PlanetDR, a scalable architecture for federated repositories supporting IMS Learning Design

Josep Blat [1], David Griffiths [1], Toni Navarrete[1], José Luis Santos [1], Pedro García[2], Jordi Pujol[2]

[1]: Universitat Pompeu Fabra
Passeig de Circumval·lació, 8. 08003. Barcelona. Spain
{josep.blat, david.griffiths, toni.navarrete, joseluis.santos}@upf.edu
Phone : (+34) 93 5422237. Fax : (+34) 93 5422517

[2]: Universitat Rovira i Virgili
Avinguda dels Països Catalans, 26. 43007. Tarragona. Spain
{pedro.garcia, jordi.pujol}@urv.net
Phone : (+34) 977 558510. Fax : (+34) 977 559710

Abstract

This paper discusses PlanetDR, whose architecture supports very large federated educational digital repositories. It is based on the implementation of current open specifications for interoperability (such as IEEE Learning Object Metadata and IMS Digital Repositories Interoperability, in its Edusource Communication Language version), and its integration with the workflow of eLearning production in the context of the Reload Learning Design editor. This integration should support better re-use of resources; some open problems for enhancing further this re-use are also discussed.

Keywords: Federated educational repositories, Interoperability of learning objects, IMS Learning Design, Pedagogy-aware services.

Introduction

Historically, the emergence of standards and open specifications indicates that a particular process or technology is maturing and has achieved a degree of commercial success. Nevertheless, in elearning settings, the adoption of standards involves a slow process for both educational institutions and commercial companies. This is because elearning specifications are typically based on predictions of what users need (often on the basis of research, eg IMS LD (IMS Global Learning 2003)) rather than on widely accepted emerging practice. Although some learning specifications are mature, such as LOM (IEEE WG12 2002) and SCORM (ADL Technical Team 2004), their widespread adoption in institutions and software packages is still a difficult and slow process.

Concerning Learning repository interoperability specifications, the problem is even greater. Whereas a plethora of specifications have been proposed distributed content repositories (for example Edutella (Nejdl, Wolf et al. 2002), PONDER (Hatala and Richards 2002), Ariadne (The Ariadne Foundation 2004)), the lack of interoperability among them hinders universal content aggregation in a single worldwide repository. As a consequence, content islands exist full of tagged LOM contents that are only reachable to small communities.

In theory, the unifying standard that should enable server interoperability is the IMS Digital Repository Interoperability specification (DRI) (IMS Global Learning 2003). The IMS Digital Repository Interoperability Group provided a functional architecture and reference model for repository interoperability. Aiming at very broad application the specification makes recommendations only at a certain level, leaving the resolution of more operational issues to the system implementers. This rather fuzzy approach leaves many open questions that work against effective widespread adoption.

Fortunately, a Canadian network repository has proposed a concrete instance of DRI called Edusource Communication Language (ECL) (Hatala and Richards 2002). PlanetDR has made a strong commitment to open standards and tools, supporting LOM and DRI, and is the first Open Source Learning repository that fully supports ECL.

Another proposal is the Learning Object Resources Interoperability Framework (LORI) (Simon, Massart et al. 2005). This has been developed in the context of the PROLEARN (PROLEARN 2005) project, and it distinguishes between core services and application services. Both of these require a common messaging infrastructure, which enables repositories to interact (XML-RPC, Java RMI, or WSDL/SOAP). In general, LORI follows a much simpler protocol than ECL and DRI, trying to avoid the complexities of XQuery. This simplicity is good for easing the implementation of LORI's SQI (Simple Query Interface) and thus lowering the burden to Digital Repositories. On the other hand, it permits less flexible queries than DRI and thus limits content access and retrieval. LORI's SQI is a widely accepted interoperability protocol in European settings in the projects ARIADNE and ELENA (ELENA 2005).

Finally, MIT's DSpace (MIT Libraries and Hewlett-Packard Company 2005) is another Open Source Learning repository that includes federation capabilities. Although Dspace is not based on open standards, MIT has attracted a large number of Universities to the Dspace Federation. In the coming years, DSpace could reach the critical mass and become a *de facto* standard in learning repositories. In conclusion, either with de facto standards like DSpace or with well-specified protocols like DRI or LORI SQI, in the next few years we hope to see how LOM content islands can become integrated in a worldwide interconnected repository network. To make this possible, we believe that more scalable and robust technologies (structured peer to peer) will be required to construct such large server federations, such as the ones presented in the PLANET architecture and presented in this paper

On the other hand, the retrieval of material in educational repositories is an isolated task in the educational workflow. This isolation might hinder the re-use of educational resources, which should be the real goal of interoperable repositories. The retrieval functionality enabled by PlanetDR would become more educationally meaningful if integrated in the process of creating learning activities. With this in mind the second part of this paper presents the work implementing its integration with the Reload (Reload 2005) editor, a reference Open Source tool for the creation of IMS Learning Design Units of Learning. In this way, resources queried and retrieved from repositories can be seamlessly incorporated into actual "lessons". IMS Learning Design (LD) (IMS Global Learning 2003) is a recent specification allowing the representation of how multiple learners and teachers can work with resources in different activities, whose context we discuss briefly in this paper.

Finally, we discuss some open problems with the wider re-use of resources in this context, both technical, such as the need for repositories to go beyond LOM based searching, fully supporting Learning Design based search of resources, and more of social nature, such as allowing for the re-use of the most successful resources.

The next section describes the PlanetDR repository in detail, while the following section introduces IMS LD and describes how Reload has been extended to deal with PlanetDR and the path to LD-aware repositories. The final section provides some conclusions.

PlanetDR content repository

The basic operation of a content repository is to provide the means for uploading resources, which are stored into a data warehouse. Later, these resources must be made accessible to registered users by allowing them to search contents by a broad variety of criteria.

When designing our own content repository, we had interoperability in mind. We chose the Planet Digital Repository (PlanetDR) to implement the ECL protocol using web services. It also complies with the DRI interoperability specification, and both these specifications are described below.

DRI and ECL

The purpose of the Digital Repositories Interoperabilty specification is to provide recommendations for interoperating between the most common repository functions. These recommendations should be implementable across services enabling them to present a common interface. DRI utilizes existing schemas, such as IMS Meta-Data, mainly based on LOM, and Content Packaging (CP) (IMS Global Learning 2004).

The DRI specification recognises that in the area of digital repositories a wide range of pre-existing content formats, implemented systems, and established practices . Consequently, its recommendations are in two categories:

- Systems reflecting established practice (e.g. utilizing Z39.50 for repository interoperability).
- Systems that are able to implement the XQuery and SOAP-based recommendations.

Focusing on the second alternative, which PlanetDR is based on, some core functions are defined as **web services**, which are exposed through the Internet. This approach, using SOAP, combined with WSDL (Web Services Description Language), allows the content server to specify what services it provides, what the inputs/outputs of these services are, and how to encode/decode requests and responses exchanged between clients and servers. These core functions are described as follows:

Search/Expose: The search reference model defines searching through meta-data associated with content exposed by repositories. Searching is performed using the XQuery protocol over XML meta-data that follows the IMS Meta-Data Schema. XQuery has a well-defined grammar, and several commercial implementations are emerging from the community. Its strengths are query-by-example and structured searches of XML documents and repositories containing IMS meta-data.

Submit/Store: The submit/store functionality refers to the way an object is moved to a repository from a given network-accessible location, and how the object will then be represented inside that repository for access. The location from which an object is moved can be another repository, a learning management system, a developer's hard-drive, or any other networked location. It is anticipated that existing repository systems may already have established means for achieving Submit/Store functions (typically FTP). The DRI specification provides no particular recommendations for legacy repository systems, but draws attention to the following weaknesses of FTP as a transport mechanism for learning objects or other assets: plain FTP provides no encryption capabilities, presents widely-recognized security flaws and does not provide means of confirming the successful delivery of assets from one networked location to another. In the case of more recently developed repositories that deal specifically with learning objects, DRI makes significant reference to the CP specification.

Request/Deliver: The *request* functional component allows users that have located a meta-data record via the Search function to access the content object or other resource described by this meta-data. *Deliver* refers to the response received from the repository which provides access to the resource.

Gather/Expose: The *gather* reference model defines repository-exposed meta-data requests, and meta-data aggregation for use in subsequent searches, or for creating a new meta-data repository. The aggregated repository becomes another entity available for *Search/Expose* functions. The *gather* component may interact with repositories either by actively requesting meta-data from a repository, or by subscribing to a meta-data notification service. This notification service may be provided by the repository itself or by an external adapter that enables messaging between the repository and other users, thus following a push-based approach.

As mentioned above, one implementation of the DRI specification is ECL. This is part of the eduSource project, whose main aim is to create a network of linked and interoperable learning object repositories across Canada. Although previous projects had informally created a distributed network that allowed the search and retrieval of educational objects between projects and organizations, there was no formal discussion of any best practice for the future. A substantial part of the project is the creation of communication protocols for sharing information as well as publishing the web services so anyone can tap their components into that pool of educational material and services.

Since the complexity of the ECL protocol might be detrimental to its adoption, an **eduSource connector** which implements the ECL protocol is provided. The connector provides a standard API to connect an existing repository to the eduSource network. The ECL protocol requires institution repositories or tools to implement connector handlers only for those services they want to expose to others, which is far simpler than implementing and deploying every service in each institution. The connector also facilitates version synchronization during the protocol evolution. Changes in the protocol itself rarely propagate to the API level. In most cases, repositories do not have to worry about the change in the protocol, they only need to update the connector with a newer version. Changes in the ECL protocol are detected by the newer version of the connector and are dealt with automatically.

Planet Digital Repository (PlanetDR)

The educational content repository which we have developed is called **PlanetDR**, and it is an implementation of the ECL protocol described above. The web services designed include a search service, a submit service, and a request service. PlanetDR includes several search types: the *quick search* function allows searching for content keywords which match any of the meta-data fields for a particular content; the *advanced search* function can be split into two additional types as well: search by main meta-data category, where any LOM meta-data field can be specified, and the *accumulated search*, which allows searching for any field, linking together conditions of different LOM categories.

One interesting feature of PlanetDR is the possibility of invoking any web services from other content servers in the eduSource network. This is easily achieved because all of these servers follow the same ECL protocol. In this case, what we call a *federated search* (a simulated gather service) can as well be conducted by linking together request results coming from all active content servers in the eduSource network.

Nevertheless, the content server itself works as a standalone server, which makes it "unaware" of other content servers in the eduSource network. There is no way of easily identifying ECL content servers which it can interoperate with. To solve this, the EduSource network linked servers by hand in a unique central location. This approach clearly hinders the scalability of the federation for a larger number of servers.

With this in mind, we have extended PlanetDR with a **federation mode** (see figure 1), which allows a plug & play decentralized management of PlanetDR compatible servers, thus guaranteeing worldwide scalability. Federation is useful for automatically detecting any new PlanetDR instance active on the network, and easing the process of inserting these new servers into each PlanetDR's local list of servers. In this way each server can join or leave the P2P federation of educational servers, and obtain a listing of all those available in the network. This feature supports certain "awareness" of which content servers are in the network. Each PlanetDR node listens to the different events that can occur (insert / remove), and this allows each instance to maintain an updated list of available servers. Later, directed searches can be sent to any of these nodes.

The overall PlanetDR federation architecture is scalable because it is built on top of the FreePastry (Rowstron and Druschel 2001) structured peer-to-peer overlay network. Furthermore, PlanetDR constructs on top of our peer-to-peer middleware called DERMI (Pairot, García et al. 2004) that provides a decentralized naming service and remote object notification mechanisms.

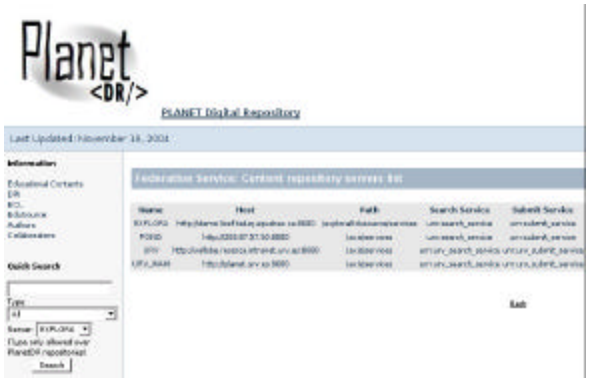

Figure 1. PlanetDR's Federation Mode

Integration of PlanetDR and content creation

IMS Learning Design, Reload and the reuse of resources in context

IMS has produced the IMS Learning Design (LD) specification to represent how multiple learners and teachers work with resources in different activities, a need not covered by SCORM, for instance. LD defines Units of Learning (UoLs) by representing how *people* carry out *activities* in an *environment* composed of learning resources and services. LD is a large and complex specification, and as a product of IMS it is by definition *an interoperability specification*, which does not constrain how eLearning applications should work, but only specifies an import and export format which they must be able to work with if they want to be LD compliant. at the risk of oversimplifying, a UoL can be seen as an interoperable lesson plan. In addition to being a file exchange format, however, LD is also an *Educational Modelling Language,* and a community of researchers has grown up who work with LD, following the lead set out in the LD Best Practice guide (IMS Global Learning Inc 2003) which proposes an *eLearning methodology* for the creation and use of UoLs. A *set of applications* has been developed to facilitate the creation and playing of UoLs using LD. All these aspects are reflected in the activities of the UNFOLD project (UNFOLD 2006) and interested readers can also find detailed information in (Koper and Tattersall 2005).

The Reload Learning Design Editor (Milligan, Beauvoir et al. 2005) available for download at (Reload 2005) is an Open Source editor of UoLs. It sets out to be a *reference implementation,* that is to say that it implements the entire specification and provides a reference point for other developers who are in doubt as to how the specification should be interpreted. The very large number of downloads from the Reload website and the number of references to it in the published literature suggest that it has been successful in this. The equivalent Open Source implementation for runtime, the "reference player", is CopperCore (Vogten, Koper et al. 2005). In its current version, Reload supports a single user on a single machine program, whose inputs, such as resources, have to be locally available, and whose output is a zip file packaged according to the CP specification.

While PlanetDR supports search and retrieval of resources based on LOM and DRI as an isolated task, a much more natural working context is to perform this task when an author is designing a UoL, and to be able to include the resources retrieved in the UoL. To this end we have extended the Reload editor to allow to search and retrieve resources from the PlanetDR repository, and include them in the workflow of UoL production. We describe next how this is done and discuss the benefits.

Connecting Reload to resources from PlanetDR

A new window has been added to the Reload editor that enables the user to specify the fields for searching the resources in the repository. These fields are name, keywords and format. According to the values provided by the user, the tool builds an XQuery statement, which is sent to the repository and executed. The results of the query are presented to the user, who can select one or more resources from the list and download them in a zip file, in order to use them in the UoL that is being designed. The zip file is a requirement related to the CP specification and contains not only the resource but also the metadata file.

The tool interacts with two of the PlanetDR web services which implement the ECL protocol, *search* and *request*. The first is called in order to send to the repository a query specified in XQuery. The query is executed on the LOM metadata files of the resources in the repository, and the service returns a string containing the list of resources that satisfy the query. Secondly the *request* service is called in order to download the resource.

Due to the very complex and varied nature of eLearning, the daunting task of providing interoperability specification has been broken into pieces, such as LOM, DRI, CP, LD, and others not mentioned in this paper. This simplifies the task of specifications implementers, and makes compliance more practicable. For the user, however, this may create difficulties, as it can cause unitary tasks (such as preparing a course module) to be divided into seemingly unrelated parts. The user needs to have these specifications transparently integrated in a workflow, and indeed in many cases the user should not be aware of the various underlying specifications. Our work integrating LD, DRI and LOM in a natural workflow of producing a UoL, shows that this is possible, and we believe this is a key step for usability.

Our work also indicates a possible path for performing this type of integration by re-using pieces of Open Source code, and gluing them together through a web services approach. A general and open architecture for eLearning, which could be based on web services is discussed by Wilson in (Wilson 2005); the SLeD project (McAndrew, Nadolski et al. 2005) has produced a prototype of such an architecture for LD making it possible for new services to be plugged in (such as searching, blogging, …).

The use of CP and zip files, probably dating from the ascendancy of CD-ROMs, makes difficult to make full use of distributed resources, which might have their own rhythm of updating, and forces local downloading, re-packaging. A much more natural perspective in the context of the Web is to link directly to the (distributed) resource(s). The way in which SLeD has built on CopperCore to provide a service layer is an indication of an approach to achieving this, which will be followed in further work. The service based approach to Learning Design established by SLeD was further elaborated in the Berlin Architecture, developed in the context of UNFOLD and being further elaborated in the TENCompetence project (TenCompetence 2006).

Further perspectives for future work

Another perspective of future work is related to fully utilising the potential of IMS LD and taking into account social use. Current approaches only use LOM, and reuse is limited to resources, but IMS LD should support reuse of pedagogy, services, etc. (Griffiths, Blat et al. 2005) points out the need of repositories providing an LD awareness. An LD-aware repository could support searches for UoLs that have been used with a certain kind of content, retrieve fragments of UoLs, or provide metadata on the use of UoLs. It is reasonable to suppose that teachers will not simply identify and use UoLs on the basis of LOM, but will also, and perhaps more importantly, base their decisions on the practice of the mass of their peers, or of individuals who they respect. Consequently popularity is one of the reasons why resources or pedagogies will attract use by others. Moreover, for the identification and refinement of successful practice it is also necessary for the history of use to be represented. As much of this as possible should be done automatically, as it has been clear for some years that most users are highly resistant to adding metadata to resources (Thomas and Griffin 1998). EduSplash (Hatala and Richards 2002) is a proposal for repository which takes into account these aspects, beyond the EduSource project. Another approach, related more to popularity, is currently being adopted by the Lionshare project (OKI Case Study 2005). Automatic analysis can show teachers which resources are popular in their area / age group / curriculum. Lionshare is using the Shibboleth system developed by Internet2 to create flexible trusted communities and in such a context it may be possible to identify the individual teachers who have been using the resources, enabling teachers to emulate the practice of their successful peers. We intend to investigate how the reworkings of UoLs are associated with the UoLs on which they have been based, to

permit browsing up and down the hierarchies of parents and children. For a discussion of how these technical innovations relate to the wider context of reuse of educational resources, please see (Griffiths 2005)

A relevant and related approach is taken in (Knight, Gaševic et al. 2005) who propose an ontology based approach to integrate learning designs and learning object content, with the aim of increasing reusability by enabling the use of a given learning design with different content. Integration of this work with that described in this paper could provide additional support for the user in identifying alternative learning resources for use in Units of Learning.

An interesting and quite different approach to the support which repositories can provide users of UoLs is provided in (Brouns, Koper et al. 2005). She suggests that it may be possible to use Latent Semantic Analysis and indexing in order to find concepts and similarities of concepts within a corpus of UoLs. The degree to which this promising idea will be practicable is not yet clear, as stated in her conclusions on the approach, setting out a number of questions for further investigation: "Can it be used to classify designs as good as well as bad practices, for example when user data, such as success or failure rates, completion time, etc are added to the analysis, or even with human classification of the design? Are acts the smallest independent units in learning designs? Are the templates sufficient for practitioners to develop new courses?"

Conclusions

In this paper we have discussed PlanetDR, which as well as being based on open specifications, such as LOM and DRI, has an architecture which can support the very large federated repositories of the future.

We have also shown and discussed the implementation and integration of the searching and retrieving facilities of such a tool into the workflow of eLearning production. We deem this to be a key to supporting the re-usability of resources, which is the ultimate goal of repositories. We have discussed some open problems in this approach to supporting re-usability both from technical and social perspectives.

The modifications which we have made to the Reload LD Editor make a contribution towards expanding the functionality and improving the usability of repositories of eLearning Resources. It is, however, clear that this first step needs to be followed up by further work along the lines of future work which are indicated above.

Acknowledgments

This work has been partially funded by the Spanish Ministry of Science and Technology through project PLANET (TIC-2003-09288-C02-00), and by the European Commission through the Framework 6 IST project TenCompetence (027087).

References

ADL Technical Team (2004). *Sharable Content Object Reference Model (SCORM) 2004 2nd Edition*, retrieved 6th October, 2005 from http://www.adlnet.org/downloads/70.cfm

Brouns, F., R. Koper, et al. (2005). A first exploration of inductive analysis approach for detecting learning design patterns. Jounal of Interactive Media in Education, (Advances in Learning Design Special Issue, eds Colin Tattersall, Rob Koper).

ELENA (2005). *Project Website*, retrieved 6th October, 2005 from http://www.elena-project.org

Griffiths, D. (2005). Print to Pixels: the Implications for the Development of Learning Resources. In (Eds.) *I2LOR 2005, second annual e-learning conference on Intelligent Interactive Learning Object Repositories*, Vancouver: November 14-18.

Griffiths, D ., J. Blat, et al. (2005). Learning Design Tools. In R. Koper and C. Tattersall (Eds.) *Learning Design: modelling and implementing network-based education & training*, Springer Verlag, 109-135.

Hatala, M. and G. Richards (2002). Global vs. Community Metadata Standards: Empowering Users for Knowledge Exchange. In (Eds.) *First International Semantic Web Conference*, Sardinia.

IEEE WG12 (2002). *IEEE Standard for Learning Object Metadata*, retrieved 6th October, 2005 from http://ltsc.ieee.org/wg12/

IMS Global Learning (2003). *IMS Digital Repositories Interoperability, Version 1*, retrieved 6th September, 2005 from http://www.imsglobal.org/digitalrepositories/

IMS Global Learning (2003). *Learning Design Specification Version 1*, retrieved 21st June, 2005 from http://www.imsglobal.org/learningdesign/

IMS Global Learning (2004). *Content Packaging Specification, 1.1.4*, retrieved 6th October, 2005 from http://www.imsglobal.org/content/packaging/index.html

IMS Global Learning Inc (2003). *IMS Learning Design Best Practice and Implementation Guide*, retrieved 29th March, 2005 from http://www.imsglobal.org/learningdesign/ldv1p0/imsld_bestv1p0.html#1501849

Knight, C., D. Gaševic, et al. (2005). Ontologies to integrate learning design and learning content. Journal of Interactive Media in Education (Advances in Learning Design. Special Issue, eds. Colin Tattersall, Rob Koper), (2005/7).

Koper, R. and C. Tattersall (2005). *Learning Design: modelling and implementing network-based education & training*. Springer.

McAndrew, P., R. Nadolski, et al. (2005). Developing an approach for Learning Design players. Jounal of Interactive Media in Education, (Advances in Learning Design Special Issue, eds Colin Tattersall, Rob Koper).

Milligan, C. D., P. Beauvoir, et al. (2005). The Reload Learning Design Tools. Journal of Interactive Media in Education (Advances in Learning Design. Special Issue, eds. Colin Tattersall, Rob Koper), (2005/07).

MIT Libraries and Hewlett-Packard Company (2005). *DSpace Federation*, retrieved from http://dspace.org/

Nejdl, W., B. Wolf, et al. (2002). EDUTELLA: A P2P Networking Infrastructure Based on RDF. In (Eds.) *WWW 2002*, Chiba.

OKI Case Study (2005). *The Repository OSID and LionShare: Leveraging standards to provide broad markets for content (Draft)*, retrieved 6th September, 2005 from http://www.okiproject.org/documents/CaseStudy-Lionshare.pdf

Pairot, P., P. García, et al. (2004). Dermi: A New Distributed Hash Table-based Middleware Framework. IEEE Internet Computing, 8(3): 74-84.

PROLEARN (2005). *Project Website*, retrieved 6th October, 2005 from http://www.prolearn-project.org

Reload (2005). *Project website*, retrieved 1st April, 2005 from http://www.reload.ac.uk/

Rowstron, A. and P. Druschel (2001). Pastry: Scalable, decentralized object location, and routing for large-scale Peer-to-Peer systems. In (Eds.) *International Conference on Distributed Systems Platforms (IFIP/ACM)*, 329--350.

Simon, B., D. Massart, et al. (2005). Learning Object Repository Interoperability Framework: 8.

TenCompetence (2006). *Project Website*, retrieved 6th March, 2006 from http://www.tencompetence.org/

The Ariadne Foundation (2004). *The two ARIADNE projects*, retrieved 6th October, 2005 from http://www.ariadne-eu.org/en/foundation/history.html

Thomas, C. F. and L. S. Griffin (1998). Who Will Create The Metadata For The Internet. First Monday, 3(12).

UNFOLD (2006). *Project Website*, retrieved 6th March, 2006 from www.unfold-project.net

Vogten, H., R. Koper, et al. (2005). An Architecture for Learning Design Engines. In R. Koper and C. Tattersall (Eds.) *Learning Design, A Handbook on Modelling and Delivering Networked Eductation and Training*, Berlin: Springer.

Wilson, S. (2005). Architectures to Support Authoring and Content Management with Learning Design. In R. Koper (Eds.) *Learning Design, a Handbook on Modelling and Delivering Networked Education and Training*, Springer, 41-62.