

面向教育的关联网络数据项目之 全面开放机制研究^{*}

王丹丹

(河南科技大学管理学院, 洛阳 471003)

摘要: 面向教育的关联网络数据项目 (LinkedUp Project) 是教育领域开放数据的先行者, 它将开放元素引入项目全过程, 实现贯穿于数据竞赛所有环节的开放机制, 充分保证项目及其成果的可持续性和可开发性。总结 LinkedUp 项目在倡导和宣传关联开放数据方面所做的努力, 从开放数据竞赛、开放评价体系、开放数据目录三个核心元素入手, 分析其开放运营管理模式; 从构建开放社区发布开放教育指南, 创造开放支撑环境促进参与宣传双向互动, 以及借助开放培训和开放合作提升数据应用能力三个方面, 总结全面开放思想在 LinkedUp 项目中的运用。通过揭示 LinkedUp 的成功经验, 为相关机构实施开放数据竞赛提供有益借鉴。

关键词: 教育; 开放数据; 关联数据; 数据竞赛

中图分类号: G250

DOI: 10.3772/j.issn.1673-2286.2015.07.008

1 引言

自2009年美国数据门户网站data.gov上线以来, 开放数据运动便在世界多国和地区陆续开展。开放数据作为现代经济活动重要的生产要素, 成为继开源软件、开放获取后又一崇尚开放共享精神的研究热点。2013年6月, 美、英、法、德、意、加、日、俄召开G8峰会, 签署G8开放数据宪章, 进一步推进开放数据运动^[1]。作为开放数据运动的产物之一, 各种形式的以创新应用为目的的开放数据竞赛在全球如火如荼地开展^[2], 已经成为释放数据能量, 鼓励更多数据开放共享, 推动创新的有效途径。欧盟第七框架计划支持行动 (EU FP7 Support Action) 资助的项目“面向教育的关联网络数据项目 (LinkedUp Project)”是教育领域数据开放的先行者, 它将开放元素引入项目的全过程, 实现贯穿于数据竞赛所有环节的开放机制, 充分保证项目及其成果的可持续性和可开发性。分析 LinkedUp 项目在倡导和宣传教育领域关联开放数据方面所做出的努力, 总结

其在构建开放数据竞争体系的过程所实现的全方位开放数据管理方式, 分析其涉及的步骤和取得的教训, 为开展数据竞赛实现创新的相关机构提供借鉴。

2 LinkedUp项目核心要素分析

LinkedUp项目旨在推进数据开放以及网络上公共、开放数据集的开发利用。项目由德国汉诺威莱布尼茨大学领导, 包含爱斯维尔、加拿大联邦学习联盟和荷兰数据归档和网络服务中心等12个利益相关的网络伙伴联盟, 同时还吸纳开放数据与数据管理、语义网和网络教育领域的知名专家组成项目咨询委员会。

LinkedUp项目的目标包括收集开放网络数据的成功案例, 即收集来自知名商业、学术和教育机构 (如 Elsevier, BBC, 或 Commonwealth of Learning) 的合适案例; 收集、评注和保存高质量的教育相关数据集, 使第三方可用; 提供完整的面向大规模网络数据应用系统的通用评价框架以及示范教育领域中开放网络数据

^{*} 本研究得到国家社科基金项目“数字图书馆用户数据资源化管理研究” (编号: 2013CTQ012) 和河南科技大学教学改革项目“基于PBL方法的文献检索课程改革研究” (编号: 2013Y-015) 资助。

应用系统。为实现这些目标, LinkedUp采取一系列行动, 包括开展数据竞赛并建立相应的评价框架, 提供描述良好的目录和可评价的数据集, 对数据实施管理, 并对竞赛的参与者、感兴趣的数据用户和一般的应用程序开发人员提供外部环境支持等。

2.1 开放数据竞赛

开展数据竞赛可以使数据所有者了解同一数据集的不同使用方法。作为鼓励创造性和创新性的一种方式, LinkedUp竞赛不仅有助于发现教育类资源、推进跨域资源的合理链接、优化并创新开放远程教育方式, 还有助于激发相关机构和人员使用关联数据的兴趣, 促进公司、大学、政府机构等共享资源。LinkedUp竞赛具有三个特色:

首先, 提供公开透明的数据资源。最近几年教育领域已经组织过不同规模的数据竞赛, 如Elsevier挑战赛^[3]、语义网服务竞赛^[4]、关联数据挖掘挑战赛^[5]、混搭挑战赛^[6], 以及汤森路透开展的面向学术研究的创新挑战赛^[7]等。纵观这些竞赛, 可以发现成功开展数据竞赛最重要的因素是数据质量, 不仅指数据自身质量和可用性, 也包括数据的一致性、相关性和可理解性。2012年亚太大学交流(UMAP)会议上因为用户不清楚数据集的用途和使用方式, 大数据竞赛工作组没有吸引到参赛人员, 也没有参赛作品提交。所以, LinkedUp竞赛通常针对对关联教育数据感兴趣的社群(学生、研究人员、新兴企业及某些行业), 对教育相关的数据进行收集和编目, 并提供数据搜索服务。

其次, 开启三阶段独立混合赛式。LinkedUp竞赛包含三次比赛共分三个阶段依次进行(如表1), 每个阶段允许新成员加入, 因此既适合全程参与的长期项目, 也适合持续时间短的小项目。参赛团队可以根据自身优势、技术实力和选择不同场次。这一阶段性持续推进的模式也给竞赛的营销和宣传创造更多的机会。每场竞赛都包含“开放任务”(对任何满足进入要求的成果均是开放的)和“重点任务”(致力于发现具体问题的解决方案)两个部分。在竞赛Veni中, 没有对这两者做区分, 只是模糊地将重点任务界定为参与者按照规定的要求提交成果, 具体而言要求所提交的成果必须提供申请人、任务情景等信息, 对问题进行陈述并提出数据集和示范解决方案。这种标准化的描述方式既向参赛者展示LinkedUp项目所希望的成果类型, 也为开

放任务的参赛者提供情景, 激发其灵感, 逐步演化为后续竞赛中的“重点任务”。在后两个阶段的竞赛Vidi和Vici中, “开放任务”和“重点任务”竞赛并行举行, 采用同一评价标准, 而重点任务则增加评判解决具体问题程度的附加标准。

表1 LinkedUp挑战赛阶段及时间

2013年3月-2013年10月	2013年11月-2014年4月	2014年5月-2014年11月
Veni竞赛		
	Vidi竞赛: 开放跟踪	
	Vidi竞赛: 重点跟踪	
		Vici竞赛: 开放跟踪
		Vici竞赛: 重点跟踪

第三, 充分考虑参赛对象间差异, 区别对待并注重平衡。LinkedUp项目要求参与者提交论文, 使用诸多学术研讨方式进行, 如征集论文、项目委员会、录用通知等。而汤森路透的创新挑战赛则使用图形而非文字表述, 如用“评判员”、“寻求解决方案”、“赢家公告”等方式进行。鉴于学术界与工业界间的诸多差异, LinkedUp采取平衡策略来兼顾两者, 在竞赛宣传方面将奖励、网络、研究问题与机会等包含进来, 在成果提交方面同时设计针对研究者的学术论文大纲和针对工业人士的表单模板。

2.2 开放评价体系

LinkedUp项目所开发的评价框架涉及教育创新、可用性、性能、数据、合法性和用户六个维度的内容, 每一个维度包含具体评价指标, 并设定相应评价方法(如图1), 是一个可用于评价大规模开放网络数据应用系统的完整评价体系。这一评价框架被成功应用于Veni中^[8]。评判者利用这6个维度和具体指标对提交的22件作品进行排序, 最终确定三个获奖者。该评价框架的特色在于:

首先, 将定量与定性评价相结合, 呈现出高度的透明性和灵活性, 不仅可用于测度大规模网络信息和数据应用系统的影响力, 也有助于识别特殊项目和作品的优势和劣势。其他组织者可以从灵活选择具体的维度和指标, 结合实际需要构筑自身评价体系。

其次, 伴随着项目的推进而动态优化升级完善。在

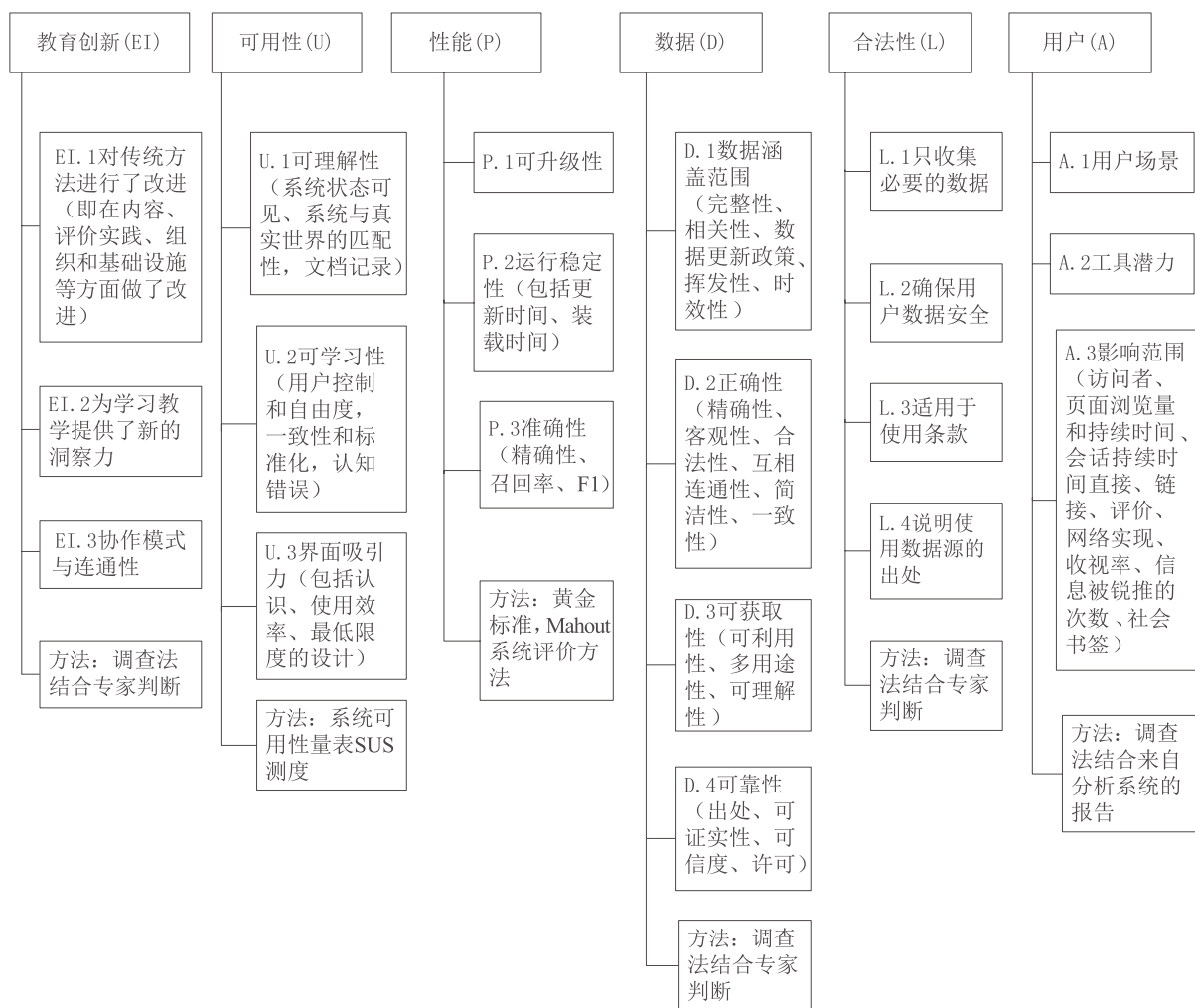


图1 LinkedUp评价框架

竞赛的每一阶段都引入评价环节, 就评价框架的有效性、可改进之处征求意见, 从而使评价框架在使用的过程中得以不断改进完善, 最终成为实用且能可持续发展的工具。

再次, 以开发工具箱为落脚点, 成为帮助数据科学家在特定领域举办竞赛活动的有利助手。工具箱包括各种格式和版本的评价框架、问题列表、评价过程和分类、最佳实践以及案例研究等, 也提供诸如经验教训、相关链接、法律参考、营销策略、评价组织攻略以及指南和模板等有用信息, 供数据再使用者、数据拥有者、竞赛组织者和数据出版商使用^[9]。

2.3 开放数据目录

为了支持竞赛参与者更广泛地使用教育领域中的开放关联数据, LinkedUp项目坚持创建并维护与教育

场景相关的数据集, 进行选择 and 分类, 构建知识库存储数据集^[10]。LinkedUp数据知识库(又称关联教育云 Linked Education Cloud), 收集并以简单可用的形式存储数据, 支持参赛者获取相关数据, 将其发展成对教育领域开放数据的开发者和实践者有用的资源^[11]。目前知识库包含22个数据结点, 共代表教育领域146个不同的数据集(包括学术课程、教育资源、研究文章、教育机构 and 设施)。

开放数据目录有多种用途: 首先是一个数据集注册表, 考虑到Datahub.io是最受欢迎的全球数据集目录注册中心, 形成关联开放数据云的核心^[12]。为了与其他正在进行的开放数据活动进行集成, LinkedUp用一组基本的元数据和分配资源对每一个数据集进行描述, 将数据目录作为Datahub.io的一部分进行创建, 在Datahub.io网站上对全球用户开放。例如, 在关联教育云中检索“大学”, 便可获得在元数据中明确提及“大

学”的数据集。对检索结果过滤和精炼后,便可筛选出以RDF或XML格式提供的数据集资源。作为开放数据目录的副产品,团队技术人员还组建“代码诊所”,便于参赛者与LinkedUp合作者分享开发过程中的具体问题,讨论可能的解决方案,确保参赛者在开发和部署系统时不受技术问题的困扰^[13]。

3 开放思想在LinkedUp项目全过程的应用

3.1 构建开放社区,激发协同效应

依托2013年扩展语义网会议(Extended Semantic Web Conference)、学习分析和知识会议(Learning Analytics and Knowledge)以及国际万维网会议,LinkedUp项目不断吸纳新的利益相关者,形成包含内容生产者、技术专家和学术界人士、数据管理社区及出版机构在内的活跃、多样、人气旺盛的社区。其目的是借助开放社区的力量,使LinkedUp的思想和成果得以传承和持续发展^[14-15]。

为了讨论教育领域中有关开放数据问题,LinkedUp项目成立开放教育工作组,关注开放知识在特定领域应用及社会效应^[16]。开放知识基金会将工作组的角色界定为特定领域的协调人,工作组由三个小组构成,即决定工作组方向和优先事项的成员组,以共享信息为目的的兴趣讨论组,由开放教育倡导者构成的咨询委员会。工作组需要与开放教育资源社区,W3C开放关联教育社群^[17]展开合作,关注开放教育所有的方面,与现有社群关联起来,发挥开放教育行动中项目、活动和人员的协同效应。

3.2 创造开放环境,实现参与、宣传双向互动

为了告知并吸引目标受众,LinkedUp定期更新网站、博客和社交媒体,发布新闻和举办面对面的活动,让受众了解竞赛等相关信息,以识别和描述目标用户。LinkedUp有LinkedUp项目网站和LinkedUp竞赛网站。项目网站包含工作组设置及其职责信息,竞赛网站是发布竞赛信息的官方工具,包含参赛人员的全部信息,以及评审流程、联系方式、应用场景、参考用例等。

为确保各相关利益群体实现有效交流,LinkedUp

构建社群合作伙伴内部交流的电子小组,面向相关合作伙伴和咨询委员会开放私人邮件列表和公共邮件列表,积极开展面向教育的开放数据、关联数据和语义网会议等活动。

为吸引公众参与筛选每场竞赛的获胜者,项目组还采用开放投票系统,进行“公众选择”,使用投票系统需进行注册、每人次一票。参赛者可以邀请朋友和同事为他们的成果投票,在宣传过程中激发人们的兴趣。正是因为有参与者的反馈和监督,才使LinkedUp项目能借助竞赛活动来扩大宣传,达到营销目的,同时客观评论也有助于通用性评价框架的完善。

3.3 借助开放培训与合作,提升数据应用能力

LinkedUp项目借助博客DevTalk定期分享基于开放关联数据构建应用系统的经验,在资源页面提供一组指向其他资源的链接,帮助开发者快速进入竞赛。同时以合同形式承诺提供支持,使参赛者与项目技术团队专家进行互动、交流,来实施基于开放网络数据的应用系统。其次,项目团队与欧盟构建关联数据课程项目(EUCLID)合作开发培训材料,主要指向现实数据集、应用场景和代码片断,开发者在自身系统上演示,提供最佳实践和操作方法。最后,为了支持竞赛,LinkedUp项目在Adobe Connect上举办网络研讨会介绍竞赛情况,解答参赛者的有关技术管理问题。

此外,LinkedUp项目尝试与其他设立类似竞赛的团队进行合作。一是与欧盟资助的开放教育竞赛合作,开放教育竞赛要求创新者提交有关开放教育领域的创新成果,投资者可以直接访问作品。二是与英国开放数据机构合作进行教育开放数据竞赛。为使用开放数据提供实质性的解决方案,并为父母教育孩子提供决策依据。

4 结语

开放教育涉及领域广泛,包含多个部门和地区。尽管LinkedUp项目主要关注教育领域中开放网络数据的使用,关注数据创建和技术层面问题,但是也需要构建更大社区,以减少障碍、支持学习。这就要求开放数据不仅要准确、有效、及时、可用、完整、与使用对象密切相关,而且还必须以合适方式进行解释,使开放数据成

为有用资源。LinkedUp数据竞赛不仅基于开放和关联数据构建,而且所有竞赛元素也被开放发布和共享。除了竞争框架可以借鉴外,竞赛过程所形成的所有成果,如网站、社交媒体、海报、传单、幻灯片、小样和出版物,以及所取得的经验和教训都可以给提供从事开放数据竞赛的人们参考,使其获益。

参考文献

- [1] 洪京一. 从 G8 开放数据宪章看国外开放政府数据的新进展[J].世界电信,2014(1):55-60.
- [2] 科研教育开放信息创新应用大赛[EB/OL].[2014-10-10].<http://www.openinnovation.cn/>.
- [3] The Elsevier Grand Challenge[EB/OL].[2014-10-10].<http://elseviergrandchallenge.com/>.
- [4] A Semantic Web Service Challenge[EB/OL].[2014-10-10].<http://www.w3.org/2005/04/FSWS/Submissions/58/HPSubmission.htm>.
- [5] CFP:Data Mining on Linked Data workshop with Linked Data Mining Challenge[EB/OL].[2014-10-10].<http://www.josemalvarez.es/web/2012/06/19/cfp-data-mining-on-linked-data-workshop-with-linked-data-mining-challenge/>.
- [6] The Mashup Challenge[EB/OL].[2014-10-10].<http://2013.eswc-conferences.org/program/co-located-events>.
- [7] The Innovation Challenge[EB/OL].[2014-10-10].<http://wokinfo.com/challenge/>.
- [8] Drachsler H,Stoyanov S,d' Aquin M,et al. An Evaluation Framework for Data Competitions in TEL[EB/OL].[2014-10-10]. http://www.l3s.de/~herder/research/papers/2014/evaluation_framework_for_data_competitions.pdf
- [9] Drachsler H, Stoyanov S, d' Aquin M,et al. An Evaluation Framework for Data Competitions in TEL[J]. Open Learning and Teaching in Educational Communities Lecture Notes in Computer Science, 2014(8719):70-83.
- [10] D' Aquin M, Dietze S, Drachsler H,et al. LinkedUp: Linking Open Data for Education[J]. Ariadne, 2014(3).
- [11] LinkedUp Dataset Catalog (or Linked Education Cloud) [EB/OL].[2014-10-10].<http://data.linkededucation.org/linkedin/catalog/>.
- [12] Datahub[EB/OL].[2014-10-10]. <http://datahub.io/>.
- [13] LinkedUp Devtalk[EB/OL].[2014-10-10].<http://data.linkededucation.org/linkedin/devtalk/>.
- [14] LinkedUp events and presentations[EB/OL].[2014-10-10].<http://linkedin-project.eu/events/> and<http://linkedin-project.eu/resources/presentations/>.
- [15] D' Aquin M, Dietze S, Drachsler H. Building the Open Elements of an Open Data Competition[J] D-Lib Magazine, 2014, 20(5/6).
- [16] Open Education Working Group[EB/OL].[2014-10-10]. <http://education.okfn.org>.
- [17] W3C Open Linked Education Community Group[EB/OL].[2014-10-10]. <http://www.w3.org/community/opened/>.

作者简介

王丹丹,女,1980年生,河南科技大学副教授,研究方向:网络化信息服务,E-mail: wangdan430806@163.com。

Research on the Open Mechanism of Linking Web Data for Education

WANG DanDan

(College of Management, Henan University of Science and Technology, Luoyang 471023,China)

Abstract: The LinkedUp Project (Linking Web data for Education) is a pioneer in the field of open education data which introduced the open elements into the whole process of the project to achieve an open mechanism throughout all aspects of the data competition, to make sure the sustainability and development of the project. Summarized the efforts of the LinkedUp project about its advocating and promoting the linking open data. From open data competition, evaluation system of open data and open data catalog, we start the analysis of the operation and management mode. Open ideas used in the LinkedUp project: constructing open education community to issue guidelines, creating supporting environment to promote participation and promotion, developing open training and open cooperation to enhance data application ability. Explored the successful experience of the LinkedUp Project, provided the beneficial reference to other open data competition.

Keywords: Education; Open Data; Linked Data; Data Competition

(收稿日期: 2015-05-14; 编辑: 雷雪)