high probability of finding a correlation because of pure chance (type I error) is too high, considering the number of possible combinations and the data flattened to their maximum value.

Qualitative data analysis

The evaluation survey gathered about 4000 words of text: justifications of answers to Likert-type scales, answers to open questions, comments, and a few field notes obtained by observing educators using the prototype.

Following the content analysis approach as in the previous studies, the text was analysed to extract and iteratively refine a list of keywords (codes) denoting the key concepts expressed. The text was then annotated (coded) with the keywords from this code list, and their frequency of occurrence used as a measure of the importance of each concept. Some of the concepts were then correlated to educators' profile characteristics. This time, given the larger amount of data, a simple CAQDAS (Computer Assisted Qualitative Data Analysis) tool (CAT - Coding Analysis Toolkit) was used to support the process.

The data happened to be quite homogeneous: a limited number of codes was sufficient to capture the concepts expressed. While this process was separately carried out for the text related to each question, most of the core concepts addressed by this research, such as "exploratory search", "domain orientation", "WC-level tasks", and "personalization", got consistently repeated in different groups. Here follows a summary of the most relevant outcomes.

Similarity by Learning Objectives

Participants were asked to rate the usefulness of the expansion by similarity based on Learning Objectives, compared to traditional metrics based on shared words (SimilarityByLO). They justified the very positive scores they assigned with three reasons: domain-orientation (15), efficiency (10) and precision (9).

Domain-orientation was clearly the most appreciated aspect, for example:

"Better precision, objectives are more domain oriented" [User20],

"It targets precisely the needs of teachers" [User4],

"It doesn't matter that words are similar: objectives should be similar" [User22].

This is an important outcome because, as discussed in the literature review, domain orientation is one of the most desirable characteristics of a specialized search engine.

Another participant commented:

"It is in line with the latest ministerial indications (learning outcomes, skills...)"
[User12].

This is also important because "utility for organizations is the ultimate measure of IS artifact relevance" (Prat et al., 2014, p.7).

Similarity in general

Five respondents justified their positive scores about the usefulness of expansion by similarity in general, that is including other similarity metrics, with the opportunity to address quality:

"Users guarantee the quality of the resources they prefer" [User3].

This is definitely consistent with the literature which identifies the quality of the resources as one of the fundamental success criteria for OERs repositories (e.g. Clements, 2016).

Interestingly, seven respondents were only moderately positive about the use of other metrics, while four respondents even mentioned explicitly that they had no interest in relying on the opinion of their peers (exploited by the relatedness metrics used as examples in the question):

“I would trust other people’s judgement only partially” [User9],

“I am not interested in other people opinions and behaviour” [User14].

These opinions explain the higher dispersion (IQR) of the corresponding scores, and are consistent with the outcomes from the task-taxonomy evaluation. Yet, as previously observed, this is not always consistent with the ideas reported in the literature.

Finally, three educators mentioned its support for WC-level tasks:

“To find the best strategy to present a topic [...]” [User22],

*“To carry out the complementary activities consistent with the main course of action
(such as remediation, reinforcement, review, etc.)” [User18].*

Supporting user tasks at WC-level is another fundamental objective of a specialized search engine, as discussed in the literature review (Section 2.8).

Clustering for overview

Participants, to motivate their high scoring on the usefulness of clustering to help educators making sense of a large volume of results, used the following main motivations: efficiency (13)

“To identify quickly the most suitable resource” [User2],

and domain/WC-orientation (7), explicitly mentioning WC-level tasks again:

“To personalize the educational activities” [User7],

*“This is helpful also to plan the development of additional educational activities”
[User23].*

Four educators explicitly mentioned the support for exploratory search, one of the key aspects of the prototype:

“It helps to focus / refine previous imprecise objectives, as it is often needed” [User15],

*“It allows you to look for variety in case your first choice of subject was very general”
[User28].*

Clustering to support educational strategies

The very positive rating of the usefulness of clustering to support search of resources targeting specific educational strategies, was largely justified (17) by its support to WC-level tasks. Participants 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” [User8],

“It simplifies personalized educational activities” [User10],

“Useful for both remediation and in-depth activities, as students require personalized support” [User12].

Support to exploratory search was mentioned again:

“Sometimes the user doesn’t know the keyword for what he/she is searching for” [User17].

Number of classes of similarity

A minor objective of the evaluation was to identify the appropriate number of clusters to be used by the prototype. The large majority of test-users (27) considered that the number of clusters used in the prototype, that is 3, was appropriate. Six participants indicated in particular that more clusters would be “dispersive”, “too fragmented”, or “complex”. For example:

“I think three classes are a good compromise - Adding more is likely to make the search platform too dispersive” [User23].

Two participants wished to have 5 classes of similarity:

“I consider that the number of classes is strictly correlated with the precision of the software in measuring similarity. I would personally prefer 5 classes” [User18].

Finally, User25 suggested, as an alternative, to rank resources in order of similarity. This is exactly the solution that was experimented with RepExp, which proved to require a different solution.

Additional functionalities

Participants were asked to provide suggestions for the two most important additional functionalities that the prototype should support.

Most participants (7) suggested including additional metadata in the resource descriptions:

“Discriminate the type of resource (test, presentation lesson plan...)” [User10],

“Include licence type and format with icons” [User3].

The prototype, in reality, already dynamically exploits all the metadata available in the dataset; unfortunately, not all metadata are specified for every resource.

Other participants (3) suggested including social and paradata to provide quality indicators:

“Educators should also include information about their personal experience” [User26]

“Perhaps some kind of mechanism that would allow feedback to be provided on how effective the OER actually was” [User17].

The possibility of selecting the language was the first priority for five Italian and Brazilian test-users:

“There should be resources in Italian” [User20].

This of course depends on the dataset, not on the prototype which is a data consumer.

Five participants indicated filtering functionalities:

“Filtering, for example by school type or age level” [User11].

This is perfectly consistent with the results from the task-taxonomy evaluation. Yet, of course, the objective of the prototype was to focus on innovative aspects not yet available in existing platforms.

Interestingly, a British test-user suggested to have the possibility of using explicitly learning objectives:

“It would be good if there were ‘a way in’ using learning objectives explicitly rather than key words” [User25].

A similar important observation, actually from the same participant, was discussed in a previous section.

Finally, two participants suggested to provide the possibility of working on search histories of previous years:

“Store and retrieve previous years search paths” [User1].

This is certainly a useful indication for a strongly domain-oriented feature supporting a WC-level task.

Relevance of the proposed scenario

Participants were asked to rate the relevance of the scenario proposed in the evaluation (ScenRelevance), that is 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 were unanimously positive, and fifteen of them appreciated again, in particular, the domain-oriented support for WC-level tasks:

“They correspond exactly to my search patterns” [User2],

“The scenarios correspond to my activities as a teacher” [User10],

“Because it fits the search workflow of teachers” [User13],

“It corresponds directly to my strategy when looking for resources” [User19].

Three of them, once again, explicitly mentioned personalization:

“They are relevant to the personalization of educational activities” [User8],

“Useful to identify resources suitable for the specific situation (personalization of educational activities)” [User9].

These outcomes are important because (as discussed in Chapter 3) establishing the relevance of the evaluation scenario is a fundamental aspect in the context of a DSR study.

Aspects educators liked most

Educators, asked about what they liked most about the prototype, indicated:

efficiency (12)

“The amount of work it can save to a teacher” [User4],

simplicity (9)

“Simple and intuitive approach” [User8],

and domain orientation, including support for WC-level tasks (9):

“Very interesting the classification by similarity for in-depth activities” [User7].

Three of them mentioned its support for exploratory oriented search activities:

*“It helps to focus / refine imprecise search objectives. It is like a Stargate to knowledge!”
[User15].*

Aspects educators disliked most

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

“Really nothing” [User15],

“Nothing, from what I can see it would be very useful in all its aspects” [User29].

Others (6) indicated some aspects of the interface, such as the use of colours, shapes, or the small fonts:

“Very rigid graphical appearance, too many rectangles” [User13],

“Text size too small” [User8].

This can be partially explained by the small screen of the tablet where the prototype was used, and by colours that were chosen to focus user's attention on specific aspects, with no intention to use them in an operational version.

Six participants lamented the exclusive use of the English language.

Finally, three participants indicated limitations in the dataset:

“At the moment it can only find a pretty limited number of resources” [Users22].

Additional comments and observations

Test-users were finally invited to leave additional comments, suggestions or critiques. Most of the comments (8) reiterated their enthusiastic feedback:

“It is really amazing” [User15],

“Great!” [User18].

Other comments reiterated previous observations, such as the need to support other languages (3).

Finally, three educators identified a minor bug: the prototype does not provide any message when resources are not found.

Field notes from the observation of participants using the system

There were some opportunities to observe directly a few participants while using the system.

Two participants experienced some initial difficulty to take full advantage of the multiple tabs used by the prototype. One of them however, after a few minutes, could use tabs very effectively to carry out multiple searches in parallel.

Another participant could easily use the system, but was annoyed by the same bug that was reported in the previous section: lack of notification when no resources are found.

In one case, the system was used to search material for a student: the teacher (math) was astonished to see the amount and quality of relevant resources that could be quickly identified, even if he (an Italian) was somewhat disappointed because all the material was in English or Spanish.

Qualitative data across different levels of experience

While there are no significant differences on the impact of participants teaching experience on quantitative data, as revealed by Kruskal-Wallis and Jonckheere-Terpstra tests, there are interesting differences emerging from the qualitative data. Indeed, experienced educators were more interested in exploiting the prototype for personalization:

*“Because it perfectly supports the need for personalization of the educational activities”
[User3, experience >20],*

“To organize personalized educational activities in the same class”

[User4, experience >20].

Less experienced participants were more interested in using the tool for initial lesson planning, to explore different strategies to propose a lesson:

“[...] most important points when you have to prepare a lesson”

[User23, experience <6],

“To find the best strategy to present a topic” [User22, experience <6].

This is partially reflected also in this explicit observation of an experienced educator:

“Very useful especially for young educators – personally I have now my consolidated material and I seldom need more” [User21, experience >20].

8.3.3 Discoverer versus Google

It is convenient, at this stage, to consider more explicitly a comparison of Discoverer with basic Google keywords search. As widely reported in the literature and confirmed by participants in this research, Google is the most frequent tool used by educators looking for OERs. 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 8.2). But when participants reported that Discoverer was very effective in reducing 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 search prototype. The actual performance of Discoverer against Google, in terms of number of resources identified, was less relevant, also because it mainly depends on the metadata available in the dataset used by Discoverer. Yet this aspect too, helped users to appreciate the advantages of the proposed strategy. The performance of Discoverer, which is identical to that of RepExp in this regard, was previously remarked, commenting on the large number of educational resources identified. Indeed, it was exactly the large number of resources identified by RepExp, which motivated the design of Discoverer.

As a first example, searching with the generic keyword “add” (Figure 8.6) 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 (Figure 8.7), returned 617 resources (Figure 8.10) following the expansion of the first one of them, and further 2854 resources following a second expansion (Figure 8.15). As another 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. This means they can be conveniently organized in clusters of different similarity, which can be explored and further expanded 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.

8.3.4 Evaluation conclusions: exploratory search and domain-oriented WC-level tasks

The first goal of this summative evaluation was to verify at what extent the new prototype overcomes the problems identified in the previous cycle, mainly linked to the large amount of excessively similar resources identified. The second general goal was to understand to what extent this last prototype properly addresses the needs of educators for a discovery and domain oriented search tool, ideally supporting them in their WC-level tasks.

An evaluation questionnaire was submitted as a structured interview or self-administered survey, to a sample of 29 educators from Italy, United Kingdom, and Brazil. Before completing the questionnaire, test-users were invited to use the prototype to carry out a few suggested tasks within a realistic scenario.

The quantitative analysis of the data collected via Likert-type scales, revealed a unanimous strong appreciation of the core features of Discoverer: the specialized search for educational material transparently starting from Google, the QBE expansion operation letting users identify similar resources starting from a sample resource, the similarity metric based on educational alignments (learning objectives), and clustering of results in expandable classes of similarity.

This level of appreciation was independent from most profile characteristics of the participants, such as sex, age, experience in education, or teaching subject. However, there were statistically significant variations across the experience teachers had with OERs. In particular, the more experienced the educators, 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.

The quantitative analysis identified also some differences across the nationality of educators. There was a slightly lower level of appreciation of the proposed metric by British educators, who considered additionally that it was slightly less difficult to use alternative techniques to obtain the same results of the prototype – compared to the opinion expressed by their Italian peers. A likely explanation, supported by some qualitative comments, is their different level of familiarity with formal learning objectives. However, both groups were generally extremely positive, and indicated they would enthusiastically recommend Discoverer to their colleagues.

The qualitative analysis clearly showed that Discoverer is strongly appreciated especially for its support to exploratory search, its strong domain orientation, and its support to WC-level tasks, such as lesson planning and personalized education. Some differences emerged across educators with different experience in education. Less experienced educators considered the prototype very useful for initial lesson planning and to explore different educational strategies. More experienced teachers tended to appreciate it most, for the possibility of personalizing education.

Evaluators also suggested to support multilingualism. The integration of quality information derived from paradata and social data was considered important by a few (3) educators, even if a considerable number of them (11) dissented. Additionally, coherently with the task-taxonomy evaluation, they suggested to integrate the functionalities of the prototype with filtering, which is already supported by some existing portals. Participants also suggested to support the possibility of working on previous year's searches, an additional very appropriate WC-level task. Finally, while the dataset used by the prototype was definitely sufficient to carry out this experimentation with a realistic scenario, a final implementation should ideally rely on a dataset with a larger coverage of the OERs available on the Internet.

Limitations

A first limitation of this evaluation is the number and representativeness of participants involved. However, the sample size is compliant with the indications from the literature for a qualitative analysis. Marshall et al. (2013) attempt to identify the ideal sample size for qualitative oriented studies, based on a secondary analysis of published research papers. They analyse the citation impact of the papers versus the samples size used. While they notice wide variations, influenced by cultural factors, they conclude recommending a sample size of 15 to 30. Fifteen was the minimum sample size of the papers with high impact, while 30 was the size limit after which the impact of papers did not show any further gain. The sample size used in this study is in the upper range of this recommendation. Concerning quantitative analysis, non-parametric statistics were used in addition to parametric statistics, and both yielded statistically significant results and meaningful effects size. Qualitative and quantitative data were mutually reinforcing, and coherent with the data collected in previous studies.

A second limitation concerns the representativeness of participants, who are mainly from the Italian higher education sector. While this was the main target of this research, the involvement of test-users from different Countries, even if in a limited number, was enlightening, as they came from contexts where formal learning objectives are commonly used, as well as from contexts where they are not.

Finally, the evaluation mainly collected reaction data from educators involved in the evaluation. While they reported that the prototype is very relevant and useful to support their high-level tasks, a more thorough evaluation could be conducted at work context level (Kules and Shneiderman, 2008) to measure its actual impact on their work.

8.4 Discoverer design and evaluation: conclusions

A final prototype, Discoverer, was designed, developed, and evaluated in this third DSR iteration. The goal of the prototype was to overcome, as far as possible, the challenges identified in the previous iteration, and to understand to what extent it could support educators in their WC-level tasks involving (exploratory) search of educational resources.

Discoverer, evolving from RepExp, replicated its original functionalities which, according to the preliminary evidence collected in the formative evaluations, were highly appreciated by educators. In particular, these include: the possibility of carrying out transparently specialized searches directly from Google SERP, the possibility of identifying resources similar to a sample resource, a similarity metric based on educational alignments. Discoverer, additionally, clusters results in three expandable classes of different degrees of similarity, to (1) offer users a quick overview of the available resources, (2) avoid locking them in the same group of resources in case of repeated expansion, and (3) let them chose the desired degree of similarity that better fits their WC-level tasks.

Clustering is carried out via the simple Jenks Natural Breaks Optimization technique, applied on the similarity vector computed by counting the educational alignments shared by the resources.

The summative evaluation of Discoverer was carried out by analysing quantitative and qualitative data collected from a sample of 29 educators, invited to carry out (autonomously or with the help of a demonstrator) the tasks proposed in the evaluation scenario. The feedback was collected by means of a questionnaire, self-administered as survey or submitted as structured interview.

The feedback collected shows that test-users were very positive about the core functionalities proposed by the prototype. The mode of all numerical scores collected via Likert-type scales, corresponded to the most positive possible value in all but one case. This was fully supported by a rich set of justifications, suggestions, and comments collected from test-users, in addition to the observation of a few of them using the prototype. These results, provided additional supporting evidence to confirm the findings of the previous formative evaluations, concerning research questions RQ4.1, RQ4.2, and RQ4.3.

The prototype was appreciated, more in particular, for (1) its strong domain-orientation, (2) its support for exploratory search, (3) its support for WC-level tasks (RQ4.5). In particular, young educators appreciated its support for lesson planning, while more experienced educators appreciated its support for specific planning of remediation, reinforcement, or in-depth activities for personalized instruction. Many of the participants spontaneously remarked that the prototype supports exactly the tasks they frequently need to carry out.

The analysis identified that the more participants were experienced in using OERs, the more they appreciated the metric based on educational alignments, and the difficulty of obtaining the same results with alternative tools. All participants were eager to recommend the prototype to their colleagues.

Finally, the evaluation also revealed opportunities to improve further the prototype. A few (3) educators proposed the integration of social and paradata to provide quality indicators, even if others (11) stated that they had limited or no interest – coherently with the findings from the task-taxonomy evaluation. A useful suggestion was to make it possible to re-elaborate previous years search sessions, to support educators in adapting their planning to necessary variations year by year. Coherently with the task analysis, some educators suggested the integration of filtering functionalities. And, as could be easily expected, educators wished to have even more metadata, and even more resources.

Figure 8.24 summarizes the findings related to this last DSR iteration, using the same conventions as in the previous two chapters. The box at the top reports the main research question driving the design and evaluation activities in this last DSR cycle. The yellow boxes at the centre indicate the key features of Discoverer (representing a form of design knowledge). Red boxes on the left report constraint knowledge, corresponding to the challenges addressed by the prototype. The green boxes correspond to the resulting new findings and the confirmation of previous preliminary findings with further supporting evidence.

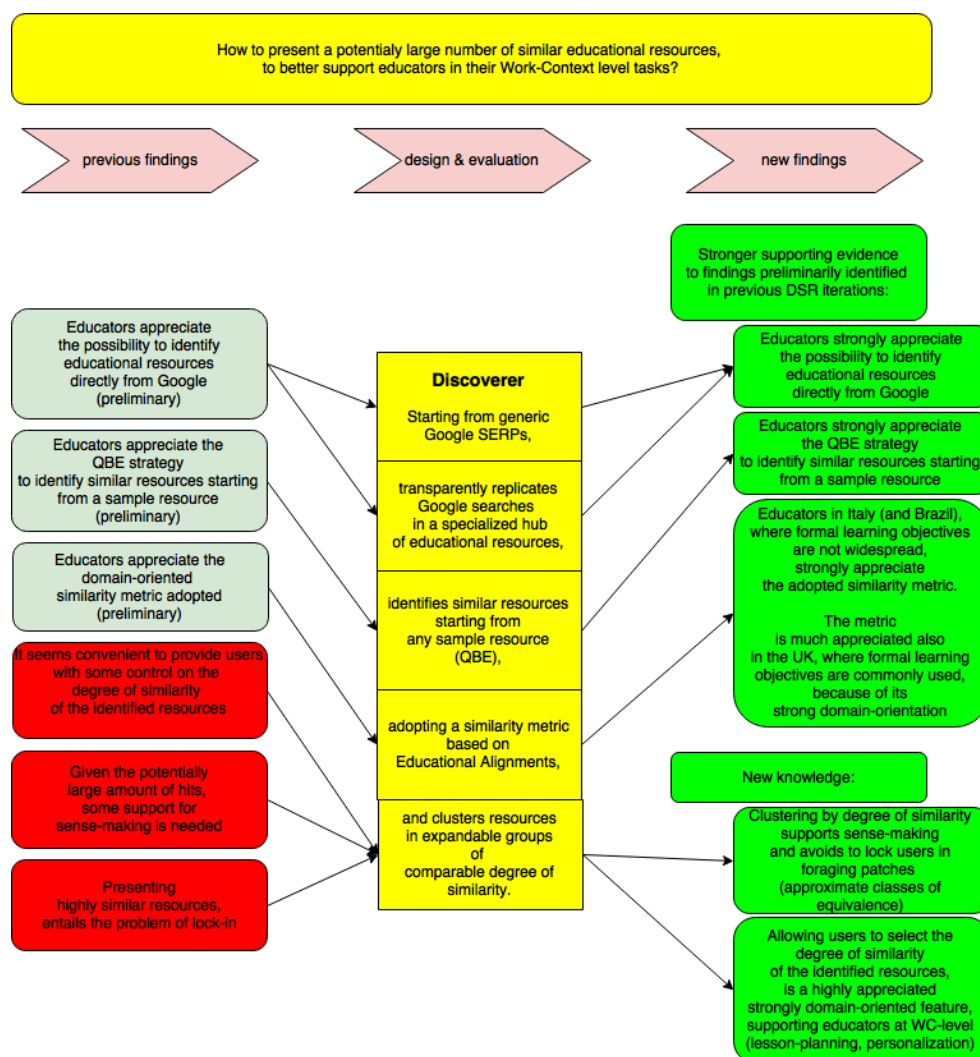


Figure 8.24 – Overview of the findings related to the design and evaluation of Discoverer.

9. Feedback from representative experts in the OER ecosystem

The previous research activities focused exclusively on educators, who were the targeted final users of Discoverer. As a complement, this additional evaluation activity aimed to collect further feedback from expert stakeholders more representative of the wider OER ecosystem. These participants were selected for having broader experience in the domain of OER use and provision.

This activity mainly contributed input to the research question:

RQ5: What barriers are restricting the deployment and uptake of OER discoverability tools and how can they be overcome?

as well as further input to evaluate the relevance of the requirements identified, and the suitability of the strategies proposed.

9.1 Method: Semi-structured interviews and thematic analysis

The objective of this activity was to collect and analyse feedback about the strategies proposed by Discoverer, its potential adoption in operational contexts, and, more in general, barriers and strategies to address related challenges in the OER ecosystem. This feedback led to suggestions for potential improvements and further research activities.

In order to cover as many dimensions as possible, participants were selected with purposeful sampling (Palinkas et al., 2013) from among OER experts and advocates who had been actively involved in shaping the OER ecosystem for several years, having different background and interests. Table 9.1 describes the participants: in order to broadly represent the various actors in the OER ecosystem, their profiles are very heterogeneous. For example, PartA had a strong background in Open Software development, and was responsible for the design, development, and deployment of an OER search engine; PartB was a leading advocate for the uptake of OERs, dealing more with policy rather than technical aspects; PartC was an experienced user and developer of open course material; PartG was the leader of a nascent standardization attempt with a strong librarian background. Their backgrounds are very synergetic: Tennant et al. (2020), for example, identify some relationships between the open source and the open scholarship movements (including OER) relevant in this context. These include fundamental goals like sharing, reuse and collaboration, important issues such as resource quality and reputation, and even shared concerns about potential conflicts with commercial players and cultural neo-colonialism.

In order to better exploit this broad set of competences, the collection of data was organized as semi-structured interviews (participants PartA, PartB, PartC, and PartD). This made it possible to orient the activity with some predefined questions, yet leaving, at the same time, maximum freedom for the

participants to address the aspects they considered most important according to their own experience, interests, and sensibility.

Participant	Role	Country
Interviewees		
PartA	Search engine designer with computing and Open Software development background.	US
PartB	Policy oriented OER evangelist and leader of open educational activities at global level.	US
PartC	Experienced university educator and manager of OER-based course material developers, with computer science background.	UK
PartD	Flipped learning evangelist with software engineering and Open Software development background.	CA
Further experts contacted via private forum and email		
PartE	Leader of LRMI and previous standardization activities.	UK
PartG	Leader of a new metadata standardization group with strong librarian background.	US
PartF	Developer and administrator of search platforms, computer science background.	DE

Table 9.1 Profile of expert stakeholders participating in this activity.

Additional feedback was collected through interactions with participants PartE and PartG, via the private Forum “CC Open Education Platform” (Creative Commons, 2017). This Forum is very much relevant in this context because it aims to discuss, among other things, the nascent search engine “CC Search” for open resources, being deployed by Creative Commons, and a new metadata standard for OERs. Finally, further feedback was collected through email interactions with an experienced designer and administrator of OER related platforms (PartF).

9.1.1 Semi-structured interviews

The first step was the preparation of supporting material for the interviews, including an outline of the questions, a consent form, and an accompanying participant information sheet. The Information sheet provided basic information about the activity: objectives, planned data collection through recorded interviews, data storage policy, confidentiality, rights of the participants. A favourable opinion was obtained from the Human Research Ethics Committee at The Open University (UK). A personalized invitation letter was subsequently sent to each participant, outlining the overall objectives while emphasizing an issue particularly relevant to their personal interests. The participants were invited to complete and return the consent form, agreeing on the planned collection and usage of the data.

The Discoverer prototype was used as discussion starter. Participants were invited to read a half page summary of the research, and to look at a sequence of commented screenshots and a brief screencast of Discoverer, before taking part in the interview. At the start of the interview, the participants were asked if there was anything else they liked to know about Discoverer and related research. A quick (three minutes) recap of the most important aspects followed: aim of the project, outcomes of the

preliminary task analysis, and key features of Discoverer. The interview continued probing for information along the following broad directions:

1. Specific feedback about Discoverer;
2. Uptake of Discoverer in an operational environment;
3. Sustainability of the OER ecosystem concerning discoverability and beyond;
4. Any other related issue not previously considered but deemed important to address.

These broad directions only served to impart an overall direction to the interviews, whose loose structure aimed to focus on the issues most related to the concerns and expertise of the interviewees. Therefore, the various topics were introduced with slightly personalized questions for each interviewee, following an examination of the relevant issues raised in their publications and professional profile, in order to stimulate their interest and elicit more relevant answers. For example, PartD was asked how Discoverer could be useful in his flipped classroom activities, PartB was solicited with questions related to OER macro challenges, and PartA was engaged by starting with aspects related with software engineering.

The interviews were conducted with each participant individually via videoconferencing. They lasted between 40 and 50 minutes, and were recorded and transcribed for subsequent analysis.

PartE, PartG, and PartF did not respond exactly to all the same questions as the participants in the interviews, but addressed only a subset of them. PartF, for example, mainly addressed the possible uptake of Discoverer in an existing operational environment, and his feedback was collected before the interviews took place. PartG only addressed issues related to metadata standards; her feedback was collected after the interviews, to further discuss issues emerged from them. PartE, on the contrary, had various opportunities, before and after the interviews, to both comment on the prototype and discuss about metadata standards.

9.1.2 Thematic qualitative analysis

The data collected through the interviews were analysed with thematic qualitative analysis. Qualitative content analysis was already used to analyse qualitative data in previous studies of this research. Yet there is a wide range of overlapping variations of “content analysis”, such as quantitative, qualitative, inductive, deductive, manifest, latent, and so on (Stemler, 2001). These variations have in common the fundamental technique of segmenting the data in chunks expressing single concepts, to be classified with a set of codes indicating recognizable meanings.

The terms codes and themes are sometimes used interchangeably; however, codes are usually structured hierarchically, and are aggregated in higher level codes that can be called “themes”. Codes and themes can be generated in different ways, always involving an iterative process of revision and validation during analysis (Ryan and Bernard, 2003). In general, Braun and Clarke (2006) propose that initial codes are generated after an initial review of the data. However, more specifically, Ryan and Bernard (2003, p. 88) observe that “the first pass at generating themes often comes from the

questions in an interview protocol”. In order to identify themes, Braun and Clarke (2006, p. 10) argue that “a theme captures something important about the data in relation to the research question”. Ryan and Bernard (2003, p. 87) suggest that “You know you have found a theme when you can answer the question, What is this expression an example of?”

In the content analysis previously carried out in this research, most codes could be anticipated from the narrowed focused prompts in the questionnaires, hence mainly following an a-priori process. In this case, the interviews were less structured, leaving more control of the discussion to the interviewees, aiming to better exploit their unique competences. This called for a more inductive (grounded) oriented approach, which targets “a process of coding the data without trying to fit it into a pre-existing coding frame, or the researcher’s analytic preconceptions” (Braun and Clarke, 2006, p. 12). This also lessened the risk of circularity in the interview prompts and undesirable researcher bias.

Figure 9.1 shows the preliminary map of codes and themes that could be envisaged from the guiding questions, hence closely reflecting them, before the interviews.

- | |
|---|
| <ol style="list-style-type: none"> 1. Discoverer evaluation; <ul style="list-style-type: none"> • Relevance of the problem; • Overall assessment; • Strengths; • Limitations; • Additional stakeholders potentially interested; • Suggestions for improvements and further research. 2. Discoverer uptake; <ul style="list-style-type: none"> • Integration in the interviewee’s workflow; • Challenges; • Suggested strategies. 3. OER sustainability; <ul style="list-style-type: none"> • Feedback on the strategies suggested in this research: <ul style="list-style-type: none"> ○ Reusable components (avoiding an unhealthy proliferation of platforms); ○ Open service-oriented infrastructure (fostering incremental synergetic contributions rather than competition); • Potential barriers. |
|---|

Figure 9.1: Preliminary map of codes and themes for the thematic qualitative analysis.

This map was only used as the starting point of the process, but – according to the methodology adopted – it was inductively and iteratively refined and validated from the data corpus during the analysis. While the first questions were more precisely related to the specific prototype, with narrowly pre-defined objectives, the follow-up questions were more open and exploratory. Therefore, the map was mainly modified especially concerning the last questions, where a more inductive oriented (grounded) approach could be applied.

Following the preliminary definition of codes and themes, the interviews were transcribed and the initial codes and themes were iteratively refined and validated. The final map of themes and codes,

resulting from the analysis, is reported in the next section. The same codemap was then used, and proved to be appropriate, to classify the feedback obtained from the additional participants PartE, PartF, and PartG. The final step discussed the findings in relation to the research objectives, supported by compelling extracts from the whole data corpus, including the interviews transcripts as well as the additional feedback from the last three participants.

9.2 Results

9.2.1 Final coding map

Following the procedure described in the methods section (9.1), the data corpus from the transcribed interviews and the additional feedback was iteratively coded. In this process, the initial map shown in Figure 9.1 was inductively refined to the new map in Figure 9.2.

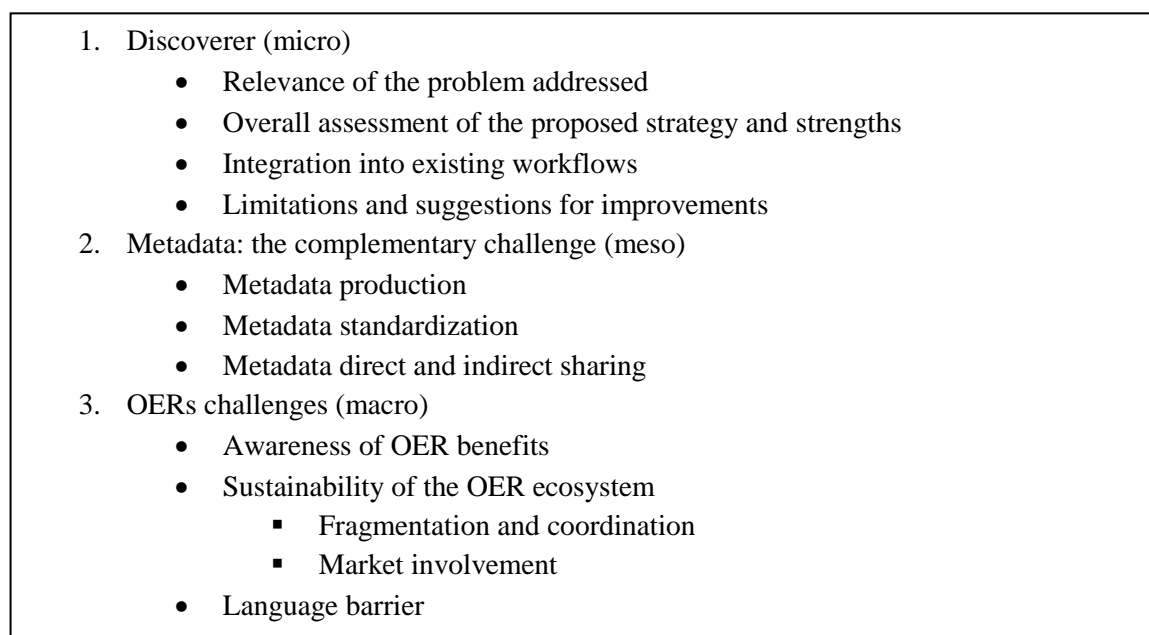


Figure 9.2: Final map of themes and codes for the thematic qualitative analysis.

Some of the codes in the preliminary map were merged. For example, “Limitations” and “Suggestions for improvements” were merged because participants preferred to implicitly address the limitations as explicit suggestions for improvements. Similarly, “Overall assessment of the proposed strategy” and “Strengths” were merged because participants provided a few generic overall comments, but contributed more useful specific comments and related motivations. The “Suggestions for further research” were integrated in the other themes as needed.

While the intention was to identify themes and codes as mutually independent as possible, there were significant overlaps among them. For example, the issue of fragmentation surfaced in more than one theme: in addition to “OERs challenges”, at a macro level, it also surfaced in “Overall assessment of the proposed strategy and strengths”, concerning fragmentation of search engines, as well as in “Metadata standardization” concerning fragmentation of metadata standards. Yet, the map was structured hierarchically, with the objective to improve its understandability. To this end, some

chunks of information from the data corpus which related to different themes were annotated with multiple codes.

A criterion to structure the map hierarchically, has been the level of abstraction of the codes and themes. “Suggestions for improvements”, for example, is more directly related to the micro-level of “Discoverer”, while “Sustainability of the OER ecosystem” is more related to the higher macro-level of “OERs challenges”.

9.2.2 Coding and data extracts examples

In this section, a representative verbatim (data extract) example is reported, for each theme and code.

Theme 1: Discoverer (micro)

The first theme concerned very directly Discoverer: the relevance of the problem addressed, Discoverer’s strengths, limitations, potential improvements, and integration into existing workflows.

Relevance of the problem addressed

Discoverability is the main challenge: finding resources, and good quality resources...

[PartA]

Overall assessment of the proposed strategy and strengths

First and foremost, I really like what you have done. [PartE]

Good idea, it is an interesting concept of similarity. [PartC]

Building on top of already existing workflow of an educator is key. [PartD]

Integration into existing workflows

In addition to teachers, it could benefit learners, especially those struggling with their studies: they could use it to find additional complementary learning material. [PartC]

Limitations and suggestions for improvements

Not providing explanations about its results: some of the users, once they become more experienced and wish to be even more efficient, will likely require this. [PartA]

Indicate, in simple terms, the level of openness of the resources identified. [PartC]

Theme 2: Metadata (meso)

At a higher abstraction level, the participants argued that metadata represent a fundamental challenge to discoverability, and mentioned in particular sub-challenges concerning production, standardization, as well as direct or indirect sharing. Direct sharing is intended here to mean the sharing of metadata, while indirect sharing designates the mutual federation of search services.

Metadata production

Nobody is going to go through the pain of (re)adding metadata to existing material. From my experience as an OA repository administrator, the barrier to add even the most basic metadata when initially publishing a creative work is huge. [PartF]

Metadata standardization

First, have people understand the need for a common schema. [PartA]

The big question is how matchable metadata is across all of those disparate sources. [PartF]

Metadata direct and indirect sharing

In fact, rather than being a listing service, for individual titles, to essentially become a listing service for repositories [PartA]

Theme 3: OERs challenges (macro)

Finally, participants argued about high level challenges for OERs, which were classified with the following codes.

Awareness of OER benefits

Research shows that OERs save student's money, increase student success, reduce dropout rate, increase course completion rates, reduce time to degree [PartB]

Sustainability of the OER ecosystem

Fragmentation and coordination

Some of those people need to pull back [from individual efforts] and put some of their efforts into starting pulling things together [PartA]

Market involvement

There are things that the State is not very good at, and there is room for the private sector [PartB]

Language barrier

The language barrier is definitely a big problem. [PartD]

9.3 Discussion

In this section, documenting the last step in the process outlined in the methods section, each theme and sub-theme is discussed in relation to the objectives of this research. This activity mainly focused on Discoverer and its potential uptake in an operational environment, but explored its wider context too.

9.3.1 Theme 1: Discoverer (micro)

Participants were first invited to examine screenshots and screencasts of Discoverer, and its key features were also briefly summarized at the beginning of the interviews. Then, the participants offered their views about Discoverer, comments about its limitations and suggestions for its improvement, as well as considerations about its possible uptake.

Relevance of the problem addressed

Coherently with the feedback from educators in the previous studies and the literature (UNESCO, 2017), participants considered the problem of discoverability as a very relevant challenge in the OER ecosystem:

OERs discoverability is definitely a relevant challenge [PartB]

PartA, having a background in search engines, initially claimed that discoverability was *the* main challenge, quickly acknowledging, however, that quality too was a fundamental issue, and concluding that discoverability was *one* of a few main challenges:

*Discoverability is the main challenge: finding resources... and good quality resources...
maybe it is one of the two or three main challenges [PartA]*

Quality of the resources is indeed another major challenge, recurring under different themes, and will be addressed under “Limitations and suggestions for improvements”.

Overall assessment of the proposed strategy and strengths

Participants, independently from their different background, were positive about the prototype. For example:

The prototype is excellent! This is a wonderful work, it is really good. I would definitely recommend it to colleagues. [PartD]

First and foremost, I really like what you have done. [PartE]

I love your approach and I think that it is a brilliant approach. [PartA]

An interviewee did not comment on it:

I don't know it enough to comment on it. [PartB]

but he added a generic positive remark:

Any tool improving the discoverability of OERs is going to be useful. [PartB]

This is actually debatable, because there is the risk that additional tools in general, such as one more search platform, might just worsen the situation in the OER ecosystem, by increasing its fragmentation. PartC pointed out, indeed, that this proliferation of platforms makes it more difficult to propose alternatives to Google:

[Other challenges are] awareness in a highly fragmented context, habits: people are used to Google, it takes effort to change habits [PartC]

In this regard, PartD praised the strategy to develop Discoverer as a reusable component on top of existing search engines, exactly with the objective to avoid this fragmentation:

You don't want to build another one [search platform], [a reusable component to reduce fragmentation is] very good [...] that is excellent! [PartD]

Along the same lines, PartA pointed out:

At some point you have too many people stirring the pot and at some time you need to be an executive chef who is sort of unifying things rather than dividing things [PartA]"

This challenge of fragmentation will be further discussed more in general under the broader theme “OERs challenges”.

Participants expressed appreciation for specific features of the prototype, such as its exploratory orientation:

Brilliant approach, really, really attractive, I love it, nice exploratory approach. [PartA]

its domain-oriented similarity metric:

Good idea, it is an interesting concept of similarity. [PartC]

or its simple interface:

I think it would be easy for them to use this tool. The interface looks easy. [PartD]

They appreciated its QBE approach, that is the possibility to issue a query starting from any sample resource:

And I really love that idea, that you have, and looking at a resource just some random web page somewhere on the internet, and I can click that button and it is going to feedback into this loop. [PartD]

as well as from any snippet in a standard Google SERP:

They are going to use a web browser, they are going to use Google: having either filters or plugins that are right on top of the search queries from Google, is going to be key. [PartD]

They also praised the importance of addressing the work context of educators:

Building on top of already existing workflow of an educator is key. [PartD]

Integration into existing workflows

While the strategy to explicitly target educators' tasks was appreciated, the interviewees suggested other contexts where the prototype could be equally useful. In addition to educators, they suggested that it could benefit stakeholders such as learners or course developers.

Coherently with the previous study, an interviewee suggested that students could conveniently use Discoverer to find remediation material:

In addition to teachers, it could benefit learners, especially those struggling with their studies: they could use it to find additional complementary learning material. [PartC]

Another interviewee, a flipped classroom evangelist, considered that Discoverer could be conveniently used by learners in that context:

I put the discovery of resources more in the hands of my [flipped classrooms] students, [...] putting that in the students hands, that would be very useful, that is wonderful.
[PartD]

Finally it was suggested that Discoverer could be useful for course material developers:

It could benefit groups of course material developers too. [PartC]

I could use it to [...] help members in my courses development team finding material to adapt and integrate in new courses. [PartC]

Limitations and suggestions for improvements

Participants mentioned explicitly only two “limitations” of Discoverer, preferring to address other limitations implicitly, by offering “suggestions for improvements”. Hence the two separate codes anticipated in the preliminary coding map were merged.

The first limitation explicitly mentioned, at micro-level, was the lack of explanation capabilities:

Not providing explanations about its results. [PartA]

In reality, explanations were not included on purpose. Indeed, explanations would necessarily expose the formal learning objectives to the users, while all but one educators previously consulted, preferred them to be invisible. This suggestion could be partially explained because the interviewee, like the only educator who suggested to make the formal learning objectives explicit, was comfortable with their use. Yet the interviewee, reacting to my justification for the lack of explanation capabilities, argued that:

Some of the users, once they become more experienced and wish to be even more efficient, will likely require this. [PartA]

This is indeed a helpful observation: the preferences and behaviour of some of the educators, might well change in response to the use of the prototype over a long period (Kules and Shneiderman,

2008). This effect could be exposed by a longitudinal evaluation of the prototype – which is a good suggestion for a further research activity.

Providing explanations was considered very important for this interviewee, who reinforced the idea formulating it also as a suggestion for improvements. In this case, possibly because of the previous discussion, he added a detail:

Provide concise but effective explanations on demand. [PartA]

Letting users control this functionality looks like a good compromise, because users with more familiarity with formal learning objectives might turn this feature on, while it could remain invisible to other users.

The second limitation explicitly indicated, was the dependence of the prototype on a specific metadata standard:

The weakness of the process is its dependence on LRMI [PartA]

The interviewee argued that, on the contrary, his approach (he developed a search engine), even if inelegant, was independent of any particular metadata standard:

We of course took the opposite approach, we just said we are going to assume no commonality, we are going to... our approach was not elegant in any way, but recognizing that there is not any elegance right now in the environment so therefore the approach probably isn't going to be very elegant, is going to be very brute force. For each repository there is going to be a different metadata structure there, and we were just going to have to deal with that, and we were willing to deal with that. So, two diverging approaches that reflect where we are. [PartA]

Discoverer, in reality, is partially independent from LRMI, because it retrieves the educational alignments (learning objectives) exploiting the so called “data services” of the Learning Registry. These data services can extract the learning objectives from metadata expressed with LRMI, as well as with other standards. This comment, however, shows that a commonly accepted metadata standard, despite LRMI being a recent development backed by Google, is still an open challenge. This will be discussed under the following theme, directly related to metadata, in Section 9.3.2.

The participants in this activity, given their long involvement with open material, were definitely more aware of the importance of openness for OERs, compared to the educators involved in the previous studies. PartC, experienced in leading teams of course developers exploiting open material, argued about the need to provide easy to understand information about the level of openness of the resources:

Indicate, in simple terms, the level of openness of the resources identified. [PartC]

Support people in quickly understanding the licensing constraints. [PartC]

A similar concern for licencing specifications to be expressed in simple terms, was indicated by PartB too:

We are thinking about possible new definitions of OER, trying to keep them simple enough and legally exhaustive. [PartB]

This concern was echoed by an educator involved in the evaluation of RepExp (Section 7.3.3), who suggested to highlight the OERs in the results, in order to discriminate them from other educational resources. The opportunity to provide simple information to clarify licencing constraints is also frequently mentioned in the literature (Anderson and Leachman, 2019). Yet Discoverer simply displays the metadata available, that may or may not include information about the level of openness. Hence the prototype could be improved, for example, by mapping these metadata to a more immediately evident visual representation of the level of openness (reusable, remixable, etc.).

Another suggestion concerned the need to provide information about the quality of the resources:

Integrate quality information. [PartA]

This concern was reinforced by another comment, previously reported, where quality was indicated as a relevant issue:

Discoverability is the main challenge: finding resources... and good quality resources... [PartA]

While the prototype, at the moment, does not display any information about the quality of the resources, it is acknowledged that quality is critical to the success of OERs repositories (Clements, 2016) and the ecosystem in general (UNESCO, 2017), and is definitely worth of further research.

Other suggestions concerned more specific aspects. These include adding filters:

Integrate the possibility to filter on text complexity. [PartC]

This can be implemented, provided that the related metadata are available in the search platform which Discoverer integrates. In case these were not available, it is still possible to compute some readability indexes, such as the Gunning Fog index or the Automated Readability Index, as approximated indicators.

Another suggestion concerned the use of different similarity metrics:

Address different meanings of similarity. [PartC]

This possibility, indeed, was foreseen and discussed since the early phases of this research.

A bilingual interviewee, working in a non-English speaking country, and fully conscious of the language barrier, suggested to integrate a translation service:

If we can help with translations, and even using auto-translations of some of these resources into the native language of the searcher, that may be helpful as well. [PartD]

Machine translation is, of course, a research field in its own. However, even an approximate translation, that can be implemented using existing services, could be acceptable for certain use cases:

You are not necessarily going to use those resources you found, you don't need to use a complete port to help your craft, but it can give you an inspiration. [PartD]

Indeed, this is a common use case for OERs: de los Arcos et al. (2016) report that 80.5% of the educators they interviewed, used OERs to get inspiration and new ideas. Language issues, being a broad challenge for OERs, will be also discussed under the third theme.

PartD, having User Experience expertise, recommended to make the prototype easy to install:

Make sure it is easy to install. Package it up so that it is really easy to use, definitely the highest priority to make it really easy. [PartD]

Indeed, the current version of Discoverer, being a prototype, is not particularly easy to install. This problem was highlighted already by a few test-users involved in previous evaluations, but could be conveniently fixed in an operational version.

Finally, it was suggested to use the search history to improve the relevance of the queries:

Are the search history to change the queries at all into your database backend? [PartD]

Indeed, the efficiency of the queries could be improved, taking into account previous searches or other information (e.g. the user profile). This information could be employed to disambiguate terms and prioritize the results most relevant for the users, for instance.

9.3.2 Theme 2: Metadata (meso)

Discoverer is totally dependent from the metadata which power it. While it demonstrates a possible use of metadata, a complementary challenge is how to get them in the first place. Participants discussed, in particular, issues related to metadata production, Metadata direct and indirect sharing, as well as to the definition of a common standard schema.

Metadata production

Discoverer focused on metadata consumption, but a complementary challenge which participants were well aware of, is metadata production:

The big challenge is metadata [PartA]

The problem is that we have all these resources with no metadata, how do we convince people, or help them, to create the metadata for the objects that they already have?

[PartD]

It is well known in the literature that authors tend to avoid to properly document their resources with rich metadata (Doctorow, 2001) and that manual curation is very resource hungry. This was echoed by the comments of the participants:

Nobody is going to go through the pain of (re)adding metadata to existing material. From my experience as an OA repository administrator, the barrier to add even the most basic metadata when initially publishing a creative work is huge. [PartF]

This is particularly evident in the case of “little” OER (Weller, 2010) produced by educators in their own time:

None of us are going to tag them properly, none of us are going to put learning objectives on this, we are just like throwing them up there and say hey, you can use it if you want, but... [PartD]

This gets to the heart of questions around whether an educational resource can be really open if it is not discoverable. It was suggested that including metadata in a revised definition of the term OER, to elevate the importance of metadata to the level of content, could foreground this. There were enthusiastic reactions from participants more directly involved in OER platforms:

[...] the very idea that the metadata needs to be elevated to equal prominence with the content itself, and I had not considered that, and I think that is really insightful. I urge you to keep pushing that idea. [PartA]

Very good point! [PartF]

Yet participants with a different background, though acknowledging the importance of metadata, were concerned that making metadata mandatorily required might disincentive potential authors to make their material openly available:

To be a proper OER you need to be accessible as well [...] but it has a danger that someone says well, then I just won't create the object itself because I can't do it. If I need to do it fully, then I just won't do it, and so there is a danger there when you make too many requirements, people just drop it. [PartD]

The concern to keep an updated definition of OER as simple as possible, was also voiced by PartB who, considering his background and role, was more concerned about legal aspects:

We do recognize the fundamental importance of metadata. And we are also thinking about possible new definitions of OER, trying to keep them simple enough and legally exhaustive. But the definition should be as easy as possible and we are more concerned about legal aspects. [PartB]

The challenge to request authors to provide metadata is acknowledged in the literature. For example, one of the experts interviewed by Atenas et al. (2014) argued that “every additional field makes it less likely OER will be deposited and released”.

As discussed in the literature review, there are three main broad techniques to obtain metadata: manual by human experts with a very time-consuming curation process; fully automatic with the use of artificial intelligence (AI); and via crowdsourcing. Coherently, the interviewees mentioned both automatic generation and crowdsourcing as alternatives to the creation of OER metadata directly by authors or curators. PartA, who had background in AI techniques, questioned:

Can we automate this? Can we apply AI to discover attributes of resources and put them into a canonical metadata form? [PartA]

PartD, who was more experienced in crowdsourcing open software, questioned:

The other thing [...] is: how do we crowdsource some of this? [PartD]

In this regard, the need for a robust trust mechanism was also highlighted:

People can just poison the data: reputation-based, communities of trust are really, really important [PartD]

Considering that metadata are annotations about educational material available on the Internet, PartD suggested to use a generic Web social annotator as a possible solution to let people annotate resources, and later collect the annotations as crowdsourced metadata:

Creating information about the web site I am viewing, but that data are stored on a hypothes.is server, such that we can map that on top of the other thing. Maybe there is a chance for that to do it in a crowdsourcing type of environment, maybe. [PartD]

This challenge is certainly a promising area for further research:

There is obviously some future excellent research for someone there, how to make that easier, how to make it happen. [PartD]

and will be followed-up in the next chapter.

Metadata standardization

In order to be usable, the metadata should comply with some kind of common schema (a “standard”), as the participants were fully aware of:

The big question is how matchable metadata is across all of those disparate sources.

[PartF]

You need to have some kind of common schema or API that you can pull information from [PartD]

However PartA considered that this awareness was not yet widespread among common people:

First, have people understand the need for a common schema. We are not there yet! [...]

Then, but second order, select the best schema. [PartA]

Discoverer was standard-agnostic, exploiting whatever metadata were available through the Learning Registry API. However, schema.org/LRMI was considered the reference model in this research. This recent standard was considered appropriate by PartA:

It should be LRMI because that is the most common, because it is Google, and you know, schema.org. [PartA]

but not by PartB:

We initially planned to follow LRMI/schema.org, but were “disappointed” that it did not meet the initial expectations. Hence we set-up a group to identify OER metadata standards for discovery, calling upon the expertise of librarians and OER search engines administrators. [PartB]

It is evident that current standardization activities, so far, have not yet solved the problem of a single commonly accepted metadata standard. And, as previously observed in this thesis (Section 2.2), a single metadata standard might not even be the best solution. However, considering the large number of existing metadata standards, developing one more standards poses the risk of further fragmenting the solution space, making it harder to solve the problem at a global scale. Improving LRMI might be a better solution than developing one more standard. With this aim, the limitations of LRMI should be precisely identified, in order to propose solutions overcoming them: this way an existing standard could be improved while avoiding to fragment even further the current solution space.

Following this line of thought, a thread in the Creative Commons Open Education Platform was opened to discuss these issues:

I assume nobody wants to develop one more standard. If necessary, an existing one might be adapted – for example through the extension mechanism foreseen in schema.org, in the case of LRMI. [Cortinovis]

I also remarked that OERs raise specific use cases, such as the need to track the multiple versions of an OER, which are continuously revised, remixed, and republished. However, schema.org/LRMI does take into account some of them already:

Certainly, OERs rise specific use cases for metadata, which may not be shared by “traditional” resources. An example is the need to track the multiple versions of an OER, which is continuously revised, remixed, and redistributed (for example in the context of “OER-enabled pedagogy” activities). Incidentally, however, schema.org/LRMI does have some properties that could be used for that, such as isBasedOn, hasPart, or sameAs.
[Cortinovis]

It was argued, in particular, that a potential new CC standard would have to be adopted first by its federated repositories, otherwise CC Search could not collect automatically the required metadata from them, as intended:

CC would need to have “its standard” adopted first by the repositories it federates, otherwise it could not collect the required metadata from them ... [Cortinovis]

Among the replies, PartE, a key player in previous metadata standardization activities, confirmed that LRMI was designed taking OERs into account:

It was designed to deal with the OER special cases, e.g. license and reuse information, as much as that can be done through metadata rather than other means. [PartE]

He also supported the suggestion to modify or extend it:

Putting together a subset of schema.org along with any specific extensions that go beyond schema.org (e.g. vocabularies for some of the properties) as an application profile is indeed something that might be valuable for a known service such as CC Search, and may even be useful in the abstract. So, I don't have much to add to what
Renato says [...] [PartE]

However PartG justified the intention of the OER Metadata Group to develop a new standard as follows:

The primary intent is to leverage librarian expertise in metadata strategies, which so far has not been applied to OER metadata [PartG]

In reality, previous standardization activities have been influenced by librarian expertise. For example, Stuart Sutton, managing director of the Dublin Core Metadata Initiative for many years, where the LRMI activities were transferred in 2014, has a strong expertise in Library and Information Studies. Yet, one of the goals of previous standardization activities was to make any standard simple enough to be used by OER authors, curators, or users, who cannot be expected to have professional librarian expertise (Barker and Campbell, 2016b). The alternative to have professional librarians

cataloguing the growing number of existing OERs, which keep being revised, remixed, and republished all the time, was considered unrealistic. Was this a mistake?

Certainly PartG, the leader of this new standardization group, was not aware of this goal to make a standard usable by average users:

Interesting. Since I wasn't a part of the non-library metadata initiatives I wasn't aware of the goal of lowest-common-denominator metadata for the average person. [...] Thanks for your comment and it explains a lot of the push back by non-librarians I've gotten! I will be able to speak to their concerns much better now. [PartG]

It was suggested that a potential positive aspect of this new proposed strategy, is that it has some potential to improve one of the current limitations of metadata, that is, their current scarce quality (Dietze et al., 2017):

Certainly that [the previous goal] will never lead to robust indexing. One of the big issues there is lack of standardized publishing models. I would never expect extensive bibliographic control methodologies to become the norm, so this is where I see potential for library collections (and library publishing presses) to have room for growth. [PartG]

Yet, these aspects are worth of further discussion in the community, so that we are all aware and fully understand the implications of these different directions. A possible solution is to develop these two strategies so that they can work synergistically. Unfortunately, for the moment, I was asked by PartG to hold-off the discussion, while they organize their leadership team.

In case one more standard is really going to be developed, I suggested to design from the start, suitable mappings/heuristics to automatically “translate” metadata between this standard and other standards (LRMI, LOM, DCMI, Learning Tools Interoperability (LTI) Resource Search, ISO/IEC_19788...). As a second step, a CC tool could be crowdsourced as open software, to perform these “translations”. This strategy could reduce the inconvenient due to this unhealthy proliferation of standards which keeps fragmenting the solution space. Here again there is room for further research activities.

Metadata direct (index-time) and indirect (query-time) sharing

The current OER ecosystem is characterized by the availability of too many OER search engines, which make it unfeasible for educators to hop from one to the other. This is one of the reasons why educators end up using Google, despite a clear awareness of its limitations for finding educational resources. A possible solution to this problem, is to provide a centralized access point, so that users could search contemporarily in all the federated search engines, with a single query. This requires some mechanism to share metadata. Sharing metadata among OER content providers and search tools is one of the recommendations from the Ljubljana Action Plan (UNESCO, 2017).

As technical approach to support this, a single federating search engine could replicate searches and collate results from multiple federated search engines, directly at query-time. These federated engines would offer their search functionalities through standardized API, allowing the federating engine to aggregate their metadata indirectly. An alternative solution is to federate metadata directly. In this case, the centralized search engine would first harvest the metadata from the platforms to be integrated, preferably again via suitable API. In this case federation would be performed at index-time. This centralized engine could then provide an integrated search service in all the collated metadata to its users.

In the context of the evaluation activities of Discoverer, an OER platform was identified, which provided to its users a curated list of about a hundred links to independent OER search engines. Users were supposed to consult the list and independently visit the search engines of their interest, manually replicating their searches in each one. Considering that the Learning Registry was already harvesting the metadata from a number of the search engines in that long list, I proposed to replace the existing curated list with a single federating engine based on the Learning Registry and Discoverer. With this solution, users could have conveniently searched through the linked engines with a single query. A technical solution was proposed, involving a minimal technical effort. Search engines not yet providing metadata to the Learning Registry could have been encouraged to do so, joining their individual efforts in a collective endeavour. Unfortunately, this proposal was dismissed by the repository administrator, because of different priorities:

Seems legit, although this should be considered a fairly big shift of focus [PartF]

Offering users a professionally curated list of search engines to explore, appears a suitable solution at an initial stage, but may not be attractive enough when it contains a large number of engines. This solution is very much similar to the manually curated list of topics that were popular in the early days of the Web, but which were soon replaced by search engines like Google.

PartB agreed, indeed, that

There are too many search platforms, and we can't expect people to hop from one to the other [PartB]

A more attractive solution, equivalent to the one proposed in this thesis, is going to be provided by CC Search, which aims to federate major search engines via suitable API:

Our solution is to provide a "one stop" that aggregates search functionalities from other platforms via "live API". [PartB]

PartA too, acknowledged the need to federate existing metadata repositories:

In fact, rather than being a listing service, for individual titles, to essentially become a listing service for repositories [PartA]

and the fundamental role of Web API to make this technically possible:

[...] believe it or not, does not have an API! The leader in open resources does not have an API: they work out of a spreadsheet, you know, that is their registry essentially. I just... that continues to boggle my mind [PartA]

Having a strong computer science background and experience in search engines design, PartA expressed astonishment that key players in the field were still transferring metadata through simple spreadsheets, rather than using API.

In addition to dealing with technical aspects, it is important to consider obstacles of a different nature, which may actually prove to be more complex challenges. In particular, the mechanism must be designed to overcome potential resistance to share (directly or indirectly) the metadata:

However some people are “territorial” with their material [...] possibly because they need to use local languages too [PartB]

Therefore, some incentives for the federated search engines to provide access to their services must be explicitly planned. In this regard, CC Search plans to reroute interested users back to the federated search platforms:

Our system then reroutes users to the original search platform, so that search platforms administrators are incentivized to provide us their live APIs [PartB]

This is indeed an excellent idea, but which should work in the opposite direction too. That is, CC Search should, in turn, provide its search services to other applications. Providing API access to their search services, could help creating a fertile common infrastructure, which third parties could exploit by providing advanced services on top of existing ones, such as the search by similarity offered by Discoverer. Therefore, it was recommended in the CC forum that CC Search makes available in turn, its search services for open material via API.

My suggestion is that you, in turn, provide open Web API to allow third parties to programmatically access your own services. This would let them provide educational services on top of yours – enriching your services rather than competing with them (developing additional separate platforms would just worsen the situation fragmenting it further). [Cortinovis]

In addition to positive comments from other participants, PartB wrote:

Good idea - and Yes - CC does plan to offer an API to CC Search so others can build on top of the data. [PartB]

Another participant privately suggested a new standard, explicitly designed to implement these services, specifically customized for OERs (IMS, 2019). All this is encouraging, because providing open search API will make it possible to integrate Discoverer and similar applications on top of the

basic search services provided by CC Search. These applications will have the opportunity to enrich these basic services, offering more advanced functionalities such as – in the case of Discoverer – higher-level pedagogically-oriented search services.

9.3.3 Theme 3: OERs challenges (macro)

While the discoverability of resources is unanimously considered a challenge in the OER ecosystem, other fundamental challenges were mentioned, which help to collocate the focused activities of this research in their broader context.

Awareness of OER benefits

PartC considered, based on his own experience managing course developers, that it is still necessary to motivate the use of OERs, and to find convincing arguments about their advantages:

You need to make it clear what the value of sharing is. [PartC]

PartA, from his perspective as a search engine designer, considered instead that this goal was already achieved:

OERs have mainstreamed. Educators have been on defence about open resources but now they say “I see the value of this” ... [PartA]

and that the challenge is now rather their discoverability:

but discoverability has not kept pace with the number of resources [PartA]

Both PartC and PartA, however, despite their different perspective, noted that the lack of a price attached to OERs, might be psychologically associated to a lack of value:

Perhaps, in certain contexts, OERs are perceived to have no value, as they have no price attached. [PartC]

Quality has always been an issue with open resources because you know in the US there is an old saying “you get what you pay for”, and so people everywhere typically believe that something free is probably not as good as something you pay for. [PartA]

PartB argued that there is now plenty of evidence, in the literature, demonstrating the advantages of OERs against traditional resources, not just because they are cheaper, but because they can help students achieve better results:

Research shows that OERs save student’s money, increase student success, reduce dropout rate, increase course completion rates, reduce time to degree [PartB]

However, as demonstrated by the living experience of PartC and PartA, this has not always, as yet, sank in the commonly established opinion – which is the reason why the Ljubljana Action Plan still recommends to raise awareness about the benefits of OERs.

Indeed, even PartB himself mentioned, among other challenges for OERs, the lack of awareness:

A lack of awareness about OERs and Open Education among educators [PartB]

which requires educators to be supported:

The need to support busy teachers and faculty to use and create open content [PartB]

Use of open formats

Finally, PartB mentioned another fundamental limitation, concerning the need to work in a common environment and with open formats:

*We need a space where we can collectively build, share, modify and update open content,
with open formats [PartB]*

Indeed, openly licensed content is not always available in open formats that other educators can easily repurpose which limits its effectiveness.

Sustainability of the OER ecosystem

The well-known challenge of sustainability of the OER ecosystem was clearly acknowledged:

The lack of sustainability of existing activities is a serious evident challenge [PartC]

Various possible reasons causing this situation were mentioned:

*[The lack of sustainability] is possibly due to short term funding, lack of coordination,
short-term planning and lack of thinking about sustainability, competition, lack of
awareness of the advantages... [PartC]*

PartA, considering its experience in developing a small-scale search engine, focused on the problem of scale:

Small initiatives at this stage cannot have a relevant impact [PartA]

*May be some of those people need to pull back and put some of their (individual) efforts
into starting pulling things together [PartA]*

This excessive proliferation of small-scale activities, indeed, was certainly useful to the expansion of the open movement, but it may be now partly responsible for an unhealthy fragmentation of the ecosystem. This exacerbates the challenge of sustainability, for which no one seems to have a solution yet (UNESCO, 2017).

Fragmentation and coordination

Fragmentation was already discussed, at micro level, under the theme “Overall assessment of the proposed strategy and strengths”, where participants praised the strategy to offer high-level reusable functionalities on top of other search engines. Fragmentation was partially discussed at meso level under the “Metadata standardization” theme too, where the need to refrain from developing too many

competitive standards was discussed. Finally, fragmentation was discussed under the theme “Metadata direct and indirect sharing”, where it was argued about the need to avoid an unhealthy proliferation of platforms, and provide an open service-oriented infrastructure fostering collaboration rather than competition.

At macro level, participants indicated various solutions that could help reducing the problem of fragmentation, increasing the sustainability of the ecosystem. The need for more coordination was frequently mentioned:

Some of those people need to pull back [from individual efforts] and put some of their efforts into starting pulling things together [PartA]

Funders of developers of OERs should require coordination efforts, among institutions and individuals [PartC]

There is a need for organization, international agreements, and well-funded key people in the open-ed community [PartA]

You need a coordination of some kind [PartC]

PartA argued that it is possible, indeed, to achieve sustainability by suitable coordination of the activities, because this was already successfully demonstrated by the Open Software communities:

Coordination of the multiple initiatives is possible. Indeed, while the top level coordinating structure is the challenge, there are examples of success in the open software, where large complex systems have been crowdsourced. [PartA]

Market involvement

While competition was mentioned as something that should be avoided,

Need to avoid competition [PartC]

the involvement of the private sector was not excluded:

I have been heavily involved in the open source movement since the nineties, and there are a lot of parallels there that if we close off commercial interest, then it is not truly open either [PartD]

On the contrary, some participants considered that there is an opportunity for collaboration between the public and private sectors:

There are things that the State is not very good at, and there is room for the private sector [PartB]

Some of us who are open zealous have this unhealthy fear of a commercial entity and we need to have some collaboration between us. [PartD]

However, it was considered important that resources developed with public money, should be made openly available to the public:

I think it is on public funded organizations to make sure we don't lose access to these resources [PartD].

The public should have access to what the public paid for [PartB]

Language barrier

The language barrier already surfaced in the discussion of the first theme, concerning possible improvements to Discoverer, but it is certainly a much broader key challenge. PartD, a bilingual educator working in a non-English speaking country, highlighted it:

The language barrier is definitely a big problem [...] and we put ourselves in silos working in single languages, so that is difficult. [PartD]

This concern was also voiced by non-English speaking educators participating in previous evaluations of the prototypes. From my personal experience as a trainer and educators in various non-English speaking countries, I can only support this fundamental concern, which is fully acknowledged in the Ljubljana Action Plan.

9.3.4 Limitations of this study

This evaluation activity did not pretend to be an exhaustive or comprehensive study, but aimed rather to collect additional feedback from expert stakeholders and with a complementary methodology to the previous studies. A first limitation of this activity was the small number of experts interviewed. However, this limitation was partially compensated by the broadness and depth of the interviewees' competences in the domain, which made them good representatives of a wider population in the OER ecosystem. Another limitation was the inevitable bias in the questions guiding the interviews. It is hoped that this bias was minimized by the efforts made to provide the maximum freedom to the interviewees to express their real thinking and address their concerns, encouraging critical perspective. The reliability of the analysis could have been improved by having additional researchers repeating the analysis on the same corpus. While this was not possible in the context of this doctoral study, the analysis was coherent with the results of the previous studies and – concerning broader aspects – with the literature, in particular with the Ljubljana Action Plan (UNESCO, 2017).

9.4 Conclusions

In this final activity, a few key actors in the OERs ecosystem were consulted, to obtain their feedback on the relevance of the problem addressed, pros and cons of the strategies proposed, challenges and suggestions for the uptake of Discoverer in an operational environment, challenges and suggestions to improve the sustainability of the OER ecosystem.

The results of this activity supported the findings from previous studies, and helped to better situate the activities of this research in its broader, very dynamic and complex context. The discussion

around limitations of the prototype, barriers to its adoption, and broader challenges in the OER ecosystem, helped to identify further research activities, which are the focus of the next chapter. The new CC search platform for open material, in particular, was identified as a promising platform where further evaluations of Discoverer could be carried out, and where further extensions could be experimented with. Indeed, if CC Search is developed with an open service-oriented architecture, it will help reducing the fragmentation of the OER ecosystem, providing a fertile context that Discoverer and other applications could build upon, enriching it with higher-level and domain-oriented search services. The full potential of applications like Discoverer, though, strongly depend on related challenges such as the availability of good quality metadata and a widespread standard for coding them – which still appear far from being solved.

It is hoped that the small sample of participants involved in this activity was compensated by their representativeness of the domain, considering their vast and varied expertise in search engines, OER platforms, policy development, standardization activities, and use of OERs for teaching and course material development.

This activity provided the additional opportunity to feed-back some recommendations from this research to the key networks shaping the evolution of the OER ecosystem. Last but not least, these interactions with key actors in the open domain, offered the opportunity to appreciate their coherently open personality, and the noble ideals motivating their behaviour.

10. Roadmap to the future

The objectives of this research were mainly to identify innovative user requirements, and suitable strategies to support them. This chapter, stemming from the experience in this research and considerations about its limitations, identifies and discusses possible further research activities.

Section 10.1 discusses the collection of further evidence to better support the generalizability of results. Section 10.2 suggests some micro-level enhancements to the current prototype, while Section 10.3 proposes extensions to the current similarity metric. The remaining sections discuss potential research in increasing broader areas, yet related to the main focus of this research. While the activities are discussed separately, for presentation clarity, many of them could be tackled in combination.

10.1 Further evaluations

The main limitation of this study is the number and representativeness of the evaluators involved: the confidence on the generalizability of results would benefit from a sample more representative of a wider population. Hence, a first line of enquiry could collect additional evidence that the requirements identified and the strategy to support them are appropriate in a broader context. Therefore, this section discusses a number of potential evaluation activities, involving comparisons with alternative baseline solutions, engaging a larger number of participants from different domain and cultural contexts, and measuring the impact at level of productivity and quality.

10.1.1 Blind comparison of Discoverer with alternative baseline solutions

The “more like this” functionality developed in this thesis, could be evaluated against alternative (baseline) solutions provided by existing search platforms. To increase the validity of these evaluations, they could be carried out through blind comparisons, that is, with participants unknowing what system they are using. 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 argued that search by similarity based on educational alignments was a much better solution than the traditional Google search based on keywords. When participants reported that Discoverer was very effective in reducing their workload – 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 (and implicitly to Google search). However, additional evaluations could be carried out to increase the confidence in the generalizability of results.

The main challenge to overcome, is that such a comparison would depend on many variables, quite difficult to control. First, the study did not uncover an existing system which explicitly supports high-level user-oriented functionalities, such as “identify further resources for in-depth educational activities”: this hinders the possibility to organize a meaningful comparison. Second, the different set of resources indexed by different systems could heavily affect the results of the comparison, and

we could end-up comparing the datasets rather than the systems. Third, even with the same corpus of resources, the results would be heavily influenced by the characteristics of the available metadata – for example the amount of alignments available versus other metadata used by the alternative system.

These are reasons, indeed, that justified the adoption of the DSR methodology in the first place. Nonetheless, the following sections describe a few potential comparison setups, discussing in more detail the challenges and possible strategies to partially overcome them.

Comparison with Google’s “more like this” functionality

Google does provide an explicit “more like this” functionality, even if not always available for all items in its SERP. Yet, data about the suitability of Google have been already collected in this thesis: in particular, participants in the evaluations lamented Google’s limitations in finding educational resources, such as irrelevant results and uninformative snippets. A blind comparison, in this case, would not be particularly enlightening.

A more meaningful comparison could be achieved by restricting Google’s results to educational resources. This could be done by exploiting structured data with specific structured search operators, available in the context of Google Custom Search. This way, results could be restricted to pages which contain specific educational schema.org/LRMI types or attributes, with the operator `more:pagemap:TYPE-NAME:VALUE`. Whenever “VALUE” is omitted, results generalize to all documents having “TYPE-NAME” specified. For example, it is possible to restrict results to items which contain the type/class `alignmentObject`, or which specify the attribute/property `LearningResourceType`, with the following syntax:

`more:pagemap:AlignmentObject`

`more:pagemap:creativework:LearningResourceType.`

However, Dietze et al. (2017, p. 290) argue that “the mere use of LRMI terms does not provide accurate indicators of whether” a resource is educational or not; for example, the LRMI property `typicalAgeRange` is frequently used by providers of adult content. While they suggest that searching for the properties `typicalAgeRange`, `interactivityType` and `learningResourceType` would represent a useful pattern to identify educational resources, searching for the existence of these properties, at the time of writing, is not supported. However, “Google is continually extending the data it extracts and how much of this data is available for use in Custom Search” (Google, 2019).

The results of a comparison of Google and Discoverer, would also heavily depend on the respective – vastly different – datasets, and would not be very meaningful. Hence, a first potentially better solution could compare Discoverer, making use of the LR dataset, with Google restricting results to the LR domain. This could be easily done with a “Topical search engine”, using the operator `site:<URL>`. At the time of writing, however, this solution does not yield meaningful results, because

many LR resources are invisible to – and not returned by – Google. This solution, however, may work in the future, once Google harvests all LR data.

Comparison with Google using Google Knowledge Graph as the common dataset

A better solution could make use of the same dataset with both Google and Discoverer. To this end, Discoverer would need to access the Google’s Knowledge Graph, via its Custom Search Control API. This API allows programs to retrieve search results in JSON format from a Google Custom Search Engine (GCSE) – via RESTful requests – similarly to what was done with the Learning Registry. Yet, even neglecting the financial implications (this solution requires Google fees to be paid), such a comparison would be again heavily influenced by the dataset, this time by the alignments metadata available compared to the data used by the alternative native Google functionality. As previously observed, there is currently a limited number of alignments metadata available in Google’s Knowledge Graph, which hinders, at the moment, the value of this comparison.

Comparison with implicit “more like this” functionalities based on crowdsourced tags

Some OER search portals provide an implicit search by similarity functionality: users can list the crowdsourced tags associated with a sample educational resource, select one of those tags, and identify other educational resources associated with that tag. This way, users can identify resources “similar” to a given one – “similar” in the sense that they share a crowdsourced tag. A blind comparison between Discoverer and the “more like this” facility available in these portals is more appropriate than the previous Google comparisons, because the target domain of both systems are exclusively educational resources. Of course, the target search portal ought to support both crowdsourced tags and educational alignments: OERCommons is an example of such a portal, particularly interesting considering the size and quality of its dataset. Yet again, results would heavily depend on the metadata available, in particular on the quality and quantity of alignments versus tag metadata available.

Comparison with implicit “more like this” functionalities based on educational alignments

A few platforms, such as the Learning Registry, support another implicit search by similarity exploiting educational alignments. In these cases, it is possible to list the educational alignments associated with a sample resource, select one of them, and request other resources associated with that single alignment.

Comparing Discoverer with such a functionality, would compare the high-level task-oriented features provided by Discoverer, to manually oriented features available in other platforms, in this case using the same dataset and metadata. Yet, this solution would definitely not remove the dependence from them. Indeed, this comparison would rather investigate the following question: how much more convenient is it to automatically consider all the available educational alignments at the same time, as in Discoverer, rather than manually working with just one of them at a time? If most resources had just a single alignment, Discoverer would provide a limited added value, reducing the worthiness of this comparison. A possible solution could be to constrain the evaluation to resources having a

significant number of alignments, even if this would reduce its realism. A better solution could cross-check the results from multiple evaluations, using datasets having different average numbers of educational alignments associated to the resources. However, this solution is not practicable at the moment, because of the limited number of datasets available, having educational alignments.

Comparison with alternative similarity metrics

Section 10.4.1 discusses the possibility to use alternative similarity metrics based on tags, topics, or other paradata and social data. Discoverer, making use of a similarity metric based on educational alignments, could be meaningfully compared with an analogous system making use of a different similarity metric. Indeed, many aspects, such as the possibility to cluster resources by comparable similarity and iteratively expand them, could be made available in both systems – and would represent controlled variables. Results, however, once again, would heavily depend on the comparative quantity and quality of the different metadata powering the alternative metrics.

10.1.2 Broader quantitative oriented evaluation by instrumenting the prototype

From a more engineering-oriented point of view, the prototype could be further tested through a replication study involving a larger number of educators, to further strengthen the research claims. Yet, engaging a large number of educators proved to be quite a challenge. For example, when I sent a first survey to a few specialized mailing lists – with a full-fledged web site to anonymously collect results, in order to remove a potential obstacle to a wide participation – I did not get a single reply. Another significant example is when a specialized external company was tasked to collect feedback from teachers, in order to develop the first version of LRMI (2013b). While they sent out paper surveys to most schools in the USA, complemented with email and web site, and while it was not necessary to install any client software as in my case, they only managed to engage just a little more than a hundred respondents. Therefore, engaging in a replication study many more test-users than the 49 total participants engaged in this research, would likely require a prohibitive effort, better spent in other, more productive, research activities.

A more viable option to collect data from a large quantity of test-users would be to instrument Discoverer, that is to insert specific code instructions, to automatically collect usage data (analytics) for subsequent analysis. This could include the possibility of prompting users (occasionally) to provide feedback such as the specific reasons to use a certain feature, or the level of satisfaction with the results obtained.

Such a solution would require a full-fledged search engine, to be sufficiently appealing to users. The prototype experimented in this research was not intended to be a complete self-standing search engine, but just a component, intended to augment the functionalities of an existing one. This validates the proposal that Discoverer should be integrated in an existing search engine. This is not necessarily straightforward, because integrating different search strategies requires a seamless movement from one to the other (Belkin et al., 1993), supporting a fluid refinement of search (Hearst and Degler, 2013), and interactive navigation sequences (Stefaner et al., 2009). However, the QBE

strategy proposed in this research was explicitly designed to support this integration, making it possible to expand resources independently from the way they were previously obtained, hence potentially simplifying the extension of any existing platform.

A first example of a search platform to be extended is Google. This is, indeed, the approach adopted in this thesis, and its suitability has been demonstrated by the consistent positive feedback collected in the three evaluations. Modifying Google's search engine to integrate Discoverer was obviously unrealistic; therefore, Discoverer was implemented as a plug-in capable to transparently intercept and modify Google pages on the client. This way, users had the false but realistic impression that the pages visualized in their browser by Discoverer, were produced by Google. However, while very much appreciated, this solution required the previous installation and configuration of the necessary software on the test-user's computers. Such a solution cannot be expected to work on a large scale, as users would unlikely be willing to install software on their computer for a test, and to grant it the required security permissions. Hence, integrating Discoverer on top of an existing platform for a large-scale evaluation, would practically require the permission to modify it. Obtaining the permission to modify an existing operational platform, however, remains a major barrier.

As an additional challenge, the hosting platform needs to make educational alignments metadata available to the prototype. This requirement restricts considerably the number of platforms that can be used, even if this can be expected to increase in the future (Dietze et al., 2017). Google and its Knowledge Graph, accessible via a Custom Search Engine, is a potential solution, previously discussed. Another suitable platform is OERCommons, as it includes a rich set of learning resources with highly curated metadata (Campbell, 2014a). The most promising alternative is CCSearch, the search engine being developed by Creative Commons, which is supposed to provide a one-stop point of access federating metadata from the largest OERs' platforms. While this search engine is being developed incrementally media by media, and has only covered text and images at the time of writing, there are definite plans to cover educational resources too. As discussed in the previous chapter, I recommended to its designers to provide access to their search engine via open Web API. This will enable third parties (like Discoverer) to offer higher-level pedagogically-oriented services, building on top of their basic services, as experimented in this research.

10.1.3 Evaluation in different domains and cultural contexts

This research identified needs among educators, mainly focusing on secondary education in Italy. This may rise concerns about the degree of universality of the identified needs among educators with different profiles. In this regard, the analysis tested but 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 levels 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 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 extensively used. While Discoverer, as it is, would be useless in contexts where alignments are not available, 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 by degrees of similarity. Further research could identify more precisely partially different requirements which might arise in different domains, such as higher education or primary school.

Even within the same domain, different organizational and cultural contexts might also call for partially different requirements. Indeed, the strategy proposed in this research aims to support high-level tasks, and is therefore strongly affected by the organizational cultures (Wilson et al., 2010). Therefore, the proposed strategy could be evaluated in contexts with different characteristics. For example, contexts where there is a certain resistance to the use of standardized learning objectives, versus contexts where there is a strong culture of competence standards, such as in professionally oriented training settings.

10.1.4 Evaluation of the impact on productivity and quality

The evaluation of the prototype mainly collected subjective data from educators, based on their brief exposure to the prototype: the evaluation was realistic concerning users, but only partially realistic concerning system and setting. Indeed, the evaluation mainly addressed the first three levels of the Information Seeking model proposed by Abraham (2013): Search, Evaluation for Selection, and Evaluation for Use; yet, it would be convenient to address the ultimate level Use too, in the specific target context. This is a widespread challenge: Abraham (2013, p. 142) argues that the actual use of the results obtained from search activities “has received scant attention, and little is known as yet about how content is used from located or found sources”.

A more comprehensive evaluation could therefore measure the actual impact of the strategies proposed in the prototype, on the productivity and quality of teachers’ activities. For example, it could measure the time to carry out representative tasks such as “find 2 more resources for remediation activities, given the one that was already used”, or “identify an alternative pedagogical approach to target the same learning outcomes as with a given resource”. This would require more realism in the prototype, which would need, as previously discussed, to be integrated in a system providing a comprehensive set of functionalities, such as filtering or expansions with additional similarity metrics. Challenges of this evaluation strategy include, as previously discussed, the need to isolate the influence of aspects related to the dataset available, and the need to take into account the influence of different organizational cultures.

10.1.5 Evaluation with different stakeholders

The overall research goal of this research, was to identify strategies to better support educators in discovering OERs. Yet, there are stakeholders other than educators who could benefit from the availability of advanced OER discovery functionalities. For example, one of the experts interviewed, as described in the previous chapter, suggested that his students engaged in flipped classrooms could conveniently use Discoverer to find additional resources starting from the very few ones he assigns them. Another expert suggested that his group of instructional designers could use Discoverer to look for open material to incorporate in their courses. Hence additional studies could nearly replicate the summative study already carried out, yet targeting different stakeholders. Discoverer could also be integrated in portals that serve wider communities, as the future CCSearch, to collect additional evaluation data from different categories of users.

10.2 Micro-level enhancements to the current prototype

While the prototypes designed and evaluated were sufficient to reach the goals of this research, there are many possible micro-level improvements that could be considered, and their impact further evaluated. Some of them are briefly discussed here.

10.2.1 Recognizing different encodings of the same educational standard

Schema.org, at the current stage, standardizes the way to express an alignment to an educational framework, but does not standardize, yet, the way to encode the educational frameworks and its educational standards. This is, indeed, one of the open issues in schema.org (Barker, 2018). As a consequence, Discoverer considers different encodings of the same educational standard as if they were different, inheriting this behaviour from the Learning Registry Data Services. As an example, the current prototype fails to recognize that the following encodings all refer to the same educational standard, in the same framework:

- S114346A,
- <http://purl.org/ASN/resources/S114346A>,
- <http://corestandards.org/Math/Content/2/G/A/1>, CCSS.Math.Content.2.G.A.1,
- 61864044228F4141AF5A894EC059DFB7.

A more effective system should be able to recognise different formats (syntactical encodings) of the same educational standard as equivalent.

10.2.2 Controlling the number of clusters

Discoverer currently clusters results in three (or less) classes of similarity: low/medium/high. The large majority of participants in the evaluation considered that this was the ideal number of classes. However, this fixed number is not necessarily the natural number of clusters in every case. For example, the similarity vector in Figure 10.1 would more naturally split in four clusters.

Similarity degree	1	1	2	4	4	5	8	8	9	9	12	12	13
-------------------	---	---	---	---	---	---	---	---	---	---	----	----	----

Figure 10.1 – Similarity vector with four clusters.

The “natural” number of clusters of a vector could be determined, for example, by optimizing the Goodness of Variance Fit over different numbers of clusters (Slocum et al., 2009). Further research could analyse the distribution of the number of natural clusters for a given dataset. If the number of alignments is quite limited, as in the case of the LR dataset used in Discoverer, three clusters is likely a good option. In case the dataset had a larger number of alignments, hence a potential larger number of natural clusters, further research could evaluate whether it is really convenient to use a fixed number versus the natural number of clusters. Other alternatives may let users select the preferred option, among a predefined fixed number, the natural number, or any desired number of clusters.

10.2.3 Parsing in-page schema.org markup

Discoverer identifies a resource as an educational resource, by checking the presence of its URL in the large Learning Registry metadata repository. As an alternative, or in addition, it is possible to parse the web page describing the resource, to extract schema.org information that helps to identify an educational resource as such. To this end, it should be sufficient, in principle, to check the presence of educational oriented (LRMI) metadata. Unfortunately, as already noted before, the presence of LRMI metadata is, at the moment, not always a good indicator of the educational value of a resource, because some of these metadata (for example typicalAgeRange) are often used incorrectly on the Web. Other more specific metadata, in particular the AlignmentObject, could potentially be better indicators of the educational nature of a resource. Yet, these metadata are still rarely used – which is the reason why Discoverer did not use, as its dataset, the Google Knowledge Graph. However, parsing in-page schema.org mark-up is a strategy that can complement the current use of the Learning Registry, which will become more effective as more LRMI metadata will become commonly available on the Web.

10.2.4 Avoiding queries producing no results, without decreasing efficiency

The evaluation of Injector showed that users were frustrated by the inability to expand resources in a SERP, due to alignments sparsity. To avoid this inconvenience, Discoverer included in its SERP only resources that could be expanded. This was a rational solution considering the objectives of the research to evaluate the suitability of the expansion operation. However, with the LR dataset, this solution excluded from the resulting SERP of the first search by keywords, about 90% of the resources identified. A more operational oriented and efficient system could avoid this waste of resources, providing visual clues to the users about the expandability of the resources. This is in line with the suggestion from the IFT to provide as many clues as possible, to allow users to take informed decisions about the convenience to change foraging patch, that is, to perform an expansion. Further research should investigate the best solution to address these aspects.

10.2.5 Server-side metasearch engine for a service-oriented architecture

The prototype is currently implemented as a Chrome Extension, a client-side technology that was selected mainly for the possibility to have full control on the user environment, for example to intercept requests, modify returned SERPs, and generate automatically requests for additional data. However, some participants in the evaluations lamented difficulties in its installation. Indeed, recent policy changes require Google's explicit approval to install Chrome Extensions from their proprietary store, or require to work in "developer" mode, which proved awkward for non-technical users. As recommended by one of the experts interviewed, it is necessary to make the system extremely easy to install. To achieve this, the prototype could be implemented server-side, and its functionalities offered as a service.

This solution would provide the opportunity to build a service-oriented architecture (Papazoglou et al. 2007), where existing search services are integrated, to offer higher-level educational services on top of them, which could be easily used or aggregated, in turn, in other applications – such as LMS. As discussed in the previous chapter, Discoverer could work on top of CC Search, a metasearch engine federating search services (Meng et al., 2002), and could in turn provide an educational-oriented higher-level service offering functionalities to identify similar resources, addressing specific educator's tasks.

10.2.6 Optimizing screen real estate with fisheye techniques

The output of the prototype has been considerably compressed by the use of clustering, which helps to provide a synthetic view of the usually large number of resources returned. Yet Discoverer still displays three resources for each of the three clusters, and each resource includes a considerable amount of metadata. A better interface should help to further reduce the density of information on the screen, possibly with techniques such as fisheye or progressive exposure of content as in Wavelens (Paek et al., 2004).

10.3 Extensions to the current similarity metric

Having established (at least with a reasonable level of confidence) the effectiveness of search by similarity based on learning objectives, further research could design and evaluate possible improvements and extensions to the metric adopted. This section explores a number of possibilities, whose relative effectiveness would again strongly depend on the datasets.

10.3.1 Normalizing generalizations and specializations in the educational framework

A resource which is aligned to a specific standard, could also be aligned (at least partially) to its higher standards in the framework hierarchy. For example, let us assume that two resources Res_a and Res_b are both aligned to the CCSS standard "Use parentheses, brackets, or braces in numerical expressions, and evaluate expressions with these symbols" as well as to its higher standard "Operations and Algebraic Thinking". The two resources, having in common these two alignments, would have similarity:

$$\text{similarity}(Res_b, Res_a) = 2.$$

On the contrary, a resource Res_c explicitly aligned only to the lower level standard, would be incorrectly considered, with the adopted metric, less similar to Res_a , compared to how much similar Res_b is to Res_a :

$$\text{similarity}(Res_c, Res_a) = 1 < \text{similarity}(Res_b, Res_a) = 2.$$

To address this, an improved metric should exclusively include in the count the more specialized alignments, excluding all the “redundant” higher level alignments. In the previous case, for example, this new metric would correctly determine that:

$$\text{similarity}(Res_c, Res_a) = 1 = \text{similarity}(Res_b, Res_a).$$

Likewise, the adopted similarity metric would not be able to match a specialized alignment with a corresponding alignment higher in the framework hierarchy. For example, resource Res_a could be aligned to CCSS standard “Use parentheses, brackets, or braces in numerical expressions, and evaluate expressions with these symbols”, while resource Res_b could be only aligned to the more general CCSS standard “Operations and Algebraic Thinking”. The adopted metric would not recognize the commonalities between the two standards, and would fail to recognize the similarity of the two resources.

To solve this problem, an improved metric should consider a resource aligned to a given standard, as implicitly aligned to all its lower standards in the hierarchy. In this case, the metric would correctly compute that:

$$\text{similarity}(Res_a, Res_b) = 1.$$

This assumes that the hierarchy in learning outcomes expresses an aggregation relationship: alignment to a higher standard implies alignment to all its sub-standards. This is different from considering an alignment to a higher standard as more introductory, compared to an alignment to lower standards. Indeed, a resource covering a certain topic at introductory level, should be aligned to a *specific* learning outcome specifying precisely the coverage at introductory level, and not to a learning outcome higher in the hierarchy.

10.3.2 Jaccard similarity: normalizing the number of alignments

The similarity metric adopted in Discoverer is not normalized and depends on the absolute number of shared educational alignments. This strategy could incorrectly favour resources with a larger number of educational alignments, because these resources would more likely have a higher number of shared alignments, compared to resources having fewer alignments. This was not the case for the dataset used in the prototype, because the range of the number of educational alignments associated to its resources was rather limited (mean 4). From a critical perspective, large variations in the number of alignments associated to the resources in the dataset poses a potential issue. In order to

control for resources that have a large number of alignments, the similarity metric could be refined as follows:

$$\text{Similarity} (Res_i, Res_k) =_{\text{def}} \frac{\text{Number of shared educational alignments between } Res_i \text{ and } Res_k}{\text{Number of total educational alignments specified for } Res_i \text{ and } Res_k}$$

This similarity metric is equivalent to the Jaccard (1901) Index, or Jaccard similarity coefficient, computed on the sets of alignments of Res_i and Res_j :

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

Other alternatives used in data mining could also be analysed. A first candidate is the widely used cosine similarity, considering its suitability for sparse data: in this case data are sparse because most educational alignments are not shared.

10.3.3 Factoring-in the alignments' level of abstraction

The standards in an educational framework are usually organized hierarchically, lower level standards being more specific than higher level ones. The similarity metric should not be based exclusively on the number of alignments, but could take into account their level of abstraction too. Indeed, given the same number of shared alignments, two resources are likely more similar if they share more specialized alignments, that is alignments at lower abstraction level. For example, two resources Res_{a1} and Res_{a2} that share the CCSS standard “Use parentheses, brackets, or braces in numerical expressions, and evaluate expressions with these symbols” (depth 3), are likely more similar than two resources Res_{b1} and Res_{b2} that share its higher-level more generic CCSS standard “Operations and Algebraic Thinking” (depth 1). Hence, a better similarity metric could weigh the level of depth of the shared alignments in the framework hierarchy, as follows:

$$\text{Similarity} (Res_i, Res_j) =_{\text{def}} \sum \text{alignment depth in the framework hierarchy of shared educational alignments between } Res_i \text{ and } Res_j.$$

In the case of the previous example, using the metric adopted in this research, we would obtain:

$$\text{similarity} (Res_{a1}, Res_{a2}) = \text{similarity} (Res_{b1}, Res_{b2}) = 1,$$

while with this new metric we would obtain, more appropriately:

$$\text{similarity} (Res_{a1}, Res_{a2}) = 3 > \text{similarity} (Res_{b1}, Res_{b2}) = 1.$$

10.3.4 Standards cross-walking: degree of equivalence and degree of alignment

Many different educational frameworks are available, and standards belonging to different frameworks might be semantically equivalent or near equivalent. Yet, the current prototype fails to recognize that a certain educational standard in a given framework is equivalent or similar to another standard in a different framework. A more effective metric could take advantage from this knowledge, which is offered by specialized LD-oriented services such as those offered by the

Achievement Standards Network (ASN, 2016). Some of these inter-framework similarities could also be mined from existing educational alignments: if the same resource is aligned to two standards in two different frameworks, we may infer that the two standards are likely related.

The use of equivalences between standards (standards cross-walking) raises two major issues. Firstly, two standards might not be perfectly equivalent, hence there is a need to model their *degree of equivalence*. Secondly, this implies, in turn, the opportunity to model also the *degree of alignment* of a resource to a standard. Standards cross-walking services such as ASN, indeed, do model the degree of equivalence among educational standards, and the degree of alignment of a resource to a standard is also modelled in highly curated portals such as OERCommons. Yet, so far, the degree of alignment cannot be modelled in schema.org, which is, indeed, one of its open issues (Barker, 2018).

If the degree of equivalence between standards and the degree of alignment between standards and resources are available, the similarity metric could weigh the number of shared alignments as follow:

$$\text{Similarity}(Res_a, Res_b) =_{\text{def}} \sum_{i,j} \deg(Res_a, Std_i) * \deg(Res_b, Std_j) * \deg(Std_i, Std_j)$$

where:

$\deg(Res_a, Std_i)$ is the degree of alignment between Resource a and Standard i ,

$\deg(Std_i, Std_j)$ is the degree of equivalence between Standards i and j ,

and, of course, $\deg(Std_k, Std_k) = 1$.

10.3.5 Similarity vector versus similarity matrix

Discoverer clusters results by applying the Jenks Natural Breaks optimization technique to its similarity vector NumStdsInCom. This is a mono-dimensional data structure where NumStdsInCom[Res_{*i*}] is the degree of similarity between the expanded resource R and Res_{*i*}, that is, according to the metric defined in this research, the number of alignments they have in common.

This method classifies in the same cluster, two resources R_i and R_j which have a similar number of alignments in common with R . Yet, this mono-dimensional similarity vector does not provide any information concerning the similarity of resources R_i and R_j among themselves, which indeed may not be similar at all. This can happen because the relationship “similar to” is not strictly an equivalence relation:

$$R_i \text{ similar_to } R \text{ AND } R_j \text{ similar_to } R \quad \not\Rightarrow \quad R_i \text{ similar_to } R_j$$

Indeed, resources R_i and R_j might get classified in the same cluster because each one of them has some educational alignments in common with R , *even if they do not have any educational alignment in common among themselves*.

As a potentially better alternative, resources could be more precisely clustered in such a way that two resources R_i and R_j would be classified in the same cluster if and only if they have a similar number of educational alignments in common with the expanded resource R , *as well as among themselves*.

This would require to take into account the similarity among similar resources too, with a bi-dimensional similarity matrix, where:

$\text{NumStdInComMatrix}[R_i, R_j] =_{\text{def}}$ number of educational alignments that R_i and R_j have in common.

Computing such a similarity matrix requires more computational resources compared to those required to compute the similarity vector used in this research, because it entails to compute the number of shared alignments between all the possible resources in the dataset. This challenge, however, could be solved by pre-computing the matrix offline, as indicated in Figure 10.2.

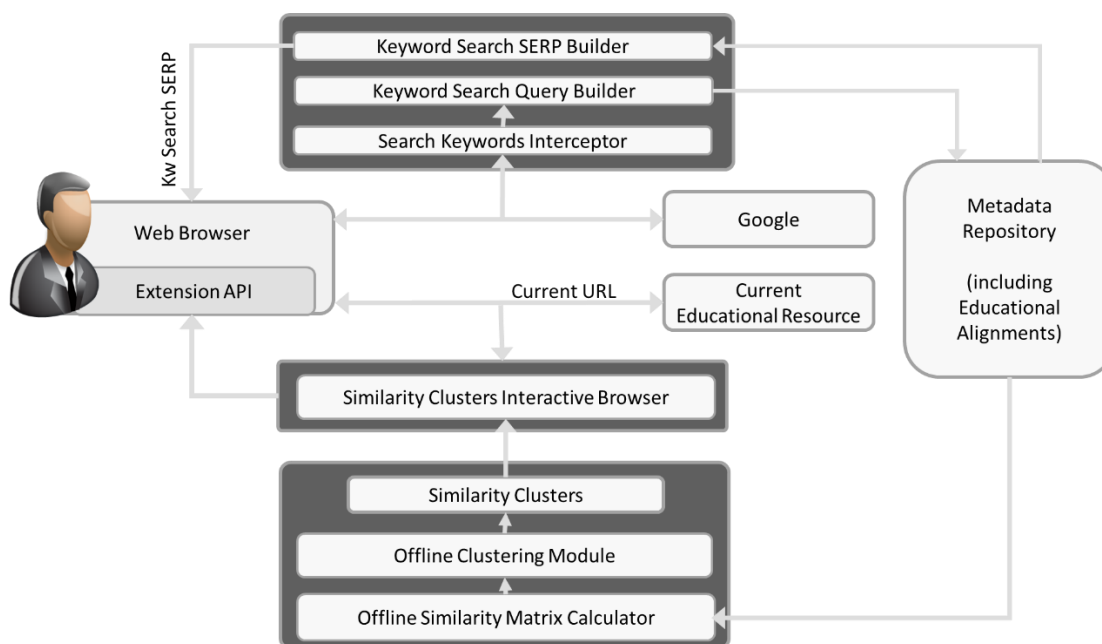


Figure 10.2 – Exploiting a similarity matrix.

Clustering would require, in this case, a technique suitable to process multi-dimensional data. The popular K-Means Clustering technique could be a suitable solution, as it is an extension of the Jenks Natural Breaks optimization (adopted in Discoverer) to the multidimensional case (Jain, 2010). The computation of the similarity matrix and clustering could be conveniently implemented offline, on a server which could then provide cluster information in real-time to a Similarity Clusters Interactive Browser client.

In this case, the major challenge is likely related to the visualization of the resulting clusters to users. It would seem natural to represent clusters on a 2D proximity map, taking advantage of the familiar spatial metaphor (Wilson et al., 2010). Regions would represent clusters, and their distance would be inversely proportional to their level of relatedness (Hall and Clough, 2013). Yet, this is not possible without considerable approximations, because the relatedness metric identified does not satisfy the triangle inequality for metric spaces. For a demonstration, suppose that resource A is similar to resource B and resource B is similar to resource C. As previously discussed, nothing can guarantee

that resource A is similar to resource C: this could require to physically place C both close to B, which is close to A, and far from A – which of course is not possible.

An alternative solution would be to use more complex visualization mechanisms such as weighted graphs, suitable to visualize these data without approximations, but requiring a definitely higher cognitive load for their interpretation. The first alternative would produce approximated mappings naturally processed at our pre-attentive stage, at visceral and unconscious level, while the second alternative would require a conscious interpretation at higher cognitive levels (Mazumdar et al., 2013).

Which solution would be preferable: a precise but complex graph structure, a simpler but approximated 2D mapping, or the even simpler mono-dimensional solution adopted in this research? There are no doubts that the more complex solutions provide more information, yet there is a need to understand the added value of this complexity to the effective support of actual educators' tasks. The experience in this research suggests that the simpler solution is likely going to be preferred by the majority of users, which is in line with results from similar experiments by other researchers. Reitsma et al. (2017) in particular, did not find significant differences in effectiveness and perceived utility between list-based and map-based representations of similar learning resources. However, the results will depend on various factors, including in particular the cognitive capabilities of the target users, the actual dataset which would affect the magnitude of the required approximations, and the added value of the more complex visualization. Further research could better investigate this new question, comparing the different solutions in the specific context of Discoverer.

10.3.6 Providing explanations, letting users prioritize educational alignments

An enhanced prototype could visualize to users, on demand, the educational alignments shared between the current resource and an identified similar resource. This is equivalent to allow users to ask for an explanation of the results of the system, which could be considered “recommendations”. This functionality was suggested for improvements and further research by one of the interviewees, as discussed in Chapter 9. Providing explanations is considered very important in recommenders, because it supports users in their decision making, and improves their satisfaction and trust (Daher et al., 2017).

Even more important, users could exploit this feature to better control their searches. Indeed, the similarity metric currently used by Discoverer does not make any distinction among educational alignments, as it has no information about their relative importance for the user. However, when presented with the list of shared educational alignments, users could indicate their relative importance, by specifying their weights. This way, users could customize and fine tune the similarity metric, in order to identify more precisely the resources more relevant to them. This new similarity metric would be computed as follows:

$$\text{Similarity}(Res, Res_a) =_{\text{def}} \sum \text{weights}_{\text{Std}_k} \text{ for every educational standard } \text{Std}_k \\ \text{shared between Res and Res}_a$$

This option is in contrast with the desire, expressed by the large majority of educators participating in this research, to make the use of alignments totally transparent to users. Yet, as suggested by one of the interviewed experts, this functionality could be offered as an advanced option to more sophisticated users, to allow them to better specify their needs. This option could satisfy the desire expressed by some evaluation participants to make the learning objectives an explicit part of the discovery process.

10.3.7 Integrating quality indicators based on social data and paradata

Educators participating in the evaluation studies of this research, showed a limited interest on the opinions or activities of their colleagues. Yet, the experts interviewed underlined the importance of providing information about the quality of the resources. This concern is widely shared in the literature: Atenas and Havemann (2014), for example, emphasize the importance of addressing quality, and Clements (2016) argues that information about the quality of the resources is a critical factor to the success of any OER repository. Explicit social data such as likes or reviews, as well as implicit paradata such as the actual use of resources, could be exploited to provide information about the quality of the resources (Drachsler et al., 2012). Therefore, these data could be used in Discoverer to rank by quality the resources within clusters of a given degree of similarity, in addition to provide non-authoritative (or third-party – independent) information.

Tracking the provenance of these data could further contribute to improve users' trust (Groth and Moreau, 2013). An enhanced prototype, for example, could allow users to weigh these data, according to their provenance, hence letting them precisely control their influence on results.

10.4 Meso-level extensions to the overall strategy

This section discusses possible extensions of this research at meso level.

10.4.1 Additional similarity metrics

The task analysis evaluation identified a number of possible similarity metrics, in addition to the adopted one based on learning objectives. In particular, participants mentioned expansion by same authors and same topics.

The expansion by same authors is, or could be, easily supported by existing search platforms. It is sufficient to identify the authors of the current resource, and perform a new search filtering by those authors. If not directly supported by the search platform, this expansion could be anyway easily carried out with a few manual operations, because the number of authors of a resource is usually very limited (frequently just a single one).

On the contrary, the expansion by subject topics is potentially more interesting. First, as discussed in the task analysis evaluation (Section 5.6), the use of subject topics to classify the resources is more widespread than the use of learning objectives. Second, topics can be annotated in schema.org with the same technique used for educational alignments, making it possible to reuse the same technical solutions used in Discoverer. Hence, all the aspects experimented or discussed in relation to the

alignment of a resource to a taxonomy of learning objectives, could be applied to the alignment of a resource to a taxonomy of subject topics. Despite these similarities, however, educators are more familiar with topics than with formal learning objectives, therefore the requirement for the discovery engine to be completely transparent to their use may not hold in this case. Consequently, as an example, the possibility to make the use of topics more explicit, to allow educators to prioritize them, which was discussed as an option in Section 10.3.6, could be offered here as a fundamental functionality. As suggested by one of the OERs' expert interviewed, specific tasks, such as the production of course material, might require specialized metrics, for example based on the combination of topics with reading difficulty indicators.

Metadata concerning educational alignments and topics are usually produced by skilled curators. Yet, it is also possible to power similarity metrics from crowdsourced data. Most platforms, in particular, allow users to tag resources. These tags could be used (Bogers, 2018) along the lines experimented with educational alignments in this thesis, using their co-occurrence as a measure of item-item similarity. That is, similarly to Discoverer, the system could rank the resources according to the number of shared tags, classify them in a few classes of different similarity, and present a few representative resources for each class. As in this research, each class could be further expanded, new resources selected, and iteratively expanded.

Social data and paradata could be used to power other proximity metrics (e.g. to find resources liked, or used, by the same users who liked or used another resource). While the educators who participated in the task-taxonomy evaluation did not show much interest in these data, it can't be excluded that they might change opinion, following a positive experience with a suitable system. The importance of social data, indeed, is frequently mentioned in the literature, for example by Atenas et al. (2014). Hence, longitudinal research could investigate the acceptance of different metrics based on classical collaborative (or social) filtering (e.g. for people who liked the resources I liked, which other resources did they like? Or what resources did my colleagues use, after a given one?).

Having multiple similarity metrics raises the additional challenge to identify the best way to integrate them in a system. Users could be allowed to select the metric of their interest, as suggested by Wilson et al (2010), they might be allowed to weigh them and use them in combination, or the system might suggest predefined sets of weights for different use cases.

10.4.2 Exploiting different types of alignments

LRMI / schema.org allows curators to specify the type of an alignment, through the property `alignmentType` that can assume values such as *teaches*, *requires*, or *readingLevel*. In this way, alignments to educational standards can be used for different purposes, not just to specify learning objectives. Indeed, learning objectives are identified by the type *teaches*, while the type *requires* is used to specify the pre-requisites of an educational resource, and *readingLevel* to specify the level of reading complexity.

The current prototype does not recognize the type of an alignment, and treats them as if they were all of type *teaches*. At the current stage, this is legitimate because the educational alignments currently available through the Learning Registry have no type specified, and are implicitly considered of type *teaches*. At broader Internet level too, the type of an educational alignment is rarely used (Dietze et al., 2017). Yet, when these metadata become available, an operational system will need to take them into account. First, the metric currently supported will have to deal exclusively with educational alignments of type *teaches*. Second, educational alignments of type *requires* might be used, as discussed in Section 10.5.2, to support a different task.

10.4.3 Clustering results of the initial keywords search

The initial keyword search of the current version of Discoverer, generates a SERP where results are not grouped in clusters. This initial phase, indeed, is just a traditional keyword search to start the whole process, aiming to produce more relevant results than Google. In this phase, the prototype does not work with learning objectives at all, so it is not possible to cluster results with the current unidimensional clustering technique because there is no reference resource from which to compute the distance.

However, clustering based on learning objectives could be carried out, at this stage too, by adopting a bi-dimensional metric. Indeed, a bi-dimensional metric would make it possible to compare the similarity of the resources among themselves, hence clustering them. As discussed in section 10.3.5, the use of a bi-dimensional metric would rise other problems, in particular the challenge of visually mapping a metric that does not satisfy the triangle inequality for metric spaces.

10.4.4 Search personalization

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.

10.4.5 Resource identifiers

The educational resources are identified, in the prototype, by their URL. This is not an ideal solution, because the same resource is often replicated in different repositories. An ideal solution would be to associate unique identifiers to the educational resources, rather than their instances (Thomas et al., 2012). Unfortunately, the URL is an identifier which is only univocally associated with each

replicated instance: hence, the metadata are associated with the instances, rather than with the resources. This makes it difficult to aggregate data related to different instances of the same resource.

This problem becomes apparent in Discoverer when it fails to expand an expanded resource. This should never happen in principle, because of the symmetric and transitive properties of the similarity relation. And yet it does happen, when the user explores an expanded resource (identified by its URL), but the browser is redirected to an updated URL. At that point, the updated URL might not be known in the dataset, so that it cannot be expanded further.

A similar problem occurs on a larger scale in schema.org. While the Semantic Web and LD consider unique identifiers as structurally fundamental (Berners-Lee, 2006), schema.org dropped the requirement of a unique identifier to identify entities. The rationale, was to facilitate the publishing process and help kicking-off the whole mechanism (Guha et al., 2016). However, this solution shifts the burden of recognizing equivalent entities from metadata producers to metadata consumers.

A tentative solution foreseen in schema.org, hence in the Learning Registry too, is the use of sameAs relationships. This relationship provides the mechanism to state the equivalence of different instances of the same educational resources (Bienkowski, 2012). Hence, metadata and paradata related to the same educational resource could be aggregated from its different instances. Yet, this relationship has not been used extensively (Guha et al., 2016). As an alternative solution, entity reconciliation techniques could be used to recognize the equivalence of different identifiers. Once identified, possibly with different techniques, the equivalence of different identifiers could be conveniently published as a service, along the lines of the site <http://sames.org>. Here, though, there would be the additional need to track potential differences between the various instances of the same resource. A service like this, addressing a very specific and widespread challenge, would have great potential for reusability – in line with the strategy of this thesis. Such a focused but complex challenge certainly presents an opportunity for further research.

10.4.6 Educational alignments as OER genes in an OER Genome Project

Educational metadata can be considered characteristics to classify educational resources. This research shows in particular that the strongly domain-oriented educational alignments are appreciated when used to precisely identify educational resources. Therefore, similarly to the music attributes used to classify songs in the Music Genome Project (Westergren, 2007), further research could investigate whether educational alignments, especially if modelled with the degree of alignment, could be considered sort of genes to capture the essence of OERs in the context of an OERs Genome Project.

10.5 Identification of additional requirements

According to the strong user orientation of this thesis, the prototype could be extended by supporting different use cases. Indeed, Discoverer, supporting the discovery of educational resources of variable degrees of similarity, provides high-level and domain-oriented functionalities such as:

- find OERs for remediation activities for some of my students,
- find OERs for in-depth activities for my most brilliant students.

Support for these tasks, as discovered by this PhD, is very much appreciated by educators. But are there other similar high-level tasks that need to be supported?

10.5.1 Discoverer as a design probe

Design probes can be used to understand user needs in complex real-world settings, by getting creative ideas from users in participatory design activities (Hutchinson et al., 2003). This methodology is particularly suitable in early phases of a design, and when the researcher needs to work in an unfamiliar domain. Therefore, the existing prototype could be used as a design probe, in order to identify additional user requirements, opening up again the design space. Users could suggest new tasks, different uses of the similarity metrics identified, or ideas for a new similarity metric, which could lead to improvements and extensions to the current prototype.

10.5.2 Identification of prerequisite resources

As a first example of a new task, educators might be interested to support the following WC-level task:

- find resources covering the maximum number of prerequisites for a given OER.

This task could be supported again with a QBE type of query, starting from the current OER as the example. The same similarity metric used in Discoverer, could identify the resources that best cover the prerequisite educational alignments of the current OER, just by working on the educational alignments with alignmentType of value *requires*, rather than *teaches*. That is, Discoverer could rank the resources according to the following scoring:

$$\text{CoveringScoring}(Res_i, Res) =_{\text{def}} |\{\text{educational alignments of type } \textit{teaches} \text{ for } Res_i\} \cap \{\text{educational alignments of type } \textit{requires} \text{ for } Res\}|$$

The resources maximizing this scoring are the sought-after resources, covering the maximum number of prerequisite learning objectives.

10.5.3 Identification of interdisciplinary resources

As a second example of a new task, educators might be interested to support the following WC-level task:

- find OERs for interdisciplinary educational activities (in relation to given learning objectives).

The similarity metric could be used to support this task, by identifying “connector hub” resources, that is, resources having high similarity with resources in more than one cluster (as discussed in Section 7.3.3). These resources play an interconnecting role between different clusters (patches in terms of IFT), bridging between resources with different set of learning objectives. They correspond

to highly interdisciplinary resources, which could be explicitly targeted because of this distinctive educational characteristic. Further user-oriented research would be needed to understand to what extent this feature would actually be appreciated, and used, by educators. Additional research would be required also to understand whether a meaningful number of such resources are actually available in the existing datasets.

10.6 Broadening the research areas

This section considers potential areas of research in increasing broader areas, yet still closely related to the main focus of this research.

10.6.1 From metadata consumption to metadata production

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 use of 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.

Educational metadata can be produced by knowledge-intensive and time consuming manual curation, by automatic or semi-automatic metadata extraction techniques based on artificial intelligence (AI), and by crowdsourcing. The metadata in the Learning Registry are crowdsourced from many actors, which make use of a variety of approaches. The data coming from the OERCommons initiative, for example, come from specialized multidisciplinary teams that align the educational resources to educational frameworks, or check the quality of data crowdsourced from a restricted set of certified users (Campbell et al., 2014). The use of AI techniques to extract educational metadata in general is a large research area (Miranda and Ritrovato, 2015). More specifically, promising attempts to extract educational alignments have been reported by Marshall et al. (2016). The educational alignments in OpenEd, for example, are originated by machine learning techniques, and then validated by a team of experts (Molnar, 2016).

As suggested by one of the experts interviewed, a generic Web social annotator is a possible solution to crowdsource metadata. This type of infrastructure allows users to independently add annotations to resources on the Web, as a new layer, without modifying the original resources, and without requiring the original resources to provide any support for these annotations, nor the permission from their owners. These infrastructures can be used to build a public information meta-layer on top of existing Web resources (including educational resources) to crowdsource non-authoritative metadata about them.

In order to impose some control on the format of the data collected, the annotation could be guided by a controlled ontology, based on metadata standards such as LRMI. This could be an interesting use case for the recently developed infrastructure *hypothes.is*, which is a solution based on open

software and open standards (Whaley, 2017). This software is particularly interesting in this context because it supports a reputation model based on peer review, which can be instrumental to support the trust of educators on the non-authoritative metadata collected.

10.6.2 New metadata standards

Despite the plethora of existing metadata standards, there are ongoing activities planning to develop new ones. A notable case is the new standard specifically oriented to OERs, being developed by the OER Metadata Group, in the context of CC activities. As personally suggested in the CC Open Education Platform, rather than developing one more standard, it is possibly more convenient to precisely identify limitations in the current LRMI standard, and propose solutions to improve it. To this end, I remarked, as an example, that OERs get continuously refined, extended and remixed, for example in the context of OER-enabled pedagogy activities, and yet at the moment it is quite challenging to track these modifications in LRMI. Even just this single use case represents an opportunity for further research with a potentially important impact, because the activities of modifying, remixing and republishing OERs are at the core of the concept of OER. But of course, there is plenty of room for other research activities aiming to identify additional use cases and solutions to address them.

Another area of potential research emerged from private conversations with the leader of the OER Metadata Group. Standards like LRMI, were explicitly designed to be simple enough to be used by OER authors, Web developers, or curators – without requiring professional librarian expertise. Yet the approach of this new OER Metadata Group is remarkably different, as their objective is, on the contrary, to leverage professional librarian expertise. This new approach addresses directly well-known challenges such as the scarce quality of the educational metadata available, but rises further questions. Is it reasonable to expect that the huge amount of existing OERs, that keep being revised, remixed, republished all the time, can be professionally curated by systems librarians? And what are the desirable and possible match points with the existing standards, to make these two strategies mutually supportive?

10.6.3 In-page authoritative metadata versus centralized open metadata

It is also convenient to reflect on the consequences of having LRMI metadata directly within web pages, as foreseen by schema.org. Embedding (authoritative) metadata directly in the web pages of the resources, requires a consistent effort to continuously crawl the whole Web, which is in the reach of only a very few companies having the required infrastructure. Furthermore, this solution privileges authoritative metadata, because it gives full control on the metadata to the resource publishers, making it difficult for third parties to provide independent metadata. The Learning Registry addressed this problem by allowing third-parties to upload metadata (concerning resources hosted somewhere else) in a centralized service, that anybody could then query through open Web API. OERCommons, as an example of a different strategy, crowdsources metadata from their community of users, with a well-controlled curation process (Campbell, 2014a). Yet unfortunately, even if OERCommons committed to make these metadata available via schema.org, and Google promised

to make the crawled schema.org metadata available via their CSE, the amount of educational metadata actually available from a single access point is still limited. This issue of in-page authoritative metadata versus centralized open metadata is therefore another area worth of further investigation.

10.6.4 OER-enabled pedagogy

There are plenty of opportunities to research on strategies to identify first, and then overcome, other barriers to the wider scale adoption of OERs – as outlined in the Ljubljana Action Plan (UNESCO, 2017). At the same time, however, it is also convenient to research on the opportunities to better exploit their unique advantages. OER-enabled pedagogy (Wiley, 2017; Wiley and Hilton, 2018), in particular, is an area with an interesting potential. OERs are not just inexpensive, possibly even higher quality replacements for alternative expensive resources: OER-enabled pedagogy aims to further leverage their unique characteristics, fostering fundamental innovations in their use. Wiley (2013) in particular, claims that the possibility to reuse, revise, remix, and redistribute open resources, makes it possible to organize activities where students learn from existing open resources by actively improving them. Making them available to future students, creates a self-fuelling cycle which will continuously improve and extend resources for future students, reducing the time wasted in “disposable” activities such as traditional assignments, and making it possible to address complex real-world problems for example related to the environment and sustainable development. Linking back to the main focus of this thesis, OER-enabled pedagogy provides additional discovery-related use cases and requirements. In particular it rises the need to keep track of the large number of modifications to the resources that these activities generate.

10.7 Conclusions

This chapter offered to future researchers some suggestions for further research, building upon the activities carried out in this thesis. The first section discussed additional evaluation activities, in different contexts and at different levels, to collect further evidence about the generalizability of the needs identified and the suitability of the solutions proposed. The most promising configuration is to integrate an instrumented service-oriented version of Discoverer, with the services that will be soon offered by the CC Search engine. The following two sections suggested possible improvements to Discoverer, from micro-level aspects such as a user-controlled number of clusters, to broader extensions to its similarity metric. These included, for example, suggestions to weigh the level of abstraction of the alignments in the educational framework, or to move from a client-side computed similarity vector to a server-side similarity matrix. The subsequent proposals concerned extensions at a higher level, including suggestions for additional user requirements and related supporting strategies. While this thesis focused on metadata consumption, the last section proposed a few ideas concerning the complementary challenges of metadata production, standardization, and distribution.

11. Conclusions

11.1 Overall aim and research questions

The overall goal of this research was to contribute solutions to the well-known challenge of OERs discoverability. Considering the complexity of the problem and reviewing extant research and development that address it, this research targeted in particular expansion by similarity tasks. This specific target was identified by a preliminary analysis of the tasks educators need to carry out in relation to OERs discoverability. The overall research question was therefore:

What are the main tasks associated with OER discovery, and how can educators be supported in performing them?

The overall question was sub-divided into the following lower-level questions. The aim was to categorize first the domain-oriented tasks and their relative importance, in order to identify those deserving further investigation, and the requirements to support them:

RQ1: What tasks do teachers need to carry out in relation to OER discovery, in the framework of the OER life cycle?

RQ2: What is the relative importance of tasks and categories identified, and which aspects deserve further investigation?

RQ3: What are the general requirements for tools supporting the aspects of interest identified?

Having preliminarily identified the general requirements, the research experimented with possible strategies to (confirm and) support them:

RQ4: Which tools could support the identified requirements, and how do teachers respond to them?

Following the Design Science Research methodology, a number of tentative solutions were iteratively prototyped and evaluated. The development and evaluation activities in each iteration identified new research questions to be addressed in the following cycle. The detailed sub-questions tackled in the various cycles were:

RQ4.1: How can an OER discovery tool be embedded into web search?

RQ4.2: How can Query By Example be added to an OER discovery tool?

RQ4.3: How can similarity based on educational alignments be added to an OER discovery tool?

RQ4.4: How can the intrinsic and alignments sparsity of web resources be addressed by an OER discovery tool?

RQ4.5: How can a large number of similar educational resources be presented by an OER discovery tool to support work context level tasks?

Finally, a last research question explored potential obstacles to the adoption of Discoverer (and similar tools) in an operational environment, and suggestions to overcome them:

RQ5: What barriers are restricting the deployment and uptake of OER discoverability tools and how can they be overcome?

11.2 Overview of the research

The overall goal of this research, was to contribute a solution to the problem of OERs discoverability. Following the preliminary identification of a specific task to be supported, a number of prototypes were developed, implemented, and evaluated, according to the Design Science Research methodology (Hevner et al., 2004). Designing and implementing the prototypes forced to develop the initial ideas in more detail, allowing the identification of roadblocks that could not be identified before (Winston, 1984), and contributing “valuable constraint knowledge” (Vaishnavi and Kuechler, 2015, p. 11) pruning the solution space. The formative evaluations in each iteration identified shortcomings and new challenges such as sparsity, lock-in, and sense-making, refining requirements and generating new research questions to drive the design of the incrementally improved prototype in the following iteration. This process of iteratively searching for a suitable solution was documented in detail, contributing to support the credibility of the results obtained (Gregor and Hevner, 2013). A final activity collected additional feedback about the research and its broader context, from prominent experts in the OER field – which supported the previous results, and helped to identify potential barriers and possible strategies for its uptake. Figure 11.1 shows the main research activities, briefly summarized in the following subsections.

What are the main tasks associated with OER discovery and how can educators be supported in performing them?

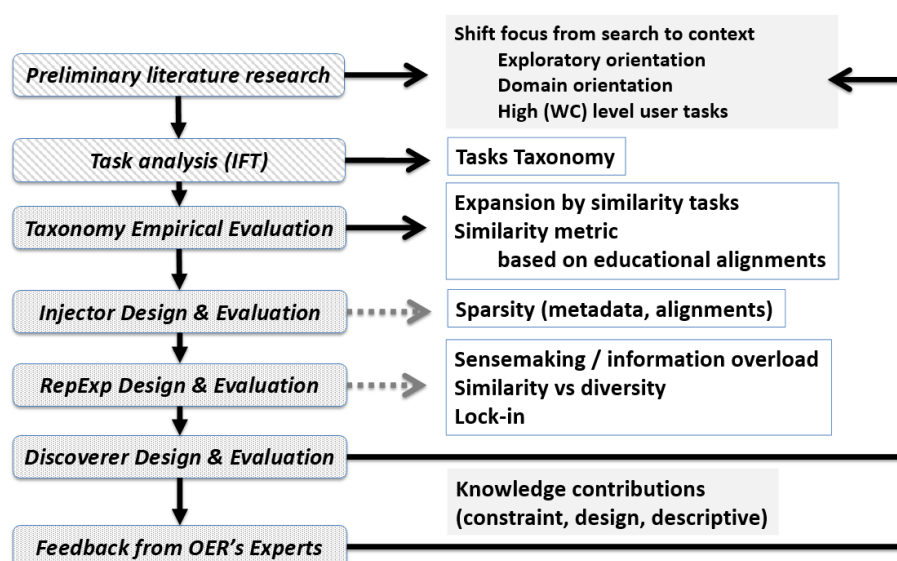


Figure 11.1 – Main research activities.

11.2.1 Preliminary literature analysis: the research gaps

The preliminary literature analysis identified the main areas of research involved, and the research gaps. In particular, the literature advocates the need to shift focus from search to the context where search problems are originated (Qu and Furnas, 2008; Wilson et al., 2010), to support:

- exploratory search,
- domain orientation,
- Work-Context level tasks.

11.2.2 OER-search Task Analysis

The next step was the analysis of educators' tasks related to OER discovery. This activity was based on a review of the scientific literature, relevant standards, and existing applications, interpreted with the Information Foraging Theory (Pirolli and Card, 1999). The result of this activity was organized as a task-taxonomy.

11.2.3 Empirical evaluation of the taxonomy of OER-search tasks

The empirical analysis of quantitative and qualitative data about the taxonomy, collected via a survey among a small (nine) sample of experienced educators, made it possible to understand habits, priorities, and an unanticipated thinking strategy of educators when looking for educational resources. These first activities preliminarily identified the general requirements and challenges to kick-off the iterative design, development, and evaluation of a sequence of prototypes.

11.2.4 Injector: design and evaluation

A first prototype, Injector, was designed and implemented. It identifies educational resources in Google SERPs (Search Engine Result Pages), where it injects educational metadata and expansion by similarity functionalities. Expansion functionalities allow educators to discover similar resources, identified via a novel similarity metric based on the number of shared alignments to educational standards (that is, with the available dataset, learning objectives).

Injector was evaluated via a discounted heuristic evaluation with a small number (four) of experienced educators. A discounted evaluation is considered appropriate for an early prototype (Nielsen, 1995). The objective was to identify possible design deficiencies, as well as to obtain preliminary feedback concerning the relevance of the problem addressed and the suitability of the proposed solution. While the evaluation showed a positive reaction by test-users towards the proposed strategies in general, it efficiently revealed the major challenge of sparsity: the modest number of educational resources within generic Google SERPs (intrinsic sparsity), and the even more limited number of resources that could be expanded (alignments sparsity).

11.2.5 RepExp: design and evaluation

The challenge of sparsity was addressed in a second prototype: RepExp. A first component of this prototype, Replicator, automatically replicates initial Google keywords searches in a specialized educational metadata repository, to increase the relevance of the starting SERP, and therefore greatly

reducing the problem of sparsity. A second component, Expander, offers users the possibility of discovering resources similar to any resource they are currently visiting, with a Query By Examples approach. The “sample” resource to be expanded, can be previously identified via Replicator, or, very important, any other way. Similar resources are identified again via the similarity metric exploiting alignments to educational standards.

Once again, a heuristic evaluation of the prototype with six representative users efficiently identified shortcomings and areas for improvements. This prototype solved the problem of sparsity generating in some cases a large number of results. However, the analysis of test-users’ feedback, again very positive concerning the general approach, identified new challenges that could not be identified before. In particular, the analysis identified the need to (1) give educators some control on the degree of similarity of the identified resources, balancing similarity and diversity to support better their high-level tasks (2) support educators making sense of a frequently large number of hits, getting a better overall picture, and (3) avoid that users get locked-in the same group of resources (foraging patch) following iterative expansions.

11.2.6 Discoverer: design and evaluation

These new challenges were addressed by the final prototype, Discoverer. This prototype presents the potentially large amount of similar resources in clusters of comparable levels of similarity, expandable under user control, with an improved interface.

Given that the design had stabilized, a heuristic summative evaluation of this final prototype was carried out with a larger sample of twenty-nine educators from three different countries. The goal was to collect additional evidence about the relevance of the tasks supported and the suitability of the solution proposed (Venable et al., 2012): a strategy to provide educational resources discovery functionalities, in particular an expansion functionality to identify “more like this” similar resources, starting directly from Google SERPs, or from any other search engine, or even from any educational resource being explored. This expansion functionality makes use of a strong domain-oriented similarity metric, based on the number of shared alignments to educational standards, rather than just generic co-occurrence of keywords. The prototype, working as a hybrid search / recommendation system under user control (Chi, 2015), offers educators the possibility of selecting the desired degree of similarity, balancing similarity and diversity to support their high-level Work-Context tasks.

11.2.7 Feedback from representative experts in the OER ecosystem

While the previous research activities focused exclusively on educators, the targeted users of Discoverer, this new activity collected further feedback from expert stakeholders, having vast experience in the domain of OER use and provision, more representative of the wider OER ecosystem. The aim was to further validate the solutions proposed by Discoverer, and understand what are the barriers and possible resolving strategies for the adoption of Discoverer, or similar tools, in an operational environment. Feedback was obtained from seven high-profile participants, through

semi-structured interviews and additional interactions via emails and forum, and analysed with thematic content analysis.

The themes were organized at three levels of abstraction. At micro level, the relevance of the problem addressed by Discoverer, the suitability of the strategies adopted, and the solutions proposed, were confirmed, offering suggestions for extensions. These include the opportunity to integrate explanations, and to provide easily interpreted information about the level of openness of the resources identified. CC Search, in particular, was identified as a suitable operational environment where Discoverer could be adopted in the future. At a meso level, the symmetrical challenges related to metadata standardization, production, and distribution, were discussed – in particular in relation to the strategies adopted in this research. At a macro level the current fragmentation of the OER ecosystem was recognized as a major challenge, and possible directions to improve its sustainability were proposed. These included the use of strategies to minimize further fragmentation while increasing reusability, as targeted in this research, and more coordination among the plethora of initiatives, including commercial activities.

The results fully supported the findings from the previous studies and – concerning broader aspects – were coherent with the literature, in particular with the Ljubljana Action Plan (UNESCO, 2017). This final activity helped to better situate this research in its broader, very dynamic, and complex context – and helped to shape a detailed chapter of this dissertation about future research activities (Chapter 10). This activity also provided the opportunity to establish promising contacts, and feedback a few ideas from this research to a key network shaping the evolution of the OER ecosystem.

11.3 Summary of findings

Gregor and Hevner (2013) discuss in considerable detail the potential knowledge contributions produced by a DSR study, focusing on prescriptive knowledge about the artificial world produced by human creativity. At the lower level of abstraction, Level 1, this research has contributed instantiations, that is prototypes that represent a situated implementation of an artefact, that “operationalize constructs, models, and methods” (March and Smith, 1995 p. 258). At a higher level of abstraction, Level 2, the research has contributed, for example, the definition of a construct in terms of a novel similarity metric, methods to compute it, and models such as the architecture of the adopted solution. At the highest abstraction Level 3, the research has contributed a coherent body of knowledge describing and justifying a solution to a problem, that Gregor and Hevner (2013) consider a design theory, one of the possible types of theory as classified by Gregor (2002).

These aspects have been addressed in detail in the previous chapters, while the most important findings, in both domains of descriptive and prescriptive knowledge, are summarized here, as answers to the original research questions. Most findings relate to more than one question, because they provide incremental answers to multiple ones. To simplify the understanding, Figure 11.2 provides an overview of the mapping between research questions and main results.

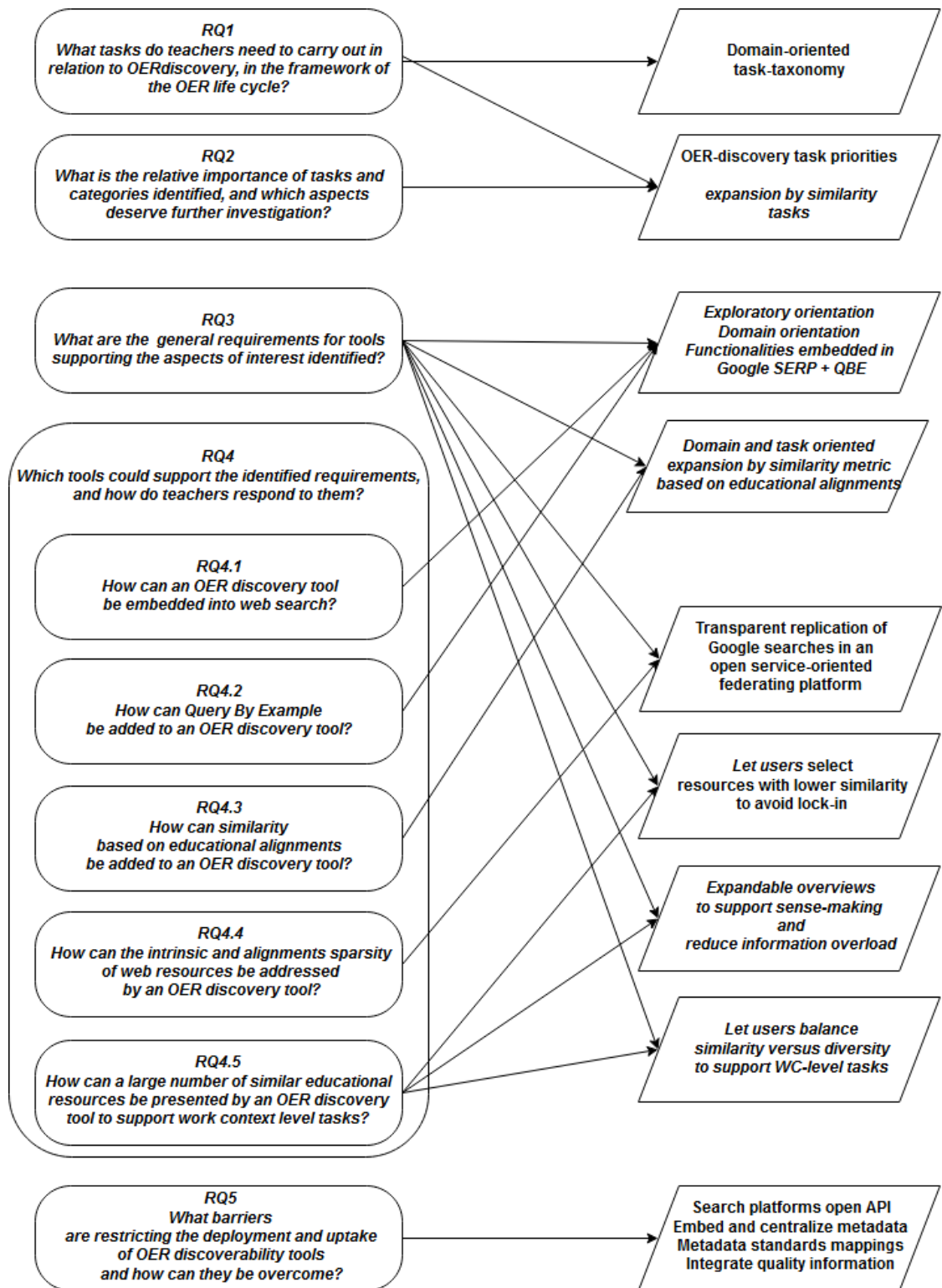


Figure 11.2 Mapping between research questions and main results.

11.3.1 RQ1: What tasks do teachers need to carry out in relation to OER discovery, in the framework of the OER life cycle?

An open domain-oriented taxonomy of OER discovery-related tasks was derived, from the analysis of the literature, existing applications, and metadata standards, interpreted with Information Foraging Theory (Pirulli and Card, 1999). This taxonomy, discussed in Chapter 4, highlights in particular the QBE (Query By Examples) “expansion” category of tasks, to discover resources related to a

previously identified resource by a relatedness metric (Wilson et al., 2010). It also highlights the importance of supporting the domain-oriented high-level WC (Work Context) tasks of educators (category “using”).

The empirical evaluation of the taxonomy with a few experienced educators, supported its validity. More interestingly, the evaluation identified new tasks in the “expansion” category, such as “Discover resources having the same learning objectives as a given sample resource”. The new tasks were implicitly foreseen in the original task-taxonomy, and could be even carried out, indirectly, in proposed or existing systems. Yet, educators were required to decompose their original natural task in non-intuitive, time-consuming, complex sequences of lower-level specialization, generalization, and navigation operations. Supporting these tasks at a declarative level, rather than requiring a complex procedural approach, fits more directly the task-oriented thinking strategy of educators.

This open taxonomy was instrumental for the following studies, but it is also a reusable component (Geisler, 2008) which constitutes a first, even if modest, contribution of this research.

11.3.2 RQ2: What is the relative importance of tasks and categories identified, and which aspects deserve further investigation?

A set of weights, quantifying the importance attributed by educators to the elements of the task-taxonomy, was obtained by means of a preliminary survey among experienced educators. It showed in particular that the QBE (Query By Examples) “expansion” by similarity category of tasks (Wilson et al., 2010) is considered important by educators, to complement the more commonly addressed “filtering” category (Hearst, 2009). The weights, discussed in Chapter 5, were collected to help deciding which tasks to address in the subsequent activities of this research. However, they also constitute a second reusable component, which could be used, for example, as a usability metric in a framework for the evaluation of applications for search/discovery of educational resources (Agarwal and Venkatesh, 2002).

11.3.3 RQ3: What are the general requirements for tools supporting the aspects of interest identified?

Preliminary general requirements were identified from the literature analysis and the taxonomy empirical evaluation. However, the answer to this question was enriched by subsequent activities, as shown in Figure 11.2.

A first requirement is to address, as much as possible, high-level and domain oriented tasks, such as planning for a lesson or finding educational resources for a remediation activity.

A second requirement is to let educators start their search for educational resources, from the familiar Google environment, that is, its SERPs (Search Engine Resource Pages). Google indeed, as widely reported in the literature (LRMI, 2013b; Abeywardena et al., 2013), and confirmed by this study, is largely the most popular starting point for, and frequently the only tool used by, most educators looking for educational resources. A possible solution could transparently replicate Google searches

on meta-search-engines federating services from different portals. This is an area where many patents have been deposited (Brette et al., 2016; Mowatt et al., 2012). A similar alternative patent-free solution successfully experimented in this research, is to access transparently a service (in this case the Learning Registry) federating metadata from different sources, including major repositories of educational metadata.

A third requirement is to support QBE style of searches, using a sample resource as the example, to find similar resources.

11.3.4 RQ4: Which tools could support the identified requirements, and how do teachers respond to them?

This research question was subdivided in sub-questions from RQ4.1 to RQ4.5, discussed below.

11.3.5 RQ4.1: How can an OER discovery tool be embedded into web search?

Test-users strongly appreciated the availability of specialized OER discovery functionalities “directly” from Google result pages, avoiding the time-consuming hopping among specialized portals. The alternative of hopping from one specialized platform to another is not acceptable anymore because, as previously happened with the Web, the situation “has grown beyond that scale of the surfable” (Wilson et al., 2010, p. 5).

11.3.6 RQ4.2: How can Query By Example be added to an OER discovery tool?

The general approach of finding resources related by various relatedness metrics with a QBE approach was considered useful by test-users to support their educational activities. The possibility to expand a sample resource, identify another resource, and expand it again iteratively, was particularly appreciated for its exploratory orientation. However, not every similarity metric was considered equally useful. Indeed, a number of educators (11 over 29) expressed partially negative or negative opinions about metrics based on social data. While this was consistent with the results from the taxonomy empirical evaluation, it should not diminish the importance of addressing these aspects. Indeed these data can be very useful to obtain information about the quality of the resources, as discussed in the literature review and confirmed by the experts consulted. Yet, the proposed metric based on educational alignments was considered of higher priority by the educators involved in the evaluation activities of this research, as there was a much firmer consensus on its benefits to support their activities.

11.3.7 RQ4.3: How can similarity based on educational alignments be added to an OER discovery tool?

A similarity metric has been defined, to identify and rank educational resources, based on the number of alignments to educational standards (the so called “schema.org / LRMI killer feature”) they share. This research shows that the metric is strongly domain-oriented, as targeted (Qu and Furnas, 2008), and definitely more appreciated by educators than generic keywords-matching (or other metrics indeed).

As this metric was used in a prototype aiming to support users in their broad WC tasks, its evaluation showed – as expected in these cases (Wilson et al., 2010) – some dissimilarities related to different organizational and cultural aspects. In particular, educators in contexts where learning objectives are not formally used, definitely appreciated the proposed resource-discovery strategy to make “transparent” use of alignments to educational standards metadata. The empirical results of this research suggest that this is because these educators are reluctant to use formal educational alignments explicitly, but they do appreciate their potential to target resources more precisely.

Yet, the use of learning outcomes and competency-based frameworks is expanding (European Commission, 2011), in education as well as by employers and professional bodies (Lane, 2017b), with the main objective to harmonize qualifications at international level (Davies, 2017). Thus, it can be expected that educators (and other stakeholders) will become increasingly familiar with formal learning objectives in the future (Adam, 2004). Additionally, as discussed in Chapter 9, some advanced users, once familiarized with the system, might likely want explanations of the reasons why resources were suggested, in order to better control the system: in this case the learning objectives will have to be exposed.

The research provided some evidence that the appreciation for this similarity metric can be generalized also to contexts where educators are, on the contrary, quite familiar with educational alignments. Arguably, these educators are inclined to characterize learning resources more in terms of their measurable learning objectives, rather than their contents descriptions as used in the past. As a result, they naturally consider two resources similar, more because they share learning objectives, rather than just because they share keywords.

The different cultural background of educators teaching different subjects such as humanities, science, or technologies, could also be expected to influence their degree of appreciation of the similarity metric proposed. Yet, no measurable differences could be identified in this regard.

Finally, it is important to note that the proposed metric supports educators at WC-level, as targeted by one of the general requirements. Indeed, it is appreciated, for example, in lesson planning, to explore alternative pedagogical strategies to reach similar educational objectives.

11.3.8 RQ4.4: How can the intrinsic and alignments sparsity of web resources be addressed by an OER discovery tool?

The strategy to identify automatically OERs directly within Google SERPs was much appreciated in principle, but proved to be practically ineffective because of “sparsity”. Indeed, there is necessarily an intrinsically limited number of educational resources available in generic Google results pages (intrinsic sparsity). Moreover, attempting to identify additional resources similar to those identified in Google SERPs with the proposed similarity metric was also not very effective. Indeed, the limited (so far, at least) availability of alignments to educational standards (alignments sparsity) for these resources, makes their expansion with the adopted metric rarely possible.

It was found that the challenge of sparsity can be effectively mitigated by transparently replicating the initial Google keywords search in specialized hubs of educational resources. While this solution does not really identify educational resources within Google SERPs, users are nonetheless largely satisfied by the possibility of obtaining a large number of relevant resources transparently, without even realizing that these are not retrieved by Google.

The amount of resources identifiable mainly depends on the dataset available in this specialized hub: the Learning Registry utilized in the prototype was the largest publicly available dataset which had the required metadata.

11.3.9 RQ4.5: How can a large number of similar educational resources be presented by an OER discovery tool to support work context level tasks?

While the direct objective of a similarity metric is to identify similar resources, it is not convenient to offer users exclusively the most similar ones. While these resources do not necessarily risk to be nearly identical, because they are similar only in terms of educational alignments, users who keep iteratively expanding similar resources may get nonetheless locked in a strongly connected group of resources. These groups correspond to near classes of equivalence, or “foraging patches” in terms of Information Foraging Theory (Pirolli and Card, 1999). This makes further expansions of resources in these groups ineffective, because it leads to the same group of resources. A solution to this problem is to offer users the possibility of selecting resources with different degrees of similarity. In this way, users can easily avoid to get locked in a patch, by expanding resources with a lower degree of similarity.

Producing potentially large result-sets, requires specific strategies to reduce information overload and support users in making sense of the available data. This can be achieved by offering users a compressed view (Alam and Sadaf, 2013), that is a small number of resources as representative as possible of the whole set. In the case of the similarity metric adopted, this can be conveniently achieved by clustering results (Jain, 2010) in a few classes of homogeneous degrees of similarity, and presenting users with a few representative resources for each group, with the possibility of expanding further each group.

Finally, the evaluation of Discoverer provided evidence that to support educators in their Work-Context tasks, in particular lesson planning and personalized instruction, it is convenient to let them control the degree of similarity of the resources identified, balancing similarity and diversity (Bradley and Smyth, 2001). Indeed, different educational goals require different degrees of similarity. For example, resources for an in-depth activity require a low similarity with resources previously used, while resources for a remediation activity require a higher similarity. This feature was particularly appreciated, especially by senior educators, for its suitability to support the personalization of educational activities.

11.3.10 RQ5: What barriers are restricting the deployment and uptake of OER discoverability tools and how can they be overcome?

The evidence collected suggests that some of the strategies adopted in this research, can support the adoption of Discoverer and similar tools in an operational environment. These include in particular the strategies of developing reusable components to be integrated on top of existing search engines, and exploiting the open services of federating platforms. Potential barriers to the adoption of these tools, as well as suggestions to overcome them, were identified. As Discoverer is a metadata consumer application, a major barrier to its adoption is linked to the symmetrical challenge of the production of metadata which are needed to power it. Another major barrier is linked to the current fragmentation of the OER ecosystem, characterized by many competing search platforms and metadata standards, which calls for strategies to support their mutually convenient coexistence.

Extend the definition of “Openness” for OERs, to encompass metadata

Metadata are instrumental to the discoverability of OERs: while metadata are somewhat “invisible” compared to content and licence, they are equally important. Arguably, OERs should not be considered really “open”, if they are not accompanied by suitable metadata to support their own discoverability. Therefore, to raise awareness about the importance of metadata, the definition of OER should be updated to explicitly encompass metadata. However, while documenting the resources with the necessary metadata could be made mandatory in some cases, it is also important to consider that this might disincentive potential authors, especially authors of “little” OER (Weller, 2010), to make their material openly available.

Reduce the ineffective proliferation of metadata standards

Metadata standards are instrumental to make the resources discoverable, but the unrelenting attempts to develop an ideally single universal standard, keeps fragmenting the existing solution space with an unhealthy amount of old and new standards.

We may well ask whether a single metadata standard will ever be developed and adopted. For example, it is unlikely that the same standard can be conveniently used by professional librarians, as well as casual developers of “little OERs”. Rather than wasting resources trying to develop this ideal universal standard, a better solution could be to identify use cases which are not satisfactorily supported by existing standards, to provide feedback for their improvement. An example use case is the need to track and document the history of modifications to OERs, which educators constantly adapt (Dietze et al., 2013), and even learners may revise and improve in the context of OER-enabled pedagogy activities (Wiley, 2017; Wiley and Hilton, 2018). At the same time, suitable mappings/heuristics could be identified to automatically “translate” among different standards. Corresponding tools could conceivably be crowdsourced as open software.

Reduce the unhealthy proliferation of search platforms

Similarly to the proliferation of metadata standards, the current proliferation of competing search platforms for educational resources reduces their benefits. This is why, as reported in the literature and confirmed by this research, most educators reluctantly revert to the use of Google.

A first solution to reduce this proliferation, experimented in this research, is by developing reusable components, fostering collaboration and incremental synergetic contributions rather than competition. A second solution, again experimented in this research, is to provide access to distributed metadata through open API of service-oriented architectures. This makes it possible to provide a one-stop centralized search service, transparently federating other search engines, like the Learning Registry or, in the future, CC Search. This can create a fertile environment where additional contributors may integrate innovative services on top of existing ones, enriching them rather than competing with them.

Provide both centralized and embedded metadata

The schema.org solution to embed educational metadata in the resources web pages, has the advantage to enable the interpretation of the semantics of the text in the resource pages, making the resources self-describing, hence more easily discoverable also by traditional search engines (Barker and Campbell, 2016a).

Yet, crawling all these metadata from an ever-increasing number of very dynamic sources, requires huge crawling efforts. While this is feasible for giants like Google, it may be impracticable for many smaller potential stakeholders. Therefore, an effective complementary solution, that this research demonstrated to be effective, is to additionally provide a centralized access to these educational metadata in one or a few metadata repositories providing basic search services. This would allow third parties to contribute additional services on top of the basic services they provide.

11.4 Implications and relevance of findings

The scientific literature advocates the need for further research on search/discovery strategies that support exploratory forms of search, are domain-oriented, and support users in their WC-level tasks. Wilson et al. (2010, p. 82), for example, claim that “there have been far fewer advances that consider the higher-level work contexts of searchers”, and advocate to “focus on the work contexts of users, and in supporting them to achieve their higher-level goals”. Search should not be considered in isolation, but in the context of high-level user tasks: supporting users at WC-level should be the ultimate goal of any specialized search application.

As discussed in this dissertation, the strategies proposed in the prototype do support WC-level tasks, in particular lesson planning for junior educators, and personalized education for more senior educators. The strategies proposed do support exploratory search too, as indicated by the evidence obtained from the evaluation of Discoverer, and the feedback from the experts interviewed.

Therefore, it is reasonable to expect that future search engines for educational resources will implement at least some part of the ideas developed in this research. This is supported by the feedback from test-users involved in the studies, by the interest expressed by the Learning Registry to implement the prototype on their infrastructure (Midgley, 2016), and by the feedback obtained from OER experts.

There is currently a limited availability of metadata describing OER, in particular with respect to educational alignments. This lack of metadata disincentivizes any effort to develop advanced search applications based on metadata. The lack of search applications exploiting metadata, in turn, disincentives any investment in producing metadata. This research, demonstrating a successful use of educational metadata, should contribute to renewed attention on the possibilities of such approaches. If the deadlock can be broken, this strategy can foster a virtuous circle of publishing and exploiting educational metadata.

11.5 Research limitations

This section summarizes and discusses the main limitations of this research. As reported in Chapter 10, considerations about these limitations led to the detailed definition of further research activities.

11.5.1 Generalizability of results

The relatively limited number and representativeness of participants involved in this research, is a limitation which may raise concerns. Certainly the collection of additional data, from a variety of settings, would improve the confidence in the generalizability of the findings. However, this research involved a total of 55 participants in the various studies. As a comparison, the initial survey to drive the development of LRMI (2013b), an endeavour with huge resources compared to this PhD, was sent to about 30,000 users taking advantage of the services of a specialized firm, but answers were received from little more than 150 respondents. Overall, the data collectively obtained from the 55 participants in this research, were very consistent among the different studies, quantitative and qualitative data were mutually reinforcing within the studies, and the results were cross-checked with prominent experts having broad expertise in the OER ecosystem.

More specifically, the first study (Task Analysis in Chapter 5) 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 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 (4 and 6) involved in the first two DSR studies (Injector in Chapter 6 and RepExp in Chapter 7) were in line with the recommendations from the literature (Nielsen, 1995), and were sufficient to reach the objective of spotting 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 of Discoverer (Chapter 8) 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 robust 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 sizes, suggesting strength of statistical claims. The last activity (experts interviews in Chapter 9), involving 7 additional participants having a broad expertise in the OER ecosystem, further confirmed the previous focused results and helped to situate them in their broader context.

11.5.2 Universality of needs among educators with different profiles

A second limitation concerns the degree of universality of needs among educators with different profiles, because the evidence to support the research findings in the first studies was mainly collected from high school educators from Italy, and to a limited extent from UK and Brazil. However, this issue was ameliorated by the additional input collected from OER experts, more representative of the broader OER ecosystem.

The analysis (Section 8.3.4) tested but did not identify any difference among educators teaching different topics, such as technical, maths or humanities. Minor differences were identified among educators from the three countries, attributed to their different familiarity with formal learning objectives. Italy and Brazil can be considered representative of a class of countries that have reduced experience in using formal learning objectives, while the UK can be considered representative of the class of countries that have more experience. 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, cannot work in 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. Yet, different domains, or even different organizational and cultural contexts in the same domain, might raise partially different requirements. Possibly, as suggested by one of the experts interviewed, they might require the adoption of different metrics, for example based on the combination of reading difficulty and topic, or on crowdsourced tags.

11.5.3 Breadth of functionalities explored

Another limitation of this research, is its focus on a specific operation, expansion by similarity, and a specific similarity metric based on educational alignments. Of course, a full-fledged system should

support a broader set of tasks, and possibly allow users to select different similarity metrics (Wilson et al., 2010).

Indeed, the need for additional functionalities, such as filtering, was already evident from the taxonomy empirical evaluation, and was explicitly advocated by a few test-users (e.g. User11 in Section 8.3.2) of the prototypes, as well as experts interviewed (e.g. PartC in Section 9.3.1). Yet, as discussed in Chapter 1, this research could not tackle the huge problem of OERs discoverability as a whole, but had necessarily to focus on a critical sub-problem, aiming to propose a reusable solution in order to avoid further fragmentation of the solution space.

Additional potentially useful similarity metrics were also identified in the task analysis. Yet the taxonomy empirical evaluation and the evaluation of the prototypes developed in this research, provided quantitative and qualitative evidence that the proposed metric was the most interesting one for the educators consulted.

11.5.4 Depth of evaluation

Educators participating in the evaluation were very positive about the features proposed in the prototype. The feedback collected from the experts interviewed, was equally positive. However, the evaluation mainly collected subjective measures from educators, and their reaction was based on a brief exposure to the prototype. Therefore, a couple of test-users suggested that they needed to use the system on their job for a while, before expressing a more definitive opinion.

This is consistent with the proposed strategy which aims to support educators in their WC-level tasks (e.g. lesson planning). Test-users suggested that the prototype is very relevant and useful to support their WC-level tasks, and some test-users explicitly expressed the opinion that it would improve the efficiency and quality of their work. However, a more thorough evaluation could attempt to measure objectively the actual impact on the work of educators.

Evaluation at this level is very challenging (Kules and Shneiderman 2008) because there are many variables influencing the output, and it is very difficult to isolate their effect (Wilson et al., 2010). Also, performances cannot be easily measured quantitatively, such as with task duration, precision or recall in Information Retrieval studies. For these reasons, these systems are frequently evaluated by collecting reaction feedback in usability studies, as it was carried out in this research. The need for usability studies is also strongly advocated by Karger (2014) in the area of the Semantic Web.

11.5.5 Data-set representativeness

A more thorough evaluation of the actual support which the prototype provides at task-level, where people would use it in their actual activities for prolonged periods of time, would require a data-set covering as many educational resources available on the Internet as possible. The data-set used in the research contained more than 600,000 resources, but only about 10% of them were annotated with the required alignments metadata. Yet, while these data cover only a portion of the educational

resources available on the Internet, they were sufficient to carry out the evaluations in a sufficiently realistic setting.

11.5.6 Open versus blind evaluation

The participants involved in the evaluations of this research were aware of their objectives and could guess the expected outcomes. This situation might have affected their reaction, so that a better set-up could have engaged participants in a blind study, to reduce potential bias. For example, Discoverer could have been compared to an alternative (baseline) solution, dividing evaluation participants in two groups ignoring which system they were using. This type of evaluation was not carried out, because an existing comparable baseline solution could not be identified (Section 10.1.1). However, most participants were totally unrelated with the researcher, and were frequently prompted to be as critical as possible, so that they had no reasons, at least intentionally, to modify their reaction to “please” him.

11.6 Future research

While this research has produced some relevant outcomes, it has opened up more opportunities for further research. These were discussed in depth in Chapter 10, and are only briefly summarized here.

The first line of enquiry proposed additional evaluations, to test the universality of the requirements identified, and the suitability of the solutions proposed. These evaluations, in addition to cover a broader sample of educators, more representative of different domain and cultural contexts, could – in particular – evaluate the impact of the strategies proposed on productivity and quality at Work-Context level. To this end, it was suggested to use, as a suitable configuration, an instrumented version of Discoverer integrated on top of CC Search, exploiting its future open API.

As a second line of enquiry, a number of micro-level enhancements to the current prototype were proposed, such as its deployment as a server-side metasearch engine in a service oriented architecture. A number of improvements and extensions to the current similarity metric were identified and discussed, such as the normalization of generalizations and specializations in the educational frameworks, the use of improved similarity measures such as Jaccard similarity, or even its extension to a bi-dimensional version. Other extensions discussed include the integration of quality indicators based on paradata and social-data, as well as the possibility to provide explanations on-demand – offering users more control on the system.

Another line of enquiry suggested meso-level extensions, such as additional similarity metrics exploiting different data or even just different types of alignments, the extension of clustering to the initial keywords-based search, as well as search personalization. The critical issue concerning the lack of universal identifiers associated to the resources (rather than to their stored instances) was discussed too.

Further research proposals concerned extensions at higher level, including opening up again the design space to identify possible additional requirements, such as finding resources for

interdisciplinary activities, or covering the maximum number of prerequisite learning objectives of a sample resource.

Finally, while this research focused on metadata consumption, plenty of research opportunities were identified in the symmetrical area of metadata production, which is a very resource-intensive process. In particular, an intriguing direction is the use of generic social Web annotators to crowdsource social data.

11.7 Concluding remarks

The overall goal of this research was to contribute to support educators in discovering open educational resources. The scientific literature advocates to target an inclusive view of search which is domain-oriented, exploratory-oriented, and supports high (Work Context) level tasks (Wilson et al., 2010). The positive feedback obtained from the evaluation of the prototypes developed in this research, and from experts in the OER ecosystem, provide evidence that some steps were made in those directions.

The positive feedback obtained, can possibly be appreciated by comparing the results obtainable from the prototype Discoverer, 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 supports (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 “This is a Stargate to knowledge!” While the research focused on a tiny area, mainly QBE expansion by similarity based on educational alignments, the results suggest that the proposed strategy is particularly appreciated for the support it provides to educators in their exploratory search activities, more specifically for lesson planning and personalization of education.

In conclusion, it is hoped that this thesis – which contributed design, constraint, and descriptive knowledge, as well as detailed suggestions for further research, aiming to improve current and future OER search applications – will contribute to the success of the OER movement.

12. References

- Abeywardena, I. S., Chan, C. S. and 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), pp. 214-237. DOI: 10.19173/irrodl.v14i4.1505.
- Abraham, A. (2013) *Information Seeking from Web-based Resources: Sensemaking Strategies and Implications for Interaction Design*, PhD thesis, The Open University, UK.
- Adam, S. (2004) 'Using Learning Outcomes', *Report for United Kingdom Bologna Seminar*, Heriot-Watt University (Edinburgh Conference Centre), Edinburgh.
- Adida, B., Birbeck, M., McCarron, S. and Herman, I. (2015) 'RDFa Core 1.1 - Third Edition 17 March 2015', *W3C Recommendation* [Online]. Available at <http://www.w3.org/TR/rdfa-core/> (Accessed 6 June 2015).
- Agarwal, R. and Venkatesh, V. (2002) 'Assessing a firm's web presence: a heuristic evaluation procedure for the measurement of usability', *Information Systems Research*, **13**(2), pp. 168-186.
- Alam, M. and Sadaf, K. (2013) 'A Review on Clustering of Web Search Result', *Advances in Computing and Information Technology. Advances in Intelligent Systems and Computing*, vol. 177, Berlin, Heidelberg, Springer.
- Al-Khalifa, H. S. and Davis, H. C. (2006) 'The evolution of metadata from standards to semantics in E-learning applications', *Proceedings of the seventeenth conference on Hypertext and hypermedia*. New York, August. ACM, pp. 69-72.
- Anderson, T. and Leachman, C. (2019) 'Strategies for Supporting OER Adoption through Faculty and Instructor Use of a Federated Search Tool', *Journal of Librarianship and Scholarly Communication*. DOI: 10.7710/2162-3309.2279.
- Andrews, P., Zaihrayeu, I. and Pane, J. (2012) 'A classification of semantic annotation systems', *Semantic Web Journal*, **3**(3), pp. 223-248.
- Ariadne (2014) *Ariadne Linked Data interface* [Online]. Available at <https://datahub.io/dataset/ariadne/resource/f7ee160c-25db-4122-8ae2-13698b344c8c> (Accessed 01 May 2016).
- Ariadne (2016) *About Ariadne* [Online]. Available at <http://www.ariadne-eu.org/> (Accessed 03 May 2016).
- ASN (2016) *Achievement Standards Network* [Online]. Available at <http://www.achievementstandards.org/> (Accessed 01 May 2016).
- Atenas, J. and Havemann, L. (2013) 'Quality assurance in the open: an evaluation of OER repositories', *INNOQUAL: International Journal for Innovation and Quality in Learning*, **1**(2), pp. 22-34.

- Atenas, J. and Havemann, L. (2014) 'Questions of quality in repositories of Open Educational Resources: a literature review', *Research in Learning Technology*, **22**(1).
- Atenas, J., Havemann, L. and Priego, E. (2014) 'Opening teaching landscapes: The importance of quality assurance in the delivery of open educational resources', *Open Praxis*, **6**(1), pp. 29-43.
- Atenas, J. (2015) 'Model for democratisation of the contents hosted in MOOCs', *RUSC Universities and Knowledge Society Journal*, **12**(1), pp. 3-14. DOI: 10.7238/rusc.v12i1.2031.
- Atkins, D. E., Brown, J. S. and Hammond, A. L. (2007) *A Review of the Open Educational Resources (OER) Movement: Achievements, Challenges, and New Opportunities*, Report to the William and Flora Hewlett Foundation.
- Baker, N. (2018) 'It's open to all, right? Who is excluded from and by open educational resources (OERs)?', *OER18 conference*. Bristol, 18-19 April [Online]. Available at <https://oer18.oerconf.org/sessions/its-open-to-all-right-who-is-excluded-from-and-by-open-educational-resources-oers-1928/> (Accessed 16 March 2019).
- Bangor, A., Kortum, P. T. and Miller, J. T. (2008) 'An empirical evaluation of the system usability scale', *International Journal of Human-Computer Interaction*, **24**(6), pp. 574-594.
- Barab, S. and Squire, K. (2004) 'Design-based research: Putting a stake in the ground', *The journal of the learning sciences*, **13**(1), pp. 1-14.
- Barker, P. (2014) 'Explaining the LRMI Alignment Object', *Sharing and learning* [Blog]. Available at <http://blogs.pjkk.net/phil/explaining-the-lrmi-alignment-object/> (Accessed 24 January 2019).
- Barker, P. (2018) 'New publication analysing LRMI metadata on the web', *Sharing and learning*, 24 April [Blog]. Available at <https://blogs.pjkk.net/phil/analysing-improving-embedded-markup-learning-resources-web/#more-1773> (Accessed 12 August 2018).
- Barker, P. and Campbell, L. (2016a) 'Technology strategies for open educational resource dissemination', *Open education: international perspectives in higher education*, Cambridge, Openbook Publishers, pp. 51-70.
- Barker, P. and Campbell, L. (2016b) 'Learning Resource Metadata on the Web', *Proceedings of the International Conference on World Wide Web (WWW '15 Companion)*. Florence, May 18-22. New York, ACM, pp. 687-688. DOI: 10.1145/2740908.2741745.
- Bates, M. J. (1979) 'Information search tactics', *Journal of the American Society for Information Science*, **30**(4), pp. 205-214.
- Bates, M. J. (1990) 'Where should the person stop and the information search interface start?', *Information Processing & Management*, **26**(5), pp. 575-591.

- Becker, J. and Niehaves, B. (2007) 'Epistemological perspectives on IS research: a framework for analysing and systematizing epistemological assumptions', *Information Systems Journal*, **17**(2), pp. 197-214.
- 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, September 18-20. Swindon, BCS Learning & Development Ltd.
- Belkin, N. J., Marchetti, P. G. and Cool, C. (1993) 'BRAQUE: Design of an interface to support user interaction in information retrieval', *Information processing & management*, **29**(3), pp. 325-344.
- Berners-Lee, T. (2006) 'Linked Data', *Design Issues*, 27 July [Blog]. Available at <https://www.w3.org/DesignIssues/LinkedData.html> (Accessed 24 July 2017).
- Berners-Lee, T. (2007) 'Giant Global Graph', *timbl's blog*, 21 November [Blog]. Available at <http://dig.csail.mit.edu/breadcrumbs/node/215> (Accessed 10 February 2014).
- Berners-Lee, T., Hendler, J. and Lassila, O. (2001) 'The Semantic Web', *Scientific American*, **284**(5), pp. 28-37.
- Bienkowski, M., Brecht, J. and Klo, J. (2012), 'The Learning Registry: Building a Foundation for Learning Resource Analytics', *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. Vancouver, April 29 - May 2. ACM, pp. 208-211. DOI: 10.1145/2330601.2330651.
- Bizer, C., Heath, T. and Berners-Lee, T. (2009) 'Linked Data – The story So Far', *International Journal on Semantic Web and Information Systems*, **5** (3), pp. 1-22.
- Bogers, T. (2018) 'Tag-Based Recommendation', in Brusilovsky P. and He D. (eds) *Social Information Access. Lecture Notes in Computer Science*, vol 10100, Springer, Cham.
- Borlund, P. (2003) 'The IIR evaluation model: a framework for evaluation of interactive information retrieval systems', *Information research*, **8**(3).
- Bradley, K. and Smyth, B. (2001) 'Improving recommendation diversity', *Proceedings of the Twelfth Irish Conference on Artificial Intelligence and Cognitive Science*. Maynooth, Ireland, pp. 85-94.
- Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology*, **3**(2), pp. 77-101.
- Brette, M., Chevalier, P., Spivak, V. and Rankov, A. (2016) *Federated search*, US Patent US9390179 B2.
- Brin, S. and Page, L. (1998) 'The anatomy of a large-scale hypertextual Web search engine', *Computer Networks and ISDN Systems*, pp. 107-117.

- Brooke, J. (1986) *System usability scale (SUS): a quick-and-dirty method of system evaluation user information*, Reading UK, Digital Equipment Co Ltd.
- Bullough, T. J., Green, A.M. and Mannis, A. (2013) 'Kritikos: creating an engineering student community search engine for learning resources', 41st *SEFI Conference*. Leuven, Belgium, 16-20 September. Available at <https://www.kuleuven.be/communicatie/congresbureau/oud2015/congres/sefi2013/eproceedings/154.pdf> (Accessed 14 March 2019).
- Butcher, N. (2015) *A basic guide to open educational resources (OER)*, Vancouver, Commonwealth of Learning. Available at <http://oasis.col.org/handle/11599/3634> (Accessed 13 June 2019).
- Byström, K. and Hansen, P. (2002) 'Work tasks as unit for analysis in information seeking and retrieval studies', *Proceedings of the 4th International Conference on Conceptions of Library and Information Science: Emerging Frameworks and Methods (CoLIS4)*. Seattle, WA, USA, July 21-25, pp. 239-252.
- Campbell, L. (2014a) 'LRMI Implementation Case Study: ISKME OER Commons', *Open World*, 3 October [Blog]. Available at <https://lornamcampbell.wordpress.com/2014/10/03/lrmi-implementation-case-study-iskme-oer-commons/> (Accessed 3 May 2016).
- Campbell, L. (2014b) 'LRMI Implementation Case Study: Gooru', *Open World*, 14 August [Blog]. Available at <https://lornamcampbell.wordpress.com/2014/08/14/lrmi-implementation-case-study-gooru/> (Accessed 3 May 2016).
- Campbell, L. (2014c) 'LRMI Implementation Case Study: Merlot', *Open World*, 15 August [Blog]. Available at <https://lornamcampbell.wordpress.com/2014/08/15/lrmi-implementation-case-study-merlot/> (Accessed 3 May 2016).
- Campbell, L. (2014d) 'LRMI Implementation Case Study: Curriki', *Open World*, 2 September [Blog]. Available at <https://lornamcampbell.wordpress.com/2014/09/02/lrmi-implementation-case-study-curriki/> (Accessed 3 May 2016).
- Campbell, L. and Barker, P. (2014) *LRMI Implementation: Overview, Issues and Experiences* [Online], Cetis. Available at <http://publications.cetis.org.uk/2014/1007> (Accessed 13 May 2015).
- Cechinel, C., Sicilia, M., Sanchez-Alonso, S. and García-Barriocanal, E. (2013) 'Evaluating collaborative filtering recommendations inside large learning object repositories', *Information Processing & Management*, **49**(1), pp. 34-50.
- Chi, E. H. (2015) 'Blurring of the Boundary between Interactive Search and Recommendation', *Proceedings of the 20th International Conference on Intelligent User Interfaces*. Atlanta, USA, March 29 - April 01. ACM. DOI: 10.1145/2678025.2700998.

- Cho, J. Y. and Lee, E. H. (2014) 'Reducing confusion about grounded theory and qualitative content analysis: Similarities and differences', *The Qualitative Report*, **19**(32), pp. 1-20. Available at <https://nsuworks.nova.edu/tqr/vol19/iss32/2>.
- Chrome (2016) *What are extensions?* [Online]. Available at <https://developer.chrome.com/extensions> (Accessed 01 May 2016).
- Clements, K. (2016) *Why Open Educational Resources Repositories Fail - the Contribution of Quality Approaches to the Success of Repositories*, PhD thesis, University of Jyväskylä.
- Clements, K. and Pawlowski, J. M. (2012) 'User-oriented quality for OER: understanding teachers' views on re-use, quality, and trust', *Journal of Computer Assisted Learning*, **28**(1), pp. 4-14.
- Cleverdon, C. W. (1960) 'ASLIB Cranfield research project on the comparative efficiency of indexing systems', *ASLIB Proceedings*, vol. 12, pp. 421-431. DOI: 10.1108/eb049778.
- Cobb, D. (2018) 'Placing the spotlight on Open Educational Resources: Global phenomenon or cultural guise?', *International Education Journal*, **17**(3), pp. 15-29.
- Cortinovis, R. (2016) 'Why no "LearningResource" type?', forum message to *Learning Resources Metadata Initiative*, 25 January. Available at <https://groups.google.com/d/msg/lrmi/itEgjKFwR4Q/HiauwulZrBgAJ>.
- Creative Commons (2017) *Open Education Platform* [Online]. Available at <https://groups.google.com/forum/#!forum/cc-openedu>.
- Creative Commons (2020) *About CC Search* [Online]. Available at <https://groups.google.com/forum/#!forum/cc-openedu> (Accessed 17 June 2020).
- Creswell, J. W. (2014) *Research design: Qualitative, quantitative, and mixed methods approaches*, 2nd edn, Sage Publications.
- Cyganiak, R., Wood, D. and Lanthaler, M. (2014) 'RDF 1.1 Concepts and Abstract Syntax', *W3C Recommendation* [Online]. Available at <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/> (Accessed 10 February 2014).
- Daher, J., Brun, A., Boyer, A. (2017) *A Review on Explanations in Recommender Systems* [Technical Report]. LORIA, Université de Lorraine.
- Daniel, J. (2012) 'Collaboration and Networking: The role of Open Educational Resources (OER)', *XX International Summit on Distance Education*. Guadalajara, Mexico, 27 November [Online]. Available at <http://sirjohn.ca/wp-content/uploads/2012/08/20121127Mexico.pdf> (accessed 28 July 2019).

- D'Aquin, M., Adamou, A. and Dietze, S. (2013) 'Assessing the educational linked data landscape', *Proceedings of the 5th Annual ACM Web Science Conference*. Paris, May 1-5. France. ACM, pp. 43-46.
- Davies, H. (2017) 'Competence-Based Curricula in the Context of Bologna and EU Higher Education Policy', *Pharmacy*, **5**(2), p. 17.
- DCMI (2012) *DC-Education Application Profile* [Online]. Available at http://dublincore.org/moinmoin-wiki-archive/educationwiki/pages/DC_2dEducation_20Application_20Profile.html (Accessed 2 May 2016).
- de los Arcos, B., Farrow, R., Pitt, R., Weller, M. and McAndrew, P. (2016) 'Adapting the Curriculum: How K-12 Teachers Perceive the Role of Open Educational Resources', *Journal of Online Learning Research*, **2**(1), pp. 23-40.
- Decreto del Presidente della Repubblica n. 275 (2009) *Autonomia delle istituzioni scolastiche*. Available at <http://archivio.pubblica.istruzione.it/argomenti/autonomia/documenti/regolamento.htm> (Accessed 31 July 2016).
- Dietze, S., Drachsler, H. and 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*, New York, Springer, pp. 47-75.
- Dietze, S., SanchezAlonso, S., Ebner, H., Yu, H. Q., Giordano, G., Marenzi, I. and Pereira Nunes, B. (2013) 'Interlinking educational resources and the web of data: A survey of challenges and approaches', *Program*, **47**(1), pp. 60-91.
- Dietze, S., Taibi, D., Yu, R., Barker, P. and d'Aquin, M. (2017) 'Analysing and Improving Embedded Markup of Learning Resources on the Web', *Proceedings of the 26th International Conference on World Wide Web Companion*. Perth, April 03-07. ACM, pp. 283-292.
- Doctorow, C. (2001) *Metacrap: Putting the torch to seven straw-men of the meta-utopia* [Online]. Available at <http://www.well.com/doctorow/metacrap.htm> (Accessed 15 April 2015).
- Dou, Z., Song, R. and Wen, J. R. (2007) 'A large-scale evaluation and analysis of personalized search strategies', *Proceedings of the 16th international conference on World Wide Web*. Banff, Alberta, May 8-12. ACM, pp. 581-590.
- Douglas, J. and Brickley, D. (2013) *From Structured Data to the Knowledge Graph* [Online]. Available at <http://www.allreadable.com/vid/google-i-o-2013-from-structured-data-to-the-knowledge-graph-542350.html> (Accessed 15 April 2015).
- Downes, S. (2003) *One Standard for All: Why We Don't Want It, Why We Don't Need It* [Online]. Available at http://zope.cetis.ac.uk/lib/media/one_standard.pdf (Accessed 2 May 2016).

Drabkin, R. (2016a) 'From silos to sharing: Why are open educational resources still so hard to find?', *EdSurge* [Online]. Available at <https://www.edsurge.com/news/2016-10-02-from-silos-to-sharing-why-are-open-educational-resources-still-so-hard-to-find> (Accessed 20 May 2019).

Drabkin, R. (2016b) Email to Renato Cortinovis, 3 June.

Drachsler, H., Greller, W., Fazeli, S., Niemann, K., Sanchez-Alonso, S., Rajabi, E., Palmér, M., Ebner, H., Simon, B., Nösterer, D., Kastrantas, K., Manouselis, N., Hatzakis, I. and Clements, K. (2012) 'D8.1 Review of Social Data Requirements', *Open Discovery Space project* [Online]. Available at https://dspace.ou.nl/bitstream/1820/4617/6/Pre-Final_ODS_D8.1_SocialDataRequirements.pdf (Accessed 16 December 2018).

Dziadosz, S. and Chandrasekar, R. (2002) 'Do thumbnail previews help users make better relevance decisions about web search results?', *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. Tampere, Finland, August 11-15. ACM, pp. 365-366.

Efron, B. (1979) 'Bootstrap methods: Another look at the jackknife', *The Annals of Statistics*, **7**(1), pp. 1-26. DOI: 10.1214/aos/1176344552.

European Commission (2011) *Using learning outcomes – European Qualifications Framework Series: Note 4*, Luxembourg, Publications Office of the European Union.

Ferguson, R. and Buckingham Shum, S. (2012) 'Social learning analytics: five approaches', *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge 2012*. Vancouver, April 29 - May 02. ACM, pp. 23-33.

Filtering (2017) Oxford Dictionaries, Oxford University Press [Online]. Available at <https://en.oxforddictionaries.com/definition/filter> (Accessed 23 Sep. 2017).

Frayling, C. (1993) 'Research in art and design', *Royal College of Art Research Papers series*, **1**(1).

Freund, L. and Wildemuth, B. M. (2014) 'Documenting and studying the use of assigned search tasks: RepAST', *Proceedings of the American Society for Information Science and Technology*. Wiley, **51**(1), pp. 1-4. DOI: 10.1002/meet.2014.14505101122.

Frické, M. (2012) 'A Logic for Organizing Information', in *Logic and the Organization of Information*, New York, Springer, pp. 121-163.

Geisler, G. (2008) 'Building Blocks For Rapid Development of Information Seeking Support Systems', *Position paper for National Science Foundation sponsored Information Seeking Support Systems Workshop 2008*, vol. 26, p. 27.

Gerhardt-Powals, J. (1996) 'Cognitive engineering principles for enhancing human-computer performance', *International Journal of Human-Computer Interaction*, **8**(2), pp. 189-211.

- GLOBE (2011) *GLOBE Metadata Application Profile Specification Document* [Online]. Available at http://globe-info.org/images/ap/globe_lom_ap_v1.0.pdf (Accessed 5 May 2016).
- GLOBE (2016) *GLOBE – Connecting the World and Unlocking the Deep Web* [Online]. Available at <http://globe-info.org> (Accessed 30 Jul. 2016).
- Google (2017) *What is Google Custom Search?* [Online]. Available at <https://developers.google.com/custom-search/> (Accessed 11 August 2017).
- Google (2019) ‘Providing Structured Data’, *Google Custom Search* [Online]. Available at https://developers.google.com/custom-search/docs/structured_data (Accessed 16 March 2019).
- Gooru (2016a) *Gooru* [Online]. Available at www.gooru.org (Accessed 01 May 2016).
- Gooru (2016b) *Gooru User Guide* [Online]. Available at <https://drive.google.com/file/d/0B1pQDKe7jWwEYW5HRkRTVjFyRDg/view> (Accessed 28 May 2016).
- Gregor, S. (2002) ‘Design theory in information systems’, *Australasian Journal of Information Systems*, **10**(1). DOI: 10.3127/ajis.v10i1.439.
- Gregor, S. and Hevner, A. R. (2013) ‘Positioning and presenting design science research for maximum impact’, *MIS quarterly*, **37**(2), pp. 337-355.
- Groth, P. and Moreau, L. (2013) *PROV-Overview – An Overview of the PROV Family of Documents* [Online], W3C Note. Available at <http://www.w3.org/TR/prov-overview/> (Accessed 15 July 2017).
- Guha, R. V., Brickley, D. and Macbeth, S. (2016) ‘Schema.org: Evolution of structured data on the web’, *Communications of the ACM*, **59**(2), pp. 44-51.
- Gurell, S. and Wiley, D. (2008) *OER handbook for educators* [Online]. Available at http://wikieducator.org/OER_Handbook/educator_version_one (Accessed 11 August 2017).
- Hall, M. and Clough, P. (2013) ‘Exploring large digital library collections using a map-based visualisation’, in *Research and Advanced Technology for Digital Libraries*, Berlin, Heidelberg, Springer, pp. 216-227.
- He, B., Patel, M., Zhang, Z. and Chang, K. C. C. (2007) ‘Accessing the deep web’, *Communications of the ACM*, **50**(5), pp. 94-101.
- Hearst, M. A. (2009) *Search user interfaces*, Cambridge, Cambridge University Press.
- Hearst, M. A. and Degler, D. (2013) ‘Sewing the seams of sensemaking: A practical interface for tagging and organizing saved search results’, *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval*. Vancouver, 3 October. ACM, p. 4.
- Heitmann, B. and Hayes, C. (2010) ‘Using Linked Data to Build Open, Collaborative Recommender Systems’, *AAAI Spring Symposium: Linked Data Meets Artificial Intelligence*. Palo Alto, March 22-24. AAAI Press, pp. 76-81.

- Herlocker, J. L., Konstan, J. A., Terveen, L. G. and Riedl, J. T. (2004) 'Evaluating collaborative filtering recommender systems', *ACM Transactions on Information Systems (TOIS)*, **22**(1), pp. 5-53.
- Herrmannova, D. and Knoth, P. (2012) 'Visual Search for Supporting Content Exploration in Large Document Collections', *D-Lib Magazine* **18**(7/8).
- Hevner, A. and Chatterjee, S. (2010) *Design research in information systems: theory and practice*, Vol. 22, Springer Science & Business Media.
- Hevner, A. R., March, S. T., Park, J. and Ram, S. (2004) 'Design science in information systems research', *MIS Quarterly*, **28**(1), pp. 75-105.
- Hobson, J. (2016) *Metadata schema for GoOpen* [Online]. Available at https://github.com/LearningRegistry/LearningRegistry/blob/goopen/LR/lr/schema/resource_lrmi_open.json (Accessed 15 Sept. 2017).
- Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B. B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversy, S., Evans, H., Hansen, H. and Roussel, N. (2003, April) 'Technology probes: inspiring design for and with families', *Proceedings of the SIGCHI conference on Human factors in computing systems*. Fort Lauderdale, April 5-10. ACM, pp. 17-24.
- IEEE (2002) *Draft standard for Learning Object Metadata* [Online]. Available at http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf (Accessed 10 April 2014).
- Iivari, J. (2007) 'A paradigmatic analysis of information systems as a design science', *Scandinavian Journal of Information Systems*, **19**(2).
- IMS (2015) *Interoperability Standards Helping to Transform the Digital Curriculum* [Online]. Available at www.imsglobal.org/sites/default/files/articles/sII17-051815.pdf (Accessed 10 June 2015).
- IMS (2019) *LTI Resource Search* [Online]. Available at <https://www.imsglobal.org/resource-search> (Accessed 30 July 2019).
- 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.
- Jaccard, P. (1901) 'Distribution de la flore alpine dans le Bassin des Drouces et dans quelques regions voisines', *Bulletin de la Société Vaudoise des Sciences Naturelles*, **37**(140), pp. 241-272.
- Jain, A. K. (2010) 'Data clustering: 50 years beyond K-means', *Pattern recognition letters*, **31**(8), pp. 651-666.
- Järvelin, K. and Ingwersen, P. (2004) 'Information seeking research needs extension towards tasks and technology', *Information Research*, **10**(1).

- Jesukiewicz, P. and Rehak, D. R. (2011) 'The learning registry: sharing federal learning resources', *Interservice/Industry Training, Simulation, and Education Conference*. Orlando, December 3-6.
- Kabel, S., De Hoog, R., Wielinga, B. J. and 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), pp. 348-362.
- Kaldoudi, E., Dovrolis, N., Konstantinidis, S. T. and Bamidis, P. D. (2011) 'Depicting educational content repurposing context and inheritance', *Information Technology in Biomedicine, IEEE Transactions on*, **15**(1), pp. 164-170.
- Kalz, M., Specht, M., Nadolski, R., Bastiaens, Y., Leirs, N. and Pawlowski, J. (2010) 'OpenScout: Competence based management education with community-improved open educational resources', *Proceedings of the 17th EDINEB Conference. Crossing Borders in Education and work-based learning*. Maastricht, FEBA ERD Press, pp 137–146.
- Kanwar, A. and Mishra, S. (2015) 'The impact of OER and MOOCs on ODL: an international perspective', *International Distance Education Development Forum*. Beijing China, 10 October [Online]. Available at <https://www.col.org/news/speeches-presentations/impact-oer-and-moocs-odl-international-perspective> (Accessed 02 June 2016).
- Karger, D. R. (2014) 'The Semantic Web and End Users: What's Wrong and How to Fix It', *Internet Computing, IEEE*, **18**(6), pp. 64-70.
- Kearney, M. (2017) *Measure Performance with the RAIL Model* [Online]. Available at <https://developers.google.com/web/fundamentals/performance/rail> (Accessed 13 Aug. 2017).
- Klo, J. (2011) *AMPlied Search extension for Chrome* [Online]. Available at <https://github.com/jimklo/AMPS-Chrome> (Accessed 01 May 2016).
- Knoth, P. (2015) *Linking Textual Resources to Support Information Discovery*, PhD thesis, Milton Keynes, The Open University.
- Krosnick, J. A., Kim, N. and Lavrakas, P. (2014) 'Survey research', in Reis H. and Judd C. (eds), *Handbook of research methods in social psychology (Second Edition)*, New York, Cambridge University Press, pp. 404-442. DOI: 10.1017/CBO9780511996481.020.
- Krosnick, J. A., Presser, S., Fealing, K. H. and Ruggles, S. (2015) *The future of survey research: Challenges and opportunities* [Online]. Stanford, The National Science Foundation Advisory Committee for the Social, Behavioral and Economic Sciences Subcommittee on Advancing SBE Survey Research. Available at http://www.nsf.gov/sbe/AC_Materials/The_Future_of_Survey_Research.pdf (Accessed 5 July 2017).
- Kuhn, T. (1962) *The structure of scientific revolutions*, Chicago, University of Chicago Press.

- Kules, B. and Shneiderman, B. (2008) 'Users can change their web search tactics: Design guidelines for categorized overviews', *Information Processing & Management*, **44**(2), pp. 463-484.
- Lane, A. (2013) 'The potential of MOOCs to widen access to, and success in, higher education study', *The Open and Flexible Higher Education Conference 2013*. Paris, 23-25 Oct 2013. EADTU, pp. 189-203.
- Lane, A. (2017a) 'Open Education and the Sustainable Development Goals: Making Change Happen', *Journal of Learning for Development - JL4D*, **4**(3), pp. 275-286.
- Lane, A. (2017b) 'The systemic implications of constructive alignment of higher education level learning outcomes and employer or professional body based competency frameworks', *Proceedings of The Online, Open and Flexible Higher Education Conference*. Milton Keynes, 25-27 October. EADTU, pp. 411-426.
- Lauesen, S. and Kuhail, M. A. (2012) 'Task descriptions versus use cases', *Requirements Engineering*, **17**(1), pp. 3-18.
- Learning Registry (2011) *Learning Registry FAQ* [Online]. Available at <http://learningregistry.org/faq/> (Accessed 28 May 2016).
- Learning Registry (2016a) *Learning Registry* [Online]. Available at www.learningregistry.org (Accessed 01 May 2016).
- Learning Registry (2016b) *Developers* [Online]. Available at <http://learningregistry.org/developers/> (Accessed 6 May 2016).
- Lee, T. M. (2017) *Understanding the Impact of Support for Iteration on Code Search*, PhD thesis, University of California.
- Leffingwell, D. (2001) 'Features, Use Cases, Requirements, Oh My!', *The Rational Edge e-zine* [Online]. Available at <http://www.ibm.com/developerworks/rational/library/content/RationalEdge/dec00/FeaturesUseCasesRequirementsOhMyDec00.pdf> (Accessed 3 May 2016).
- Leppink, J. (n.d.) *Do Likert scales like being averaged? Let's ask them!* [Online]. Available at: <https://www.maastrichtuniversity.nl/news-events/newsletters/article/HDpIb2dvqSEckXtCFerwIw> (Accessed 3 Jul. 2019).
- LinkedUp (2014) *Linking Web Data for Education* [Online]. Available at <http://linkedup-project.eu/> (Accessed 02 May 2016).
- Lockley, P. (2011) *pgogy/learning_registry_chrome* [Online]. Available at https://github.com/pgogy/learning_registry_chrome (Accessed 01 May 2016).
- LRMI (n.d.) *Use Cases* [Online]. Available at <http://lrmi.dublincore.net/the-specification/use-cases/> (Accessed 3 May 2016).

LRMI (2013a) *The Smart Publisher's Guide to LRMI tagging* [Online]. Available at http://www.lrmi.net/wp-content/uploads/2013/03/LRMI_tagging_Guide.pdf (Accessed 5 March 2014).

LRMI (2013b) *LRMI Survey Report* [Online]. Available at <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* [Online]. Available at <http://lrmi.dublincore.net/2013/06/12/lrmis-killer-feature-educationalalignment/> (Accessed 2 May 2016).

LRMI (2014) *LRMI Project Description* [Online]. Available at <http://wiki.lrmi.net/LRMI+Project+Description> (Accessed 5 April 2014).

Luo, T., Hostetler, K., Freeman, C., and Stefaniak, J. (2020) 'The power of open: benefits, barriers, and strategies for integration of open educational resources', *Open Learning: The Journal of Open, Distance and e-Learning*, vol. 35, no. 2, pp. 140-158. DOI: 10.1080/02680513.2019.1677222.

Mackenzie, N. and Knipe, S. (2006) 'Research dilemmas: Paradigms, methods and methodology', *Issues in Educational Research*, **16**(2), pp. 193-205.

Madan, A. and Dubey, S. K. (2012) 'Usability evaluation methods: a literature review', *International Journal of Engineering Science and Technology*, **4**(2), pp. 590-599..

Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H. and Koper, R. (2011) 'Recommender systems in technology enhanced learning', in Ricci, F., Rokach, L., Shapira, B. and Kantor, P.B. (eds) *Recommender systems handbook*, Boston, Springer, pp. 387-415.

March, S. T. and Smith, G. F. (1995) 'Design and natural science research on information technology', *Decision support systems*, **15**(4), pp. 251-266.

Marchionini, G. (2006) 'Exploratory search: from finding to understanding', *Communications of the ACM*, **49**(4), pp. 41-46.

Marshall, B., Cardon, P., Poddar, A. and 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), pp. 11-22.

Marshall, R., Reitsma and C. Samson (2016) 'Unraveling K-12 standard alignment; Report on a new attempt', *Proceedings of the IEEE/ACM Joint Conference on Digital Libraries (JCDL)*. Newark, NJ, USA, June 19-23. ACM, pp. 55-58.

Mazumdar, S., Petrelli, D., Elbedweihy, K., Lanfranchi, V. and Ciravegna, F. (2015) 'Affective graphs: The visual appeal of Linked Data', *Semantic Web*, **6**(3), pp. 277-312.

- McNamara, N. and Kirakowski, J. (2006) 'Functionality, usability, and user experience: three areas of concern', *Interactions*, **13**(6), pp. 26-28.
- Meng, W., Yu, C. and Liu, K. L. (2002) 'Building efficient and effective metasearch engines', *ACM Computing Surveys (CSUR)*, **34**(1), pp.48-89.
- Midgley, S. (2016) Email to Renato Cortinovis, 18 March.
- Miller, R. B. (1968) 'Response time in man-computer conversational transactions', *Proceedings AFIPS Fall Joint Computer Conference*. San Francisco, 9-11 December. New York, ACM, pp. 267-277. DOI: 10.1145/1476589.1476628.
- Miranda, S. and Ritrovato, P. (2015) 'Supporting learning object repository by automatic extraction of metadata', *Journal of e-Learning and Knowledge Society*, **11**(1) pp. 43-54.
- MIUR (2010) 'Decreto Interministeriale 211 del 7 ottobre 2010', *Gazzetta Ufficiale* n. 291. Available at <http://banner.orizzontescuola.it/decreto-interministeriale-211-del-7-ottobre-2010-indicazioni-nazionali-per-i-licei.pdf> (Accessed 12 July 2016).
- Molich, R. and Nielsen, J. (1990) 'Improving a human - computer dialogue', *Communications of the ACM*, **33**(3), pp. 338-348.
- Molnar, M. (2016) 'Flood of Open Education Resources Challenges Educators', *Education Week*, **35**(17), p. s8. Available at <http://www.edweek.org/ew/articles/2016/01/13/flood-of-open-education-resources-challenges-educators.html> (Accessed 17 July 2017).
- Morato, J., Sánchez-Cuadrado, S., Ruiz-Robles, A. and Moreiro-González, J. (2014) 'Visualización y recuperación de información en la web semántica', *El profesional de la información*, **23**(3), pp. 319-329.
- Mowatt, D., Krishnamoorthy, A., Patwardhan, N., Narula, P. and Laird-McConnell, T. (2012) *Federated search*, US Patent US 8321444 B2.
- Muller, M. J., Matheson, L., Page, C. and Gallup, R. (1998) 'Methods & tools: participatory heuristic evaluation', *Interactions*, **5**(5), pp. 13-18.
- Murphy, C. (1995) *Curriculum-Enhanced MARC (CEMARC): A New Cataloging Format for School Libraries* [Online], Available at files.eric.ed.gov/fulltext/ED399952.pdf (Accessed 3 May 2016).
- Nielsen, J. (1993) *Response Times: The 3 Important Limits* [Online]. Available at <https://www.nngroup.com/articles/response-times-3-important-limits/> (Accessed on 18 July 2017).
- Nielsen, J. (1994) 'Guerrilla HCI: Using discount usability engineering to penetrate the intimidation barrier', in Bias, R. G. and Mayhew D. J (eds) *Cost-justifying usability*, Orlando, FL, USA, Academic Press, Inc., pp. 245-272.
- Nielsen, J. (1995) 'Applying discount usability engineering', *Software, IEEE*, **12**(1), pp. 98-100.

- Nilsson, M. (2010) *From interoperability to harmonization in metadata standardization: designing an evolvable framework for metadata harmonization*, PhD thesis, Stockholm, KTH.
- Nielsen, J. (2012) *Usability 101: Introduction to usability* [Online]. Available at <https://www.nngroup.com/articles/usability-101-introduction-to-usability/> (Accessed 2 May 2016).
- Nielsen, J. (2013) *How many test users in a usability study?* [Online]. Available at: <https://www.nngroup.com/articles/how-many-test-users/> (Accessed on 18 July 2018).
- Norman, G. (2010) 'Likert scales, levels of measurement and the "laws" of statistics', *Advances in health sciences education*, **15**(5), pp. 625-632.
- NSDL (2015) *National Science Digital Library* [Online]. Available at <https://nsdl.oercommons.org/> (Accessed 3 May 2016).
- OERCommons (2016) *OERCommons* [Online]. Available at www.oercommons.org (Accessed 01 May 2016).
- Offermann, P., Levina, O., Schönherr, M. and Bub, U. (2009) 'Outline of a design science research process', *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*. Philadelphia, May 07-08. New York, ACM. DOI: 10.1145/1555619.1555629.
- Offersgaard, L. and Hansen, D. H. (2016) 'Facilitating Metadata Interoperability in CLARIN-DK', *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož, Slovenia, 23-28 May. European Language Resources Association (ELRA), pp. 2510-2515.
- Okada, A., Mikroyannidis, A., Meister, I. and Little, S. (2012) 'Colearning - collaborative networks for creating, sharing and reusing OER through social media', *Proceedings of Cambridge 2012: Innovation and Impact - Openly Collaborating to Enhance Education*. Cambridge, UK, 16-18 April. Milton Keynes, The Open University, pp. 60-70.
- Open University (2016) *Linked Data from The Open University* [Online]. Available at <http://data.open.ac.uk/> (Accessed 01 May 2016).
- OpenEd (2016a) *OpenEd* [Online]. Available at www.opened.com (Accessed 01 May 2016).
- OpenEd (2016b) *GoOpen, LearningRegistry, and Schema.org* [Online]. Available at <http://blog.opened.com/post/140812366424/goopen-learningregistry-and-schemaorg> (Accessed 01 May 2016).
- OpenEd (2016c) *Learning Registry to OpenEd importer* [Online]. Available at <https://github.com/openedinc/lr-oe> (Accessed 01 May 2016).

- Paek, T., Dumais, S. and Logan, R. (2004) 'WaveLens: A new view onto internet search results', *Proceedings of the SIGCHI conference on Human factors in computing systems*. Vienna, 24-29 April. ACM, pp. 727-734. DOI: 10.1145/985692.985784.
- Palinkas, L., Horwitz, S., Green, C., Wisdom, J., Duan, N., and Hoagwood, K. (2013) 'Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research', *Administration and policy in mental health*, **42**(5), pp. 533-544. DOI: 10.1007/s10488-013-0528-y.
- Papazoglou, M. P., Traverso, P., Dustdar, S. and Leymann, F. (2007) 'Service-oriented computing: State of the art and research challenges', *Computer*, **40**(11), pp.38-45.
- Peppers, K., Tuunanen, T., Rothenberger, M. A. and Chatterjee, S. (2007) 'A design science research methodology for information systems research', *Journal of management information systems*, **24**(3), pp. 45-77.
- Petre, M. and Rugg, G. (2010) *The unwritten rules of PhD research*, McGraw-Hill Education (UK).
- Petrides, R. (2013) *The selling of Open Educational Resources* [Online], Opensource.com. Available at <https://opensource.com/education/13/7/oer-selling-metadata> (Accessed 20 May 2019).
- Phelan, C. and Wren, J. (2005) *Exploring reliability in academic assessment* [Online]. Available at <https://www.uni.edu/chfasoa/reliabilityandvalidity.htm> (Accessed 04 Aug. 2016).
- Piedra, N., Chicaiza, J., Lopez, J., Martinez, O. and Caro, E. T. (2010) 'An Approach for Description of Open Educational Resources based on Semantic Technologies', *Proceedings of the IEEE EDUCON 2010 Conference*. Madrid, 14-16 April. IEEE, pp. 1111-1119.
- Pirolli, P. and Card, S. K. (1999) 'Information foraging', *Psychological Review*, **106**(4), pp. 643-675.
- Pirolli, P. (2009) 'Information Foraging', *Encyclopedia of Database Systems*, Boston, MA, Springer.
- Porter, A., McMaken, J., Hwang, J. and Yang, R. (2011) 'Common core standards the new US intended curriculum', *Educational Researcher*, **40**(3), pp. 103-116.
- Prat, N., Comyn-Wattiau, I. and Akoka, J. (2014) 'Artifact Evaluation in Information Systems Design-Science Research-a Holistic View', *Proceedings of PACIS 2014*. Chengdu, China, 24-28 June. Available at <https://aisel.aisnet.org/pacis2014/23/>.
- Pruitt, S. L. (2014) 'The Next Generation Science Standards: The Features and Challenges', *Journal of Science Teacher Education*, **25**(2), pp. 145-156.
- Pyke, G. H. (1984) 'Optimal foraging theory: a critical review', *Annual review of ecology and systematics*, pp. 523-575.
- Qu, Y. and Furnas, G. W. (2008) 'Model-driven formative evaluation of exploratory search: A study under a sensemaking framework', *Information Processing & Management*, **44**(2), pp. 534-555.

- Rattray, J. and Jones, M. C. (2007) 'Essential elements of questionnaire design and development', *Journal of clinical nursing*, **16**(2), pp. 234-243. DOI: 10.1111/j.1365-2702.2006.01573.x.
- Reichheld, F. F. (2003) 'The one number you need to grow', *Harvard business review*, **81**(12), pp. 46-55.
- Reitsma, R. F., Hsieh, P.-H., Diekema, A. R., Robson, R., and Zarske, M. (2017) 'Map- or list-based recommender agents? Does the map metaphor fulfill its promise?', *Information Visualization*, **16**(4), pp. 291-308. DOI: 10.1177/1473871616669193.
- Riley, J. (2010) *Glossary of metadata standards* [Online]. Available at http://www.dlib.indiana.edu/~jenlrile/metadatamap/seeingstandards_glossary_pamphlet.pdf (Accessed 02 May 2016).
- Ryan, G. W. and H. Bernard, R. (2003) 'Techniques to identify themes', *Field methods*, **15**(1), pp. 85-109.
- Sacco, G. (2000) 'Dynamic taxonomies: a model for large information bases', *Knowledge and Data Engineering, IEEE Transactions on*, **12**(3), pp. 468-479.
- Sarwar, B., Karypis, G., Konstan, J. and Riedl, J. (2001) 'Item-based collaborative filtering recommendation algorithms', *Proceedings of the 10th international conference on World Wide Web 2001*. Hong Kong, 1-5 May. ACM, pp. 285-295.
- Schema.org (2013) *Frequently Asked Questions* [Online]. Available at <http://schema.org/docs/faq.html> (Accessed 10 June 2015).
- Schraefel, m. c. (2009) 'Building Knowledge: What's beyond Keyword Search?', *Computer*, **42**(3), pp. 52-59.
- Sclater, N., Peasgood, A. and Mullan, J. (2016) *Learning analytics in higher education* [Online]. Bristol, JISC. Available at <https://www.jisc.ac.uk/reports/learning-analytics-in-higher-education> (Accessed 20 May 2017).
- Sengupta, S. (2013) 'Facebook unveils a new search tool', *The New York Times*, January, 14.
- Shear, L., Means, B., and Lundh, P. (2015) *Research on Open: OER Research Hub Review and Futures for Research on OER* [Online], Menlo Park, SRI International. Available at <https://www.hewlett.org/wp-content/uploads/2016/08/OERRH%20Evaluation%20Final%20Report%20June%202015.doc> (Accessed 30 July 2019).
- Silveira, I. F. (2016) 'OER and MOOC: The need for openness', *Issues in Informing Science and Information Technology*, **13**, pp. 209-223. Available at <http://www.informingscience.org/Publications/3478>.

- Singhal, A. (2001) 'Modern information retrieval: A brief overview', *IEEE Data Eng. Bull.*, **24**(4), pp. 35-43.
- Singhal, A. (2012) 'Introducing the knowledge graph: things, not strings', *Official Google Blog* [Blog]. Available at <http://googleblog.blogspot.co.uk/2012/05/introducing-knowledge-graph-things-not.html> (Accessed 24 January 2019).
- Slocum, T. A., McMaster, R. B., Kessler F. C. and Howard H. H. (2009) *Thematic Cartography and Geovisualization*, Third Edition, Upper Saddle River, New Jersey, Pearson Education.
- Smyth, B. and McClave, P. (2001) 'Similarity vs. Diversity', *International conference on Case-Based Reasoning Research and Development*. Vancouver, July 30 - August 2. Heidelberg, Springer, pp. 347-361.
- Stefaner, M., Ferré, S., Perugini, S., Koren, J. and Zhang, Y. (2009) 'User interface design', in Sacco, G. M. and Tzitzikas, Y. (eds). *Dynamic Taxonomies and Faceted Search*, Springer, pp. 75-112.
- Stefanut, T., Gorgan, D., Dovrolis, N., Kaldoudi, E. and Dietze, S. (2012) 'Creating Medical-related Learning Resources using Repurposing Procedures and Semantic Web Functionalities', *Proceedings of IADIS eSociety*. Berlin, 10-13 March. IADIS Press, pp.52-58.
- Stemler, S. (2001) 'An overview of content analysis', *Practical assessment, research & evaluation*, **7**(17), pp.137-146.
- Sue, V. M. and Ritter, L. A. (2007) 'Designing and developing the survey instrument', in *Conducting online surveys*, Thousand Oaks, Sage, pp. 59-88. DOI: 10.4135/9781412983754.n5.
- Sumner, T., Ahmad, F., Bhushan, S., Gu, Q., Molina, F., Willard, S., Wright, M., Davis, L. and Janée, G. (2005) 'Linking learning goals and educational resources through interactive concept map visualizations', *International Journal on Digital Libraries*, **5**(1), pp. 18-24.
- 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), pp. 81-107.
- Sutton, S. A. and Mason, J. (2001) 'The Dublin Core and metadata for educational resources', *Proceedings of the International Conference on Dublin Core and Metadata Applications*. Tokyo, 24-26 October. NII, pp. 25-31.
- Tang, H., Lin, Y.-J. and Qian, Y. (2020) 'Understanding K-12 teachers' intention to adopt open educational resources: A mixed methods inquiry', *British Journal of Educational Technology*. DOI: 10.1111/bjet.12937.
- Tang, T. Y. and McCalla, G. (2009) 'The pedagogical value of papers: a collaborative-filtering based paper recommender', *Journal of Digital Information*, **10**(2).

Tennant, J., Agarwal, R., Baždarić, K., Brassard, D., Crick, T., Dunleavy, D. J., ... Yarkoni, T. (2020, March 6). 'A tale of two 'opens': intersections between Free and Open Source Software and Open Scholarship' [Online]. DOI: 10.31235/osf.io/2kxq8 (Accessed 15 June 2020).

Thomas, A., Campbell, L.M., Barker P. and Hawksey, M. (2012) *Into the Wild – Technology for Open Educational Resources*, Bolton, University of Bolton.

Timpany, G. (2015) *Reaching the Constant Sum* [Online]. Available at <http://survey.cvent.com/blog/customer-insights-2/using-constant-sum-questions> (Accessed 2 May 2016).

Tlili, A., Jemni, M., Khribi, M.K., Huang, R., Chang, T.W., and Liu, D. (2020) 'Current state of open educational resources in the Arab region: an investigation in 22 countries', *Smart Learning Environments*. DOI: 10.1186/s40561-020-00120-z.

TREC (2016) *Text Retrieval Conference* [Online]. Available at <http://trec.nist.gov/> (Accessed 3 May 2016).

UNESCO (2012) *UNESCO Paris OER Declaration* [Online]. Available at http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/Events/Paris%20OER%20Declaration_01.pdf (Accessed 24 January 2019).

UNESCO (2017) *Ljubljana Action Plan* [Online]. Available at <https://www.oercongress.org/woerc-actionplan/> (Accessed 7 January 2018).

Ushakova, S. (2015) *Usability of Metadata Standards for Open Educational Resources* [Online]. Available at <https://helemuseum.wordpress.com/2015/10/02/usability-of-metadata-standards-for-open-educational-resources/> (Accessed 19 July 2019).

Vaishnavi, V. and Kuechler, W. (2015) *Design Science Research in Information Systems* [Online]. Available at <http://www.desrist.org/design-research-in-information-systems/> (Accessed 5 May 2017).

Van Rijsbergen, C. J. (1979) *Information Retrieval*, 2nd edn, Newton, Butterworth-Heinemann.

Venable, J., Pries-Heje, J. and Baskerville, R. (2012) 'A comprehensive framework for evaluation in design science research', *Proceedings of the International Conference on Design Science Research in Information Systems*. Las Vegas, 14-15 May. Springer, pp. 423-438.

Verbert, K., Drachsler, H., Manouselis, N., Wolpers, M., Vuorikari, R. and Duval, E. (2011) 'Dataset-driven research for improving recommender systems for learning', *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*. Banff, Alberta, Canada, February 27 - March 01. ACM, pp. 44-53.

Weller, M. (2010) 'Big and little OER', *Proceedings of the OpenED2010: Seventh Annual Open Education Conference*. Barcelona, Spain, 2-4 Nov 2010. Available at <http://oro.open.ac.uk/24702/>.

- Westergren, T. (2007) *The music genome project* [Online]. Available at <http://pandora.com/mgp> (Accessed 23 June 2017).
- Whaley, D. (2017) *Annotation is now a web standard* [Online]. Available at <https://web.hypothes.is/blog/annotation-is-now-a-web-standard/> (Accessed 20 December 2018).
- White, R. W., Drucker, S.M., Marchionini, G. and Hearst, M. (2007) 'Exploratory search and HCI: designing and evaluating interfaces to support exploratory search interaction', *Proceedings of CHI'07 Extended Abstracts on Human Factors in Computing Systems*. San Jose, CA, USA, April 28 - May 03. ACM, pp. 2877-2880.
- Wieringa, R. J. (2014) *Design science methodology for information systems and software engineering*, Springer.
- Wikipedia (2016a) *Discoverability* [Online]. 29 March 2016. Available at <https://en.wikipedia.org/wiki/Discoverability> (Accessed 02 May 2016).
- Wikipedia (2016b) *Findability* [Online]. 20 February 2016. Available at <https://en.wikipedia.org/wiki/Findability> (Accessed 02 May 2016).
- Wildemuth, B. M. and Freund, L. (2009) 'Search tasks and their role in studies of search behaviors', *Third Annual Workshop on Human Computer Interaction and Information Retrieval*. Washington DC, October 23, pp. 17-21.
- Wildemuth, B. M. and Freund, L. (2012) 'Assigning search tasks designed to elicit exploratory search behaviors', *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval 2012*. Cambridge, California, 04 October. ACM. DOI: 10.1145/2391224.2391228
- Wiley, D. (n.d.) *Defining the "Open" in Open Content and Open Educational Resources* [Online]. Available at <http://opencontent.org/definition/> (Accessed 11 May 2019).
- Wiley, D. (2013) 'What is Open Pedagogy?', *Iterating toward openness* [Blog]. Available at <https://opencontent.org/blog/archives/2975> (Accessed 12 November 2018).
- Wiley, D. (2017) 'OER-Enabled Pedagogy', *Iterating toward openness* [Blog]. Available at <https://opencontent.org/blog/archives/5009> (Accessed 12 November 2018).
- Wiley, D. and Hilton, J. (2018) 'Defining OER-enabled Pedagogy', *International Review of Research in Open and Distance Learning*, **19**(4). DOI: 10.19173/irrodl.v19i4.3601
- Wilson, M. L., Kules, B., Schraefel, M. C. and Shneiderman, B. (2010) 'From keyword search to exploration: Designing future search interfaces for the web', *Foundations and Trends in Web Science*, **2**(1), pp. 1-97.

Wilson, M., Schraefel, M. C. and White, R. W. (2009) 'Evaluating Advanced Search Interfaces Using Established Information-Seeking Models', *Journal of the American Society for Information Science and Technology*, **60**(7), pp. 1407-1422.

Winston, P. H. (1984) *Artificial intelligence*, Reading, Mass., Addison-Wesley.

Wolpers, M., Memmel, M., Giretti, A. (2009) 'Metadata in Architecture Education - First Evaluation Results of the MACE System', in: Cress, U., Dimitrova, V., Specht, M. (eds) *Learning in the Synergy of Multiple Disciplines. EC-TEL 2009. Lecture Notes in Computer Science*, vol 5794., Berlin, Heidelberg, Springer. DOI: 10.1007/978-3-642-04636-0_13.

Woodruff, A., Faulring, R. R. A., Morrison, J. and Pirolli, P. (2001) 'Using thumbnails to search the web', *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Seattle, March 31- April 5, ACM, pp. 198-205. DOI: 10.1145/365024.365098.

W3C (2014) *JSON-LD 1.0* [Online]. Available at <https://www.w3.org/TR/json-ld/> (Accessed 3 May 2016).

W3C (2015) *Open Linked Education Community Group* [Online]. Available at <https://www.w3.org/community/opened/> (Accessed 2 May 2016).

Yergler, N. (2010) 'Search and Discovery: OER's Open Loop', *Proceedings of OpenEd 2010*. Barcelona, 2-4 November 2010. Universitat Oberta de Catalunya, pp. 603-612. Available at http://www.issuelab.org/resource/search_and_discovery_oers_open_loop (Accessed 20 February 2014).

Yergler, N. (2011) *The Future of DiscoverEd* [Online]. Available at <http://labs.creativecommons.org/2011/04/11/the-future-of-discovered/> (Accessed 4 March 2014).

Yu, H. Q., Dietze, S., Li, N., Pedrinaci, C., Taibi, D., Dovrolls, N., Stefanut, T., Kaldoudi, E. and Domingue, J. (2011) 'A linked data-driven & service-oriented architecture for sharing educational resources', *Proceedings of the ESWC2011: 8th Extended Semantic Web Conference - Linked Learning 2011: 1st International Workshop on eLearning Approaches for the Linked Data Age*. Herakleio, May 29-30.

Zaid, Y. A. and Alabi, A.O. (2020) 'Sustaining Open Educational Resources (OER) initiatives in Nigerian Universities', *Open Learning: The Journal of Open, Distance and e-Learning*. DOI: 10.1080/02680513.2020.1713738.

Zamir, O. and Etzioni, O. (1999) 'Grouper: a dynamic clustering interface to Web search results', *Computer Networks*, **31**(11), pp. 1361-1374.

Zimmerman, J. and Forlizzi, J. (2014) 'Research through design in HCI', in Olson J. and Kellogg W. (eds) *Ways of Knowing in HCI*, New York, Springer, pp. 167-189. DOI: 10.1007/978-1-4939-0378-8_8.

Zimmerman, J., Stolterman, E. and Forlizzi, J. (2010) 'An analysis and critique of Research through Design: towards a formalization of a research approach', *Proceedings of the 8th ACM Conference on Designing Interactive Systems*. Aarhus, Denmark, 16-20 August. ACM, pp. 310-319. DOI: 10.1145/1858171.1858228.

Appendix A: Task analysis for the sub-categories Publishing, Repurposing, and Collaborating

Task analysis for the sub-categories Publishing, Repurposing, and Collaborating

A.1 Goal of Annex A

The OER Search/Discovery Task Analysis in the main text, reports the analysis of the Searching and Using sub-categories, more directly relevant to this research, as well as the top level categories, to provide them a meaningful context. Scope of this Annex, a complement to the main text, is to briefly discuss the remaining top-level categories: Publishing, Repurposing, and Collaborating.

A.2 Publishing: subcategories and related tasks

Publishing is at the heart of the concept of OER, and the essential starting point of their life cycle. Making the resources openly available, however, is not sufficient. Indeed, the considerable amount of published resources risks to create a “success disaster” (Atkins et al., 2007), if not complemented by the availability of suitable metadata to support their discovery.

At one extreme, metadata concerning Open Educational Resources are available, but are not open at all. OpenEd, as an example, created metadata for open resources freely available on the Internet, in particular alignments to educational frameworks, but legitimately considered these metadata its own intellectual property.

There are other similar cases, where some metadata are available but little effort is made to share them openly (Campbell and Barker, 2014), so that they are kept in isolated siloes. OERCommons is a notable example of an organization with a well-organized process to produce curated metadata, taking advantage of the active collaboration of resource providers, specialized curators, as well as trusted members of their community (Campbell, 2014a). They include articulated metadata comprising quality information about the resources, such as quality of explanation, quality of technological interactivity, quality of instructional and practice exercises, as well as alignments to educational standards. OERCommons is committed, in principle, to publish openly some of its metadata according to LRMI/schema.org, making them available for Web-scale search. Yet as an example, at the time of writing, searching in the OERCommons portal for "addition", Material Type: "Simulations" produces 181 hits, while searching via a Google Custom Search (CSE) with the equivalent query parameters "addition+more:pagemap:creativework-learningresourcetype:simulations" just produces 1 hit. Technically, this is due to the specification of those metadata as orphan “itemprop”, which are not properly embedded within an “itemscope”. This is regarded as a syntactical ambiguity in the specification of schema.org, and makes these metadata invisible and impossible to discover by a Google CSE.

These problems suggest an updated definition of Open Educational Resources, which could explicitly mandate the open availability of related metadata. Concerning tasks, therefore, the category Publishing should not just be concerned about making available the educational material, or a

reference (e.g. a URL) to educational material available somewhere else, but it could be also concerned about publishing related metadata, to support resource discoverability. These tasks are reported in Table A.1.

Category: Publishing [WC level]	
Sub-category	Sample tasks
None	Submit an OER, or its reference. Specify metadata for the submitted OER.

Table A.1 – Category Publishing.

A.3 Repurposing: subcategories and related tasks

Repurposing is a very common task carried out by educators, who keep adapting and integrating existing resources (Dietze et al., 2013). For example, de los Arcos et al. (2016) report that 85.5% of the educators they interviewed had adapted OERs to their needs. Repurposing is needed for many reasons, among others to cope with different languages, different cultures, different pedagogical approaches, different educational levels, different technologies, or different target people (Kaldoudi et al., 2011). Despite its importance, however, there is limited compelling evidence about formal OER repurposing even in large programs such as UK OER, and it is difficult to track it at web scale as there are very few data available (Thomas et al., 2012).

Repurposing is relevant for searching, as educators may be interested to know which resources were derived from a given one, the resource a given one was derived from, how a resource was modified, or how many derivatives are available (Drachsler et al., 2012). Hence, there should be suitable mechanisms to assist educators in repurposing existing resources, with the goal of reusing their efforts, while keeping track of every derivation. These mechanisms should require a minimum extra effort on the part of educators, to incentivize them to republish their repurposed resources.

Repurposing resources is obviously more easily supported by portals that integrate authoring facilities. In Gooru, for example, “remix” is a fundamental element in their overall strategy: repurposed resources (mainly aggregation of component resources in “collections”) can be published and shared as new resources, which can be discovered by other users, with minimal additional effort. It is also possible to split back existing collections in their constituent parts, an operation that was also mentioned in the second LRMI (n.d.) use case.

Other initiatives acknowledge the importance of repurposing and provide different mechanisms to support it. OpenScout is an example of specialized federation portal which is not built around any specific authoring environment, yet it pro-actively suggests and proposes for discussion, references to tools that can be used for repurposing, as well as documented case studies and best practices.

mEducator developed an ontology to identify repurposed resource as well as to document the whole process, supporting repurposing of existing material and searching among derived resources (Stefanut et al., 2012). Schema.org foresees metadata properties to identify the parts a resource is composed of, or the resource it is derived from. Gooru simply advises users to make sure that the description of new derived collections contain basic information and metadata such as learning objective, narrative, questions, standard and teacher tips as appropriate; new collections are then manually checked before publishing.

The fundamental tasks concerning repurposing identifiable in the previous cases are therefore: modify, aggregate, split the resources, document which resources were modified, and how. Table A.2 reports these tasks, organized in the two subcategories Modify and Document.

Category: Repurposing [WC level]	
Sub-category	Sample tasks
Modify	<ul style="list-style-type: none"> • Split existing resources in smaller constituent components. • Modify existing components or resources. • Aggregate existing components or resources as a new resource.
Document	<ul style="list-style-type: none"> • Specify metadata related to the new resource obtained, on the basis of pre-existing metadata of the modified resources. • Specify metadata about the repurposing process such as resources involved, repurposing relationships, and description of changes applied.

Table A.2 – Category repurposing.

A.4 Collaborating: subcategories and related tasks

This category refers to another class of important WC level tasks strongly related to search activities, where teachers engage in collaboration tasks such as “rate a resource” or “send a message to a peer”.

Data associated with these socially oriented activities can be conveniently exploited for searching. A notable example is FaceBook Graph Search, announced by Zuckerberg in 2012 (Sengupta, 2013), that relies on likes and other connections to provide a mixed search / personalized recommendation engine capable of answering chained requests such as “restaurants liked by your friends who live in London and work for a particular company”. This is potentially interesting for OER search, as it would make it possible, as an example, to ask for “OERs used by trusted colleagues who teach a given subject at a certain level”.

These tasks are frequently related to the fundamental quest for quality, a ubiquitous concern. As an example, a user might want to gather quality indicators about a resource by exploring its reviews, or

might want to rank resources according to their rating. This matter is explicitly addressed, for example, by Clements and Pawlowski (2012), and it is the underlying paramount concern in many other situations, for example the Irma use case (Drachsler et al., 2012). Yet, as mentioned in the FAQ of the Learning Registry, “quality is an opinion”. Hence, the provenance of these quality indicators is fundamental. As an example, a teacher might wish to discover resources used by a more experienced colleague because he/she trusts him/her to use appropriate materials.

Tasks related to this Collaborating category can be arranged in the two subcategories Contribute and Communicate.

The sub-category Contribute refers to situations where educators explicitly engage in a collaboration task. For example, when they decide to write a review, rate a resource (Drachsler et al., 2012), rate the alignment of a resource (OERCommons), share annotated bookmarks or playlists (Gooru), or propose additional non authoritative metadata for a resource (OERCommons). These tasks generate so called Social Data that can be used for search. This is in contrast to situations where user actions generate implicit traces (paradata), such as usage metrics, that can also be exploited for search.

The sub-category Communicate includes commonly supported individual and group communication tasks such as send a message to a user or a group, or subscribe to a group. The support for these communication oriented tasks is widespread in most portals, and its advantages are widely agreed upon in the literature. OpenScout for example, emphasizes community interaction functionalities. Okada et al. (2012) discuss the contribution of collaborative networks to foster creation, sharing and reuse of resources. The strong support for a community is considered a strategic priority especially by teachers from industrialized countries characterized by abundance of OERs, as revealed by the survey among teachers reported by Drachsler et al. (2012).

Table A.3 summarizes the subcategories and some representative tasks of the category Collaborating.

Category: Collaborating [WC level]	
Sub-category	Sample tasks
Contribute	<ul style="list-style-type: none"> • Rate or “like” a resource. • Tag a resource. • Contribute non authoritative metadata.
Communicate	<ul style="list-style-type: none"> • Post a message to a user or a group. • Contact users of a resource.

Table A.3 – Category Collaborating.

Appendix B: Task Analysis Survey

This Appendix contains the Task Analysis Survey as it was proposed to the participants in the Task Analysis Empirical Evaluation.

Open Educational Resources for Educators

Task Analysis Survey

Overview

Goal of this survey is to collect information related to tasks educators need to accomplish in relation to their use of Open Educational Resources (OERs), with a view to develop an evaluation framework for OERs discovery applications. Tasks of interest in this context include searching but extend to wider educators activities in relation to OERs. Important, “searching” in this context privileges “discovery”, that is exploratory search for potentially interesting material whose existence is not previously known.

This survey is expected to take some 30 minutes to complete.

This survey is available in three different modalities:

- self-administered by filling-in this document to be sent via email;
- self-administered via Web (available at: <https://goo.gl/poR1t2>), that can be filled-in also anonymously if desired;
- as a structured interview, including via telephone or Skype.

Background

OERs discoverability is widely recognized as instrumental in realizing the huge potential of the OER movement. Briefly, while a huge number of high quality OERs is available, finding - or more properly discovering - them is still quite challenging. A clear identification of the tasks that OERs discovery applications need to support, is the first essential step to properly design, evaluate or compare their functionalities, interfaces, as well as metadata schemas.

For these reasons, an analysis of the tasks educators need to carry out in relation to OERs is being carried out. The tasks have been identified from an analysis of the research literature and existing systems, suitably validated or enriched according to the Information Foraging Theory behavioural model, which leverages similarities between strategies of users looking for information (“informavores”) and foragers hunting for food. As an example, the forager strategy "discover potentially interesting prays, following the footprints of other expert foragers" can be directly transposed to the informavore strategy: "discover potentially interesting OERs by checking resources actually used by other users". This can in turn suggest corresponding user tasks such as "identify OERs used by trusted colleagues".

According to various research studies, most educators use traditional search engines such as Google to carry out their searches for OERs. These keywords-oriented search engines, while extremely powerful, do not provide yet the capabilities to precisely filter resources, for example, by type of

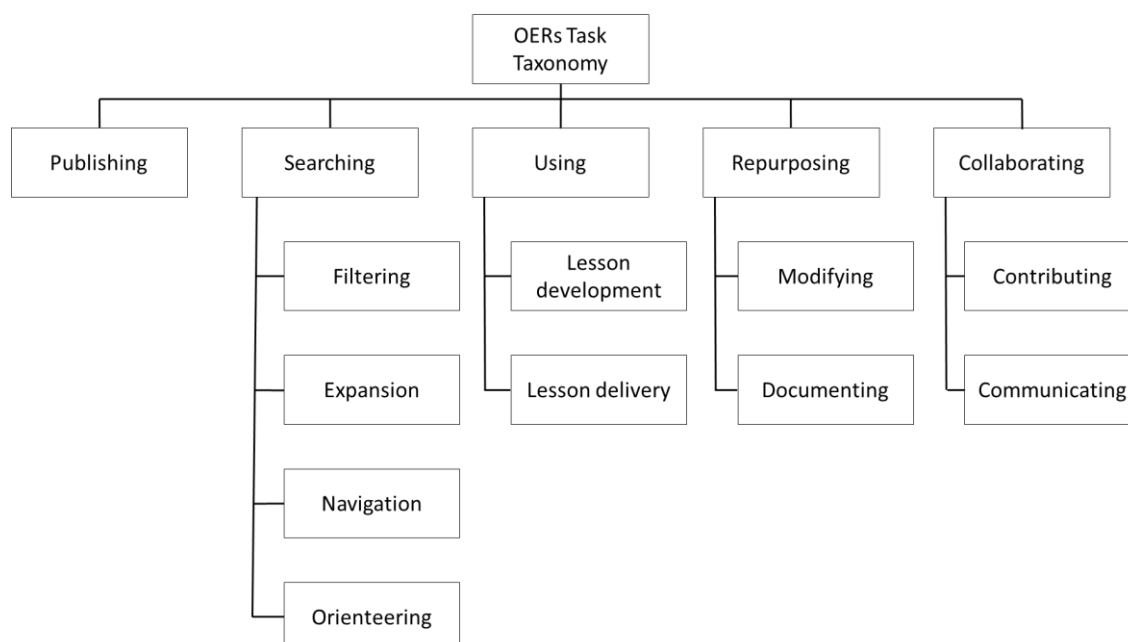
activity (simulation or presentation versus assessment test) or target audience (an 8 years old student versus a high school student).

Specialized search portals such as Curriki, OERCommons, Gooru or Merlot, just to mention a few, or emerging metadata standards such as LRMI/schema.org, do support powerful filtering capabilities, allowing educators to identify, for example, resources aligned to a specific educational standard.

Furthermore, additional “expansion” functionalities suggested by the Information Foraging Theory could let educators identify other resources starting from previously identified ones, such as resources often used together with a given one, resources recommended by people who liked a given one, or even just resources used by a trusted educator.

Objective of this survey is to collect evidence directly from experienced teachers, about the various professional tasks they feel that need to be supported, in relation to OERs search/discovery, their organization in categories, and their perceived relative importance. This should make it possible to qualitatively understand and possibly quantitatively measure the difference in the perceived importance of being able to search an educational resource on the basis of characteristics such as the type of interactivity, compared to search resources that have been used by trusted colleagues or that have been used together with certain others.

The survey is structured according to the following task taxonomy, whose top-level categories reflect the OERs life cycle: publishing new resources, searching, using, repurposing existing resources, and collaborating with peers about resources. While answering the survey it may be useful to come back to this general overview diagram:



Instructions

For each group of tasks or categories of tasks presented, you are invited to allocate a fixed number of points (100) as a numerical indicator of the relative importance you think they have in the context

of your professional activities. For example, if you think an item is twice as important as another, you can assign it twice as much points, but the total number of allocated points for each group must always add up to 100.

At every stage your comments, motivations of your choices, as well as any suggestions to add, modify, or reorganize tasks and categories, are highly appreciated. Please feel free to write in English, Italian, French, or Spanish.

Once the survey is completed, please send it to renato.cortinavis@open.ac.uk.

Privacy and data management

The survey does not collect any sensitive data, except very basic and optional demographic information. Following anonymization, the research data will be made available in open research data repositories. If necessary, any participant will be referred to by a pseudonym in any publications arising from the research.

Contacts

For any question or to request an alternative submission modality, please contact the principal investigator at renato.cortinavis@open.ac.uk or his Research Department:

Knowledge Media Institute

The Open University

Walton Hall

Milton Keynes

MK7 6AA

United Kingdom

Tel: +44 (0)1908 653800

Fax: +44 (0)1908 653169

<http://kmi.open.ac.uk/email/kmi-support>

Thank you for your collaboration!

Demographic and general information section

Note: Surname, First Name and Email, even if not mandatorily required, would be much appreciated in case the need arises to contact you for clarifications.

Surname: [.....] First Name: [.....] Email: [.....]

What is your age?

[30 or under / 31-40 / 41-50 / 51 or older]

What is your gender?

[Male / Female]

What best describes the highest level of education you have completed?

[Secondary school / Bachelor's degree / Master's degree / Doctoral degree]

For how many years have you been teaching?

[5 or less / 6-10 / 11-20 / 21 or more]

In which Country are you teaching?

[.....]

What best describes the educational level you are teaching at?

[Primary education (5-10) / Secondary education (11-18) / University]

What best describes the subjects you are teaching?

[Science / Math / Technology / Humanities]

How frequently (times per year) do you need to look for educational material on the Internet?

[0-5 / 6-20 / 21-50 / more than 50]

Which portals do you generally use to search for educational resources (examples of portals could be generic ones such as Google or specialized ones such as OERCommons)?

[.....]

Please indicate your three most important concerns related to tasks involving OERs (open educational resources):

[.....]

[.....]

[.....]

Tasks and categories section

TOP LEVEL CATEGORIES		
Category	Explanation	Importance (allocated points)
Publishing	Publishing (or sharing) refers to the task of making available an OER, possibly obtained by repurposing a pre-existing one, to other users.	[...]
Searching	Searching lumps together the most specific tasks related to OERs search and discovery. It is subdivided in the following categories: filtering, expansion, navigation, and orienteering.	[...]
Using	Using refers to the tasks related to the use of existing OERs, that is in lesson planning and lesson delivery.	[...]
Repurposing	Repurposing refers to the tasks related to the modification (adaptation) of existing resources.	[...]
Collaborating	Collaborating refers to a class of tasks related to search activities, involving collaboration with peers. Examples of tasks: “post a review about a resource”, “rate a resource”, or “subscribe to a group of interest”.	[...]
TOTAL (must be 100):		[...]

Comments or motivations for the choice:

[.....]

Suggestions to modify, add or reorganize tasks:

[.....]

Category Searching and related subcategories

Category: SEARCHING		
Category	Explanation	Importance (allocated points)
Filtering	Filtering refers to the possibility of adding restrictions to a query, that is restricting the search results, for example to resources tagged with a given tag, classified under a given topic, targeted to a given audience, rated above a given threshold.	[...]
Expansion	Expansion refers to the possibility of extending currently identified resources, by finding additional related ones. This allows users, in other words, to find new resources starting from already identified ones, which act as examples. Possible tasks include finding additional resources aligned to similar educational standards as a selected one, additional resources “frequently used together” with a given one, additional resources used by the same educator, or additional resources “liked by the same users” who also liked a sample one.	[...]
Navigation	Navigation refers to ancillary tasks such as going back to previous states of the current search via breadcrumb mechanisms, span search activities over multiple separate sessions, explore topics or educational standards taxonomies.	[...]
Orienteering	Orienteering refers to ancillary tasks for gathering information related to the organization of the data, or about their consistency. For example: check how a known resource is classified, explore metadata shared by a number of selected resources, explore various types of relationships among resources.	[...]
TOTAL (must be 100):		[...]

Comments or motivations for the choice:

[.....]

Suggestions to modify, add or reorganize tasks:

[.....]

Subcategory: SEARCHING / FILTERING		
Task	Explanation	Importance (allocated points)
Filter (show) resources related to given free keywords	This corresponds to the classical free keywords search available in traditional search engines such as Google.	[...]
Filter resources tagged with given free tags	Tags are free keywords associated by users to resources (as, for example, in Flickr).	[...]
Filter resources classified under a given topic of a predefined subjects taxonomy	To facilitate searching, resources can be identified starting from a generic hierarchical classification (taxonomy) of educational topics.	[...]
Filter resources aligned to a given skill or educational standard	To facilitate searching, resources can be identified starting from specific educational standards (such as the Common Core State Standards in the US).	[...]
Filter resources classified with other metadata (such as audience, learning resource type, educational use, author)	Resources can be searched making use of rich descriptive metadata such as audience (primary school, high school, undergraduate) or learning resource type (simulation, assessment, presentation, lab exercise, etc.).	[...]
Filter resources with rating above a given value	Limit resources to those having a minimum rating (assigned by other users).	[...]
Filter (or rank) resources according to usage metrics	Resources can be filtered (or maybe ranked) for example according to (possibly recent) number of visits, downloads, or times used.	[...]
Filter resources recently published	Recently published resources would not have been yet rated or used, and could be difficult to find: this function provides the possibility of surfacing them anyway.	[...]
TOTAL (must be 100):		[...]

Comments or motivations for the choice:

[.....]

Suggestions to modify, add or reorganize tasks:

[.....]

Subcategory: SEARCHING / EXPANSION		
Task	Explanation	Importance (allocated points)
Expand current results by automatically translating search terms	Discover additional resources classified with equivalent metadata expressed in different languages.	[...]
Expand current results by automatically expanding search terms with synonyms	Discover additional resources classified with metadata having the same meaning as the currently specified ones.	[...]
Discover additional resources aligned to standards equivalent to those of selected resources (QBE) among the current results (standards “crosswalking”)	Select a resource (or more) (the “Examples” in this Query By Example – or QBE), and discover additional resources aligned to a standard equivalent to the standard that the selected resource is aligned to.	[...]
Discover additional resources related to selected resources (QBE), by the relatedness metrics: standard similarity	Select a resource (or more) among the currently identified ones (QBE), and discover additional resources aligned to a standard approximately equivalent to the standard that the selected resource is aligned to.	[...]
Discover resources related to selected resources (QBE) by the relatedness metrics: explored together	Select a resource (or more) among the currently identified ones (QBE), and discover additional resources which have been explored together with the selected one, by other users.	[...]
Discover resources related to selected resources (QBE) by the relatedness metrics: togetherness	Select a resource (or more) among the currently identified ones (QBE), and discover additional resources frequently used together with the selected one, by other users.	[...]
Discover resources related to selected resources (QBE) by the relatedness metrics: likedness	Select a resource (or more) among the currently identified ones (QBE), and discover additional resources liked by the same users who liked the selected one.	[...]

Discover resources related to selected resources (QBE) by the relatedness metrics: common ancestor in repurposing history	Select a resource (or more) among the currently identified ones (QBE), and discover additional resources which have been repurposed from the same (ancestor) resource as the selected one.	[...]
Discover resources related to selected resources (QBE) by the relatedness metrics: content semantic similarity	Select a resource (or more) among the currently identified ones (QBE), and discover additional resources which have a strong content semantic similarity to the selected one.	[...]
TOTAL (must be 100):		[...]

Comments or motivations for the choice:

[.....]

Suggestions to modify, add or reorganize tasks:

[.....]

Subcategory: SEARCHING / NAVIGATION		
Task	Explanation	Importance (allocated points)
Navigate to previous statuses of the current search through breadcrumb mechanisms	Easy navigation to previous searches, via widely used dynamic breadcrumb mechanisms.	[...]
Navigate query and exploration history spanning multiple sessions	Easy navigation to previous searches, including searches carried out in separate sessions (for example in the last week).	[...]
Explore the topics taxonomy: navigate to more general, more specialized, or related categories	Exploration of hierarchically structured topics or categories, which are used to index / classify the resources.	[...]
Explore the alignment standards taxonomy: navigate to more general or more specialized standards	Exploration of hierarchically structured standards taxonomies (educational, skills or achievement standards), where resources are aligned to.	[...]
Navigate to equivalent, or similar, standards taxonomies	Navigation from a standards taxonomy to other equivalent or nearly equivalent standards taxonomies – allowing users to expand their search to resources aligned to other equivalent or similar standards.	[...]
Preview the effect of every potential action before committing	This is to provide users with a preview of every possible action in each context, before they commit to it.	[...]
TOTAL (must be 100):		[...]

Comments or motivations for the choice:

[.....]

Suggestions to modify, add or reorganize tasks:

[.....]

Subcategory: SEARCHING / ORIENTEERING		
Task	Explanation	Importance (allocated points)
Semantic zooming: gather additional information about a selected resource	Additional information about a resource could be obtained, for example, by hovering on its title.	[...]
Explore (collective) metadata related to a selection (QBE) of resources among the current results	Selecting a subset of resources (the usual examples in a Query By Examples), could visualize, for example, their common metadata (“categorized overviews”) as well as consistency indicators such as Numerical Volume Indicators or equivalent graphical representation with bars.	[...]
Explore tag clouds	Tag clouds could be explored, in this context, for sense making (i.e. to get a feeling of the kind of resources available).	[...]
Explore existing relationships among resources	Relationships among resources could include “used together”, “liked by the same users”, “repurposed from the same ancestor”, “semantic similarity”.	[...]
Explore the organization of the dataset (its schema)	This is to get a feeling about how the resources are organized and indexed.	[...]
Gather information about dataset consistency	This is to get a feeling about the volume of resources available (the extension of the dataset) in each category.	[...]
TOTAL (must be 100):		[...]

Comments or motivations for the choice:

[.....]

Suggestions to modify, add or reorganize tasks:

[.....]

Category Publishing

Category: PUBLISHING		
Task	Explanation	Importance (allocated points)
Submit (publish) a resource, or its reference	This task refers to the actual publishing of an OER in a public portal.	[...]
Specify metadata for the submitted resource	Whenever a new OER is published, it is necessary to accompany it with suitable metadata (for example learning objectives, topics, educational level, level of interactivity) so that other users will be able to find it.	[...]
Suggest additional metadata for an existing resource	This refers to the possibility for a user to add additional metadata to enrich metadata already available for a published resource. For example, a user might add metadata related to the alignment of the resource to an additional educational standard.	[...]
TOTAL (must be 100):		[...]

Comments or motivations for the choice:

[.....]

Suggestions to modify, add or reorganize tasks:

[.....]

Category Using and related subcategories

Category: USING		
Subcategory	Explanation	Importance (allocated points)
Lesson planning	This category is related to lesson planning tasks making use of OERs. Example of tasks include: organize and annotate collection of resources in a notes taking application, plan teaching activities with different delivery strategies, organize personalized remedial activities on the basis of student performances.	[...]
Lesson delivery	This category is related to lesson delivery tasks. Examples of tasks include: delivery of learning resources with different strategies (for example whole class instruction, small groups, flipped classrooms), tracking learners collecting analytics about their performances.	[...]
TOTAL (must be 100):		[...]

Comments or motivations for the choice:

[.....]

Suggestions to modify, add or reorganize tasks:

[.....]

Subcategory: USING / LESSON PLANNING		
Task	Explanation	Importance (allocated points)
Organize and annotate collection of resources in a bookmarking service or notes taking application	This task refers to the organization of search results in personally annotated collections.	[...]
Plan teaching activities with different delivery strategies	Possible delivery strategies include for example whole class instruction, small groups, individual learning, flipped classrooms, assessment.	[...]
Organize personalized remedial activities on the basis of student performances	Student performances results can be exploited to organize personalized remedial activities.	[...]
Collect formative evaluation analytics to improve the organization of the educational material	Evaluation analytics can be used for formative evaluation, that is to improve the teaching process including the organization of the educational material.	[...]
TOTAL (must be 100):		[...]

Comments or motivations for the choice:

[.....]

Suggestions to modify, add or reorganize tasks:

[.....]

Subcategory: USING / LESSON DELIVERY		
Task	Explanation	Importance (allocated points)
Delivery of learning resources with different strategies	Possible delivery strategies include whole class instruction, small groups, individual learning, flipped classrooms, and assessment.	[...]
Track learners collecting analytics about their performances	These tasks are related to summative evaluation, that is a final evaluation of students' performances.	[...]
TOTAL (must be 100):		[...]

Comments or motivations for the choice:

[.....]

Suggestions to modify, add or reorganize tasks:

[.....]

Category Repurposing and related subcategories

Category REPURPOSING		
Subcategory	Explanation	Importance (allocated points)
Modifying	Tasks related to the modification of existing resources, such as extracting their components, modifying or integrating them.	[...]
Documenting	Tasks related to the documentation of a new resource obtained by repurposing existing ones.	[...]
TOTAL (must be 100):		[...]

Comments or motivations for the choice:

[.....]

Suggestions to modify, add or reorganize tasks:

[.....]

Subcategory: REPURPOSING / MODIFYING		
Task	Explanation	Importance (allocated points)
Split existing resources in smaller constituent components	It may be desirable to be able to extract and use just a part of an existing resource.	[...]
Modify existing components or resources	It may be desirable to be able to modify an existing resource, to adapt it to a particular context.	[...]
Integrate existing components or resources as a new resource	It may be desirable to be able to create a new resource by integrating existing ones.	[...]
TOTAL (must be 100):		[...]

Comments or motivations for the choice:

[.....]

Suggestions to modify, add or reorganize tasks:

[.....]

Subcategory: REPURPOSING /DOCUMENTING		
Task	Explanation	Importance (allocated points)
Specify metadata related to a new resource obtained modifying an existing one, on the basis of pre-existing metadata of the modified resource	It may be desirable to be able to specify metadata for a new repurposed resource, taking advantage from the pre-existing metadata of the original resource that was modified, rather than recreating all the metadata ex-novo.	[...]
Specify metadata about the repurposing process, such as resources involved, repurposing relationships, description of changes applied	It may be desirable to be able to specify metadata tracking the repurposing history of a resource. This would make it possible, for example, to search for all derived resources of a given one, or to obtain a list of modifications to a given resource.	[...]
TOTAL (must be 100):		[...]

Comments or motivations for the choice:

[.....]

Suggestions to modify, add or reorganize tasks:

[.....]

Category Collaborating and related subcategories

Category COLLABORATING		
Subcategory	Explanation	Importance (allocated points)
Contributing	This sub-category refers to situations where educators explicitly engage in collaboration tasks. Examples include: write a review, rate a resource, rate the contribution of another user, rate the alignment of a resource, tag a resource, share annotated bookmarks or playlists.	[...]

Communicating	This sub-category includes traditionally supported individual and group communication tasks such as send a message to a user or a group, or subscribe to a group.	[...]
TOTAL (must be 100):		[...]

Comments or motivations for the choice:

[.....]

Suggestions to modify, add or reorganize tasks:

[.....]

Subcategory: COLLABORATING / CONTRIBUTING		
Task	Explanation	Importance (allocated points)
Rate or “like” a resource	Users can indicate that they “like” a resource, or express a rating reflecting its perceived quality.	[...]
Tag a resource	Associate a free keyword (tag) to a resource.	[...]
Write a review	Write a review about a resource.	[...]
Contribute non authoritative metadata	Users (not just the author) can contribute additional metadata to a resource, such as an additional alignment to an educational standard.	[...]
Rate a contribution of another user	A user rates the contribution of another user such as a review, or a standard alignment.	[...]
Share annotated bookmarks or playlists	Users might share the results of their search activities that could prove useful to other users, for example annotated list of resources.	[...]
Share a manual re-ranking of the result-set	The automatic ranking of the system could be modified by a user to reflect his perceptions – this crowdsourced information could then be used by the system to optimize its ranking strategies.	[...]
TOTAL (must be 100):		[...]

Comments or motivations for the choice:

[.....]

Suggestions to modify, add or reorganize tasks:

[.....]

Subcategory: COLLABORATING / COMMUNICATING		
Task	Explanation	Importance (allocated points)
Post a message to a user or a group	Generic community interaction functionalities, but in the specific context of OERs.	[...]
Create / subscribe to a group of interest	Tasks related to participation in groups of interests about OERs.	[...]
Follow a user	This allows followers to be notified about relevant activities of the user followed.	[...]
Contact users of a resource	This makes it possible to get in touch with actual users of a given resource.	[...]
Manage the user profile	The profile would allow users to manage personal information, including to present themselves to other users.	[...]
Set automatic notifications	Automatic push notifications notify users (via email or sms) of new messages or events they subscribed to.	[...]
Integrate communication with other social networks	Examples of other social networks are Twitter, Facebook, or LinkedIn.	[...]
TOTAL (must be 100):		[...]

Comments or motivations for the choice:

[.....]

Suggestions to modify, add or reorganize tasks:

[.....]

Final section

How would you judge the overall completeness of this task analysis, on the following scale?

1 (very low)	2	3	4	5	6	7 (very high)
...

Do you feel that there are important tasks that have not been considered in this survey?

[Yes/no]

If you answered yes to the previous question, please indicate which tasks:

[.....]

Do you feel that this survey mentioned important tasks that are not supported by the portals you currently use to search for OERs?

[Yes/no]

If you answered yes to the previous question, please indicate which tasks:

[.....]

Please add any other comment / consideration that you feel relevant:

[.....]

Thank you very much for your appreciated collaboration,

Renato Cortinovia

Appendix C: Task-Taxonomy Survey: weights for tasks and categories

This appendix contains the weights for tasks and categories in the preliminary Task Taxonomy resulting from the analysis of the literature and existing portals, assigned by the participants in the Task Analysis Empirical Evaluation. They were mainly collected to engage participants in critical thinking, and were used to select the tasks to focus on in this research. However, they are reported here as they represent potentially useful information, that could be reused (Geisler, 2008), for example, as weights in the evaluation of search portals (Agarwal and Venkatesh, 2002).

Task-Taxonomy Survey: weights for tasks and categories

Top level categories:

Publishing	Searching	Using	Repurposing	Collaborating
20	20	23	23	14
5	35	30	20	10
25	30	30	10	5
10	30	30	20	10
5	30	35	25	5
20	20	30	15	15
10	40	20	20	10
22	19	23	18	18
14	30	21	21	14
Grand Total				
Mean				
15	28	27	19	11
StdDev				
7	7	5	4	4

Category Searching:

Search/Filtering	Search/Expansion	Search/Navigation	Search/Orientteering
27	19	27	27
30	25	25	20
35	25	20	20
30	30	30	10
30	20	35	15
35	30	15	20
30	30	20	20
30	20	20	30
28	28	20	24
Grand Total			
Mean			
31	25	24	21
StdDev			
3	5	6	6

Category Searching / Filtering:

Filter with free keywords	Filter with free tags	Filter by topic of predefined taxonomy	Filter by alignment to educational framework	Filter by other metadata (audience, educational use...)	Filter by rating above threshold	Filter by usage metrics	Filter recently published resources
11	13	10	15	17	11	13	10
5	15	10	10	10	17	15	18
20	20	10	10	10	20	5	5
20	10	20	20	10	10	0	10
20	20	15	5	15	10	5	10
10	5	25	15	10	15	10	10
15	5	20	5	20	10	15	10
9	11	11	15	14	13	13	14
15	15	15	9	8	11	15	12
Grand Total							
Mean							
14	13	15	12	13	13	10	11
StdDev							
5	6	5	5	4	4	6	4

Category Searching / Expansion:

Expand by keywords translation	Expand by keywords synonyms	Expand by equivalent alignment	Expand by standard similarity	Expand "explored together"
8	9	12	12	12
5	10	20	10	15
20	20	15	15	0
10	10	10	10	10
0	10	10	15	0
15	20	10	5	10
0	10	20	20	0
10	10	11	10	11
16	14	14	11	9
Grand Total				
Mean	9	13	14	12
StdDev				
7	4	4	4	6

Expand by "frequently used together"	Expand by likedness	Expand by common ancestor	Expand by content similarity	
12	13	14	8	
20	5	5	10	
5	0	10	15	
10	10	10	20	
30	20	5	10	
10	10	15	5	
20	20	5	5	
13	13	11	11	
11	7	7	11	
Grand Total				
Mean	15	11	9	11
StdDev	8	7	4	5

Category Searching / Navigation:

Navigate to previous statuses by breadcrumbs	Navigate search history spanning multiple sessions	Explore topics taxonomy	Explore alignments standards taxonomy	Navigate to equivalent or similar standards	Preview effect of every action
16	16	18	16	16	18
20	25	15	10	25	5
25	20	20	5	10	20
15	20	10	15	20	20
20	30	10	0	20	20
10	10	25	25	20	10
20	5	25	15	15	20
15	15	18	17	15	20
18	20	15	15	15	17
Grand Total					
Mean	18	18	17	13	17
StdDev					
4	8	6	7	4	5

Category Searching / Orienteering:

Semantic zooming	Explore collective metadata	Explore tag clouds	Explore other relationships among resources	Explore dataset schema	Gather information about dataset consistency
20	17	12	17	17	17
30	10	5	25	20	10
40	10	10	20	15	5
30	0	30	20	20	0
20	20	20	15	15	10
25	10	25	20	15	5
20	20	10	20	15	15
15	14	17	20	17	17
21	13	18	18	15	15
Grand Total					
Mean	25	13	16	19	17
StdDev					
8	6	8	3	2	6

Category Publishing:

Publish a resource	Specify metadata for the submitted resource	[Suggest additional metadata for existing resource]
36	36	28
50	30	20
40	30	30
30	40	30
20	35	45
35	50	15
20	30	50
40	32	28
30	37	33
Grand Total		
Mean	33	36
StdDev		
10	6	11

Category Using:

Using / Lesson planning	Using / Lesson delivery
43	57
50	50
50	50
70	30
50	50
40	60
60	40
45	55
60	40
Grand Total	
Mean	52
StdDev	
10	10

Category Using / Lesson Planning:

Organize and annotate collections	Plan different delivery strategies	Personalized remedial activities	Evaluation analytics to improve material
22	27	27	24
20	30	40	10
10	30	40	20
10	30	30	30
20	30	30	20
15	20	35	30
15	30	30	25
22	26	26	26
24	31	24	21
Grand Total			
Mean			
18	28	31	23
StdDev			
5	3	6	6

Category Using / Lesson Delivery:

Delivery with multiple strategies	Learners performances tracking
44	56
60	40
50	50
70	30
70	30
50	50
60	40
47	53
62	38
Grand Total	
Mean	
57	43
StdDev	
10	10

Category Repurposing:

Repurposing / Modifying	Repurposing / Documenting
50	50
60	40
40	60
60	40
70	30
60	40
60	40
50	50
67	33
Grand Total	
Mean	
57	43
StdDev	
9	9

Category Repurposing / Modifying:

Split resources in components	Modify	Integrate
33	34	33
30	40	30
30	50	20
20	40	40
10	60	30
50	30	20
25	50	25
27	35	38
22	45	33
Grand Total		
Mean	27	43
		30
StdDev		
	11	10
		7

Category Repurposing / Documenting:

Specify metadata for repurposed resource	Specify metadata about repurposing process
53	47
65	35
50	50
40	60
40	60
70	30
35	65
50	50
43	57
Grand Total	
Mean	50
	50
StdDev	
	12
	12

Category Collaborating:

Collaborating / Contributing	Collaborating / Communicating
50	50
50	50
30	70
20	80
70	30
50	50
35	65
56	44
67	33
Grand Total	
Mean	48
	52
StdDev	
	17
	17

Category Collaborating / Contributing:

Rate or Like	Tag	Review	Contribute non authoritative metadata	Rate user contribution	Share bookmarks or playlists	Share re-ranking
14	18	14	14	13	13	14
15	15	15	10	15	20	10
15	10	10	25	0	30	10
20	10	10	30	10	10	10
20	15	15	15	10	15	10
25	30	10	10	5	10	10
20	10	10	25	15	15	5
14	12	14	17	17	17	9
15	12	12	18	18	15	10
Grand Total						
Mean	18	15	12	18	11	16
StdDev						
4	6	2	7	6	6	2

Category Collaborating / Communicating:

Post a message	Subscribe to a group	Follow user	Contact user	Manage personal profile	Set notifications	Integrate with other social networks
11	11	13	18	16	13	18
20	20	15	10	5	25	5
10	10	20	10	10	10	30
20	15	10	25	10	10	10
30	25	10	20	0	15	0
5	25	20	15	10	10	15
20	20	5	25	5	20	5
13	16	15	16	15	13	12
12	12	15	20	9	20	12
Grand Total						
Mean	16	17	14	18	9	15
StdDev						
8	6	5	6	5	5	9

Appendix D : Injector Heuristic Evaluation Form

This Appendix reports the heuristic evaluation form proposed to the participants in the discounted evaluation of the prototype Injector.

Injector heuristic evaluation form

Demographic and general information section

Note: Surname, First Name and Email, even if not mandatorily required, would be much appreciated in case the need arises to contact you for clarifications.

Surname: [.....]

First Name: [.....]

Email: [.....]

What is your age?

[30 or under / 31-40 / 41-50 / 51 or older]

What is your gender?

[Male / Female]

In which Country do you work?

[.....]

Are you an educator?

[Yes / No]

If yes:

For how many years have you been teaching?

[5 or less / 6-10 / 11-20 / 21 or more]

What best describes the educational level you are teaching at?

[Primary education (5-10) / Secondary education (11-18) / University]

What best describes the subjects you are teaching?

[Science / Math / Technology / Humanities]

Evaluation section

Imagine that you, an educator, are looking for educational resources on a certain subject. You start your search in Google with the keywords “*solar system*”, “*tree diagrams and figure probabilities of events*”, “*modeling linear relationships*”, or “*advanced fire simulation*”. You found a potentially interesting resource in Google results, however you would like to explore additional similar resources.

Now carry out this task with the support of the Injector prototype, and consider how you could carry it out with the tools you normally use. In this context, please answer the following questions.

Functionalities

(1) The discovery oriented functionality (based on similar educational alignments) supported by the prototype is:

1 (totally useless)	2	3	4	5	6	7 (very useful)
...

(2) The relevance / similarity of the suggested resources is:

1 (very weak)	2	3	4	5	6	7 (very strong)
...

(3) Finding the suggested resources with alternative techniques (you currently use) would be:

1 (very easy)	2	3	4	5	6	7 (very difficult)
...

(4) What other related primary functionalities should the prototype support?

[.....]

[.....]

Usability

(5) The system speaks the users' language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms:

1 (fully disagree)	2	3	4	5	6	7 (fully agree)
...

(6) I consider the "transparency" of the tool (exploiting educational standards without the need to explicitly manipulate them):

1 (totally useless)	2	3	4	5	6	7 (very useful)
...

User Experience

(7) I think that I would like to use this system frequently:

1 (fully disagree)	2	3	4	5	6	7 (fully agree)
...

(8) I think this system automates unwanted workload:

1 (fully disagree)	2	3	4	5	6	7 (fully agree)
...

General

(9) How relevant is the proposed task in relation to your searches for educational resources:

1 (totally irrelevant)	2	3	4	5	6	7 (very relevant)
...

(10) Please provide one (or more) positive comment(s):

[.....]

[.....]

(11) Please provide one (or more) negative comment(s):

[.....]

[.....]

Thank you very much for your appreciated collaboration,

Renato Cortinovis

Appendix E : RepExp Heuristic Evaluation Form

This Appendix reports the heuristic evaluation form proposed to the participants in the discounted evaluation of the prototype RepExp.

RepExp: the prototype

RepExp (Replicator-Expander) is a prototype supporting the discovery of (open) educational resources, aiming to experiment with an innovative **subset** of functionalities of a full-fledged (open) educational resources discovery application. It consists of a pair of Chrome browser extensions: “Replicator” and “Expander”.

Replicator intercepts generic keyword-based Google searches and **replicates them in a specialized educational network**, producing a custom results page with educational resources and related (LRMI) educational metadata. The aim is to offer educational resources discovery functionalities directly and **transparently from Google** search results pages, being it the search engine most commonly used by educators.

Expander allows a user who has found an interesting educational resource, to **discover additional “similar” (“more like this”) resources**. Most applications available identify similar resources considering the **co-occurrence of keywords**. On the contrary, the similarity metric defined in the prototype considers two resources **the more similar, the more specific learning objectives they have in common**. With this strategy, the prototype makes transparent use of sometimes large numbers of learning objectives (alignments to formal educational standards) without the need for users to be acquainted with (or even aware of) them. This makes it possible to reuse the considerable efforts in aligning resources to educational standards, very high-quality curated data, in Countries where educational standards are not formally used, where different educational frameworks are used, or whenever educators are not interested in explicitly handling them.

RepExp: where to find it

There is a short description of the prototype and a meaningful commented sequence of screenshots available at: <https://github.com/renatomario/RepExp/blob/master/README.md>

There is also a simple short (three minutes) screencast of the latest version of the prototype (RepExp), including hints to its predecessor (called Injector), available at: <http://people.kmi.open.ac.uk/renato/>

The sequence of screenshots and the screencast should already give you a pretty good idea of the prototype, and make it possible to fill-in the RepExp Evaluation Form.

As an alternative, if you prefer so, you can contact me and I will be pleased to demonstrate you the prototype live via Skype – so that I can answer any question you might have.

Finally, if you prefer to try-out the prototype on your own, it is necessary to install the two “extensions” in a Chrome browser. All the necessary files of the extensions are available via GitHub at <https://github.com/renatomario/RepExp>. They should be downloaded in a local directory and loaded in Chrome as explained at: <https://developer.chrome.com/extensions/getstarted#unpacked> .

RepExp: Evaluation

Following the exploration of the prototype and/or its documentation, kindly fill-in the Evaluation Form available in the following pages. For any question or comment, please do not hesitate to contact me at: Renato.Cortinovis@open.ac.uk

Thank you very much for your appreciated collaboration,

Renato Cortinovis

Replicator-Expander Heuristic Evaluation Form

Demographic and general information section

Note: Surname, First Name and Email, are **not** mandatorily required, but would be much appreciated in case the need arises to contact you for clarifications.

Surname: [.....]

First Name: [.....]

Email: [.....]

What is your age?

[30 or under / 31-40 / 41-50 / 51 or older]

What is your gender?

[Male / Female]

In which Country do you work?

[.....]

Are you an educator?

[Yes / No]

If yes:

For how many years have you been teaching?

[5 or less / 6-10 / 11-20 / 21 or more]

What best describes the educational level you are teaching at?

[Primary education (5-10) / Secondary education (11-18) / University]

What best describes the subjects you are teaching?

[Science / Math / Technology / Humanities]

Evaluation section

Imagine that you are looking for educational resources on a certain topic. You start your search in Google with keywords of your choice (for example “algebra”). You are not fully satisfied with the **generic** results obtained, so that you need to search for **more specific educational resources and more relevant data** with other techniques. Among the new results obtained, you identify some interesting resource, but you would like to **discover additional similar resources** (e.g. for remediation, reinforcement or in-depth activities).

Please carry out this task with the support of the Replicator-Expander prototype, and consider how you could carry this out with the **tools you normally use**. In this context, please answer the following questions.

Please feel free to write in English, French, Spanish, or Italian.

Functionalities

(1) The prototype lets you search in **specialized** educational resources networks, **directly and transparently from Google** results pages (or indeed from any other search engine, or even from any educational resource you previously identified). How do you compare this strategy to support the discovery of educational resources, to the alternative of having multiple, separate, specialized search portals? Transparently starting from Google results pages is:

1 (much better)	2	3	4	5	6	7 (much worst)
--------------------	---	---	---	---	---	-------------------

Could you indicate a reason?

(2) The prototype lets you identify “similar” resources, characterized by having the maximum number of learning objectives in common. Thinking for example to the need to identify additional resources for remediation activities, how do you compare the similarity metric used in the prototype, based on shared learning objectives, to the commonly available metrics based on shared words? Similarity based on learning objectives is:

1 (much better)	2	3	4	5	6	7 (much worst)
--------------------	---	---	---	---	---	-------------------

Could you indicate a reason?

(3) The prototype lets you find resources by similarity. How useful in your job is the possibility of discovering “similar” resources **in general** (therefore also with potentially different metrics such as, for example, “frequently used together”, “used by the same educator”, “liked by the same users”, etc.)?

1 (very useful)	2	3	4	5	6	7 (totally useless)
--------------------	---	---	---	---	---	------------------------

Could you indicate a reason?

(4) The prototype lets you to find “similar” resources, that is, resources which **share the same learning objectives**. Discovering this type of **similar** resources with alternative techniques (you currently use) would be:

1 (very easy)	2	3	4	5	6	7 (very difficult)
------------------	---	---	---	---	---	-----------------------

Usability

(5) The system speaks the users’ language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms:

1 (fully disagree)	2	3	4	5	6	7 (fully agree)
-----------------------	---	---	---	---	---	--------------------

(6) I consider the “transparency” of the tool (exploiting educational standards without the need to explicitly manipulate them):

1 (totally useless)	2	3	4	5	6	7 (very useful)
------------------------	---	---	---	---	---	--------------------

User Experience

(7) I think that I would like to use this system frequently:

1 (fully disagree)	2	3	4	5	6	7 (fully agree)
-----------------------	---	---	---	---	---	--------------------

(8) I think this system automates unwanted workload:

1 (fully disagree)	2	3	4	5	6	7 (fully agree)
--------------------------	---	---	---	---	---	--------------------

General

(9) How **relevant** is the proposed task in relation to your searches for educational resources (for example to find educational material for reinforcement, remediation, or in-depth educational activities):

1 (totally irrelevant)	2	3	4	5	6	7 (very relevant)
------------------------------	---	---	---	---	---	-------------------------

(10) Please provide one (or more) positive comment(s) about the prototype:

[.....]

[.....]

(11) Please provide one (or more) negative comment(s):

[.....]

[.....]

(12) Please feel welcome to leave any further comment, suggestion, or critique:

[.....]

[.....]

Thank you very much for your appreciated collaboration,

Renato Cortinovis

Appendix F: Discoverer Heuristic Evaluation Form

This Appendix reports the heuristic evaluation form proposed to the participants in the summative evaluation of the last prototype Discoverer.

Discoverer evaluation survey

Goal

Goal of this survey is to collect data for the evaluation of Discoverer, a research prototype supporting the discovery of open educational resources.

Privacy and data management

The survey does not collect any sensitive data, except very basic and optional demographic information. However, gathered data will be treated confidentially and only used for research purposes.

Data collected will be initially stored in a private database for further analysis, as well as on encrypted portable devices for backup. Following anonymization, the data may be made available in open research data repositories. If necessary, any participant will be referred to by a pseudonym in any publications arising from the research.

Contact information

For any additional question, please contact me at Renato.Cortinovis@open.ac.uk, or my Research Department:

Knowledge Media Institute
The Open University
Walton Hall
Milton Keynes
MK7 6AA
United Kingdom

Tel: +44 (0)1908 653800
Fax: +44 (0)1908 653169
<http://kmi.open.ac.uk/email/kmi-support>

Thank you for your collaboration.

Human Research Ethics Committee approval: HREC/2017/2517/Cortinovis

Discoverer: the prototype

Discoverer is a prototype supporting the discovery of (open) educational resources, aiming to experiment with an innovative *subset* of functionalities of a full-fledged (open) educational resources discovery application.

As a first basic functionality, Discoverer automatically replicates keyword-based Google searches in a *specialized hub of educational resources*, producing a custom results page with educational resources and related educational metadata. The aim is to offer educational resources discovery functionalities directly and *transparently from Google* search results pages, being it the search engine most commonly used by educators.

As a second main functionality, Discoverer allows a user, starting from an existing educational resource previously identified via Discoverer, Google, Yahoo or any other way, to *discover additional “similar” (“more like this”) resources*. While many generic applications available identify similar resources simply considering the *co-occurrence of keywords*, Discoverer is strongly domain-oriented, and considers two resources *the more similar, the more learning objectives* (more precisely alignments to educational frameworks) they have in common.

The prototype, making transparent use of sometimes large numbers of learning objectives (alignments) without the need for users to be acquainted with (or even aware of) them, makes it possible to reuse the considerable efforts in aligning resources to educational standards, very high-quality curated data, in Countries where educational standards are *not* formally used, where *different* educational frameworks are used, or whenever educators are *not* interested in explicitly handling them.

The resources identified are *clustered in groups of comparable degree of similarity*, supporting sense-making by offering a compressed view of the available resources, and giving educators control on the level of similarity desired, that might differ according to the intended educational use of the resources: for example remediation (higher similarity required) versus in-depth activities (lower similarity required).

Discoverer: where to find it

A short description of the prototype and a meaningful commented sequence of screenshots is available at:

<https://github.com/renatomario/ERD3/blob/master/README.md>

A short screencast of the latest version of the Discoverer prototype, is available at:

https://github.com/renatomario/ERD3/raw/master/screencasts/Discoverer_English.mp4

You are also welcome to contact me, and I will be pleased to demonstrate you the prototype live via Skype – so that I can answer any question you might have.

Finally, if you prefer to try-out the prototype on your own, it is necessary to install an “extension” in a Chrome browser. It is necessary to download the directory “Discoverer” from

<https://github.com/renatomario/ERD3>

in a local directory, and load it in Chrome as explained at:

<https://developer.chrome.com/extensions/getstarted#unpacked>

Discoverer Heuristic Evaluation Form

Demographic and general information section

Note: Surname, First Name and Email, are **not** mandatorily required, but would be much appreciated in case the need arises to contact you for clarifications.

Surname:

First Name:

Email:

What is your age?

[30 or under / 31-40 / 41-50 / 51 or older]

What is your gender?

[Male / Female]

In which Country do you work?

For how many years have you been teaching?

[5 or less / 6-10 / 11-20 / 21 or more]

What best describes the educational level you are teaching at?

[Primary education (5-10) / Secondary education (11-18) / University]

What best describes the subjects you are teaching?

[Science / Math / Technology / Humanities]

Please indicate your knowledge about Open Educational Resources (OERs):

[expert / good knowledge / some knowledge / never heard of OERs before]

How often do you use OERs for teaching?

[Very often / Occasionally / Never]

Which tool / repository do you currently use to find educational resources?

.....

Evaluation section

As a first scenario, imagine that you are looking for educational resources. You start your search in Google with keywords of your choice (for example “algebra”, “probability”, or “solar system”). You are not fully satisfied with the **generic** results obtained, so that you need to search further for **more specific educational resources and more relevant educational data**. Among the new results obtained, you identify some interesting resource, but you would like to **discover additional similar resources**.

As a second scenario, you already have an educational resource, you may be already using it, but you would like to **find similar resources to be used, for example, for remediation, reinforcement, or in-depth activities**.

Please carry out the tasks in these scenarios with the support of the Discoverer prototype, and consider how you could carry them out with the **tools you normally use**. In this context, please answer the following questions.

Please feel free to write in English, Italian, French, or Spanish – full sentences or even just single words.

Functionalities

(1) The prototype lets you search in *specialized* educational resources hubs, *directly and transparently from Google* results pages (or indeed from any educational resource you previously identified with any other technique). How do you compare this strategy to support the discovery of educational resources, to the alternative of having multiple, separate, specialized search portals? Transparently starting from Google results pages is:

1 (much worse)	2	3	4	5	6	7 (much better)
-------------------	---	---	---	---	---	--------------------

(2) The prototype lets you identify “similar” resources, whose similarity is evaluated taking into account the *number of learning objectives in common*. How do you compare this similarity metric, based on *shared learning objectives*, to the commonly available metrics usually based on *shared words*? Similarity based on learning objectives is:

1 (much worse)	2	3	4	5	6	7 (much better)
-------------------	---	---	---	---	---	--------------------

Could you indicate one or more reasons?

(3) The prototype lets you find resources by similarity. How useful in your job is the possibility of discovering “similar” resources *in general* (therefore including with potentially different metrics such as, for example, “frequently used together”, “used by the same educator”, “liked by the same users”, etc.)?

1 (not at all useful)	2	3	4	5 (very useful)
--------------------------	---	---	---	--------------------

Could you indicate one or more reasons?

(4) The prototype lets you find “similar” resources, that is, which *share the same learning objectives*. Discovering this type of *similar* resources with alternative techniques (you currently use) would be:

1 (very easy)	2	3	4 (neutral)	5	6	7 (very difficult)
------------------	---	---	----------------	---	---	-----------------------

(5) The prototype groups (clusters) results in classes of homogeneous similarity, letting users select resources with the desired degree of similarity. How much useful is this feature to help users *making sense* (getting an overall idea) of the sometimes large volume of returned resources, in comparison to the alternative of simply visualizing the most similar resources?

1 (not at all useful)	2	3	4	5 (very useful)
--------------------------	---	---	---	--------------------

Could you indicate one or more reasons?

(6) The prototype groups (clusters) results in classes of homogeneous similarity, letting users select resources with the desired degree of similarity. How much useful is this feature for educators searching for similar resources, considering for example that the learning objectives of resources for *remediation activities* should be very similar to those of resources previously used, while the learning objectives of resources for *in-depth* activities should not be too similar?

1 (not at all useful)	2	3	4	5 (very useful)
--------------------------	---	---	---	--------------------

Could you indicate one or more reasons?

(7) The prototype groups (clusters) results in three classes of similarity: Maximum, Average, and Minimum. Do you think it should use a different number of classes, and why?

(8) What are, in your view, the two most important additional tasks that the prototype should support?

Usability

(9) The system speaks the users' language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms:

1 (fully disagree)	2	3	4 (undecided)	5	6	7 (fully agree)
-----------------------	---	---	------------------	---	---	--------------------

(10) The “*transparency*” of the tool, exploiting learning objectives and educational frameworks without the need to explicitly manipulate them, is:

1 (not at all useful)	2	3	4	5 (very useful)
--------------------------	---	---	---	--------------------

User Experience

(11) I would like using a tool like this to find educational resources:

1 (fully disagree)	2	3	4 (undecided)	5	6	7 (fully agree)
-----------------------	---	---	------------------	---	---	--------------------

(12) I think this system automates unwanted workload:

1 (fully disagree)	2	3	4 (undecided)	5	6	7 (fully agree)
-----------------------	---	---	------------------	---	---	--------------------

(13) I would recommend a tool like this to a colleague:

1 (fully disagree)	2	3	4 (undecided)	5	6	7 (fully agree)
-----------------------	---	---	------------------	---	---	--------------------

General

(14) How *relevant* are the scenarios proposed in this evaluation, in relation to your searches for educational resources as an educator?

1 (not at all relevant)	2	3	4	5 (very relevant)
----------------------------	---	---	---	----------------------

Could you indicate one or more reasons?

(15) What do you like most about the prototype?

.....
.....

(16) What do you dislike most about the prototype?

.....
.....

(17) Please feel welcome to leave any further comment, suggestion, or critique:

.....
.....

Thank you very much for your appreciated collaboration,

Renato Cortinovia

Additional field notes about the use of the prototype when the questionnaire is administered as structured interview:

.....