

# Du terrain au tableau de bord : l'écosystème {shinyODK}

David Carayon\*

## Résumé (max 300 mots)

La collecte de données sur le terrain constitue une étape critique dans de nombreuses disciplines scientifiques : écologie, agronomie, sciences de l'eau, épidémiologie, où les conditions d'acquisition imposent souvent de travailler hors connexion, sur des supports mobiles, dans des contextes variés et parfois contraints.

Open Data Kit (ODK) s'est imposé comme une solution de référence pour répondre à ces besoins : formulaires XLSForm flexibles, collecte offline sur smartphone / tablette via ODK Collect, et centralisation des données sur un serveur ODK Central. Cependant, une fois les données collectées, leur exploitation (visualisation, contrôle qualité, reporting) nécessite des développements complémentaires rarement intégrés au flux de travail. Nous présentons shinyODK, un écosystème articulant ODK et RShiny autour d'une chaîne de traitement reproductible et automatisée.

L'architecture repose sur un socle central composé de PostgreSQL (stockage relationnel pérenne), du package R {ruODK} (Mayer 2021) (extraction programmatique des soumissions depuis ODK Central) et de mécanismes d'orchestration (cron jobs, pipelines CI/CD) assurant la synchronisation régulière des données. En aval, des applications Shiny (Chang et al. 2025) exploitent ces données pour offrir des tableaux de bord interactifs, des outils de contrôle qualité en temps quasi-réel et la génération automatisée de rapports Quarto (Allaire et al. 2022).

Cette approche modulaire présente plusieurs avantages : découplage entre collecte et analyse, permettant à chaque composant d'évoluer indépendamment ; traçabilité complète du flux de données ; reproductibilité du pipeline grâce à l'outillage R et CI/CD ; et accessibilité des résultats pour des utilisateurs non-programmeurs via les interfaces Shiny.

L'écosystème a été déployé avec succès dans plusieurs projets de recherche à INRAE. shinyODK illustre comment la combinaison de briques open source matures (ODK, PostgreSQL, R/Shiny) permet de construire des chaînes de données complètes, de la saisie terrain jusqu'au livrable analytique, en restant intégralement dans un cadre de science ouverte.

**Mots-clefs (3 à 5) :** ODK - RShiny - PostgreSQL

## Développement

La collecte de données sur le terrain pose un défi récurrent dans les sciences environnementales : les agents opèrent souvent en zones non couvertes par un réseau, sur des protocoles variés, et les données collectées doivent ensuite être centralisées, contrôlées et restituées à des publics hétérogènes. Si ODK (Hartung et al. 2010) répond efficacement à la problématique de collecte (formulaires XLSForm versionnés, saisie offline sur smartphone, synchronisation sur ODK Central) le chaînon entre la donnée brute et son exploitation analytique reste généralement artisanal, reposant sur des exports manuels et des traitements ad hoc peu reproductibles.

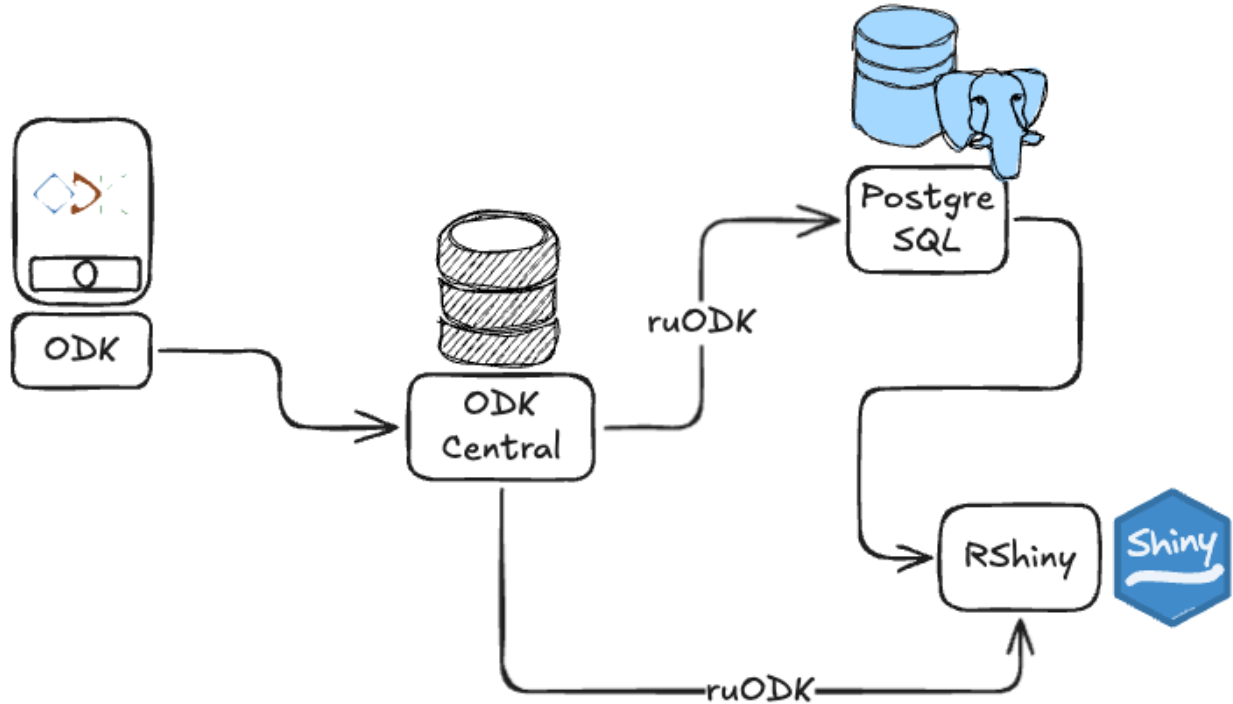
L'écosystème shinyODK propose une architecture en trois couches pour combler ce manque :

- Collecte (ODK) : les formulaires sont conçus en XLSForm et déployés sur un serveur ODK Central. Les agents de terrain saisissent les données via ODK Collect, y compris en mode totalement déconnecté. La synchronisation s'effectue automatiquement dès le retour en zone connectée, avec horodatage et traçabilité des soumissions.

---

\*INRAE UR EABX, david.carayon@inrae.fr

- Middleware (PostgreSQL, ruODK, cron/CI-CD): les soumissions sont extraites programmiquement d'ODK Central via le package R {ruODK} (Mayer 2021), puis injectées dans une base PostgreSQL qui assure le stockage relationnel pérenne et le contrôle d'intégrité. Des tâches cron ou des pipelines CI/CD (GitLab CI) orchestrent cette synchronisation à intervalle régulier, garantissant la fraîcheur des données sans intervention humaine. Cette couche gère également les transformations nécessaires : nettoyage, jointures, calcul d'indicateurs dérivés.
- Restitution (RShiny, Quarto): des applications Shiny (Chang et al. 2025) se connectent directement à PostgreSQL pour proposer des tableaux de bord interactifs, des interfaces de contrôle qualité permettant de détecter anomalies et valeurs manquantes en temps quasi-réel, et des modules de génération automatique de rapports Quarto (Allaire et al. 2022) à destination des commanditaires ou partenaires.



Le découplage entre ces trois couches est un choix architectural central : chaque composant peut évoluer, être remplacé ou mis à l'échelle indépendamment. Un formulaire ODK peut être modifié sans impacter les dashboards ; une nouvelle application Shiny peut être ajoutée sans toucher au pipeline d'ingestion. L'ensemble du flux est versionné (Git), testé (CI/CD) et documenté, garantissant la reproductibilité. L'écosystème a été déployé sur plusieurs projets à INRAE, confirmant sa robustesse en conditions réelles de production.

## Références

- Allaire, J. J., Charles Teague, Yihui Xie, and Christophe Dervieux. 2022. "Quarto." Zenodo. <https://doi.org/10.5281/zenodo.5960048>.
- Chang, Winston, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert, and Barbara Borges. 2025. *Shiny: Web Application Framework for r*. <https://CRAN.R-project.org/package=shiny>.
- Hartung, Carl, Adam Lerer, Yaw Anokwa, Clint Tseng, Waylon Brunette, and Gaetano Borriello. 2010. "Open Data Kit: Tools to Build Information Services for Developing Regions." In *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*. ICTD '10. ACM. <https://doi.org/10.1145/2369220.2369236>.
- Mayer, Florian Wendelin. 2021. "ruODK: An r Client for the ODK Central API." Zenodo. <https://doi.org/10.5281/zenodo.5559164>.