

ONTOSTATS : UN OUTIL POUR INSTRUMENTALISER LES RESSOURCES EDUCATIVES EN STATISTIQUES

Jean-Marc Meunier

¹ Université Paris 8, Laboratoire Paragraphe, 2 rue de la liberté, F-93200 Saint-Denis, France. jmeunier@univ-paris8.fr

Résumé. Enseigner les statistiques suppose que les outils, conceptuels et matériels utilisés acquiert le statut d'instrument. C'est également le cas des ressources pédagogiques tant pour les étudiants que pour les enseignants. Une condition nécessaire est la structure des connaissances. Dans le projet Ontostats, nous avons utilisé l'ontologie STATO développée par l'Université d'Oxford pour indexer un dépôt de ressources dans une plateforme sous Omeka-S. Dans cet article, nous adoptons une perspective instrumentale pour en explorer les avantages, les perspectives, mais aussi les limites.

Mots-clés. Enseignement des statistiques ; instrumentalisation, ontologie, ressource éducative, Environnement informatique d'apprentissage humain

Abstract. Teaching statistics implies that the tools, concepts and materials used acquire the status of an instrument. This is also the case for educational resources for both students and teachers. A necessary condition is the structure of knowledge. In the Ontostats project, we used the STATO ontology developed by Oxford University to index a resources repository in a platform under Omeka-S. In this article, we adopt an instrumental perspective to explore its advantages, perspectives, but also its limitations.

Keywords. Teaching statistics; instrumentalization, ontology, educational resource, Human learning computer environment

1. Introduction

Comme dans bien des domaines nécessitant l'acquisition d'un savoir-faire, les statistiques articulent deux logiques : celle relative au fonctionnement (les fondements théoriques des statistiques) et celle relatives à la mise en œuvre (quel test choisir et comment le mettre en œuvre) (Richard, 1983; Sander, Meunier, & Bosc-Miné, 2004). Cette coordination s'appuie sur l'organisation des concepts. Or cette structure est très difficile à acquérir par les étudiants (Hahn & Stoltz, 2014; Quilici & Mayer, 1996, 2002). Pour y parvenir, les enseignants ont recours à différentes approches pédagogiques. Armatte (2010) en distingue quatre : (i) une approche formelle, hypothético-déductive (centrées sur l'enseignement des axiomes) ; (ii) une approche inductive ou expérimentale, centrée sur des simulations et l'exploration de données ; (iii) une approche orientée problèmes s'appuyant sur la logique du champ social ou des problèmes de la vie quotidienne ; (iv) l'approche historique qui consiste à présenter aux étudiants les suites de problématisations qui ont conduit à développer telle ou telle méthode. On peut ajouter une cinquième, procédurale celle-ci, qui consiste à se focaliser sur le quand et comment faire, donc très dépendante de l'outil d'analyse, notamment logiciel.

L'expert parvient à articuler plus ou moins ces cinq approches, mais il est rare d'avoir le temps, dans le cadre d'un même cours, de les présenter toutes. En fonction de la discipline et du rôle que les statistiques jouent dans celle-ci, un enseignant pourra en privilégier une plutôt qu'une autre. Offrir des voies d'accès aux autres approches est un des enjeux, mais pas le seul de l'usage des ressources éducatives libres et éditoriales (REL) pour l'enseignement des statistiques comme dans bien d'autres disciplines. Les REL constituent naturellement un prolongement du cours. C'est ce qu'on fait en proposant une bibliographie ou une sitographie complémentaire. Pourtant, à en croire la littérature sur le sujet, même cet usage basique des REL ne semble pas aller de soi pour tous les enseignants

(Atkins, Brown, & Hammond, 2007; Baumgartner, 2014). Rééditorialiser des ressources, voire exploiter une archive à des fins d'apprentissage encore moins. Internet regorge pourtant de documents divers utilisables à des fins d'enseignement (tutoriels vidéos ou non, ebooks, site Internet, article de presse, articles scientifiques ; base de données). Une des ambitions du projet Ontostats est d'explorer les conditions à la fois techniques, mais aussi cognitives et documentaires de faciliter cette réutilisation des ressources libres et éditoriales. Le problème est à la fois psychologique (quels usages dans quelles conditions et par qui ?) ; documentaires (quelles ressources ; quelle indexation, quelle valorisation ; quelle éditorialisation etc.) et informatiques (moyens de diffusion et d'exploitation, mais surtout d'alignement des deux précédentes problématiques).

2 Vers une instrumentalisation des ressources pédagogiques

Notre approche se situe résolument dans la lignée de l'approche instrumentale développée par Rabardel (1995) et formalisé pour la situation d'apprentissage par Rézeau (2002) dans le modèle du carré pédagogique. Au triangle didactique classique reliant le savoir, l'apprenant et l'enseignant, Rézeau propose un quatrième pôle constitué par les instruments qui peuvent être des supports, des activités, des tâches ou tout type de matériels, mais aussi des instruments psychologiques, tels que le langage, les concepts. Rabardel (1995) distingue l'outil (objet matériel donné à un sujet) et l'instrument (l'objet et les schèmes d'action associés) qui est construit au cours d'un processus, la genèse instrumentale. Il souligne le double rôle de ces instruments dans les situations d'enseignement : permettre la médiatisation du savoir, autrement dit la mise en forme des notions et contenus à transmettre. Ce processus correspond à ce que Rabardel (1995) nomme une instrumentalisation, c'est-à-dire une mise en forme pour servir d'instrument. Les instruments servent aussi d'intermédiaire entre le savoir et l'apprenant. Ce processus appelé de médiation vise à doter l'apprenant d'instruments pour apprendre. La médiation est le rôle premier des enseignants. Elle correspond à un processus d'instrumentation c'est-à-dire la construction et l'adaptation des schèmes d'utilisation des instruments par l'apprenant.

Dans l'enseignement des statistiques, comme dans bien d'autres, nous avons plusieurs types d'outils à notre disposition. Les premiers sont conceptuels et correspondent aux concepts et au cadre théorique sur lequel nous nous fondons ? Viennent ensuite sont les outils techniques utilisés pour produire les analyses (calculatrice, logiciel etc.), mettre en forme les résultats et produire les commentaires. Enfin les exemples que nous utilisons (références à des recherches réelles, jeux de données réels ou fictifs), les exercices que nous produisons, les activités mises en place pour étayer l'apprentissage sont autant d'outils que nous instrumentons à des fins d'apprentissage. De ce point de vue, les ressources pédagogiques libres ou éditoriales constituent donc des instruments pour l'enseignant qui les crée car elles s'insèrent dans une démarche singulière pour lesquels l'enseignant dispose des schèmes d'utilisation. Leur production est inhérente à toute activité de préparation d'un cours. Pour qui souhaitent les réutiliser, ce ne sont que des outils qui, pour être opérationnels doivent être instrumentalisés. C'est sans doute ce qui rend si difficile leur réutilisation notamment dans une démarche d'autoformation ou de réutilisation par d'autres enseignants tels que l'envisagent les promoteurs des REL. Pour y parvenir, au moins deux conditions sont nécessaires :

- 1) Les décontextualiser, c'est l'objet des dépôts de ressources tels que les UNT, et les indexer pour les rattacher à une portion du savoir qui indique quand elles sont utilisables et à quelles conditions.
- 2) Il faut ensuite les recontextualiser dans une situation de classe en les insérant dans une situation didactique.

La décontextualisation est d'autant plus facile que la granularité de la ressource est faible. Cette première condition nous a conduit aussi à chercher un outil d'indexation pour identifier les ressources sans l'aide du contexte de production. Les bibliothèques en utilisent plusieurs (par exemple la classification Dewey, la norme SupLOMFR) cependant aucun ne permet un alignement des ressources avec la structure des connaissances. D'où l'idée d'utiliser une ontologie. Il s'agit de la description d'un domaine de connaissance à l'aide d'un ensemble de concepts hiérarchisés et mis en

relation les uns avec les autres. Cette voie présente plusieurs intérêts que nous développerons plus loin.

La deuxième condition dépend des compétences techniques de l'enseignant et de ses habitudes de travail et éventuellement l'accompagnement techno-pédagogique dont il dispose, notamment le soutien d'ingénieurs pédagogiques. Elle suppose la coordination de plusieurs instruments (conceptuel, technique, didactique). C'est en ce sens que Trouche (2004, 2005) parle d'orchestration instrumentale à propos de la scénarisation pédagogique. Celle-ci se définit par une configuration didactique (les outils, le public visé) et ses modes d'exploitation (objectifs, paramétrage de l'outil, type de tâche). C'est par exemple utiliser des jeux de données pour entraîner des étudiants à reproduire des procédures d'analyse, utiliser des exercices ou des QCM pour favoriser une autoévaluation continue des étudiants. Or ces exercices et ces jeux de données, loin d'être des à côté ou des suppléments, sont constitutifs de la situation d'enseignement. Nous en produisons plusieurs chaque année, parfois en nous inspirant de ce qu'a fait un collègue. Cependant nous disposons de peu d'outils pour les partager, voir en suivre les différentes versions. Nous pouvons bien sûr les mettre sur Internet, mais leur identification et l'évaluation de leur pertinence par les étudiants peuvent parfois s'avérer compliquées, d'où l'idée de développer un dépôt collaboratif permettant de mettre à disposition des collègues les ressources et de travailler collégialement à leur indexation et leur évaluation. La mise à disposition des ressources est une condition nécessaire mais pas suffisante à la réutilisation des ressources. Les freins techniques sont connus (granularité ; interopérabilité, coûts de production, droits d'auteurs ; voir Baumgartner, 2014), les aspects psychologiques le sont un peu moins (Terras, Ramsay, & Boyle, 2013). Ils sont à la fois liés aux compétences informationnelles, aux compétences dans le domaine cible, ici les statistiques, et aux capacités de résolution de problèmes. C'est parce que les technologies du web sémantique peuvent permettre d'apporter, au moins partiellement, des solutions à ces trois séries de problèmes que nous avons choisi d'explorer l'utilisation d'une ontologie pour l'indexation des ressources pédagogiques libres et éditoriales

3. L'ontologie Stato

Dans le cadre d'un projet baptisé Ontostats (<http://www.ontostats.univ-paris8.fr/>), soutenu par le ministère de l'enseignement supérieur, de la recherche et de l'innovation français, nous avons traduit en français l'ontologie des concepts statistiques STATO (<http://stato-ontology.org/>) développée par l'université d'Oxford (Gonzalez, Rocca-Serra, Burke, & Sansone, 2014). Cette ontologie comporte 678 classes et 53 propriétés. Elle appartient à la catégorie des ontologies de domaine car elle est orientée vers la description d'un domaine de connaissances.

L'intérêt d'une telle ontologie est de permettre le partage des connaissances, non seulement sur les concepts, mais aussi sur les relations qu'ils entretiennent, la variété des termes alternatifs, les traductions possibles. L'ontologie est en fait une structure de concepts qu'un vocabulaire permet d'étiqueter. Cela permet une description multilingue du domaine de connaissance, mais surtout des inférences notamment pour ce qui est habituellement implicite dans le discours. Elle concourt ainsi à l'élaboration de systèmes d'aide à la décision (voir par exemple les cas de requêtes sur le site STATO, cf. supra). Les ontologies présentent également de nombreux avantages pour l'indexation des ressources pédagogiques par rapport à l'utilisation d'une norme (Bouzeghoub & Elbyed, 2006). Cependant, c'est surtout pour construire des dispositifs d'aide à l'apprentissage que les ontologies sont prometteuses. Elles fournissent un modèle conceptuel pour l'étude des difficultés d'apprentissage. Une ontologie peut ainsi être utilisée pour concevoir des parcours de formation en adaptant la structure des séquences pédagogiques (Guo & Chen, 2007), pour concevoir des systèmes de recommandation de ressources basés sur les relations entre les concepts (Shen & Shen, 2005), et, lorsque l'ontologie est associée à l'analyse des traces d'apprentissage, à la personnalisation des cours (Castro & Alonso, 2011) et à l'évaluation des apprenants (Romero, North, Gutiérrez, & Caliusco, 2015).

STATO articule trois sous-ensembles de concepts qui décrivent respectivement les buts (objective specification), les entités (information content entity) et les procédures (data transformation). Chacun

de ces ensembles est relié aux autres permettant ainsi de rendre compte de l'organisation générale de l'activité (relation entre les buts), de la représentation de la tâche (buts et entités) et de l'identification de la ou des procédures applicables ainsi que de ses prérequis et résultats. Une telle structuration permet d'espérer son utilisation pour la description des compétences (Meunier, 2019). Pour l'instant, l'ontologie STATO a été utilisée pour indexer un ensemble de ressources éducatives libres, notamment celles développées par les universités numériques thématiques françaises. L'ensemble des notices descriptives de ces ressources sont disponibles dans l'open data du ministère de l'enseignement supérieur de la recherche et de l'innovation (<https://data.enseignementsup-recherche.gouv.fr/>). Les notices de ces ressources ont été rassemblées dans une archive numérique grâce au protocole OAI-MPH.

4. La plateforme Ontostats

La plateforme Ontostats a été développée à l'aide du système de gestion de contenu (Content management system ou CMS) Omeka-S. Elle est accessible à cette adresse : <http://www.ontostats.univ-paris8.fr/omk/s/consulter-les-ressources/page/Accueil>. Ce système est dédié à la création de plateforme de diffusion et d'éditorialisation de documents. Très utilisé dans les bibliothèques pour créer des plateformes de diffusion des collections documentaires, ce CMS a connu récemment une évolution essentielle pour notre projet avec une version orientée vers le web sémantique, une gestion multisites des contenus et multiutilisateurs. Le système gère schématiquement trois types de d'objets : les vocabulaires qui servent à indexer les ressources, les items et les collections d'items qui correspondent aux contenus gérés dans l'archive et les moyens de publication.

Dans Omeka-S les contenus sont les notices décrivant les documents et les médias qui y sont attachés (vidéo, images, textes, sons etc.). Ces contenus sont décrits par des vocabulaires. Dans notre cas, nous avons utilisés les descripteurs Dublin Core, SupLOMFR et les concepts de l'ontologie Stato, exceptées les classes sans label. Deux types de contenus ont été rassemblés dans l'archive Ontostats : (i) les ressources pédagogiques provenant des UNT ou collectées par moi et (ii) les fiches descriptives des concepts de l'ontologie. Le choix de considérer les concepts comme des contenus présente deux intérêts. Le premier est de pouvoir consulter les notices des concepts. Celles-ci contiennent des définitions en anglais et en français. La liste des termes alternatifs pour chaque concept et les relations qu'ils entretiennent avec les autres concepts. En ouvrant la fiche des tests paramétriques, on sait par exemple que ce concept a deux sous-classes, les tests de Mann-Withney et de Kruskal-Wallis.

Le second intérêt est de pouvoir faire des requêtes sur l'ontologie pour savoir par exemple quels tests d'hypothèses permettent d'évaluer l'association entre deux variables catégorielles (voir pour d'autres exemples de requêtes le site de Stato : <http://frog.oerc.ox.ac.uk:8080/stato-app/queryCases.jsp>). Ces deux exemples illustrent bien ce qui fait l'intérêt de la version sémantique d'Omeka : établir des relations entre les contenus. Non seulement les contenus sont indexés par un vocabulaire, mais il est possible d'explorer le vocabulaire et sa structure et de relier les contenus entre eux. Ouvrir la fiche concept de la moyenne permet également d'avoir la liste des ressources pédagogiques traitant de la moyenne. Ces contenus organisés en collections. Un contenu peut appartenir plusieurs collections. Il est donc facile de réaliser sa propre sélection ? Le système permet surtout de limiter le parcours de l'utilisateur (les étudiants) à cette collection afin que sa recherche ne soit pas « polluée » par des ressources jugées non pertinentes par l'enseignant. Enfin le système facilite le partage des contenus avec d'autres collègues, la création et le suivi des versions d'une même ressource grâce aux relations entre les contenus. La mise à jour des contenus peut être faite par lot grâce à des imports via des CSV notamment.

Omeka-S est également orientée vers une gestion multiutilisateurs avec une gestion des différents rôles et surtout une gestion multisites permettant à partir d'une seule plateforme et une seule archive de créer autant de sites de diffusion qu'on souhaite. Il est alors possible de créer son propre site avec le CMS en puisant dans l'archive de ressources puis de le publier sur Internet. Le site est compatible avec tous types de terminaux : smartphones, tablette ou ordinateurs (voir par exemple :

<http://www.ontostats.univ-paris8.fr/omk/s/logicielsStats/page/Introduction>). Nous travaillons actuellement à un export des sites vers un livre électronique (format epub) pour permettre une consultation hors connexions, au moins de certains contenus.

5. Conclusion et perspectives

Le projet a démarré il y a deux ans. C'est donc un projet très jeune. Nous avons bien avancé sur les aspects informatiques. La plateforme est opérationnelle, même si de nombreux aspects doivent encore être travaillés. Le nombre de ressources indexées est pour l'instant modeste et un travail d'évaluation est à mener. La plateforme offre la possibilité d'annoter les ressources, mais il faut pouvoir aller plus loin, notamment dans la perspective de développement d'un moteur de recommandation. L'ontologie est sans doute discutable et doit être améliorée. Les définitions qu'elle contient, les relations entre les concepts, les concepts manquants ou les concepts inutiles sont autant de sujets qu'ils seraient intéressants de travailler collectivement. L'indexation doit également être approfondie. Celle que nous avons fait repose pour le moment sur la simple recherche de mots-clés dans les descriptifs des ressources. Enfin la livraison récente de la plateforme ne nous a pas encore permis de tester son usage dans un contexte de classe. Toutes ces pistes de travail nécessitent la constitution d'une communauté autour du projet dont la faiblesse constitue pour l'instant le principal point faible du projet. Les perspectives sont pourtant réelles.

Nous avons évoqué en introduction les différentes approches de l'enseignement des statistiques. Offrir aux étudiants cette pluralité d'approches est un enjeu important face à la réduction des volumes horaires consacrés aux enseignements de méthodologie, mais aussi face à la massification et à la diversification de nos publics. Le développement des logiciels de statistiques rend cette approche d'autant plus nécessaire qu'il fait perdre le contrôle conceptuel permis autrefois par le calcul à la main. La même réflexion s'applique aux aspects épistémologiques des statistiques trop souvent laissés de côté. Faute de pouvoir tout produire et d'avoir le temps de tout enseigner, partager les ressources pédagogiques est l'unique voie pour offrir cette pluralité de points de vue, voire d'aligner plusieurs versions d'un même cours, entendu ici comme un ensemble de concepts et de procédures.

L'évaluation est une autre piste d'utilisation de l'ontologie. Celle-ci permettant de faire des inférences, il est possible de décliner des exercices permettant de travailler sur celles-ci notamment celles qui sont implicites ou de construire des questions à choix multiples en puisant les alternatives à la bonne réponse parmi des distracteurs probables parce que liés sémantiquement. Il est assez facile de faire le lien avec la possibilité de décrire les compétences à l'aide de l'ontologie dont nous avons parlé précédemment. Notre conception de la notion de compétence, est une conception dynamique (Buléa Bronckart & Bronckart, 2005) qui situe la capacité d'agir dans l'articulation entre différents types de savoirs. Ceux-ci permettent de se construire une représentation de la situation, de comprendre les buts d'une situation d'analyse et les procédures applicables, mais surtout de trouver des schèmes permettant d'utiliser les outils dont le sujet dispose. Cette perspective laisse entrevoir la possibilité d'instrumentaliser l'ontologie pour élaborer des parcours pédagogiques.

Références

- Armatte, M. (2010). Le rôle de l'histoire dans l'enseignement de la statistique. *Statistique et Enseignement*, 1(2), 23-47.
- Atkins, D. E., Brown, J. S., & Hammond, A. L. (2007). A Review of the Open Educational Resources (OER) Movement: Achievements, Challenges, and New Opportunities, 84.
- Baumgartner, P. (2014). Improving Reusability of OER. In D. Hutchison, T. Kanade, B. Steffen, D. Terzopoulos, D. Tygar, G. Weikum, ... C. P. Rangan (Éd.), *2nd Information and Communication Technology - EurAsia Conference (ICT-EurAsia)* (Vol. LNCS-8407, p. 277-285). Bali, Indonesia: Springer. https://doi.org/10.1007/978-3-642-55032-4_27

- Bouzeghoub, A., & Elbyed, A. (2006). *Ontology Mapping for Learning Objects Repositories Interoperability*. In *Lecture notes in computer science* (p. 794-797). SpringerSpringer, Berlin, New York.
- Buléa Bronckart, E., & Bronckart, J.-P. (2005). Pour une re-définition de la compétence comme processus dynamique. In *Cahiers de la section des sciences de l'éducation* (Vol. 104, p. 187-132). Publication de la section des sciences de l'éducation. Consulté à l'adresse <https://archive-ouverte.unige.ch/unige:81539/ATTACHMENT01>
- Castro, F., & Alonso, M. A. (2011). *Learning Objects and Ontologies to Perform Educational Data Mining* (p. 532-536). CSREA Press.
- Gonzalez, A., Rocca-Serra, P., Burke, O., & Sansone, S.-A. (2014). *Stato - Statistics Ontology*. Consulté 13 février 2018, à l'adresse <http://stato-ontology.org/>
- Guo, W.-Y., & Chen, D.-R. (2007). An Ontology Infrastructure for an E-Learning Scenario. *International Journal of Distance Education Technologies*, 5(1), 70-78.
- Hahn, C., & Stoltz, G. (2014). *Savoir académique, savoirs pratiques : tensions et recherche d'équilibre* — Le cours de statistique à HEC Paris et à ESCP Europe. *Statistique et Enseignement*, 4(2), 19-52.
- Meunier, J.-M. (2019). On-demand training in a professional environment from a repository of open educational resources. Présenté à ICDE Lifelong Learning summit : Lifelong learning for lifelong employability, Lillehammer, Norway. Consulté à l'adresse <https://app.oxfordabstracts.com/events/552/submissions/62537/abstract-book-view>
- Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, 88(1), 144-161. <https://doi.org/10.1037/0022-0663.88.1.144>
- Quilici, J. L., & Mayer, R. E. (2002). Teaching students to recognize structural similarities between statistics word problems. *Applied Cognitive Psychology*, 16(3), 325-341. <https://doi.org/10.1002/acp.796>
- Rabardel, P. (1995). *Les hommes et les technologies; approche cognitive des instruments contemporains*. Armand Colin. Consulté à l'adresse <https://hal.archives-ouvertes.fr/hal-01017462>
- Rézeau, J. (2002). Médiation, médiatisation et instruments d'enseignement : du triangle au « carré pédagogique ». *ASp*, (35-36), 183-200. <https://doi.org/10.4000/asp.1656>
- Richard, J. F. (1983). *Logique du fonctionnement et logique de l'utilisation*. Consulté à l'adresse <http://hal.inria.fr/inria-00076356>
- Romero, L., North, M., Gutiérrez, M., & Caliusco, L. (2015). Pedagogically-Driven Ontology Network for Conceptualizing the e-Learning Assessment Domain. *Journal of Educational Technology & Society*, 18(4), 312-330.
- Sander, E., Meunier, J. M., & Bosc-Miné, C. (2004). Approche ontologique et navigation dans un E.I.A.H: Le cas de l'enseignement des statistiques. *Revue STICEF*, 11. Consulté à l'adresse http://sticef.univ-lemans.fr/num/vol2004/sander-08/sticef_2004_sander_08.htm
- Shen, L.-P., & Shen, R.-M. (2005). Ontology-based learning content recommendation (English). *International journal of continuing engineering education and life-long learning*, 15(3-6), 308-317.
- Terras, M. M., Ramsay, J., & Boyle, E. (2013). Learning and Open Educational Resources: A Psychological Perspective. *E-Learning and Digital Media*, 10(2), 161-173.
- Trouche, L. (2004). Environnements Informatisés et Mathématiques: quels usages pour quels apprentissages? *Educational Studies in Mathematics*, 55(1-3), 181-197. <https://doi.org/10.1023/B:EDUC.0000017674.82796.62>
- Trouche, L. (2005). Construction et Conduite des Instruments dans les Apprentissages Mathématiques: Nécessité des Orchestrations. *Recherches en Didactique des Mathématiques*, 25(1), 91-138.