



HAL
open science

IDEATools: Un environnement informatique pour l'automatisation et le reporting de données IDEA4

David Carayon

► **To cite this version:**

David Carayon. IDEATools: Un environnement informatique pour l'automatisation et le reporting de données IDEA4. [Rapport Technique] INRAE. 2021, pp.34. hal-02983957v2

HAL Id: hal-02983957

<https://hal.inrae.fr/hal-02983957v2>

Submitted on 2 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

IDEATools : Un environnement informatique pour l'automatisation et le reporting de données IDEA4

28 janvier 2021 – Version 2.0

INRAE

Institut national de recherche pour
l'agriculture, l'alimentation et l'environnement



David Carayon | Unité de recherche ETBX

Document distribué sous licence CC-BY



<https://creativecommons.org/licenses/by/4.0/fr/legalcode>

Sommaire

Préambule	i
1 Introduction	1
2 Outils informatiques liés à la méthode	3
2.1 Outils pour l'approche par les dimensions	3
2.2 Outils pour l'approche par les propriétés	5
2.3 Réflexion sur ces premiers outils	7
3 Introduction du logiciel R dans l'écosystème IDEA	9
3.1 Transition vers R : Lecture des calculateurs	9
3.2 Production de graphiques élaborés	12
3.3 Développement des analyses de groupe	13
3.4 Standardisation et compilation des résultats	13
4 Le package R IDEATools	17
4.1 Présentation d'IDEATools	17
4.2 Stratégie de diffusion et d'intégration continue	18
4.3 Mise à disposition d'IDEATools aux utilisateurs	19
5 Perspectives de développement	23
5.1 Les analyses de groupe	23
5.2 Continuer l'informatisation de la méthode	23
Bibliographie	27

Préambule

La méthode IDEA est une méthode d'évaluation de la durabilité des exploitations agricoles. Dans sa quatrième version, les concepteurs de la méthode ont introduit de nouveaux concepts nécessitant le développement de nouveaux outils informatiques afin de rendre la méthode opérationnelle et utilisable par le plus grand nombre. Le présent rapport présente les travaux réalisés au cours de l'année 2020 sur ces outils et en particulier les développements en langage R qui ont mené à la création d'un package et d'une interface web interactive. Enfin, nous discutons des perspectives de développements informatiques qui restent à explorer.

Ce travail est financé par le projet de recherche CASDAR ACTION.

Chapitre 1

Introduction

La méthode IDEA, pour **I**ndicateurs de **D**urabilité des **E**xploitations **A**gricoles, est une méthode permettant d'évaluer la durabilité globale d'une exploitation agricole. Elle est conçue comme un outil transparent et pédagogique pour enseigner la durabilité de l'agriculture et son évaluation, mais aussi pour accompagner les démarches de progrès des agriculteurs dans la diversité des voies possibles de la durabilité.

L'évaluation s'appuie sur 53 indicateurs, mobilisant des données quantitatives et qualitatives collectées et calculées à l'échelle de l'exploitation. La méthode analyse et qualifie les pratiques agricoles et activités de l'exploitation au regard de ses impacts sur l'environnement, de sa contribution au développement socio-territorial et de sa performance économique globale.

L'agriculture durable est une agriculture écologiquement saine, économiquement viable, socialement juste et humaine. Elle se caractérise par sa multifonctionnalité, qui contribue à la fourniture de services environnementaux globaux et à son ancrage territorial. Une exploitation agricole durable est définie comme une exploitation viable, vivable, transmissible et reproductible ([Landais, 1998](#)), qui inscrit son développement dans une démarche sociétale responsable.

Ainsi, les 53 indicateurs d'IDEA sont agrégés selon deux grilles de lecture :

- par les 3 dimensions d'une agriculture durable
- par les 5 propriétés d'une exploitation agricole durable

Pour plus d'informations sur le cadre conceptuel et théorique de la méthode, vous pouvez consulter sa publication de référence ([Zahm et al., 2019](#)).

Chapitre 2

Outils informatiques liés à la méthode

2.1 Outils pour l'approche par les dimensions

Tout comme pour les précédentes versions d'IDEA, l'approche dite par les dimensions a été rendue opérationnelle au travers d'un calculateur Excel (Figure 2.1). Ce fichier au format .xls présente l'avantage de rassembler au sein d'un même espace :

- Un formulaire imprimable dédié à la collecte de données
- Un formulaire de saisie informatique
- Différents onglets de résultats connectés avec le formulaire de saisie

Ce calculateur permet donc de saisir les données collectées puis d'obtenir les scores calculés pour les 53 indicateurs, 13 composantes et 3 dimensions. Certains graphiques simples (diagrammes en barres) sont ensuite proposés.

Dans le cadre du CASDAR ACTION, le comité scientifique a émis la volonté de construire une base de données compilant les résultats de diagnostics IDEA tout en proposant un certain nombre d'analyses automatisées. Ce chantier a débouché sur la création d'une plateforme web <https://idea4.agro-bordeaux.fr/> adossé à une base de données PostgreSQL, l'ensemble étant développé et maintenu par Bordeaux Sciences Agro (BSA).

Afin de rendre possible la communication entre calculateur Excel et application web, un onglet spécifique intitulé "renvoi BDD" a été ajouté dans le calculateur (Figure 2.2). Celui-ci va récupérer les données relatives aux items et certaines métadonnées, puis proposer un export au format .json au travers d'une macro VBA.

IDEATools : Un environnement informatique pour l'automatisation et le reporting de données IDEA4



FIGURE 2.1 – Capture d'écran du calculateur Excel développé pour la saisie des données et la restitution des résultats de l'approche par les dimensions.

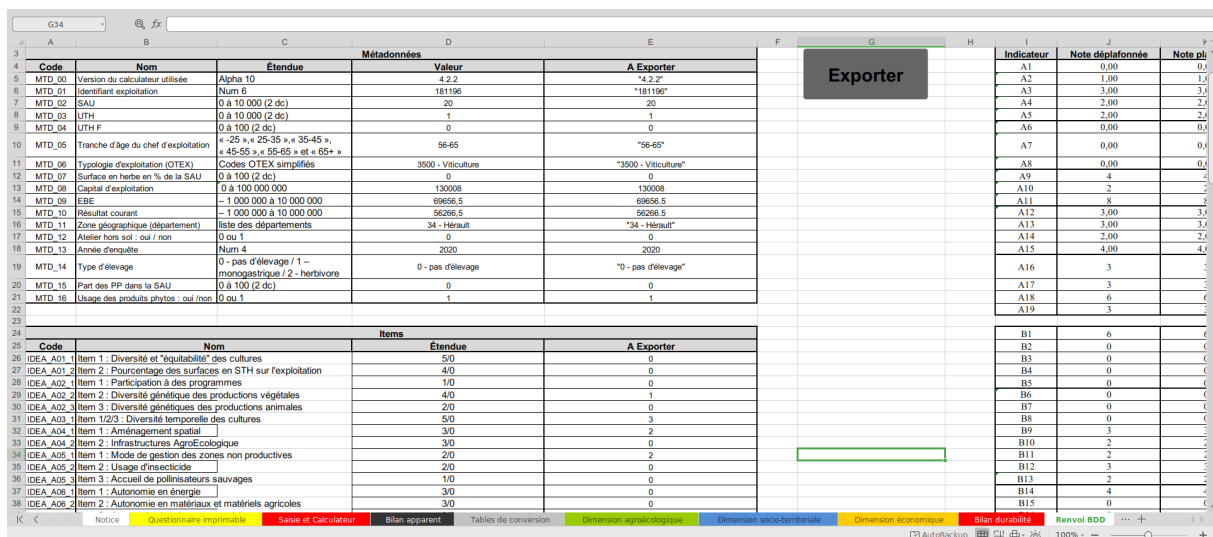


FIGURE 2.2 – Capture d'écran de l'onglet "renvoi BDD" visant à faciliter l'export de données vers d'autres outils informatiques.

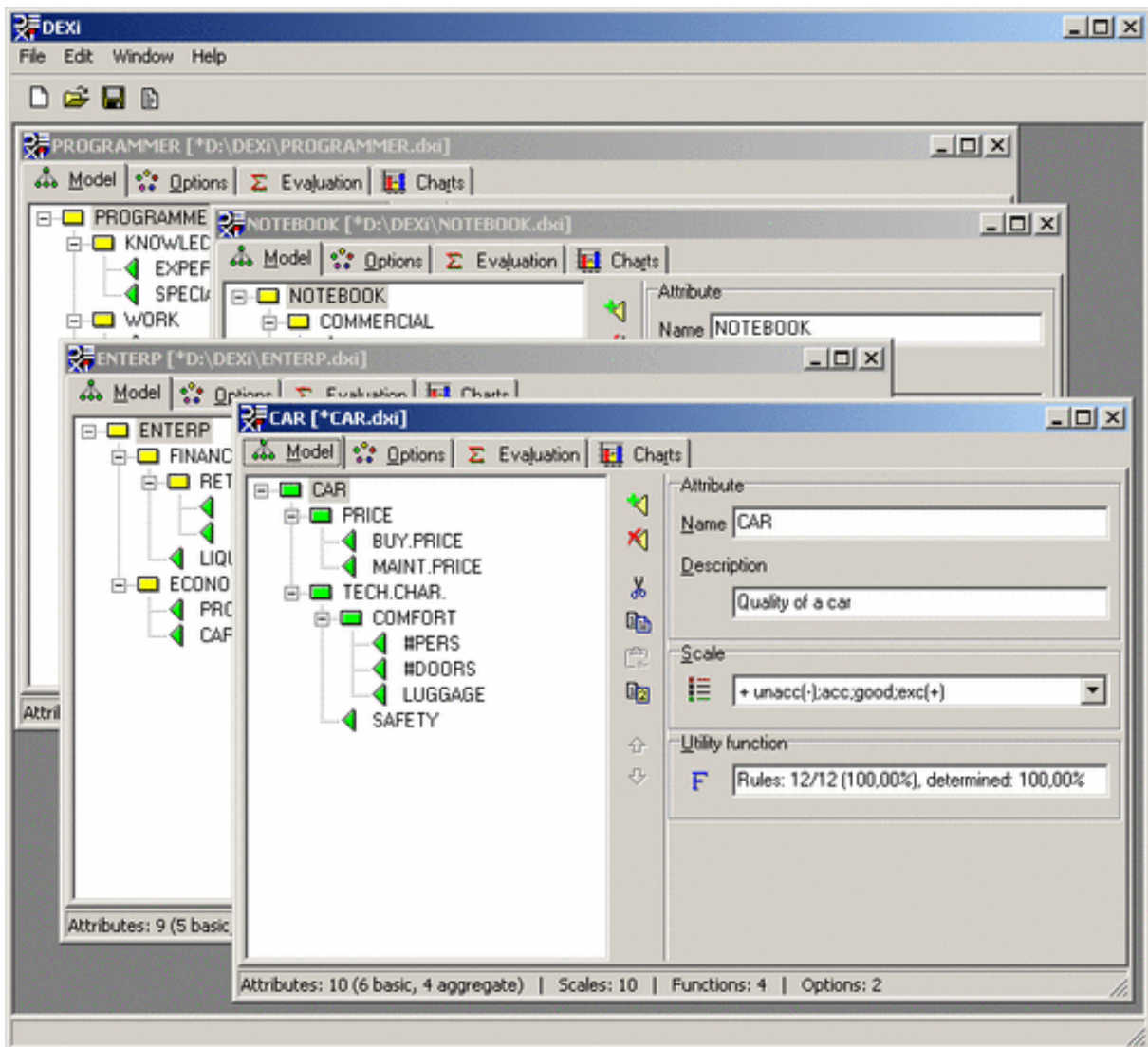


FIGURE 2.4 – Capture d'écran de l'interface du logiciel DEXi utilisé pour la construction des tables de contingence.

L'opérationnalisation de cette agrégation a été rendue possible par le logiciel DEXi (Bohanec, 2008) dans lequel les auteurs de la méthode IDEA ont, pour chaque nœud d'agrégation, défini une "table de contingence" répertoriant les résultats de l'agrégation (allant de "Très défavorable" à "Très favorable") pour chaque combinaison possible en amont de l'agrégation (Figure 2.4). A partir des agrégations obtenues, les résultats doivent ensuite être présentés sous forme d'arbres "éclairés", c'est-à-dire une représentation de la carte heuristique avec chaque nœud coloré selon son évaluation. Cette étape a été opérationnalisée via un second calculateur Excel, sur lequel il fallait copier les résultats issus de DEXi puis colorer manuellement les cases (Figure 2.5).

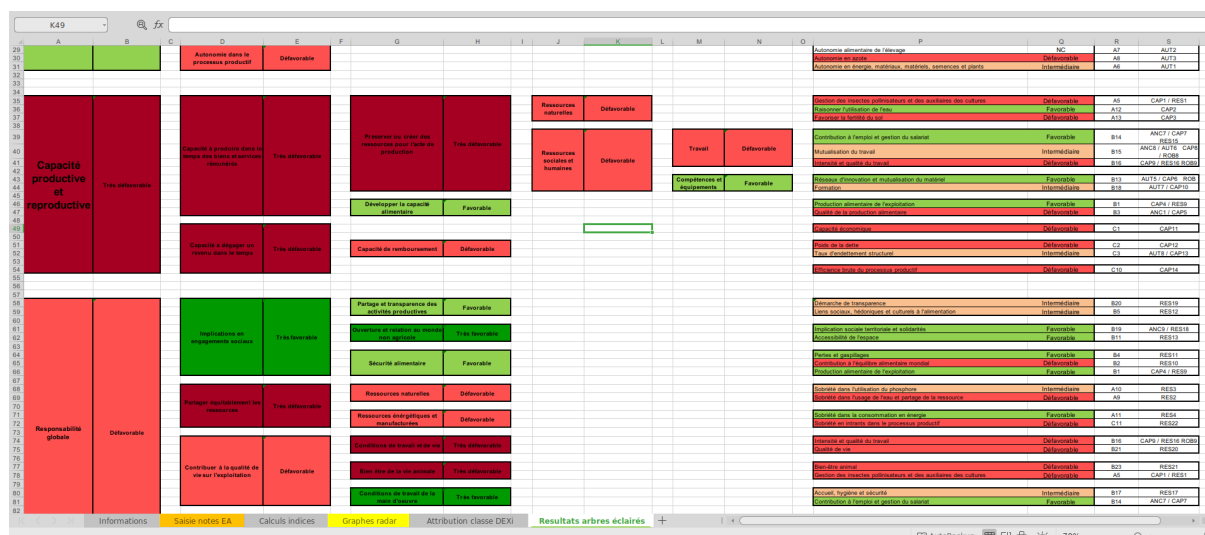


FIGURE 2.5 – Capture d’écran de l’ancien applicatif Excel utilisé pour illustrer les résultats de l’approche par les propriétés.

2.3 Réflexion sur ces premiers outils

La méthodologie présentée aux utilisateurs pour l’approche par les propriétés jusqu’en septembre 2019 était donc la suivante :

1. Réaliser l’enquête
2. Saisir les données dans le calculateur : obtention des résultats “dimensions”
3. Copier les résultats d’un onglet spécifique vers le logiciel DEXi
4. Manipuler DEXi pour déclencher l’évaluation
5. Re-copier manuellement les résultats vers un deuxième calculateur
6. Colorer manuellement les cases

L’opérationnalisation de cette méthodologie, bien que fonctionnelle, s’est rapidement révélée très délicate pour son usage par le plus grand nombre (enseignants, élèves, conseillers agricoles). Elle demanderait, pour l’analyse de plusieurs diagnostics, plusieurs heures d’actions répétitives et d’allers et retours entre deux logiciels (avec à chaque fois un risque d’erreur humaine/technologique) pour le diagnostic d’une seule exploitation.

Concernant l’approche par les dimensions, bien que le calculateur développé soit pratique et fonctionnel, le logiciel Excel reste limité sur certains aspects :

- Graphiques plus élaborés
- Analyses spécifiques (ex : Analyse de groupe)
- Traitement de 20, 50, 100 exploitations en simultané
- Reporting automatisé

De ces différents constats est née la volonté de créer un outil automatisé dédié à la méthode IDEA permettant de limiter le risque d'erreur, de diminuer la pénibilité du diagnostic et d'augmenter les possibilités de rendus proposés.

La solution technologique proposée pour relever les défis techniques apportés par cette nouvelle version d'IDEA est le logiciel R ([R Core Team, 2020](#)). Ceci pour différentes raisons, notamment :

- R est open source : c'est un programme libre et gratuit.
- R est populaire : c'est un logiciel beaucoup utilisé par les universitaires mais aussi dans le privé ou le milieu industriel.
- R est un langage de programmation : le travail repose sur des scripts (ou algorithmes), ce qui permet une automatisation efficace et garantit la reproductibilité des résultats.
- Les membres du comité scientifique de la méthode IDEA en charge du développement de ces outils sont des statisticiens. R est donc leur outil de travail.

Les chapitres suivants sont dédiés à la description du travail réalisé sous R afin de répondre à ces besoins.

Chapitre 3

Introduction du logiciel R dans l'écosystème IDEA

3.1 Transition vers R : Lecture des calculateurs

Le premier défi à relever dans l'informatisation de la méthode est la lecture des calculateurs. Il s'agit en effet de construire des scripts pouvant automatiquement piocher les données nécessaires aux calculs IDEA dans un fichier Excel comportant une dizaine d'onglets, dont la forme n'est pas encore définitivement arrêtée (ce qui impacte donc le ciblage de cellules particulières dans les scripts) pour ensuite les convertir dans un format compatible avec le logiciel R mais aussi avec d'autres systèmes informatiques tels que des bases de données SQL.

Le calculateur Excel IDEA permet la saisie des données de l'exploitation, obtenues après enquête, dans un onglet "Saisie et Calculateur" (Figure 3.1).

Cet onglet comporte un peu plus de 1000 lignes et a été structuré de manière très pédagogique pour guider les utilisateurs dans sa complétion. Le principal inconvénient de cette approche est qu'il n'a pas été pensé, à l'époque de son développement, pour une compatibilité avec des scripts d'extraction automatique. De ce fait, la modification de la forme de cet onglet à chaque mise à jour du calculateur rendait impossible, dans le laps de temps donné pour le développement, de créer une solution informatique permettant de récupérer les informations nécessaires au calcul des indicateurs IDEA. Au vu de la diversité de versions du calculateur déjà existantes et qui restent encore à développer au fil des évolutions de la méthode, le développement de processus d'import automatisés, efficaces et fiables a représenté un des plus grands défis de ce travail.

IDEATools : Un environnement informatique pour l'automatisation et le reporting de données IDEA4

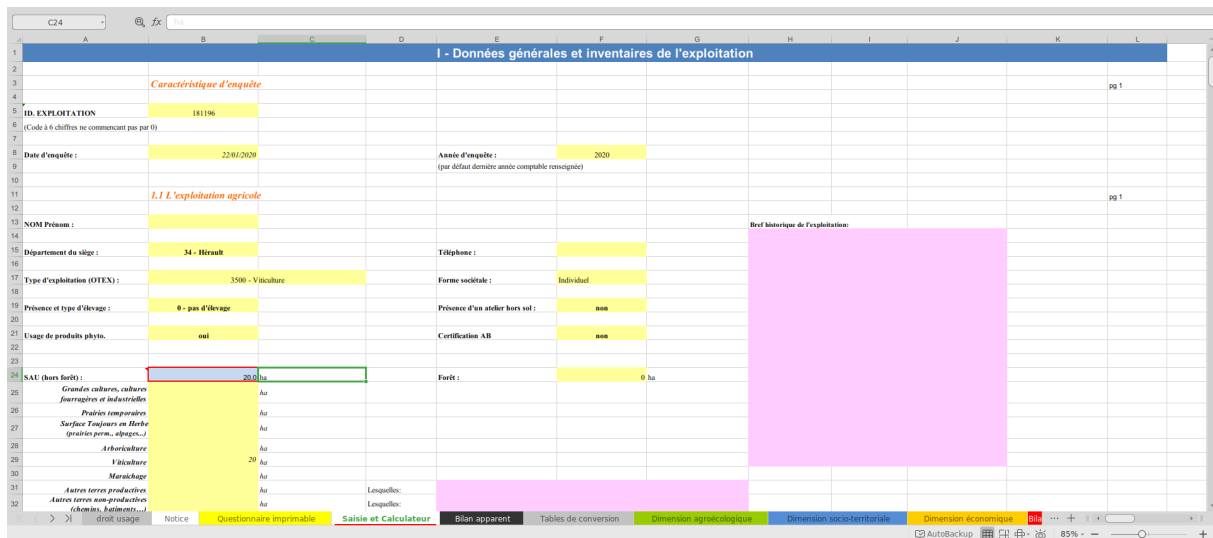


FIGURE 3.1 – Capture d'écran de l'onglet "Saisie et Calculateur".

Deux stratégies majeures d'extraction ont été implémentées en fonction de l'ancienneté du calculateur. Une première solution, la plus fiable, est basée sur un onglet rajouté dans les versions les plus récentes du calculateur qui compile toutes les données nécessaires au calcul des indicateurs. La seconde solution, plus hasardeuse, va tenter de piocher ces données aux travers des autres onglets de saisie et résultats (dont la forme évolue plus ou moins fortement au fil des versions). Les fonctionnalités implémentées à ce jour permettent d'extraire des données quasi-systématiquement pour des calculateurs postérieurs au 01/01/2019. Le support de fichiers JSON a également été facilement implémenté puisque ce format de sortie du calculateur était pensé à l'origine pour l'alimentation d'une base de données.

Les paquets R ici utilisés sont `{readxl}` pour la lecture de fichiers Excel (.xls ET/OU .xlsx), `{jsonlite}` pour la lecture des fichiers .json, puis les packages de l'écosystème `{tidyverse}` pour la manipulation et mise en forme des données extraites.

3.1.1 Comment remplacer DEXi ?

Une fois les données importées dans un format manipulable sous environnement R, l'objectif était de reproduire l'évaluation produite par DEXi en langage R. Pour cela, la première étape de transfert de scores (numériques) vers des modalités (qualitatives) a été implémentée sous forme de "SI.. ALORS..." avec une table répertoriant, pour chacun des 53 indicateurs, les seuils de chaque modalité.

Ensuite, chacune des 46 tables de contingences construites sous DEXi devait être rapatriée sous un format lisible par R. Fort heureusement, le logiciel DEXi fournit une possibilité d'export des "règles de décision" au format CSV. Le travail a donc consisté en la traduction de ces fichiers en tables de données R, puis à l'élaboration de scripts

Chapitre 3. Introduction du logiciel R dans l'écosystème IDEA

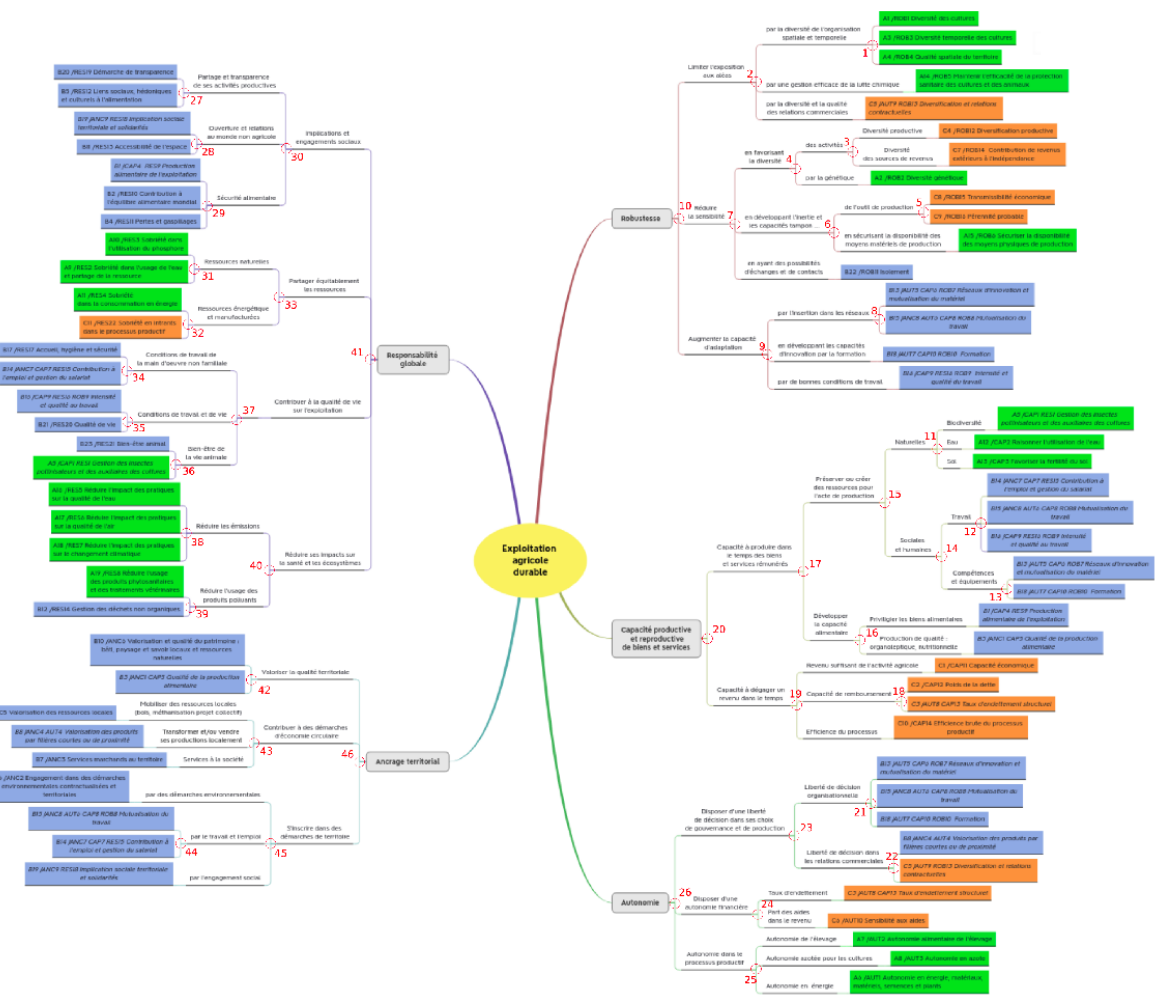


FIGURE 3.2 – Carte heuristique des propriétés avec la numérotation des nœuds ayant été utilisée dans les scripts R.

basés sur des jointures (près de 1000 lignes de script) pour permettre, à partir de 53 scores numériques, d'obtenir l'évaluation qualitative de chacun des nœuds de la carte heuristique. Chacun des nœuds a ici été identifié selon la numérotation présentée sur la Figure 3.2.

Une fois cette évaluation réalisée, il reste encore à la restituer de façon graphique par ce qui a été nommé des "arbres éclairés". Le tracé de graphiques à la structure si complexe et précise, avec des intitulés de nœuds parfois très long, a demandé un lourd travail de recherche et veille technologique pour identifier une solution viable et automatisable. La solution finalement retenue s'appuie sur le langage de balisage SVG, utilisé par des logiciels de dessin vectoriel tels que Inkscape.

Les modèles sont d'abord dessinés en "blanc" sur Inkscape, ce qui permet une liberté totale de mise en page et de taille de texte.

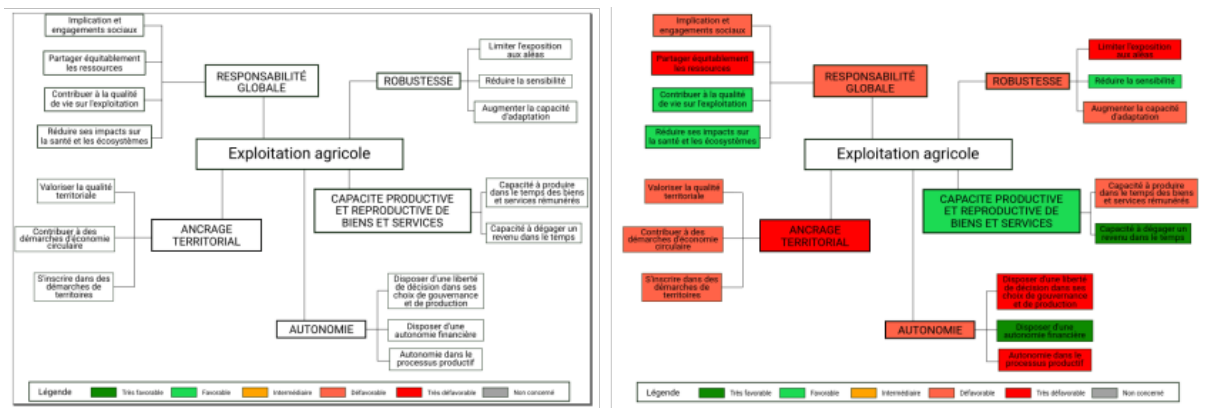


FIGURE 3.3 – (Gauche) Modèle utilisé pour l'arbre éclairé global synthétisant les résultats pour les 5 propriétés. (Droite) Ce même modèle après coloration programmatique.

Puis, un algorithme à base de “Rechercher/Remplacer” va être appliqué sur le code source de ce modèle (le SVG étant de structure XML, similaire au HTML par exemple) afin de remplacer le blanc par la bonne couleur correspondant à l'évaluation (Figure 3.3).

Le paquet R `{rsvg}` permet ici la manipulation de ce code source SVG “modifié” pour le transformer au format .png et/ou .pdf.

Au terme de ce premier travail, l'algorithme développé sous R permet d'importer les données d'un calculateur et de produire ces arbres éclairés en **quelques secondes**.

3.2 Production de graphiques élaborés

Une fois les algorithmes R nécessaires à la restitution opérationnelle des arbres éclairés développés, le conseil scientifique a décidé de continuer à exploiter les possibilités offertes par le langage R en recherchant notamment la production de graphiques plus élaborés que ceux proposés par le calculateur excel.

Ce travail s'est beaucoup appuyé sur le retour d'utilisateurs, au fil de l'eau ou au travers de questionnaires spécifiques, afin d'identifier les principaux besoins exprimés par les utilisateurs en fonction de leur utilisation de la méthode (enseignement, conseil, recherche, etc.). De nombreuses réflexions propres à la data-visualisation ont également été menées en ce qui concerne les diagrammes "radar" ou "araignée", très souvent critiqués car ce type de représentation introduit un biais de lecture en fonction de l'ordre des intitulés (Albo et al., 2015). Dans ce travail, nous leur avons préféré des diagrammes en barre classiques auxquels nous avons appliqué une rotation polaire pour conserver un aspect circulaire (Figure 3.4).

Ces nouveaux graphiques sont exclusivement produits à partir du paquet R `{ggplot2}`.

Chapitre 3. Introduction du logiciel R dans l'écosystème IDEA

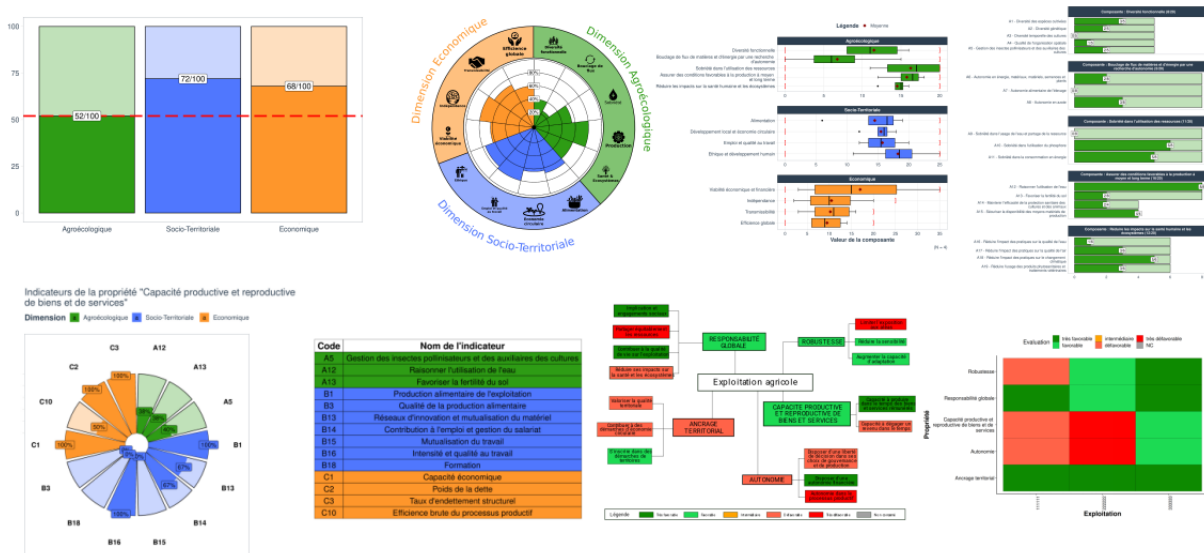


FIGURE 3.4 – Exemple de sorties graphiques automatiques développées dans le cadre de ce travail.

3.3 Développement des analyses de groupe

Au travers de l'automatisation du traitement des données IDEA4, l'utilisation de scripts R permet également d'envisager des analyses multiples et/ou de groupes dans un délai temporel tout à fait raisonnable. En effet, à titre d'exemple, le code actuellement développé permet de traiter les données de 40 exploitations (donc 40 calculateurs) **en moins d'une minute**.

Des travaux sont encore en cours sur le sujet des analyses de groupe avec IDEA, ayant notamment pour but d'établir l'échelle d'analyse et les représentations les plus pertinentes. Des représentations génériques ont cependant déjà été proposées telles que des figures présentant la distribution des notes des dimensions sous forme de boîte à moustache (ou boxplot) ou encore des matrices de couleur représentant l'évaluation de chaque exploitation par les propriétés au sein d'un collectif (Figure 3.5).

3.4 Standardisation et compilation des résultats

Une fois l'ensemble des états de sorties élaborés, il restait à définir des supports permettant d'assembler les résultats d'une exploitation sous forme de document synthétique, plutôt qu'un ensemble de fichiers image .png ou .pdf. Le conseil scientifique de la méthode n'a pas souhaité implémenter des conseils "automatisés" basés sur des seuils puisque tout l'intérêt d'IDEA réside dans les commentaires apportés aux résultats par un expert. Ces rapports automatisés consistent seulement en l'addition des



FIGURE 3.5 – Exemple de sorties automatiques proposées pour les analyses de groupe dans le cadre de ce travail.

Chapitre 3. Introduction du logiciel R dans l'écosystème IDEA

graphiques de résultats mis les uns à la suite des autres, avec un minimum de description pour la compréhension et l'apposition de logos / métadonnées sur l'analyse.

Afin de répondre aux besoins d'impression et/ou de visualisation simple, des rapports aux formats PDF et HTML ont d'abord été développés puis proposés aux utilisateurs. Ensuite, il semblait utile pour certains utilisateurs de pouvoir disposer d'un fichier librement éditable afin d'ajouter, par exemple, directement leurs commentaires sur les résultats. Des sorties au format DOCX (Word) et au format ODT (Libre Office) ont donc été implémentées. Une sortie au format XLSX (Excel) a ensuite été rajoutée pour les utilisateurs souhaitant travailler directement avec leur données et souhaitant réaliser d'autres analyses. Enfin, une sortie au format PPTX (Powerpoint) a été rajoutée afin de fournir un support de présentation prêt à l'emploi pour une situation de conseil ou pour une restitution rapide des résultats. L'ensemble de ces développements sont illustrés par la Figure 3.6.

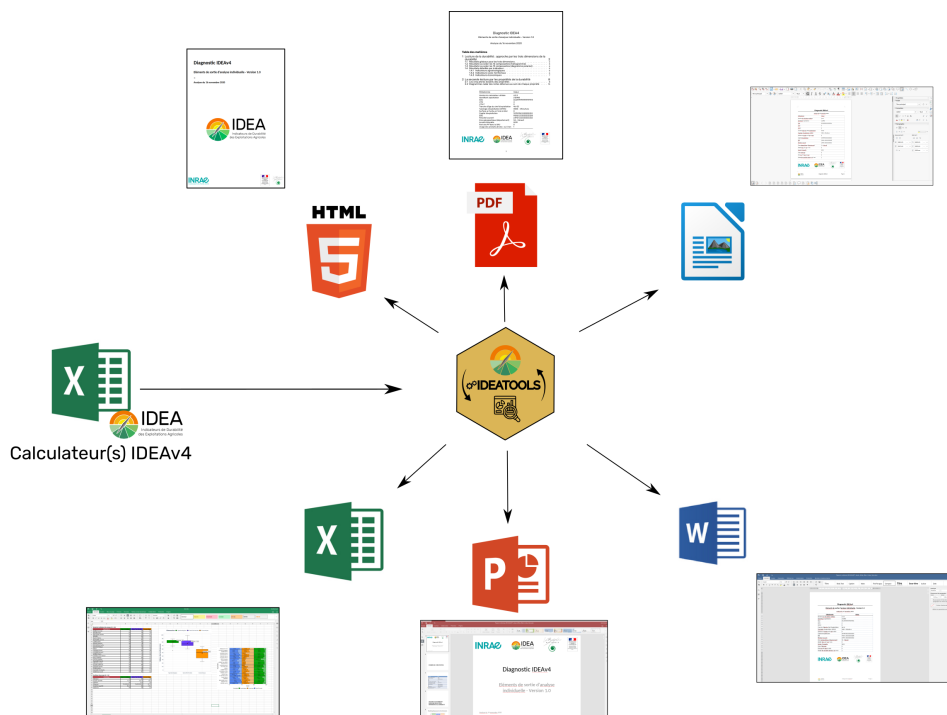


FIGURE 3.6 – Illustration des formats de restitutions possibles pour les diagnostics IDEA.

A l'issue de ces développements, près de 4500 lignes de code R ont été écrites au travers d'une dizaine de fichiers. A ce stade, il s'avérait important de compiler, synthétiser, documenter et améliorer la reproductibilité des processus par des utilisateurs autres que le conseil scientifique de la méthode. Ce travail de compilation s'est traduit par la création d'un package R, décrit plus en détail dans le chapitre 4.

Chapitre 4

Le package R IDEATools

4.1 Présentation d'IDEATools

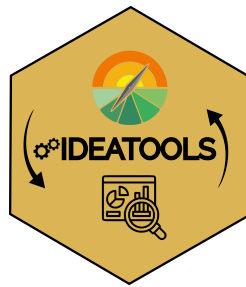


FIGURE 4.1 – Logo d'IDEATools

IDEATools est un package R dédié à la méthode IDEA4, visant à fournir aux utilisateurs des outils pour le traitement, l'automatisation et le reporting de diagnostics IDEA (Carayon, 2020).

Un package R est une collection de fonctions R réunies au sein d'un environnement bien documenté, dans lequel toutes les dépendances à d'autres packages R sont bien tracées et qui s'inscrit dans une logique de suivi de version. Tous ces éléments assurent une reproductibilité des résultats par tout utilisateur du package.

De plus, il s'agit d'un format de diffusion très efficace puisque tous les packages R sont open-source. Ainsi, par la formulation de ce package, l'ensemble de ces travaux sont librement partagés avec la communauté des utilisateurs, qui sont libres de les utiliser et/ou de les modifier sous condition du respect de la licence de diffusion GPL-3.

L'ensemble des règles de décisions / choix arbitraires décidés par le comité scientifique de la méthode (sur la base d'avis d'experts et de revues bibliographiques) y sont affichés en toute transparence puisque ce travail s'inscrit pleinement dans une démarche de science ouverte.

En résumé, un diagnostic IDEA individuel devant être restitué aux formats DOCX (Word) et XLSX (Excel) via le package IDEATools s'écrit en langage R comme dans la Figure 4.2.



```
> library(IDEATool)
> diag_idea(
+   input = path,
+   output_directory = tempdir(),
+   type = "single",
+   export_type = "report",
+   prefix = "Exploitation_A",
+   dpi = 300,
+   report_format = c("docx", "xlsx"),
+   quiet = FALSE
+ )

— Début du diagnostic individuel IDEA4 —
i Nombre de rapports demandés : 2
— Import du calculateur —
✓ Calculateur bien importé (0.01s)
— Calcul des indicateurs IDEA4 —
✓ Les 53 indicateurs ont bien été calculés (1s)
— Tracé des graphiques —
✓ Les graphiques ont été construits sans erreur (2s)
— Production des 2 rapports demandés —
— Production du rapport Microsoft Word...
i Le rapport a été exporté à l'adresse '/tmp/Rtmp928qkS/2021-01-15/Exploitation_A/Rapport_individuel_Exploitation_A.docx' (10s)
— Production du rapport Microsoft Excel...
i Le rapport a été exporté à l'adresse '/tmp/Rtmp928qkS/2021-01-15/Exploitation_A/Rapport_individuel_Exploitation_A.xlsx' (10s)
```

FIGURE 4.2 – Extrait de code utilisé pour la réalisation d'un diagnostic IDEA4 pour l'exploitation "A" visant à produire un rapport Word et un document Excel.

La documentation complète du package ainsi que ses différents guides d'utilisation sont présentés dans la section suivante.

4.2 Stratégie de diffusion et d'intégration continue

Le package R IDEATools est aujourd'hui hébergé sur la plateforme GitHub à l'adresse <https://github.com/davidcarayon/IDEATools> et sera bientôt soumis pour archivage au dépôt R officiel et international du CRAN. Toujours dans une démarche de transparence, un site web dédié au package a été développé et met à disposition une grande quantité de documentation expliquant en détail la construction et le

fonctionnement du package : <https://davidcarayon.github.io/IDEATools/> (Figure 4.3).

The screenshot shows the IDEATools website homepage. The header is dark blue with navigation links: 'Fonctions', 'Articles', and 'News'. The main content area is white. On the left, the IDEATools logo is displayed, followed by a description: 'IDEATools est un package R dédié à la méthode IDEA4, visant à fournir aux utilisateurs des outils pour le traitement, l'automatisation et le reporting de diagnostics IDEA.' Below this is the 'Installation & Prérequis' section, which includes a note about the official CRAN publication and provides two code snippets for installation and loading. The 'Dev status' sidebar on the right shows: 'License: Full license, GPL-3'; 'Developers: David Carayon (Author, maintainer)'; 'Dev status: lifecycle: maturing, Package version: 2.0, licence: GPL-3, R-CMD-check: passing, codecov: 0%'. A note at the bottom mentions that 'Roboto' and 'Rubik' fonts are required for the package.

FIGURE 4.3 – Capture d'écran de la page d'accueil du site web dédié au package IDEATools.

Le choix de la plateforme GitHub permet également d'ajouter une étape de vérification de la qualité du code et de son interopérabilité entre les différents systèmes d'exploitation (Windows, Mac, Linux) via l'utilisation du service d'intégration continue **GitHub Actions**. A ce jour, le package IDEATools a passé avec succès tous les tests de vérification.

4.3 Mise à disposition d'IDEATools aux utilisateurs

Bien que le package R {IDEATools} soit aujourd'hui opérationnel, son utilisation nécessite la maîtrise du logiciel R, qui reste un outil principalement utilisé dans un milieu universitaire et/ou certaines entreprises. La dernière étape du processus d'information de la méthode IDEA consistait donc en l'encapsulation de ces outils R sous forme d'interfaces graphiques "presse-bouton".

Toujours dans l'écosystème R, le paquet {shiny} permet de construire de très belles applications web à partir d'un code R. Shiny est un cadre conceptuel de développement

(ou framework) permettant le développement d'applications interactives pour le web tout en profitant de la puissance calculatoire et des nombreuses bibliothèques graphiques de R. De ce fait, les outils IDEATools peuvent être directement embarqués dans une application web, complétés par des visualisations inédites rendues possibles par le passage dans le monde du web. Ce passage d'IDEATools sous forme d'application web a été entrepris à deux reprises.

4.3.1 Module de production des arbres éclairés

Une partie des outils développés dans le package IDEATools a d'abord été mobilisée dans la construction d'une application web basée sur la technologie `{shiny}`, spécifiquement dédiée à la production d'arbres éclairés et qui a été intégrée à la plateforme WEB-IDEA développée par Bordeaux Sciences Agro <http://idea4.agro-bordeaux.fr>.

L'application est intégrée au site web via une balise HTML `<iframe>` et est hébergée sur le même serveur que ce site web via une instance de ShinyServer (Figure 4.4).

En fonction de l'utilisateur connecté, l'application restreint la sélection de dossiers aux dossiers appartenant à l'organisme de l'utilisateur. Une fois les filtres choisis et la recherche exécutée, une connexion lecture-seule à la base de données derrière le site web est effectuée, les données sont manipulées et transformées pour devenir identiques à un fichier JSON sortant du calculateur Excel, puis les modules d'IDEATools permettent de produire les arbres éclairés. Ceux-ci sont présentés directement dans l'interface puis téléchargeables via une archive ZIP.

4.3.2 L'application ShinyIDEA

Les travaux de développement web ont été poursuivis dans le but de construire une application complète restituant la totalité des outils développés dans IDEATools par des simples boutons à presser.

Cette application dont le nom de code est "ShinyIDEA" est aujourd'hui diffusée au grand public via le lien <https://outils-idea.inrae.fr/>.

Cette application permet en quelques clics d'insérer un ou plusieurs calculateurs IDEA, d'observer directement sur l'application un certain nombre de résultats et/ou statistiques, puis de télécharger l'ensemble des résultats au format souhaité (Figures 4.5, 4.6 et 4.7).

Chapitre 4. Le package R IDEATools

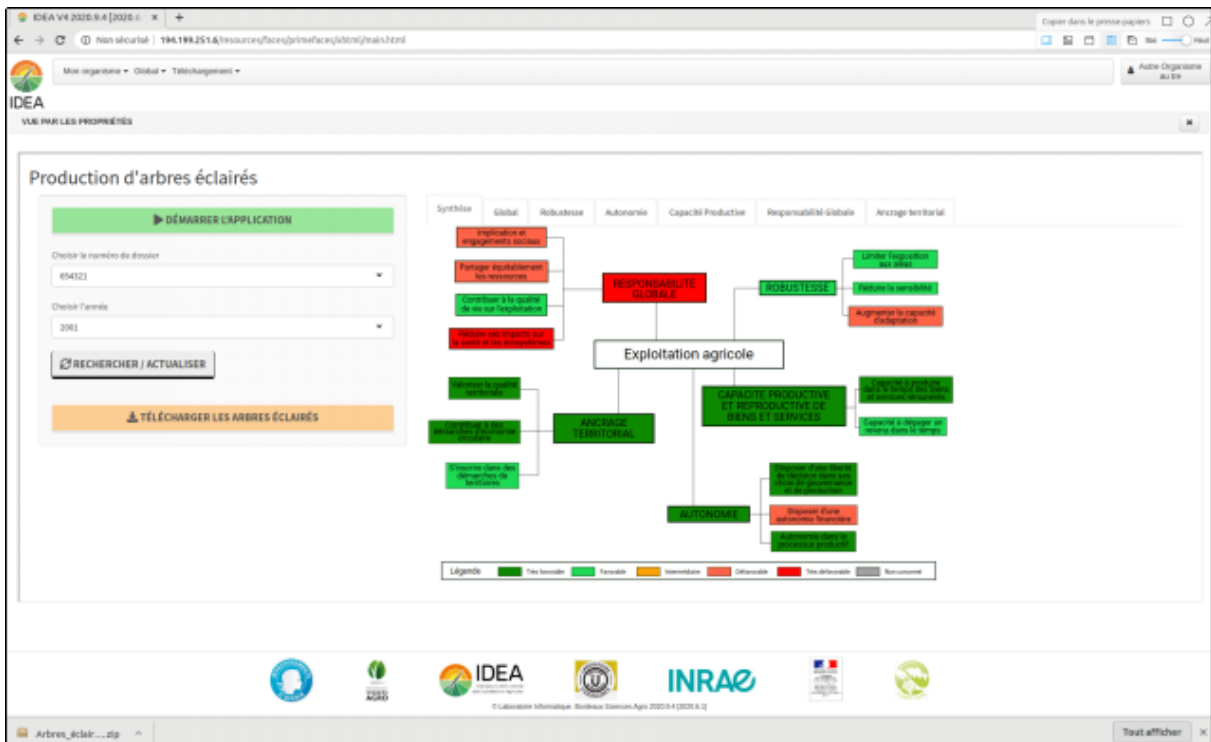


FIGURE 4.4 – Module de production des arbres éclairés implémenté directement dans l’interface du site développé par Bordeaux Sciences Agro.

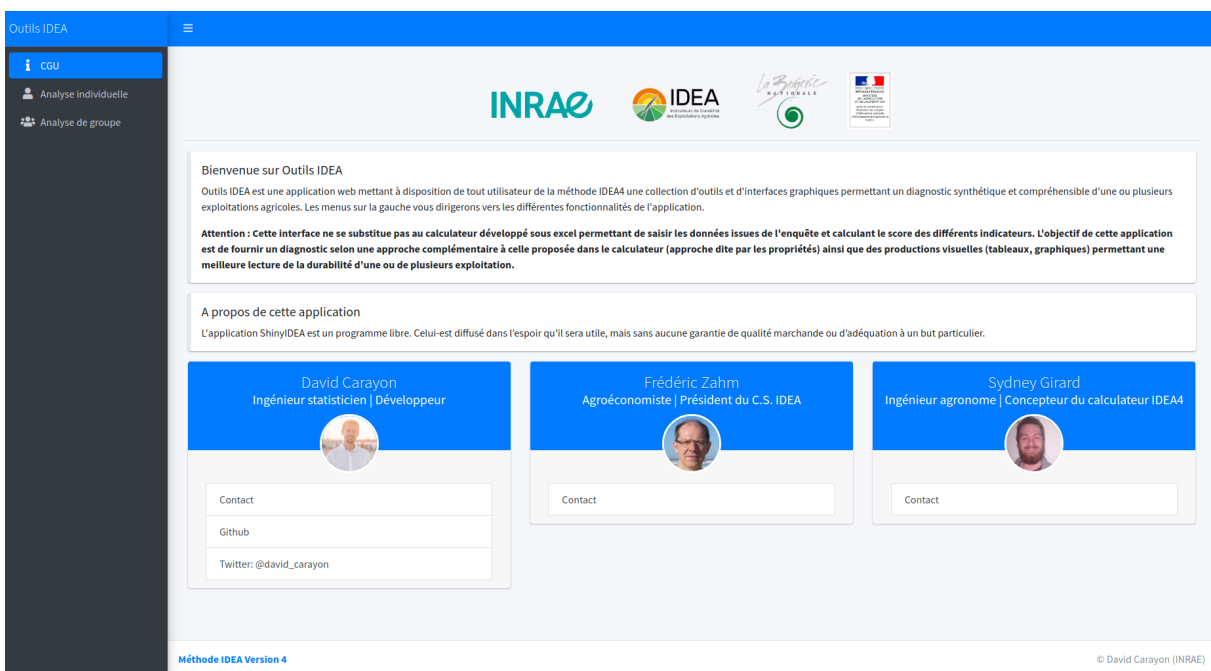


FIGURE 4.5 – Page d’accueil de l’application ShinyIDEA, présentant les principales fonctionnalités de l’application ainsi que l’équipe de développement.

IDEATools : Un environnement informatique pour l'automatisation et le reporting de données IDEA4

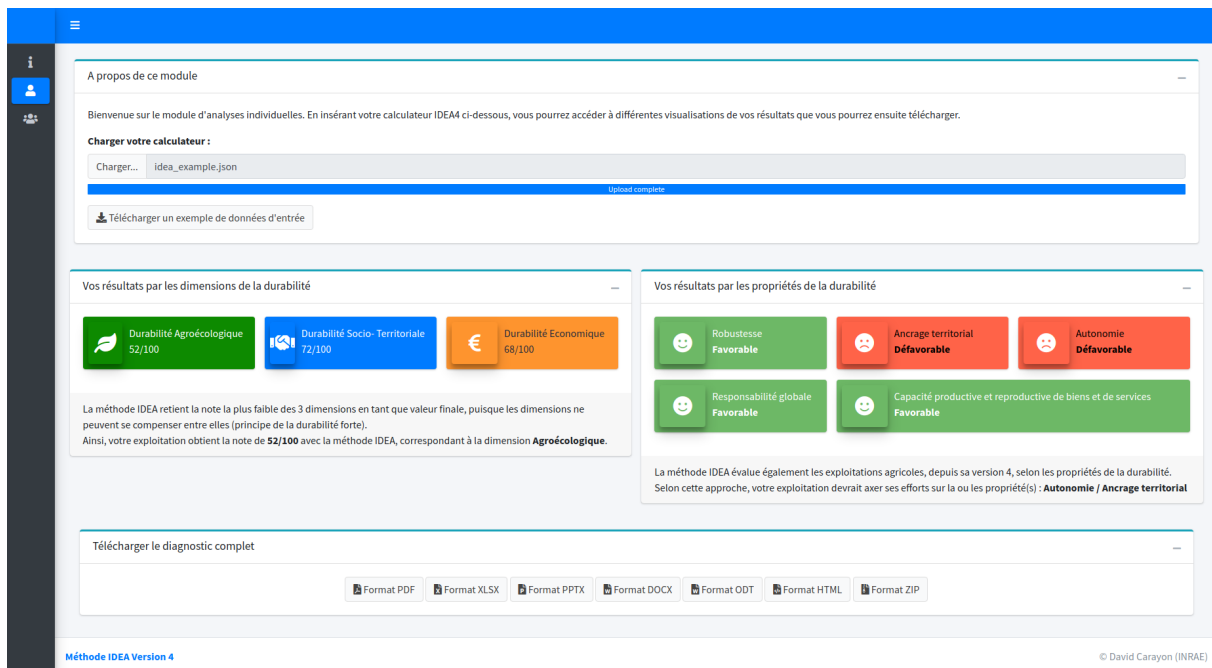


FIGURE 4.6 – Capture d'écran présentant un exemple d'analyse individuelle sur ShinyIDEA. L'utilisateur a pu insérer son calculateur, obtenir des statistiques simples, puis télécharger ses résultats au format qu'il souhaitait.

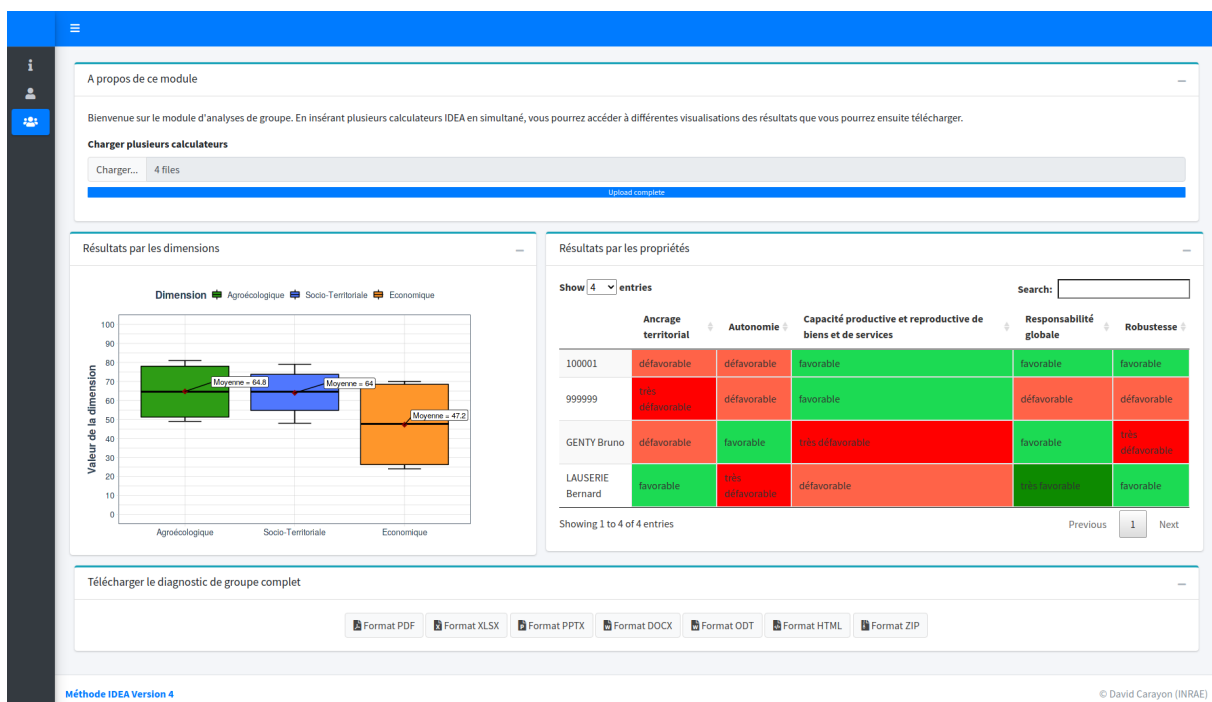


FIGURE 4.7 – Capture d'écran présentant un exemple d'analyse de groupe sur ShinyIDEA. L'utilisateur a pu insérer 4 calculateurs, obtenir des statistiques simples, puis télécharger ses résultats au format qu'il souhaitait.

Chapitre 5

Perspectives de développement

5.1 Les analyses de groupe

Comme nous le précisons dans le chapitre 4, le travail sur les analyses de groupe n'a consisté à ce jour qu'en le développement de code et de certaines sorties graphiques servant de base pour la réflexion. Les prochains mois/années de développement des outils informatique de la méthode IDEA permettront la mise en place de travaux spécifiques dédiés à ce type d'analyse afin de réfléchir aux représentations pertinentes (échelle de rendu choisie, type de graphique, de comparaison, etc.).

5.2 Continuer l'informatisation de la méthode

Les outils informatiques dédiés à la méthode IDEA4 existant à ce jour ne collectent en réalité que des données déjà pré-transformées et/ou calculées (items, indicateurs). Aucun outil n'existe actuellement pour extraire la faramineuse quantité d'informations saisies par l'utilisateur dans "Saisie et Calculateur". En effet, cet onglet n'a pas été, au moment de sa conception, prévu dans une optique de "moissonnage automatique de données". La structuration de cet onglet évolue assez fortement au fil des nouvelles versions du calculateur et il comporte de nombreuses entrées sous forme de tableaux et de listes déroulantes qui complexifient la requête de cet onglet.

En partant de ce constat, deux trajectoires de développement ont été identifiées pour espérer récupérer ces données qui présentent un grand intérêt pour la recherche :

1. Abandonner ou revoir complètement le calculateur, au profit d'un formulaire de saisie beaucoup plus standardisé (type formulaire web)
2. Considérer le calculateur Excel comme pérenne et indissociable de la méthode IDEA4 et creuser les travaux d'extraction de l'onglet "Saisie et Calculateur"

Le conseil scientifique de la méthode est attaché à la conservation du calculateur Excel qui est la seule solution pour réaliser un diagnostic complet de manière autonome et sans connexion internet. C'est donc la seconde option qui sera à privilégier.

La stratégie envisagée à ce jour serait de fractionner cet onglet en une multitude de "briques" élémentaires, chaque brique correspondant à un input utilisateur (un tableau, une liste déroulante, un choix oui/non, une valeur numérique, etc.). Puis, des scripts d'extraction (dans un langage de programmation à définir) seraient développés pour chacune de ces briques. L'ensemble des scripts serait inséré dans une routine ETL (Extract, Transform, Load) par exemple via le logiciel Talend. Par ce procédé, l'algorithme d'extraction sera beaucoup plus robuste et moins soumis aux variations dans la structure de cet onglet. En cas d'erreur d'import, il sera donc très facile de localiser le problème.

Une fois ce levier technique levé, il sera tout à fait envisageable de créer une base ouverte de données nationale de la durabilité évaluée avec la méthode IDEA4, sous forme de Data Warehouse par exemple (Figure 5.1), puis d'envisager des ponts entre cette base et d'autres bases de données liées à des méthodes d'évaluation au travers, par exemple, de la plateforme MEANS (<https://www6.inrae.fr/means/>). De cette manière, la méthode IDEA s'inscrirait pleinement dans une démarche "Big-" et "Open-" Data.

L'ensemble de ces travaux d'informatisation de la méthode seront initiés au travers d'un projet à venir en 2021 nommé POLIDeA pour "POLitiques publique et Indicateurs de Durabilité en Agriculture. Pour une « montée en puissance » de l'usage d'indicateurs de durabilité des exploitations agricoles dans l'action publique agroenvironnementale et la construction d'une stratégie Big et Open Data sur données durabilité à INRAE".

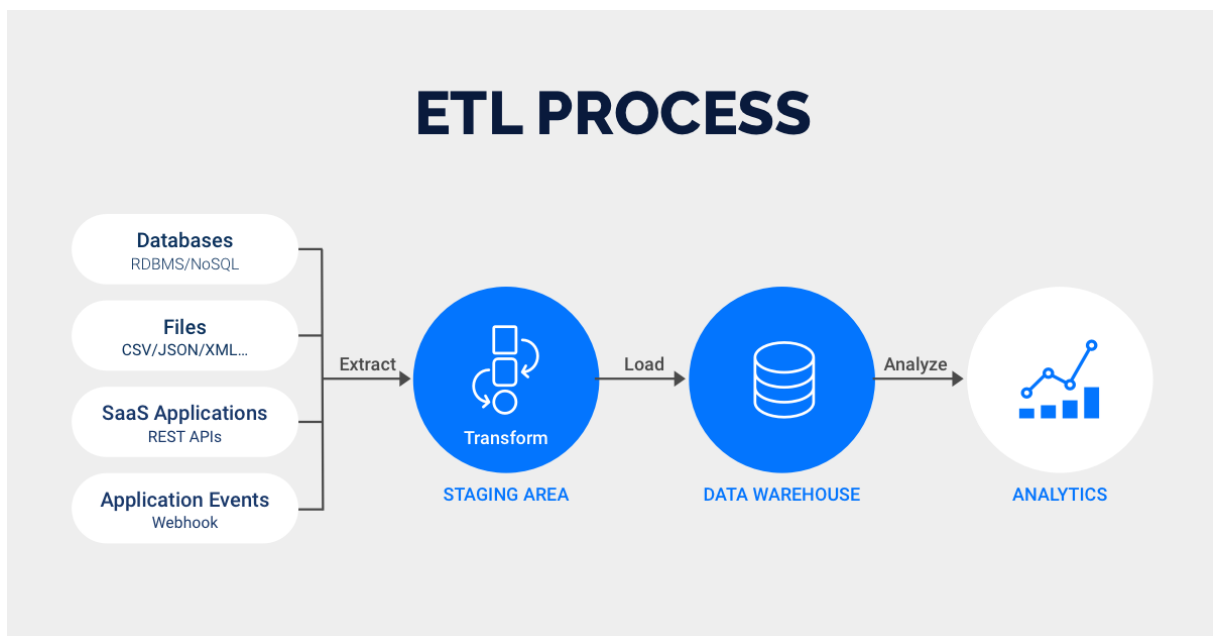


FIGURE 5.1 – Illustration d'un procédé Extract, Transform, Load théorique et de mise à disposition de ces données via un Data Warehouse (Illustration ©Kevin Bartley).

Bibliographie

Etienne Landais. *Agriculture durable : les fondements d'un nouveau contrat social?* 1998.

Frédéric Zahm, A Alonso Ugaglia, Jean Marc Barbier, Héloïse Boureau, Bernard Del'Homme, Mohamed Gafsi, Pierre Gasselin, Sydney Girard, Laurence Guichard, Chantal Loyce, et al. *Évaluer la durabilité des exploitations agricoles : la méthode idea v4, un cadre conceptuel combinant dimensions et propriétés de la durabilité.* 2019.

Marko Bohanec. *Dexi : Program for multi-attribute decision making user's manual.* *Ljubljana, Slovenia : Institut Jozef Stefan, 2008.*

R Core Team. *R : A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2020. URL <https://www.R-project.org/>.

Yael Albo, Joel Lanir, Peter Bak, and Sheizaf Rafaeli. *Off the radar : Comparative evaluation of radial visualization solutions for composite indicators.* *IEEE transactions on visualization and computer graphics*, 22(1) :569–578, 2015.

David Carayon. *IDEATools : A Collection of Tools Dedicated to the IDEA4 Method,* 2020. R package version 2.0.1.



ETBX – Environnement, Territoires et Infrastructures

50, avenue de Verdun

33612 CESTAS Cedex

+33 (0)5 57 89 08 00

Rejoignez-nous sur :



**Institut national de recherche pour
l'agriculture, l'alimentation et l'environnement**

INRAE
la science pour la vie, l'humain, la terre