



SI-DEAP: Système d'information des données environnementales actuelles et passées

Réponse expérimentale à des besoins
récurrents.

– Sylvie Damy – Anne Vignot – Jean-Daniel Tissot –
Journée bases de données 21 Mars 2013

Contexte et constats : Laboratoire Chrono-E(cologie)nvironnement

- Production accrue de données biotiques et abiotiques issues de nombreuses disciplines.
- Production continue de la donnée (question du devenir patrimonial).
- Accroissement des échanges entre les communautés scientifiques internationales. Deux types d'utilisateurs : producteur et modélisateur
- Hétérogénéité des outils et des formats
- Difficulté à pérenniser les données et les outils (en local comme en international)
 - Trop d'outils inféodés à une personne et à sa conception ou à un organisme (financ.).
 - Problème de transmission, de pérennité et d'accessibilité (données restreintes - monoposte)
 - Evolution des supports, langages, systèmes



Notion du cycle de vie des données et des bases de données en laboratoire

- *Outre, les préoccupations :*

- *de mettre à disposition, le plus rapidement, les données collectées et produites*
- *de garantir leur qualité et leur validation*

***Indispensable d'éviter de perdre les données, ainsi que leur contexte ;
d'anticiper la caducité des outils.***

- *Enjeu immédiat et dans la durée : la pérennité.*

- les données produites sont-elles accessibles aux collègues, à la communauté, mais aussi aux successeurs.*
- Est-ce que les données seront encore accessibles dans 20 ans et quelles valeurs auront-elles ?*

*– Sylvie Damy – Anne Vignot – Jean-Daniel Tissot –
Journée bases de données 21 Mars 2013*

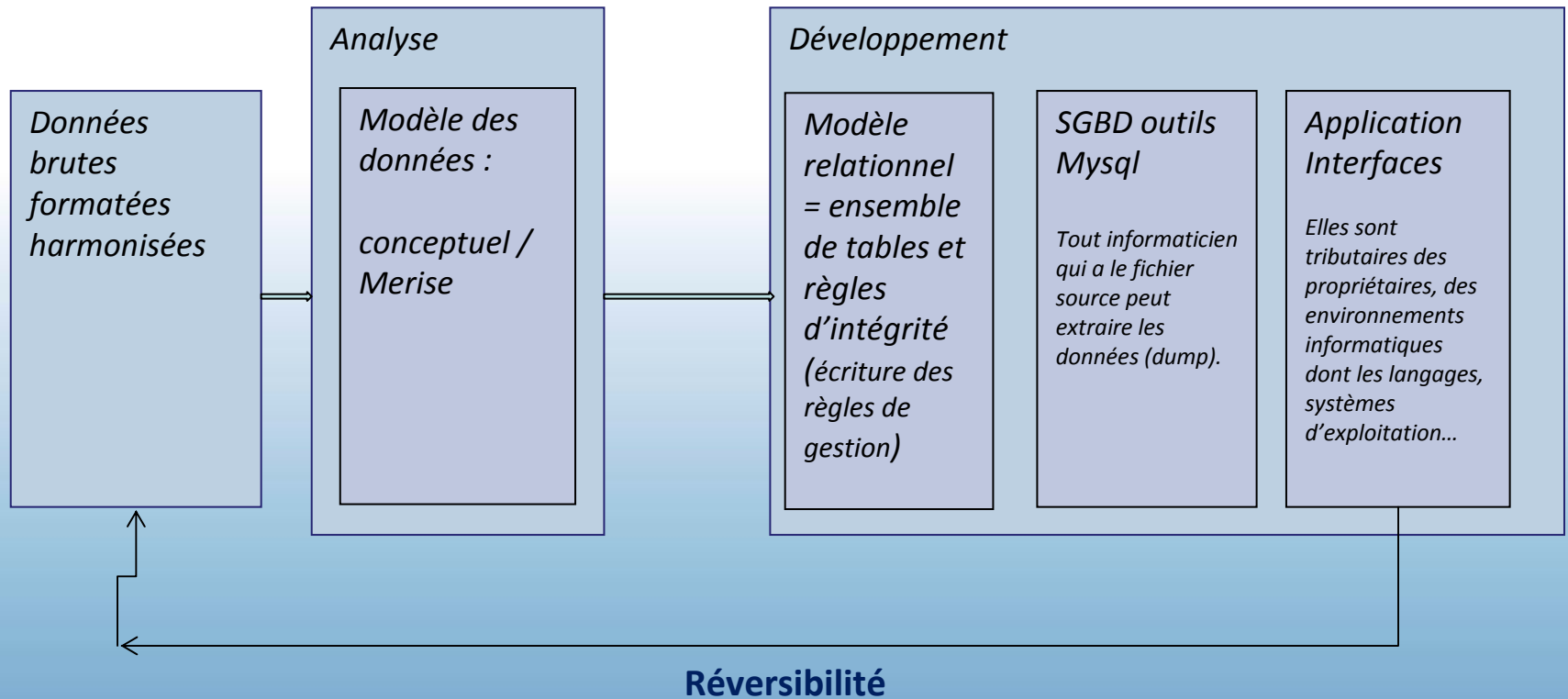
Garant de la pérennité, le principe de l'indépendance

Les données doivent rester indépendantes de l'outil. L'application n'est pas une fin en soi :

- il doit toujours y avoir la possibilité d'extraire l'information brute = format d'exportation standard de la donnée la plus initiale possible = REVERSIBILITE
- de connaître leur organisation = modèle des données
- Elles doivent être accompagnées des métadonnées = informations sur leur cycle de production, leur stockage (format, lieu, etc.), auteurs, ...

Ce qui n'empêche pas de faire les choix techniques qui doivent permettre l'évolutivité

Rappel des étapes pour le développement de l'outil :



- 1- stocker dans la base de données grâce au SGBD (ici relationnel, par exp. Mysql)
- 2- traiter, ou en interne ou en externe, la donnée grâce aux applications (en interne par exp. langage Php et javascript, ajax)
- 3- Environnement multi-développeurs (SVN)

Choix d'un Système d'information

C'est-à-dire un protocole selon la méthode présentée, accompagné d'un environnement technique et humain.

On a opté pour **développer** les outils de gestion, pour éviter les écueils des outils marchands : volume des données, coûts, inadaptation, évolutions, diversité des langages, des formats...

On choisit de garder la **maitrise de l'information**, mais de préparer la donnée à l'exportation (l'importation) vers les bases de données internationales.

On privilégie la saisie par des listes **référentiel**, pour éviter les redondances. Des outils d'auto-implémentations sont opératoires.

On ne réinvente pas les outils existants de traitement de la données (secondaire, tertiaire), on choisit les connexions, et selon le niveau de développement : l'interopérabilité.

Choix d'un Système d'information

Ce système intègre une ou plusieurs bases de données, mais dans un environnement partagé avec un vocabulaire commun...

On met un outil web à disposition pour générer une culture partagée de bases de données.

- Consultable à partir de tout ordinateur
- Consultable de toute plateforme : Windows – Linux – Mac

Importance des interfaces pour une plus grande réussite : plus l'outil est ergonomique et s'inscrit dans une **logique métier**, mieux il sera utilisé.

L'administration des données, du système... est organisée

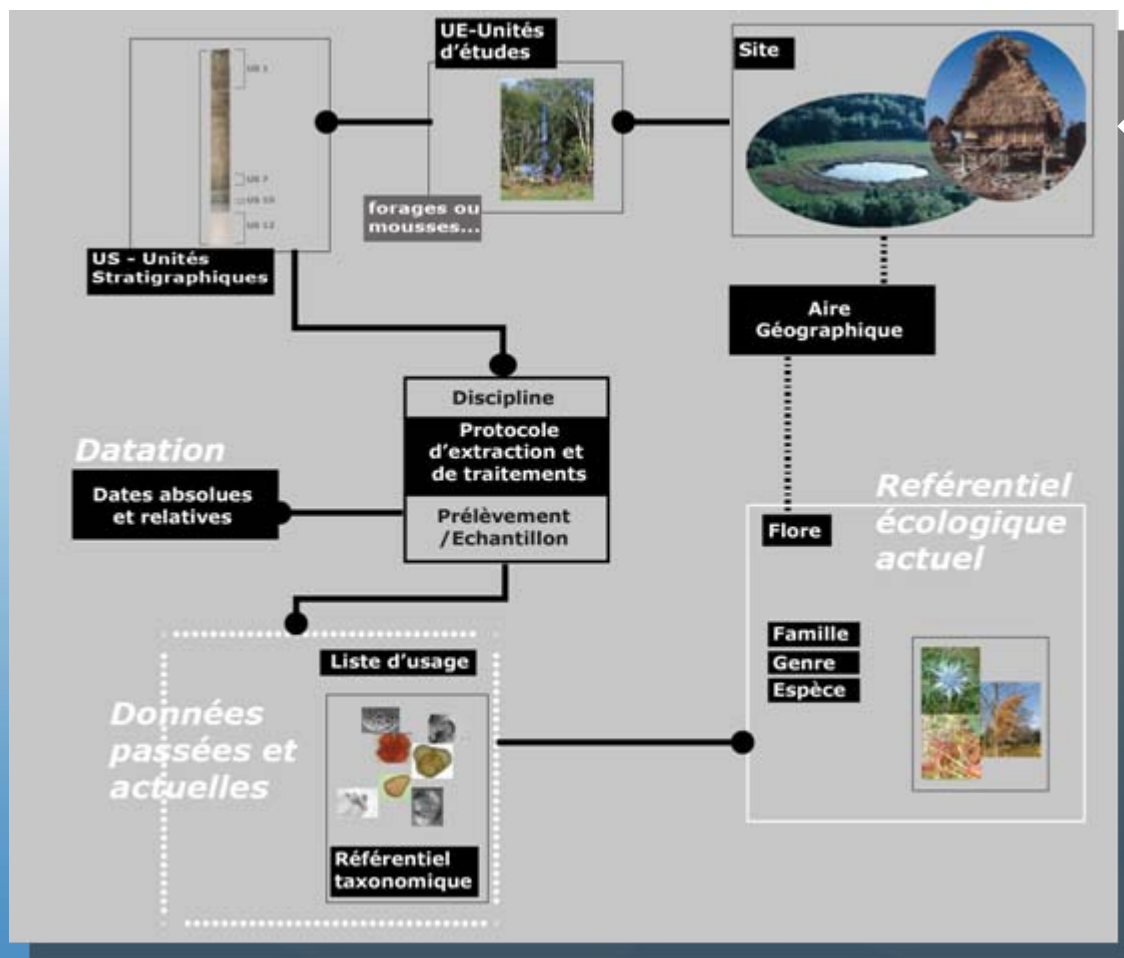
On insère un protocole de **sécurisation** : Code d'accès - Profil utilisateur

- Gestion de traitements « lourds » : transaction pour garantir la sécurité
- Système de sauvegarde par le serveur

Le système intègre des procédures de **validation** : la validation de la donnée est obligatoire (elle passe par deux niveaux : le chercheur = validation scientifique, l'administrateur = validation de la cohérence)

– Sylvie Damy – Anne Vignot – Jean-Daniel Tissot –
Journée bases de données 21 Mars 2013

SI-DEAP : un choix pour un système interne ouvert (??)



Exemple à partir de l'intégration des données polliniques actuelles dans SI-DEAP, système d'information pour la compilation de ces données environnementales (actuelles et passées)



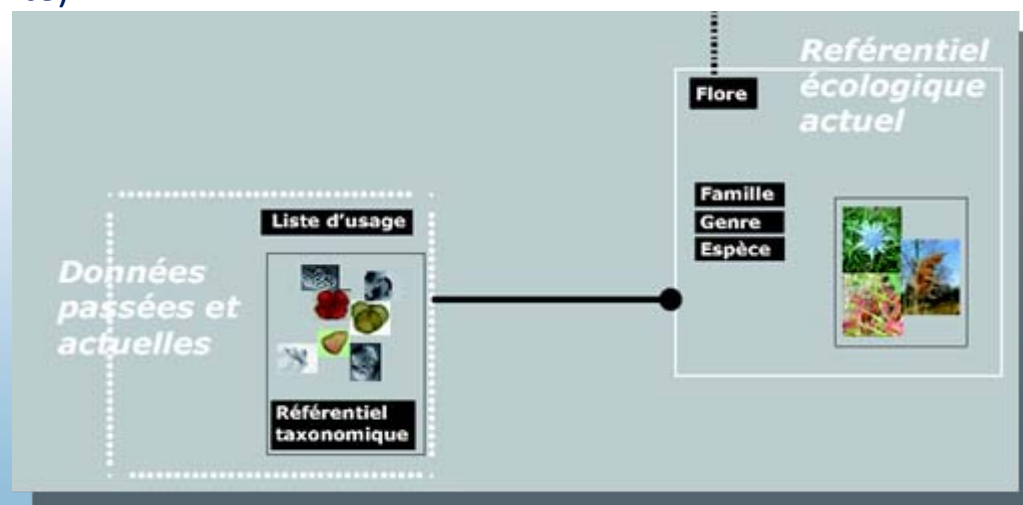






Cohérence du vocabulaire et de la donnée

Le système élaboré doit permettre d'éviter la polysémie et donc le risque de redondance ... Il a fallu définir le vocabulaire du domaine selon un vocabulaire accepté par une communauté scientifique. Nous tendons vers une **ontologie** (étape première pour garantir l'interopérabilité des bases et l'évolutivité).



Pour interdire l'hétérogénéité taxonomique infradisciplinaire et interdisciplinaire, nous avons développé un outil de gestion de la synonymie et homonymie taxonomique plutôt que d'imposer une taxonomie universelle réductrice. Nous avons intégré le caractère évolutif des classifications

– Sylvie Damy – Anne Vignot – Jean-Daniel Tissot –
Journée bases de données 21 Mars 2013

Niveau de précision exigé et exigeant

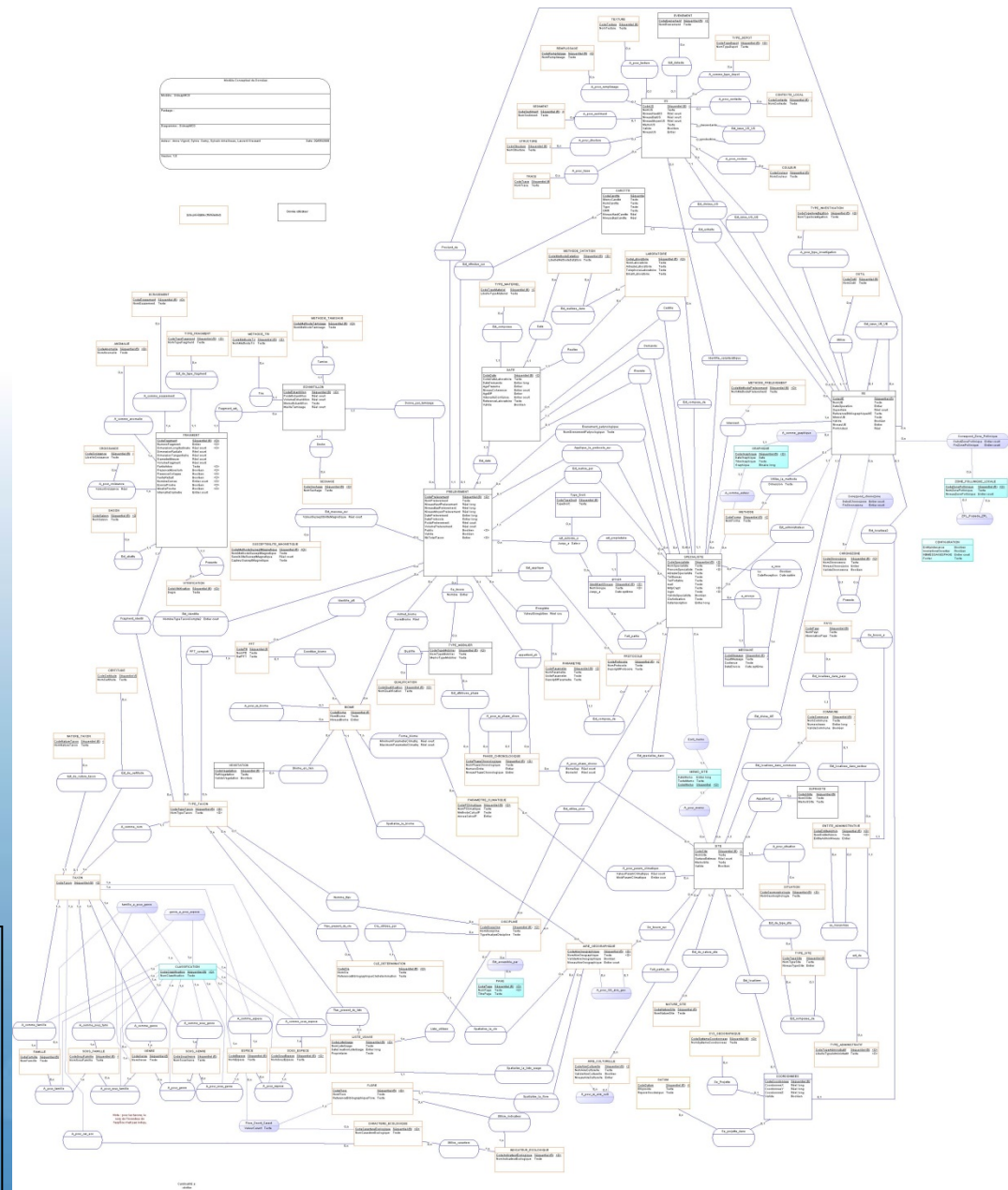
Décrire le domaine par une représentation schématique.

Affiner la description et sa transcription à des **niveaux fins de précision** (granularité) qui garantissent un requête maximisé (multicritère), une qualité de la donnée, sa réversibilité, etc.

Cette finesse permet une **évolutivité** du modèle et donc des outils

Par contre, si le requête est « infini », il est complexe.

Modèle relationnel = 141 tables
Permet des évolutions à tout moment
Dont l'ajout de nouvelles
Disciplines (données) et de nouveaux
protocoles



Les écueils

Nous devons développer des outils/interfaces qui permettent à tout utilisateur de construire des requêtes complexes.

Cela demande que les utilisateurs développent aussi des compétences en requêtage. (idem logiciel R...)

Le temps et les moyens à consacrer à son développement (écriture de code), à la gestion des données.

Quand l'outil doit-il être opérationnel ?

- Utilisation ponctuelle ou au contraire sur le long terme

Obsolescence de l'outil : évolution des langages, des systèmes (SGBD, SE)

Merci

L'application SI_DEAP : Exportation

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Profondeur	4	6	8	10	11	12	13	14	15	16	18	20	22	24	2
2	Poids															
3	Volume															
4	Abies	12	15	17	16	41	46	82	53	31	23	53	30	41	15	2
5	ACER	0	1	0	0	0	0	1	0	3	0	2	2	0	2	
6	AESCLUS	0	1	0	0	0	0	1	0	0	0	0	0	0	0	
7	Alnus	3	1	3	1	3	8	4	5	4	11	11	7	15	3	
8	Betula	21	6	17	6	4	2	3	1	8	12	9	13	7	7	
9	BUXUS	1	1	0	0	0	2	0	2	1	0	1	1	0	0	
10	CALLUNA	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
11	CANNABIS/H	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
12	CARPINUS	13	1	7	4	2	3	0	3	2	1	1	2	0	0	
13	CASTANEA	5	2	2	2	7	4	6	3	4	2	1	5	2	1	
14	CORYLUS	25	13	24	38	24	33	35	32	55	83	116	82	66	82	8
15	EVONYMUS	0	0	0	2	0	0	0	0	0	0	0	0	0	0	
16	FAGUS	19	19	23	10	13	10	11	11	7	4	3	6	3	7	
17	FRAXINUS	10	5	9	9	3	2	1	3	2	1	1	1	6	3	
18	HERERA	0	1	0	0	0	0	0	0	1	0	0	2	0	1	
19	HIPPOPHAE	0	0	0	0	0	0	0	0	0	0	0	1	0	1	
20	JUGLANS	1	1	0	1	2	0	1	1	1	3	2	2	2	0	
21	JUNIPERUS	1	0	3	4	1	1	1	1	1	2	0	3	1	1	
22	PICEA	43	73	58	16	74	54	107	75	23	31	66	29	34	23	
23	PINUS	44	78	78	50	63	77	85	68	65	73	81	79	75	72	
24	PISTACIA	0	1	1	0	0	0	0	0	0	0	0	0	0	0	
25	QUERCUS	33	31	28	38	36	23	28	35	63	56	59	45	41	43	4
26	ROSACEAE	1	0	0	1	0	0	0	0	0	0	0	1	0	1	
27	SALIX	6	2	1	4	4	3	3	4	2	5	1	1	5	1	
28	TILIA	3	2	1	2	1	0	1	2	6	3	14	11	11	3	
29																
30																
31																

	A	B	C	D	E	F
1	Nom du Site	Site de Joux				
2	Code du Site	14				
3	Memo site					
4	Coordonnées	x :	46,866294	y :	6,385746	z :
5						
6	Code UE	13				
7	Nom UE	Joux				
8	MEMO UE					
9	Profondeur	10				
10	code spécialiste	3				
11	Nom spécialiste	DAMY				
12	Prenom spécialiste	Sylvie				
13	Mail spécialiste	sylvie.damy@univ-fcomte.fr				
14	Reference bibliographique					
15	Code liste d'usage					
16	Nom liste d'usage					
17	Code exportateur	3				
18	nom exportateur	DAMY				
19	prenom exportateur	Sylvie				
20	mail exportateur	sylvie.damy@univ-fcomte.fr				
21	nom destinataire	essaiu damy				
22	Date export	17-02-2012				
23	commentaire	Test pour exportation du 17/02/2012 par S. DAMY -- prÃ©sentation OSU				
24						
25	Généré par :	Laboratoire Chrono-Environnement UMR 6249				
26						
27						

Feuille Excel avec deux onglets pour deux types d'informations