

CRAWLSPEC: une solution pour la gestion des données spectrales

Maxime Joissains, Philippe Vermeulen, Audrey Pissard, Juan Antonio Fernández Pierna, Vincent Baeten
Et toutes les équipes impliquées dans les analyses spectrales

Contact: m.joissains@cra.wallonie.be

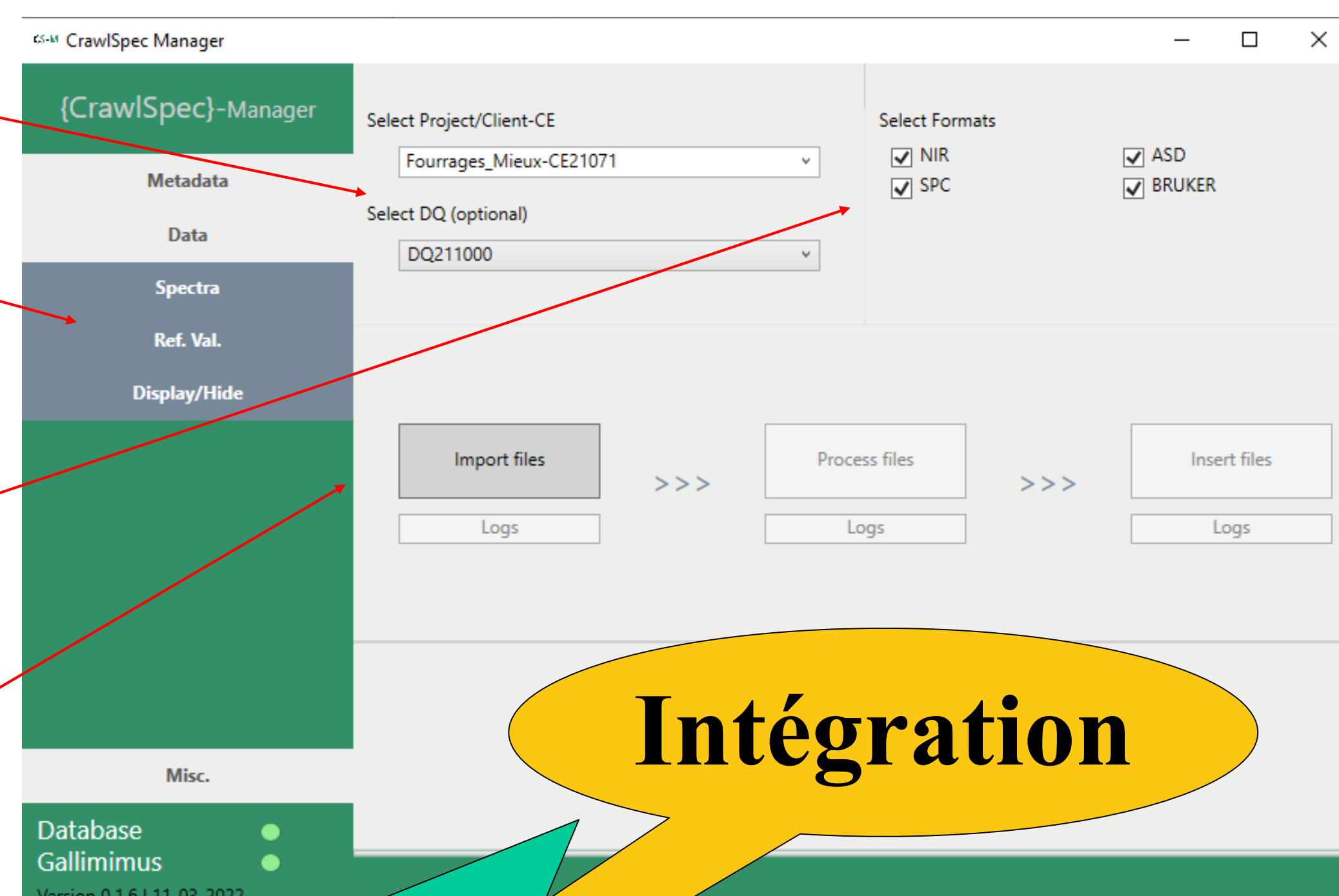
Afin de centraliser et de faciliter l'accessibilité aux données spectrales issue de la spectroscopie vibrationnelle et aux métadonnées correspondantes, le CRA-W développe une solution originale dénommée **CrawlSpec**:

Centralised Reference Analytical Web-accessible Library of SPECTra.

CrawlSpec regroupe les outils nécessaires pour l'alimentation et l'exploitation de la base de données en se basant sur 4 fonctionnalités principales : la *centralisation*, l'*intégration*, l'*extraction* et l'*exploitation* des données.

L'étape préliminaire est la mise en place de la **sémantique** pour identifier les différents objets repris dans la base de données, à savoir : les identifiants des données, les catégories de produits ainsi que les paramètres analytiques qualitatifs et quantitatifs. Ensuite, une **structure de fichier MS Excel** permet de centraliser, par projet/client, les données de référence qualitatives et quantitatives, issues des différents laboratoires et unités d'expérimentation internes et externes à l'institution.

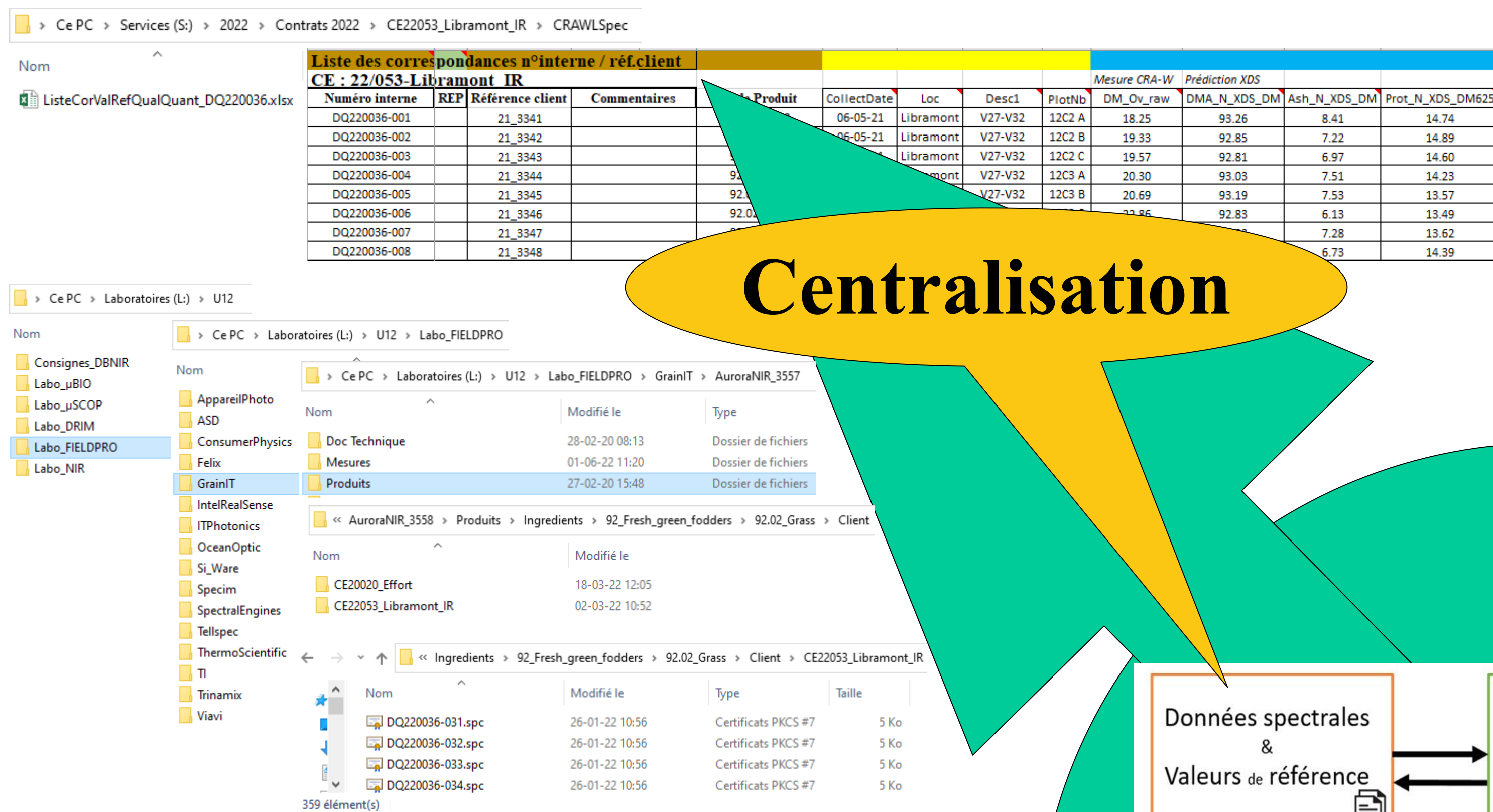
- Importation par projet/client/ liste d'échantillons
- Intégration de données de référence et de spectres
- Format MS-Excel pour les données de référence
- Divers formats propriétaires pour les données spectrales
- Contrôle de l'intégrité des données



Intégration

CrawlSpec Manager est une application informatique locale visant à intégrer les informations dans une base de données centralisée et sécurisée, tout en vérifiant leur intégrité.

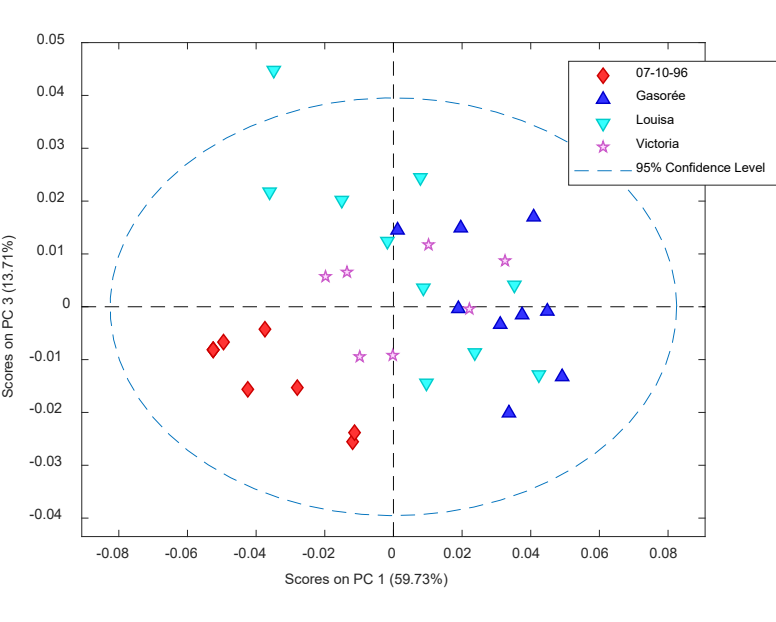
Centralisation



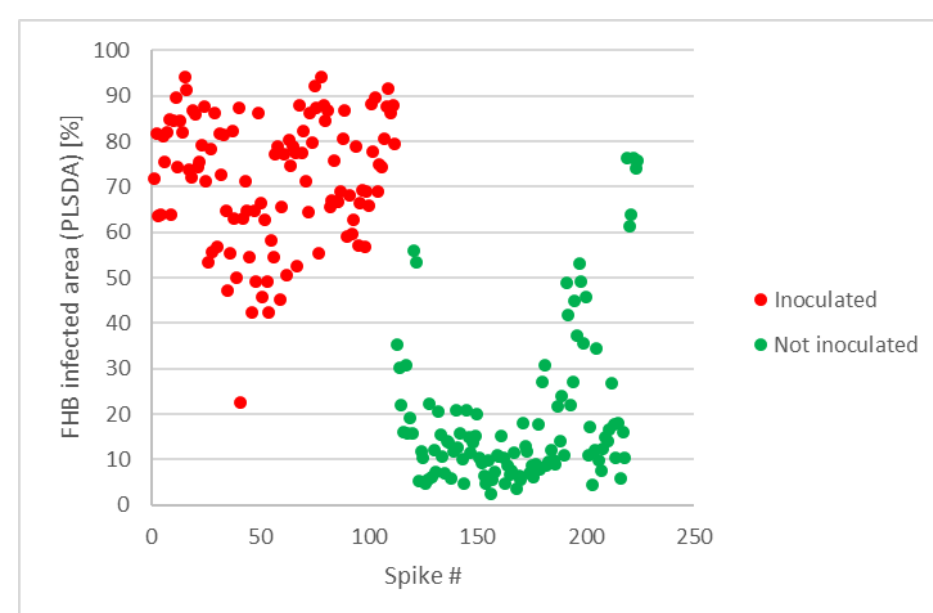
Tous les fichiers de données spectrales acquises en laboratoire ou sur site sont stockés sous format propriétaire et dans une **arborescence** structurée en fonction du type d'instrument, de la catégorie de produit et du projet.

Exploitation

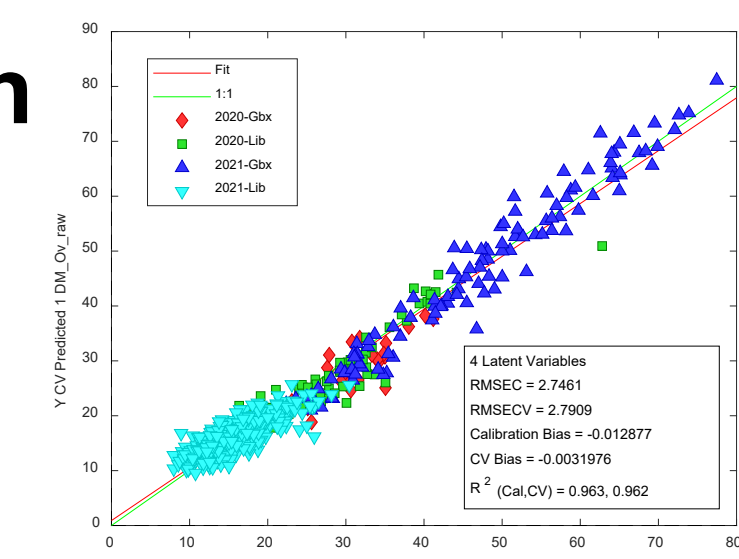
ACP



Discrimination



Régression



Les résultats sont exportés sous différents **formats de fichier** suivant les spécifications de l'utilisateur. CrawlSpec permet de pérenniser l'**exploitation** des données pour le développement de modèles de régression et de discrimination.

Extraction

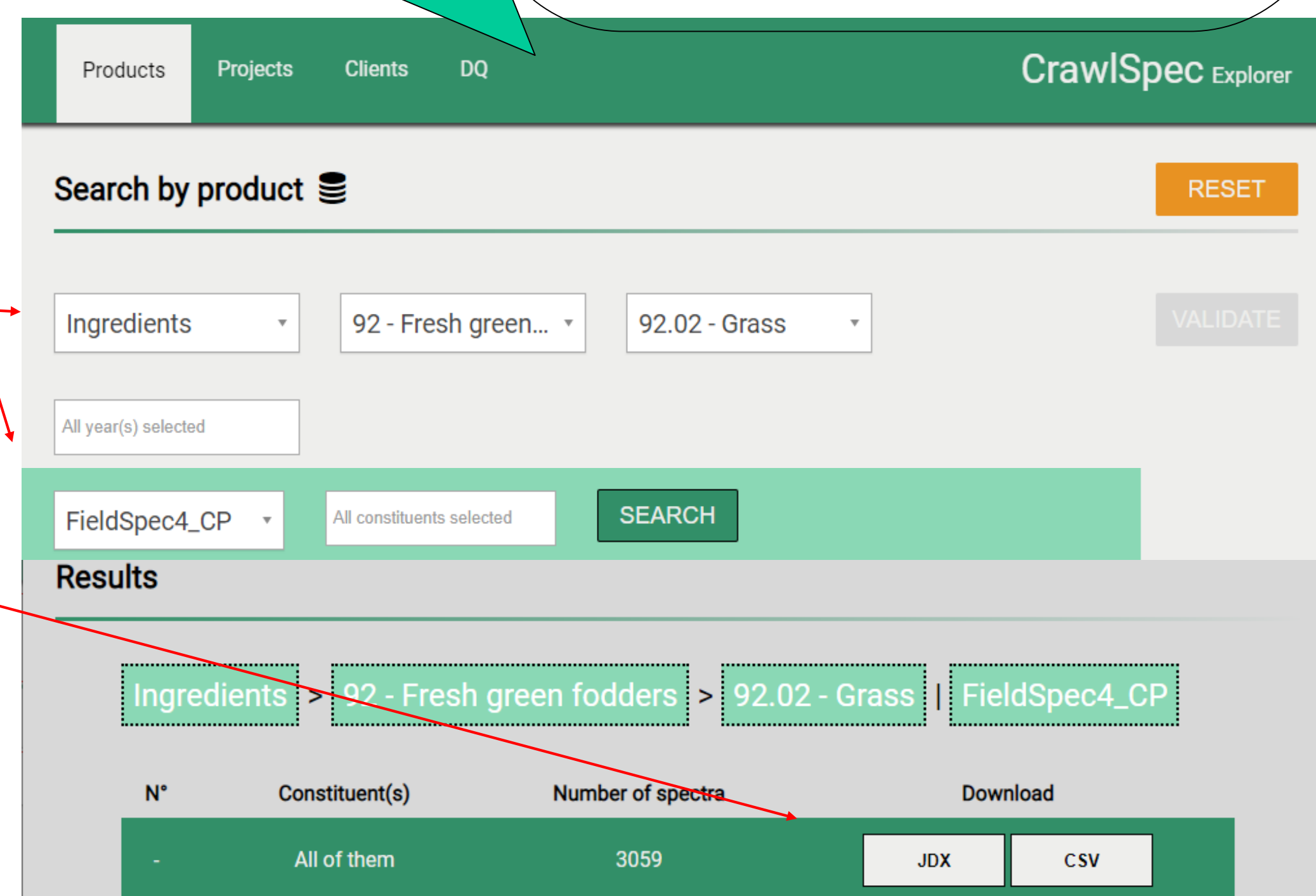
Id Données qualitatives et quantitatives

DQ	DATE	PRODUCT LIV2	PRODUCT LIV1	DM_OV	DM_A	DM_N	DM_P	DM_F	DM_D	DM_S	DM_T	DM_U	DM_V	DM_W	DM_X	DM_Y	DM_Z	DM_Ash	DM_N_XDS_DM	DM_P_XDS_DM	DM_T_XDS_DM	DM_U_XDS_DM	DM_V_XDS_DM	DM_W_XDS_DM	DM_X_XDS_DM	DM_Y_XDS_DM	DM_Z_XDS_DM
DQ1805200001	17-08-18	Fresh silages	Pre-wilted grass silage	48	95.34	13.41	29.1	33.14	3.4	32.73	3.25	15.52	32.4	35.95	31.06	35.69	32.17										
DQ1805470101	29-08-18	Fresh silages	Pre-wilted grass silage	25	95.04	11.38	36.11	39.85	2.35	62.61	0.41	8.67	61.72	66.14	62.81	64.01	59.65										
DQ18054701201	29-08-18	Fresh silages	Pre-wilted grass silage	32	93.09	12.25	31.24	35.95	2.4	53.48	3.57	9.44	68.97	69.5	70.32	71.24	67.13										
DQ1807080201	02-11-18	Fresh silages	Pre-wilted grass silage	42.03	93.12	16.18	29.12	33.81	2.12	50.61	3.81	8.59	68.4	75.31	73.14	69.26	70.78										
DQ1901700201	09-04-19	Fresh silages	Pre-wilted grass silage	82.89	94.62	13.33	25.97	29.98	3.59	55.56	9.17	9.85	48.39	61.43	69.89	50.87	66.63										
DQ19020001001	17-04-19	Fresh silages	Pre-wilted grass silage	82	95.1	11.93	31.28	38.23	5.33	60.91	4.03	7.81	46.08	53.29	53.36	48.27	49.69										
DQ2109410401	10-01-22	Fresh silages	Pre-wilted grass silage	35.49	92.87	0	24.9	27.98	2.47	46.96	0	8.51	0	75.5	0	0	78.65										
DQ2109410501	10-01-22	Fresh silages	Pre-wilted grass silage	19.31	98.92	0	40.2	44.28	4.36	72.88	0	5.81	0	58.95	0	0	42.97										

Données spectrales

400	402	404	406	408	410	412	414	416	418
0.72276	0.76987	0.814206	0.8550887	0.893627	0.9205058	0.9455805	0.9654555	0.9812391	0.9941828
0.70898	0.75783	0.806952	0.851592	0.891186	0.9252033	0.9539584	0.9779732	0.9981766	1.015058
0.69947	0.74961	0.798692	0.8454777	0.8829355	0.9170523	0.9459496	0.9700488	0.9898863	1.006445
0.54011	0.59046	0.6382215	0.6817296	0.7198548	0.7521038	0.7787075	0.8004446	0.8182814	0.8331442
0.61341	0.65825	0.6994133	0.7350999	0.7662287	0.7918949	0.8119407	0.8274704	0.8394756	0.8491197
0.76027	0.80946	0.8553555	0.8979444	0.9370957	0.9630345	0.9850939	1.002284	1.0153304	1.025918
0.75458	0.80363	0.8500775	0.8942471	0.9331375	0.9666755	0.9947799	1.017946	1.037041	1.053063
0.78858	0.83622	0.8819811	0.9241675	0.9614433	0.993376	1.020348	1.042262	1.060555	1.075955

- Extraction des données par
 - produit & instrument
 - projet/client
 - Liste d'échantillons
- Fichiers de sortie en format jdx/csv incluant les données de référence qualitatives et quantitatives et les spectres



Cet outil de gestion de données spectrales permettra d'accompagner les secteurs agricoles et alimentaires dans la mise en œuvre de l'Industrie 4.0. Elle contribuera à partager l'expertise du CRA-W dans le domaine des capteurs optiques et de la modélisation.