Donner de l'R aux SHS

Rzine est un projet du Collège international des sciences territoriales (CIST, Fédération de recherche) qui réunit une vingtaine d'unités et plusieurs disciplines autour des sciences territoriales. Son objectif est de contribuer à l'apprentissage et à la pratique de R en SHS, en favorisant le partage, la production et la diffusion de ressources. Ce projet collaboratif et interdisciplinaire s'étend au-delà des équipes partenaires du CIST et tente de répondre à un besoin commun à de nombreuses unités de recherche de l'InSHS.

R, un outil polyvalent et interdisciplinaire au service de la science ouverte et reproductible

C'est en 1992 que Robert Gentleman et Ross Ihaka de l'université d'Auckland se lancent dans le projet de développement du langage de programmation pour l'analyse de données R. Ils cherchent alors à développer un environnement pour le calcul scientifique plus performant que l'existant. Trente ans plus tard, R s'est imposé comme une référence en matière d'analyse de données. Sa communauté de développeurs / développeuses et d'utilisateurs / utilisatrices connaît une croissance impressionnante depuis plusieurs années. D'abord adopté par les sciences dures, R s'est peu à peu imposé comme une référence en sciences humaines et sociales à partir des années 2010. Aujourd'hui, la pratique de ce langage est enseignée à l'université dans de nombreuses disciplines (sociologie, économie, géographie, démographie, archéologie…) et son utilisation dépasse le seul domaine de la recherche.

R est gratuit, *open source* et fonctionne sur les principaux systèmes d'exploitation. À la différence d'un logiciel « clic-bouton », chaque opération est facilement archivable, partageable et reproductible. Cet outil permet ainsi de s'inscrire pleinement dans le courant de la science ouverte et de la recherche reproductible, soutenu par le CNRS. Grâce à sa structure modulaire, le potentiel de développement de ce langage est immense car chaque utilisateur / utilisatrice peut mettre à disposition de toutes et tous des extensions supplémentaires appelées *packages* (Figure 1). Ainsi, l'ensemble des méthodes d'analyse et de représentation des données utilisées par les différentes disciplines de SHS sont exécutables en langage R.

L'inscription du langage R comme l'un des projets du mouvement du logiciel libre, ainsi que le soutien financier d'acteurs majeurs du secteur privé assurent son développement, sa fiabilité et sa popularité à long terme. C'est dans ce contexte qu'un groupe d'ingénieur(e)s et de chercheurs et chercheuses a initié le projet Rzine pour contribuer au développement de la pratique de ce langage pour l'analyse de données, fédératrice pour les disciplines de sciences humaines et sociales.

Rzine.fr : favoriser la diffusion et le partage de connaissances sur la pratique de R en SHS

L'étude de la pratique de R et de sa communauté d'utilisateurs et utilisatrices dans le monde francophone a permis de dresser un bilan positif en matière de ressources disponibles et de pluridisciplinarité. Si la communauté R francophone est indéniablement active et productive, plusieurs axes d'amélioration ont pu être recensés :

- l'accessibilité aux ressources n'est pas toujours aisée pour un débutant ;
- la diffusion de ressources se limite parfois à un groupe d'utilisateurs, une structure ou une discipline ;
- deux grands types de ressources s'opposent : la documentation généraliste, orientée vers la technique et la documentation spécialisée dans un domaine, peu accessible et transposable pour les autres.

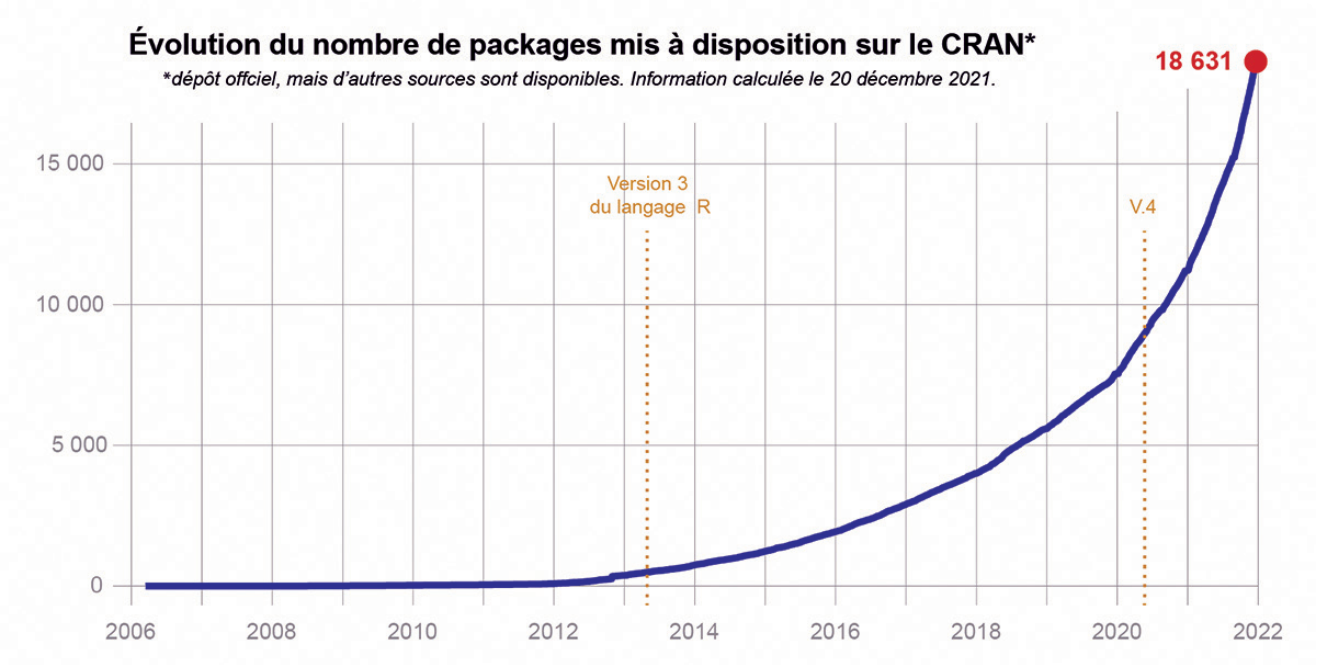

Figure 1 : L'évolution du nombre de *packages* (extensions) mis à disposition au cours du temps traduit la popularité croissante du langage R

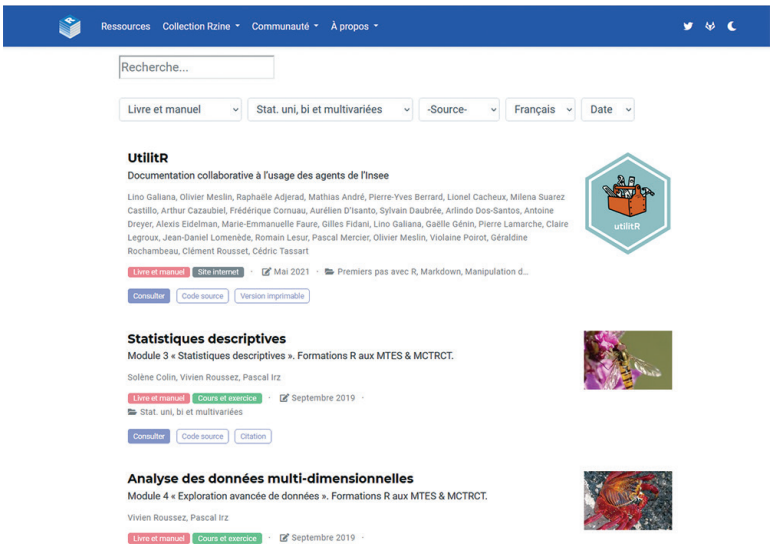

Figure 2 : Plus de 300 ressources sont référencées et accessibles sur rzine.fr

Figure 3 : Impression d'écran de la première publication Rzine (Lambert, 2021), qui porte sur la réalisation d'une carte publiée plusieurs fois, et maintenant reproductible

Le site collaboratif Rzine.fr a été développé à partir de ce constat. Il référence quatre grands types de contenus :

- des ressources en libre accès. Plus de 300 sont référencées (Figure 2),
- des projets qui favorisent la pratique de R,
- des auteurs de documentation ou de formations,
- des actualités pour la communauté d'utilisateurs et utilisatrices.

Rzine.fr s'efforce de dessiner les contours de la communauté R en SHS. Il donne aux débutants un accès simple à l'information et orienté vers une utilisation autonome. Il offre aux utilisateurs et utilisatrices de niveau intermédiaire la possibilité d'étendre leurs pratiques et de s'ouvrir à d'autres méthodes parfois issues d'autres disciplines. Enfin, il s'agit d'un espace de diffusion pour les utilisateurs avancés, auteurs de développements ou de documentation.

Ce site repose sur des technologies qui facilitent son administration décentralisée et collaborative. À terme, une plus grande ouverture aux contributions externes est envisagée. L'un des futurs chantiers du projet sera la simplification et l'automatisation de la soumission de contenus.

La collection Rzine : valoriser et transmettre ses méthodes de travail avec R

Le projet Rzine a également pour objectif de contribuer à la science ouverte et à la pratique de la recherche reproductible. Pour cela, une collection de publications *open source*, ouverte à toutes et tous et soumise à une évaluation transparente par les pairs vient d'être créée. L'objectif est d'offrir un espace de publication permettant le partage d'une méthode ou d'un type d'analyse reproductible, dans un périmètre pluridisciplinaire. Il ne s'agit pas d'articles scientifiques mais de supports méthodologiques ancrés dans une thématique, qui présentent du code de programmation de manière intelligible pour les différentes disciplines de SHS. Chaque publication fait l'objet d'une relecture thématique et technique, qui assure sa pérennité et son intérêt pour les différentes disciplines. Les publications acceptées sont éditées en format HTML (Figure 3), associées à un DOI et référencées sur HAL. Leur stockage et diffusion sont assurés par le site rzine.fr.

Des publications au format *notebook*

Le choix du notebook comme format de publication de ces articles méthodologiques est inhérent au contenu qu'ils présentent. Parfois nommés « calepins électroniques » en français, les *notebooks* sont très utilisés en science des données, que ce soit pour la production de documentation, l'enseignement ou le travail collaboratif. Ils permettent de combiner des sections en langage naturel et des sections en langage informatique, et peuvent s'enregistrer sous différents formats (Figure 4). Le *notebook* permet ainsi d'appliquer le paradigme de la programmation lettrée préconisé par Donald Knuth dans les années 1970, qui est au cœur du concept de la collection Rzine.

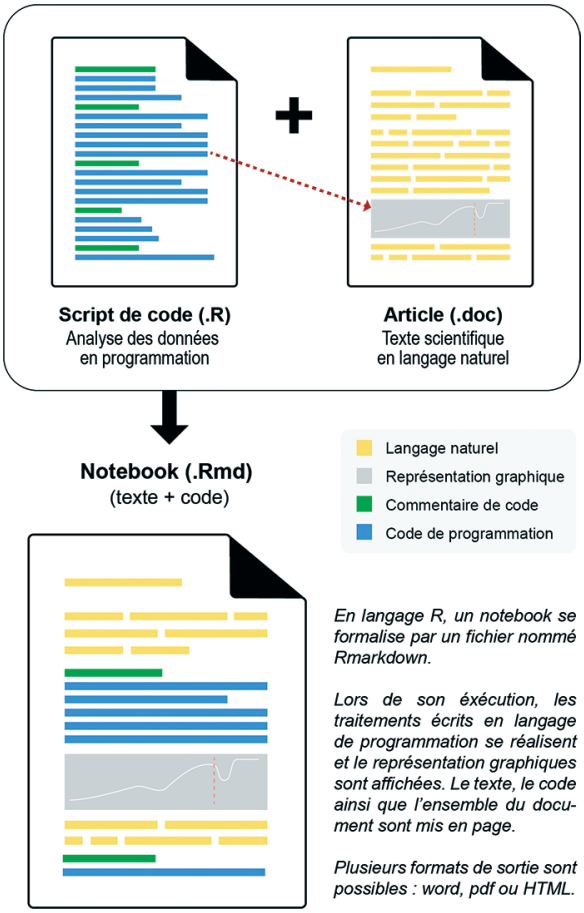

Figure 4 : En permettant la combinaison de texte, de code et de ses résultats, le *notebook* permet de faire de la programmation lettrée

À l'image des travaux et de la coopération entre la TGIR Huma-Num et le Centre de recherche interuniversitaire sur les humanités numériques - CRIHN (université de Montréal), les humanités numériques s'intéressent particulièrement à l'écriture numérique et les initiatives d'innovation en matière de format de publication apparaissent peu à peu. La revue *Programming Historian*, lancée lors de la *Digital Humanities conference* de 2012, fait partie des succès qui placent le *notebook* parmi les formats de publication de référence. Permettant l'intégration de données et de leur traitement à l'écriture scientifique (Figure 5), ce format semble idéal pour répondre aux objectifs de la collection Rzine.

Bien que la pratique de la programmation lettrée avec R soit accessible pour un novice, l'utilisation d'un logiciel de gestion de versions décentralisé (Git) sur lequel repose l'ensemble du processus d'évaluation et de publication de la collection Rzine demande des connaissances plus approfondies. C'est pourquoi Rzine contribuera, par des formations et l'encadrement des auteurs de publication, à la montée en compétence collective et participative autour de la programmation lettrée et des systèmes de versionnage. Au-delà de sa vocation à offrir un espace de publication innovant, ce projet participe, à son échelle, au développement des humanités numériques en sciences humaines et sociales.

L'animation et l'évolution de Rzine repose aujourd'hui sur l'implication d'un groupe de personnes issues de sept unités de recherche différentes. Son avenir dépend de la communauté d'utilisateurs et d'utilisatrices interdisciplinaire qu'il réussira à rassembler et à impliquer dans cette aventure.

Comité éditorial Rzine

- Laurent Beauguitte, CNRS, Géographie-cités (UMR8504)
- Marion Gentilhomme, Université Paris 1 Panthéon-Sorbonne, Collège international des sciences territoriales (CIST, FR2007)
- Timothée Giraud, CNRS, Réseau interdisciplinaire pour l'aménagement du territoire européen (RIATE, UAR2414)
- Violaine Jurie, Université de Paris, Géomatique, Télédétection, Cartographie (Géotéca)
- Grégoire Le Campion, CNRS, Passages (UMR5319)
- Marion Le Texier, Université Paul Valéry - Montpellier 3, Laboratoire de Géographie et d'Aménagement de Montpellier (LAGAM)
- Hugues Pecout, CNRS, Collège international des sciences territoriales (CIST, FR2007)
- Sébastien Rey-Coyrehourcq, Université de Rouen Normandie, Identité et différenciation de l'espace, de l'environnement et des sociétés (IDEES, UMR6266)
- Ronan Ysebaert, Université de Paris, Réseau interdisciplinaire pour l'aménagement du territoire européen (RIATE, UAR2414)

contact&info

- Hugues Pecout
CIST
contact@rzine.fr
- Pour en savoir plus
https://rzine.fr

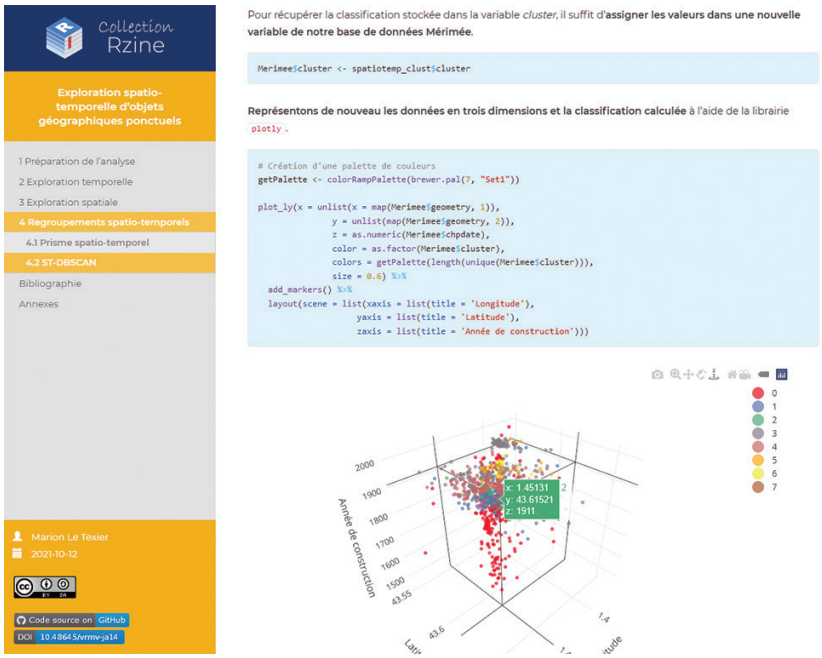

Figure 5 : Impression d'écran d'un *notebook* Rzine (Le Texier, 2021). Le texte, le code et son résultat se succèdent, dans une mise en page soignée facilitant la lecture