

Global Biodiversity Information Facility Switzerland

<http://www.gbif.ch>

Réalisation du nœud informatique GBIF-CH

Situation au 31.08.2005



F. Burri ¹, M. Bouzelboudjen ²

¹ Collaborateur Scientifique, Service Informatique et Télématique de l' Université de Neuchâtel
E-mai. francois.burri@unine.ch, Tel. +41 (0)32 718 20 32, Fax. + 41 (0)32 718 20 00

² Responsable GIS & Formation, Service Informatique et Télématique de l' Université de Neuchâtel &
Chargé de Cours
E-mai. mahmoud.bouzelboudjen@unine.ch, Tel. +41 (0)32 718 20 28, Fax. + 41 (0)32 718 20 00

Nœud informatique GBIF-CH

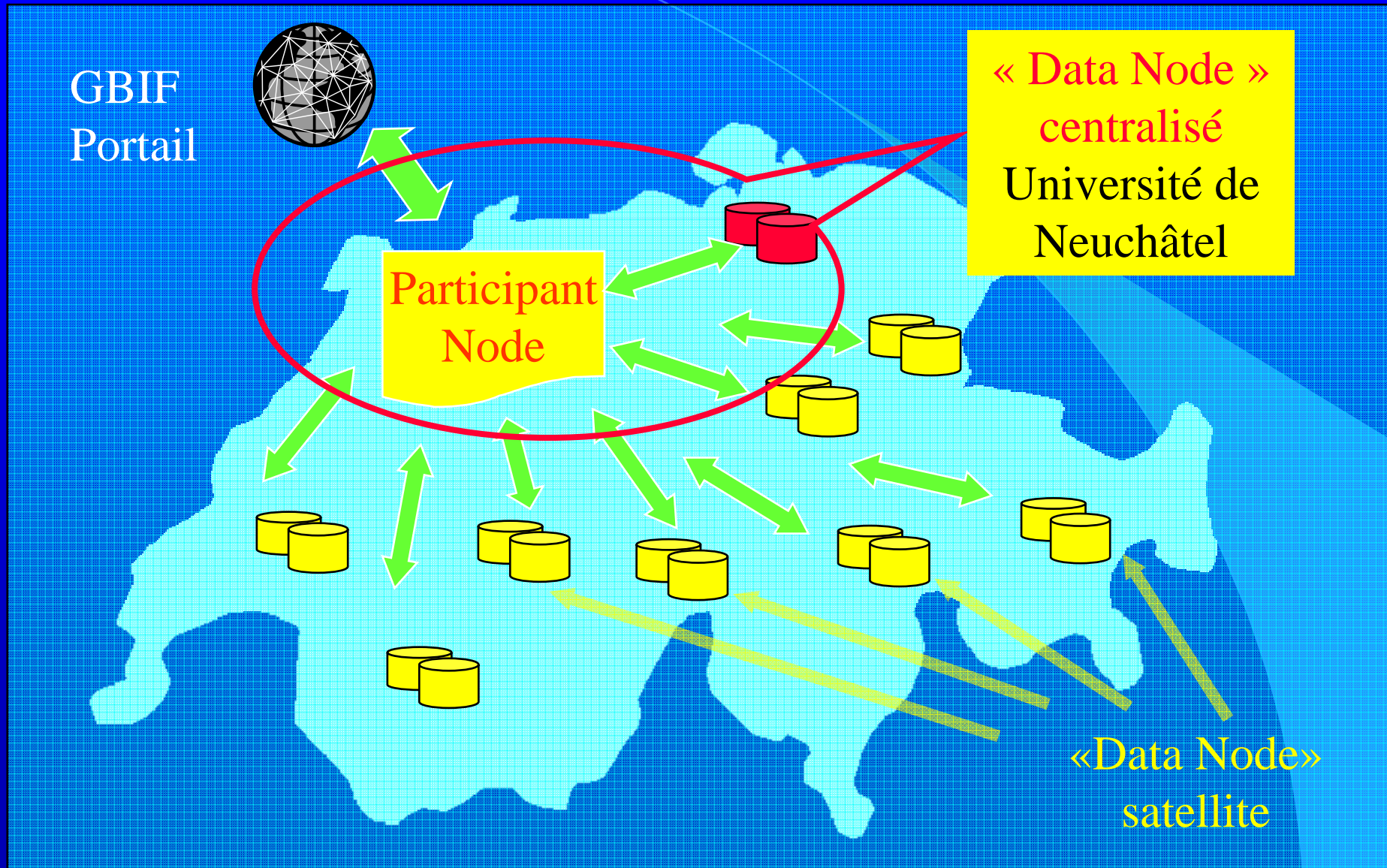
Thèmes abordés

- Rappel succinct de l'objectif et des activités
- Réalisation du portail permettant la publication d'informations en rapport avec GBIF Suisse , <http://www.gbif.ch> (ouverture le 31 mai 2005)
- Installation et configuration de BioCASE (mai-août 2005)
- Etude des problématiques d'intégration de données fortement hétérogènes (juin-juillet 2005) et proposition de concept de validation-normalisation (août 2005)
- Modélisation de la base de données centralisées GBIF-CH et élaboration du concept de validation-normalisation des données (juin-août 2005)

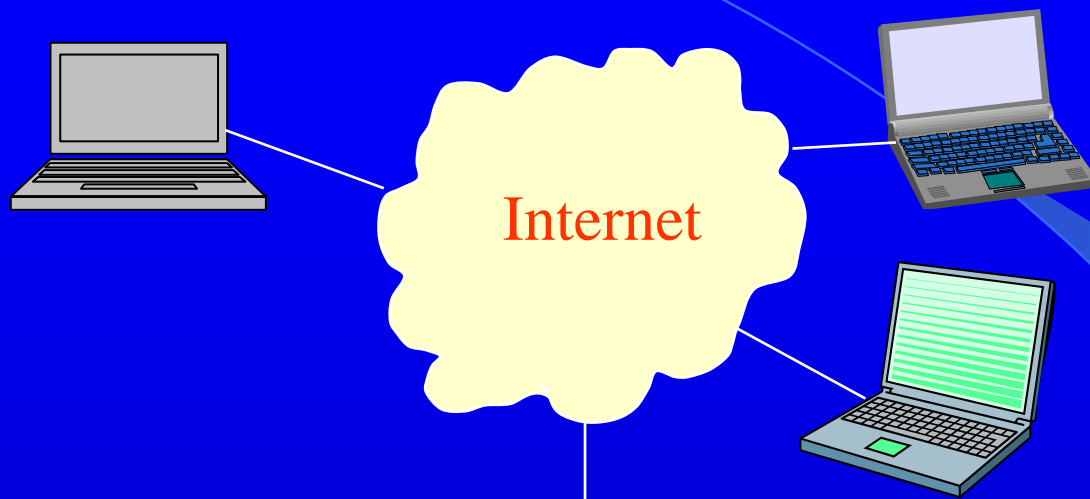
Tâches réalisées

- Rappel succinct de l'objectif et des activités au 1.9.2004
- Réalisation du portail permettant la publication d'informations en rapport avec GBIF Suisse (ouverture en 31 mai 2005)
- Installation et configuration de BioCAsE (mai-août 2005)
- Etude des problématiques d'intégration de données (juin-août 2005) qui a débouché sur la proposition de concept de validation et normalisation (août 2005)
- Modélisation de la base de données centralisées GBIF-CH et élaboration du concept de validation et normalisation des données (juin-août 2005)

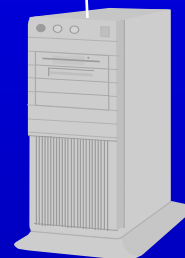
Topologie des «Data Nodes»



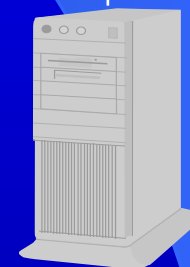
Rappel : Architecture et activités au 1.09.2004 (dernière séance GBIF-CH)



- Installation Hardware : Serveur Linux
- Installation Software : Oracle, DIGIR, Python, Zope, ...
- Intégration de données test du musée d'histoire naturelle de Neuchâtel
- Intégration de données test aux formats texte et Excel
- Publication des données sur Intranet via les outils DIGIR (voir présentation ppt, séance GBIF-CH à Berne, le 01.09.04, www.gbif.ch)



**Serveur
d'applications**



**Serveur de base
de données**

Portail Internet

Data Node centralisé

Evolution du cahier des charges depuis le 1.9.2004

- Nouvelle réorientation du projet par l'abandon des outils DIGIR (comité GBIF-CH)
- Nouvelle demande pour l'intégration des données au format ABCD (BioCAsE) pour permettre l'intégration d'images

Tâches réalisées

- Rappel succinct de l'objectif et des activités au 1.9.2004
- Réalisation du portail permettant la publication d'informations en rapport avec GBIF Suisse (ouverture en 31 mai 2005)
- Installation et configuration de BioCAsE (mai-août 2005)
- Etude des problématiques d'intégration de données (juin-août 2005) qui a débouché sur la proposition de concept de validation et normalisation (août 2005)
- Modélisation de la base de données centralisées GBIF-CH et élaboration du concept de validation et normalisation des données (juin-août 2005)

Publication d'informations en rapport avec GBIF suisse (sous CMS Jahia)

<http://www.gbif.ch>

The screenshot shows the GBIF.ch website with a dark blue header. The logo 'GBiF.ch' is prominent on the left. Navigation links for 'plan du site' and 'Contacts GBIF Suisse' are on the right. The main content area is titled 'Global Biodiversity Information Facility Switzerland' and features a central image of bird specimens and a book. A sidebar on the left contains menu items like 'GBIF International', 'GBIF Suisse', and 'Noeud informatique GBIF Suisse'. A right sidebar lists 'Actualités' with dates and 'Questions fréquentes'. The footer includes contact information and an 'authentification' link.

Deutsch | English

plan du site | Contacts GBIF Suisse

Global Biodiversity Information Facility Switzerland

► GBIF International
► GBIF Suisse
► Noeud informatique GBIF Suisse

► Accès aux données
Site GBIF International
Site GBIF Suisse -en construction-

► Téléchargement de documents
Appel à la soumission de projets
Exposés et rapports

► Glossaire

► Page d'accueil GBIF.ch

Global Biodiversity Information Facility Switzerland

L'Université de Neuchâtel réalise le noeud informatique

► Actualités
31 mai 2005 : Mise en ligne du site GBIF Suisse
Appel à la soumission de projets 2005-2007

► Agenda
31 août 2005 : séance de la commission scientifique GBIF Suisse

► Questions fréquentes

► Liens

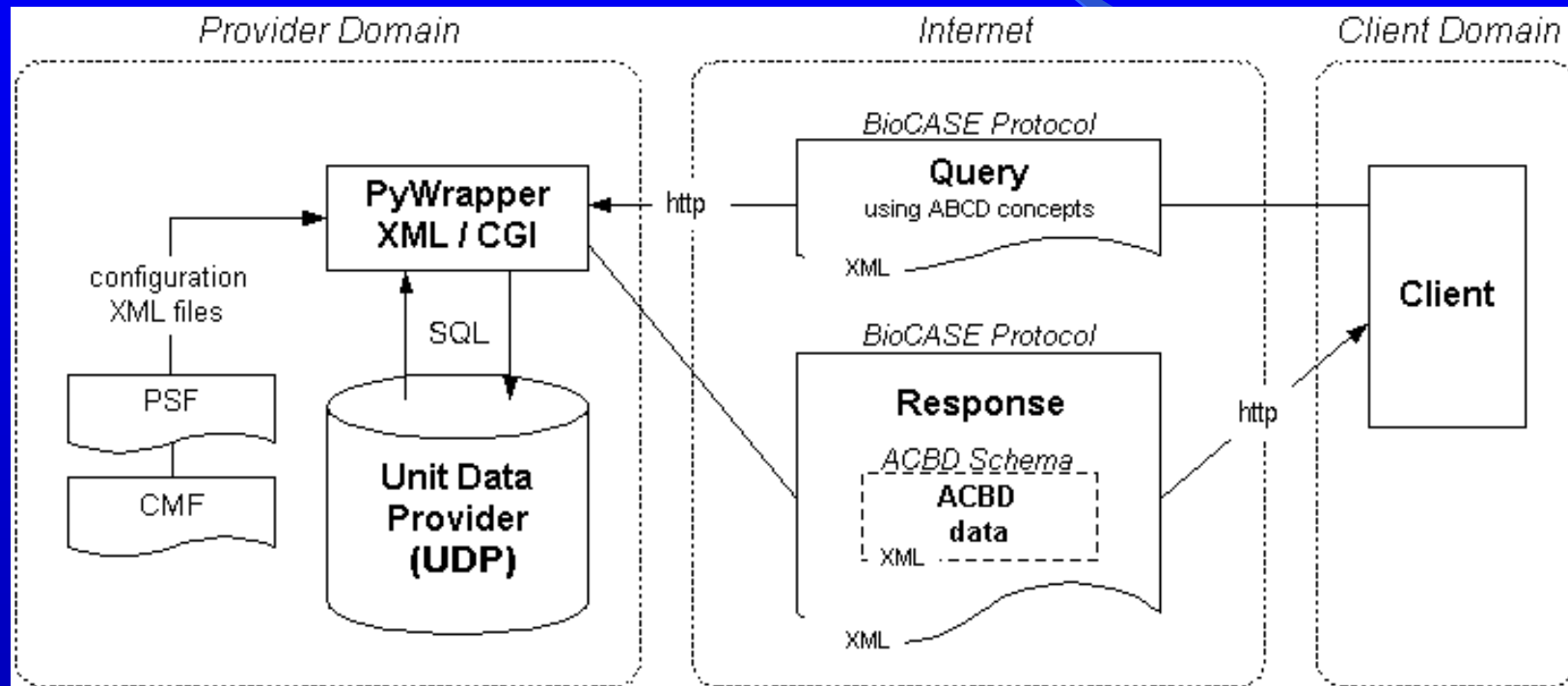
Global Biodiversity Information Facility (GBIF) Switzerland - contact.gbif@unine.ch
webmaster.gbif@unine.ch

authentification

Tâches réalisées

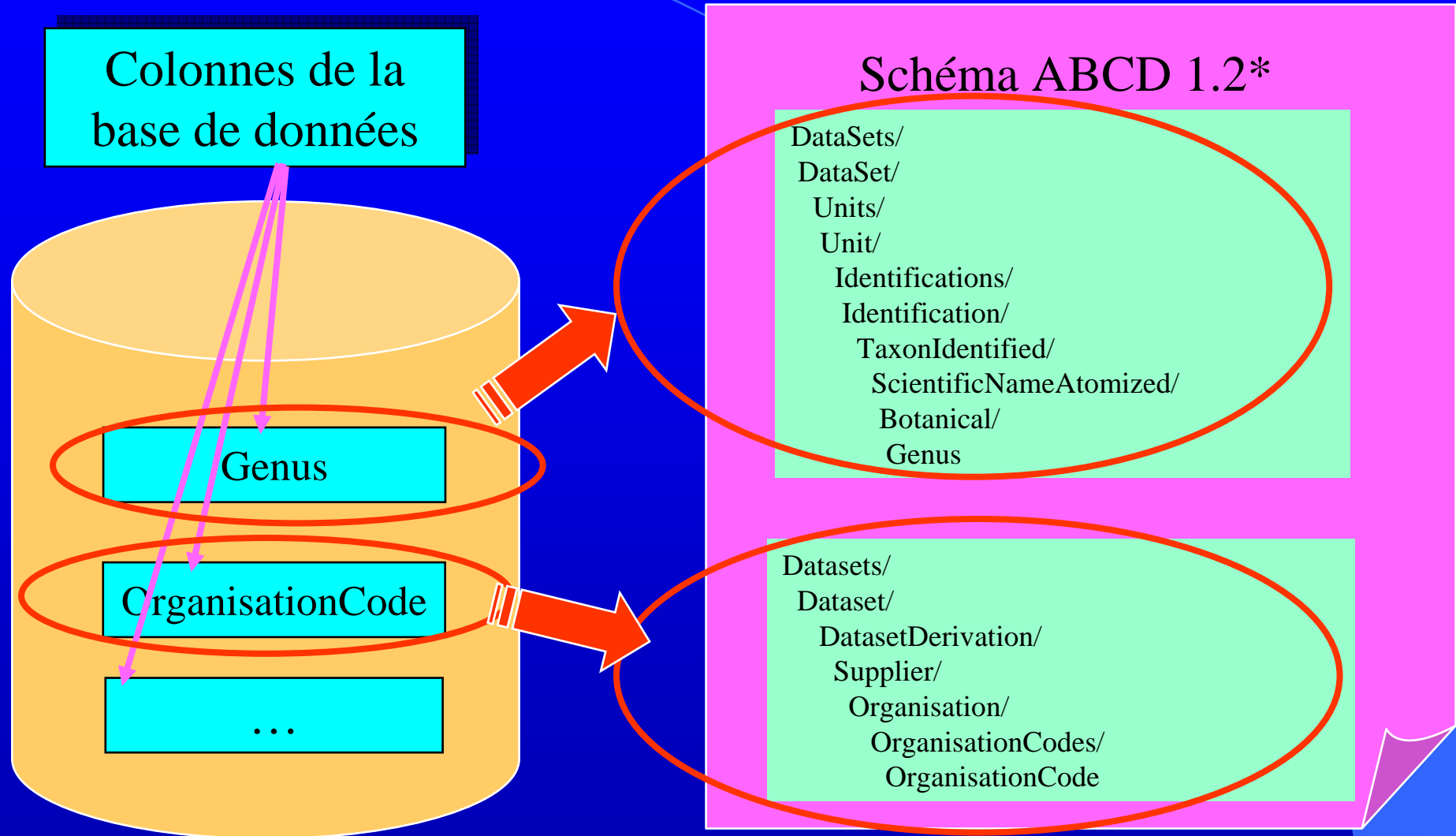
- Rappel succinct de l'objectif et des activités au 1.9.2004
- Réalisation du portail permettant la publication d'informations en rapport avec GBIF Suisse (ouverture en 31 mai 2005)
- **Installation et configuration de BioCAsE (mai-août 2005)**
- Etude des problématiques d'intégration de données (juin-août 2005) qui a débouché sur la proposition de concept de validation et normalisation (août 2005)
- Modélisation de la base de données centralisées GBIF-CH et élaboration du concept de validation et normalisation des données (juin-août 2005)

Concept BioCAsE *



* <http://www.biocase.org>

Mapping entre les « data » et le modèle ABCD



(*)<http://www.bgbm.org/scripts/ASP/TDWG/frame.asp>

Interface de « mapping » de BioCAsE

The screenshot shows a web browser window titled "Local Config Tool - Microsoft Internet Explorer" displaying the BioCAsE Mapping Editor. The browser's address bar shows the URL: `http://norma.unine.ch/biocase/configtool/mapping.cgi?dsa=gbifdat1&id=568062386134022&schema=cmf_ABCD_1.20.xml&action=revert`. The main content area displays a tree view of data sets and their mappings. A modal dialog box titled "Mapping Editor" is open, showing a form for defining a mapping. The form includes fields for "Literal1", "DB Attr1" (with a dropdown menu showing "V_GBIFCH" and "PERSONNAME" and a "text" type selector), "Literal2", and "DB Attr2" (with a dropdown menu showing "--select--" and a "text" type selector). The background tree view shows a hierarchy of data sets including "DateSupplied", "OrganisationName", "PersonName", "SourceInstitutionCode", "SourceLastUpdatedDate", "SourceName", "CountryName", "LatitudeDecimal", and "LongitudeDecimal". Each data set has a "Mapping1" field with a value and a type, and a collapse/expand icon.

Exemple d'interrogation de la base USNEA (Standard Biocase Query Tool)

Local Query Tool - Microsoft Internet Explorer

Fichier Edition Affichage Favoris Outils ?

Précédente Rechercher Favoris OK Liens >>

Adresse <http://norma.unine.ch/biocase/querytool/main.cgi?dsa=gbifdat1>

Local Query Tool *gbifdat1*

A simple user interface to query your local pywrapped databases.

Start > QueryTool > QueryForm

★ **Your datasource wrapper:** <http://norma.unine.ch/biocase/pywrapper.cgi?dsa=gbifdat1>

★ **Query the datasource:**
Use the asterisk (*) as a wild card (e.g. Pli* will find all names starting with Pli, Plinia cubensis * will find the names of subspecies etc.).

Scientific Name:	<input type="text" value="%n%"/>
Higher Taxon:	<input type="text"/>
Genus:	<input type="text"/>
Country:	<input type="text"/>
Gathering Date ISO:	< <input type="text"/>
Latitude:	< <input type="text"/>
Longitude:	< <input type="text"/>

★ **Group result by:**
Scientific Name

★ **Other available data standards:**
ABCD 1.20

Debugging:
wrapper URL: <http://norma.unine.ch/biocase/pywrapper.cgi?dsa=gbifdat1>

Terminé Internet

Résultat de l'interrogation

Local Query Tool - Microsoft Internet Explorer

Fichier Edition Affichage Favoris Outils ?

Précédente Rechercher Favoris

Adresse <http://norma.unine.ch/biocese/querytool/grouping.cgi> OK Liens

Local Query Tool *gbifdat1*

A simple user interface to query your local pywrapped databases.

Start » QueryTool » GroupingList

Scientific Name~%n%

★ **Your search retrieved 18 hits:**

Follow a link to see the list of individual records found.

- Usnea - cavernosa - Tuck.
- Usnea - ceratina - Ach.
- Usnea - cornuta - Korb.
- Usnea - diplotypus - Vain.
- Usnea - filipendula - Stirt.
- Usnea - florida - (L.) F. H. Wigg.
- Usnea - fulvovireagens - (Rasanen) Rasanen
- Usnea - glabrata - (Ach.) Vain.
- Usnea - glabrescens - (Vain.) Vain.
- Usnea - hirta - (L.) F. H. Wigg.
- Usnea - intermedia - (A.Massal.) Jatta
- Usnea - lapponica - Vain.
- Usnea - longissima - Ach.
- Usnea - scabrata - Nyl.
- Usnea - silesiaca - Motyka
- Usnea - subfloridana - Stirt.
- Usnea - substerilis - Motyka
- Usnea - wasmuthii - Rasanen

★ **Group result again by:**

You can group this result again by any another concept below.

---None---

Diagnostics:

Internet

Détail Projet Usnea – cavernosa –Tuck

Local Query Tool - Microsoft Internet Explorer

Fichier Edition Affichage Favoris Outils ?

Précédente Recherche Favoris OK Liens »

Adresse <http://norma.unine.ch/biocase/querytool/recordlist.cgi#CJB-CJB/G>

- CJB/G (CJB) 1 unit(s)
- CJB/G (CJB) 1 unit(s)
- CJB/G (CJB) 1 unit(s)
- CJB/G (CJB) 1 unit(s)

Click on UnitId to get details. [Create Map!](#) |

CJB/G (CJB)
Last update: 2005-08-30|

UnitId	Preferred Name(s)	Country
CJB/G/20236	Usnea - cavernosa - Tuck.	Switzerland

CJB/G (CJB)
Last update: 2005-08-30|

UnitId	Preferred Name(s)	Country
CJB/G/20237	Usnea - cavernosa - Tuck.	Switzerland

CJB/G (CJB)
Last update: 2005-08-30|

UnitId	Preferred Name(s)	Country
CJB/G/20238	Usnea - cavernosa - Tuck.	Switzerland

[Create Map!](#)

CJB/G (CJB)
Last update: 2005-08-30|

UnitId	Preferred Name(s)	Country
CJB/G/20239	Usnea - cavernosa - Tuck.	Switzerland

[Create Map!](#)

CJB/G (CJB)

Internet

Tâches réalisées

- Rappel succinct de l'objectif et des activités au 1.9.2004
- Réalisation du portail permettant la publication d'informations en rapport avec GBIF Suisse (ouverture en 31 mai 2005)
- Installation et configuration de BioCAsE (mai-août 2005)
- Etude des problématiques d'intégration de données (juin-août 2005) qui a débouché sur la proposition de concept de validation et normalisation (août 2005)
- Modélisation de la base de données centralisées GBIF-CH et élaboration du concept de validation et normalisation des données (juin-août 2005)

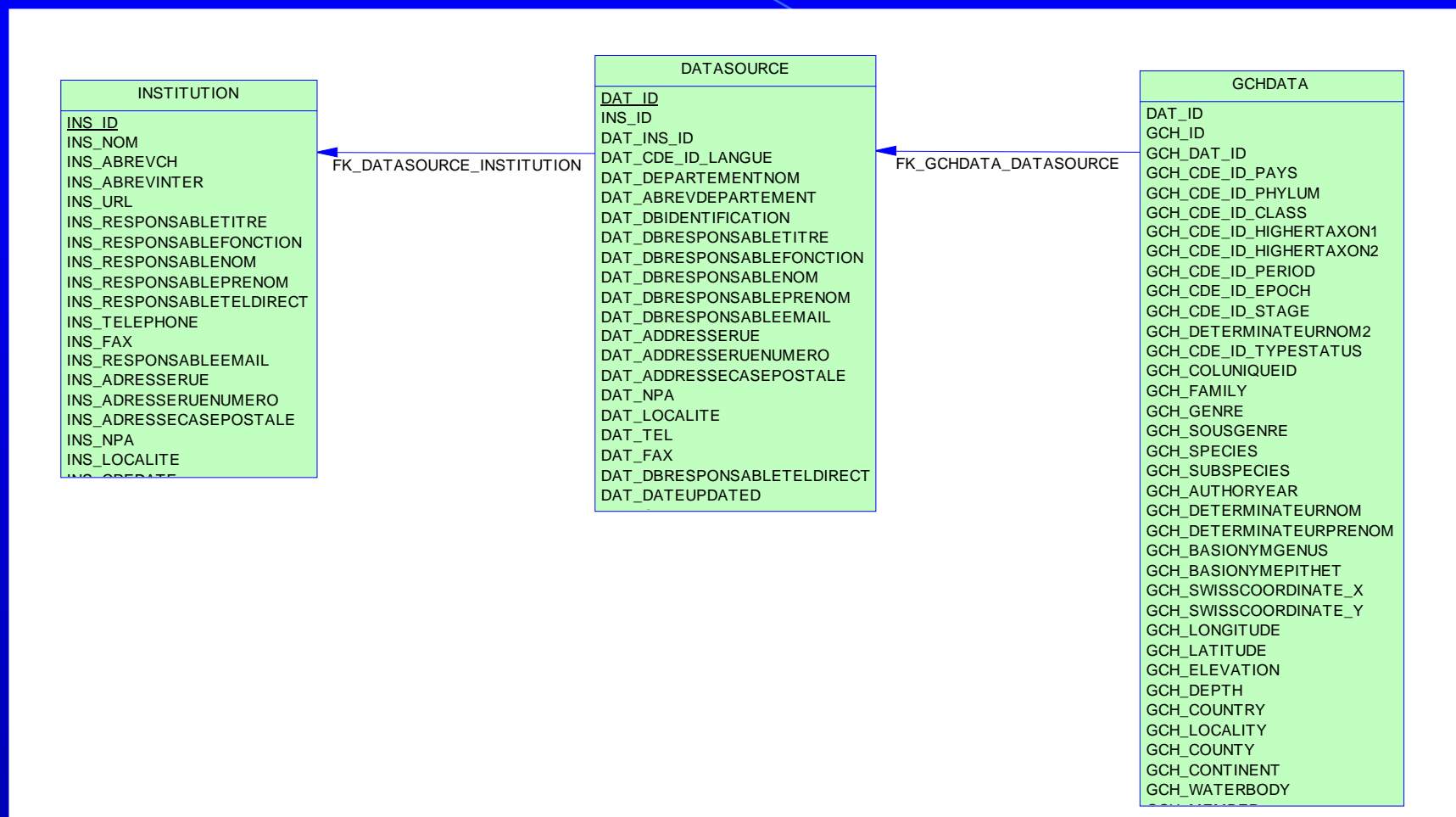
Etude des problématiques d'intégration de données

- Analyse des extraits de base de données à disposition :
 - DatenBank Sammlung Pinuz, Paläontologisches Institut und Museum der Universität Zürich (reçue en juin 2005)
 - Base de données du département de géologie et paléontologie, musée d'histoire naturelle de Genève (reçue le 16 août 2005)
 - Base de données du projet USNEA du Conservatoire et Jardin botanique de Genève (reçue le 25 août 2005)
 - ...
- Identification des problèmes :
 - Formats hétérogènes de données à description variable
 - Pas de normalisation
 - Police de caractères
 - ...

Tâches réalisées

- Réalisation du portail permettant la publication d'informations en rapport avec GBIF Suisse (ouverture en 31 mai 2005)
- Installation et configuration de BioCAsE (mai-août 2005)
- Etude des problématiques d'intégration de données (juin-août 2005) qui a débouché sur la proposition de concept de validation et normalisation (août 2005)
- Modélisation de la base de données centralisées GBIF-CH et proposition d'un concept de validation - normalisation des données (juin-août 2005)

Aperçu simplifié du modèle GBIF-CH *



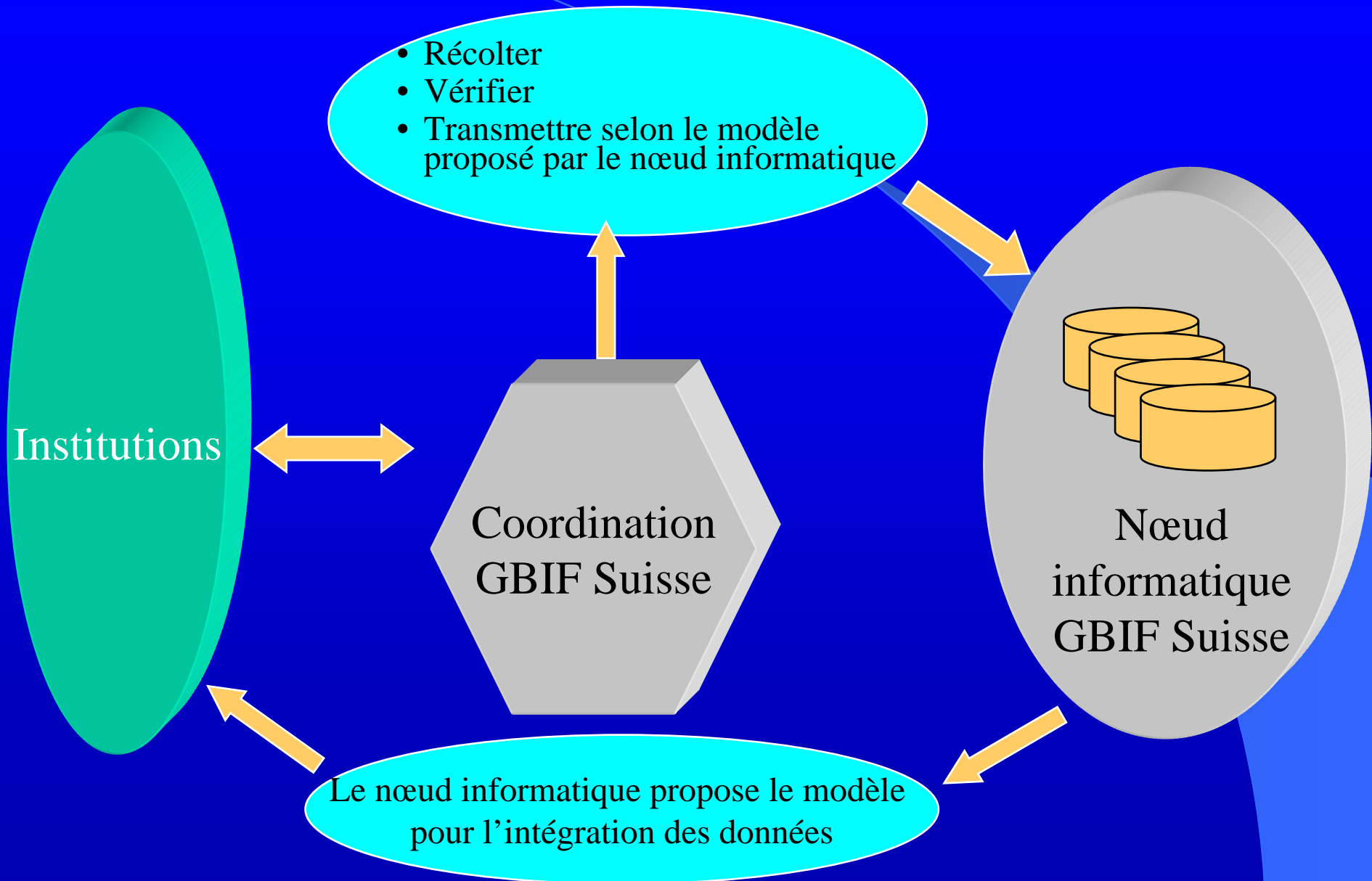
* Extrait de 3 tables du modèle élaboré

Les 51 colonnes de la table GCHDATA

1. GCH_COLUNIQUEID
2. GCH_CDE_ID_PAYS
3. GCH_CDE_ID_PHYLUM
4. GCH_CDE_ID_CLASS
5. GCH_CDE_ID_ORDER
6. GCH_FAMILY
7. GCH_CDE_ID_HIGHTAXON1
8. GCH_CDE_ID_HIGHTAXON2
9. GCH_GENRE
10. GCH_SOUSGENRE
11. GCH_SPECIES
12. GCH_SUBSPECIES
13. GCH_AUTHORYEAR
14. GCH_DETERMINATEURNOM
15. GCH_DETERMINATEURPRENOM
16. GCH_DETERMINATEUR
17. GCH_CDE_ID_TYPESTATUS
18. GCH_BASIONYMGENUS
19. GCH_BASIONYMEPITHET
20. GCH_ORIGINALLABEL
21. GCH_ORIGINALLABELCONTENT
22. GCH_SWISSCOORDINATE_X
23. GCH_SWISSCOORDINATE_Y
24. GCH_LONGITUDE
25. GCH_LATITUDE
26. GCH_ELEVATION
27. GCH_DEPTH

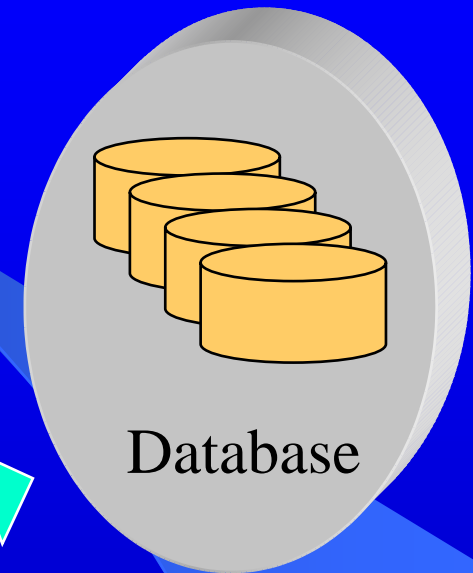
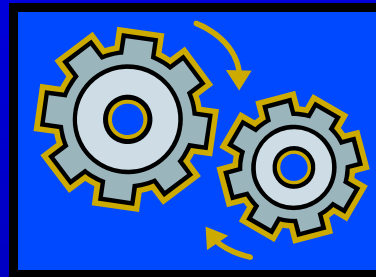
28. GCH_COUNTRY
29. GCH_LOCALITY
30. GCH_COUNTY
31. GCH_CONTINENT
32. GCH_WATERBOBY
33. GCH_CDE_ID_PERIOD
34. GCH_CDE_ID_EPOCH
35. GCH_CDE_ID_STAGE
36. GCH_MEMBER
37. GCH_FORMATION
38. GCH_ZONE
39. GCH_INDIVIDUALCOUNT
40. GCH_MALECOUNT
41. GCH_FEMALECOUNT
42. GCH_ADULTECOUNT
43. GCH_LARVECOUNT
44. GCH_NYMPHECOUNT
45. GCH_ANATOMICALDESCR
46. GCH_COLLECTOR
47. GCH_DAYCOLLECTED
48. GCH_MONTHCOLLECTED
49. GCH_YEARCOLLECTED
50. GCH_REFBI
51. GCH_DATELASTMODIFIED

Processus d'intégration de données



Proposition du nœud informatique pour le concept de validation et de normalisation *

	A	B	C	D	E	F	G	H	I
1	DEPT	NO_INVENTA	PHYLUM	CLASSE	ORDRE	FAMILLE	GENRE	SOUS_GENRE	ESPECE
2	GEPI	20000	Mollusca	Cephalopoda			Berriasella		pc
3	GEPI	200001	Mollusca	Cephalopoda					
4	GEPI	200002	Mollusca	Cephalopoda			Neocomites		neocomiensis
5	GEPI	200003	Mollusca	Cephalopoda			Acanthodiscus		malbosi
6	GEPI	200004	Mollusca	Cephalopoda			Berriasella		aff. callisto
7	GEPI	200005	Mollusca	Cephalopoda			Thurmannites		thumanni
8	GEPI	200006	Mollusca	Cephalopoda			Berriasella		cf. pontica
9	GEPI	200007	Mollusca	Cephalopoda			Berriasella		pontica
10	GEPI	200008	Mollusca	Cephalopoda			Berriasella		cf. callisto
11	GEPI	200009	Mollusca	Cephalopoda			Berriasella		aff. callisto
12	GEPI	200010	Mollusca	Cephalopoda			Neocomites		neocomiensis
13	GEPI	200011	Mollusca	Cephalopoda			Perisphinctes		aff. ponticus
14	GEPI	200012	Mollusca	Cephalopoda			Perisphinctes		aff. tiziani
15	GEPI	200013	Mollusca	Cephalopoda			Ammonites		hollandrei
16	GEPI	200014	Mollusca	Cephalopoda			Ammonites		complanatus
17	GEPI	200015	Mollusca	Cephalopoda			Oppelia		flexuosa
18	GEPI	200016	Mollusca	Cephalopoda			Berriasella		aff. pontica
19	GEPI	200017	Mollusca	Cephalopoda			Ochetoceras		gr. canaliculata
20	GEPI	200018	Mollusca	Cephalopoda			Berriasella		aff. pontica
21	GEPI	200019	Mollusca	Cephalopoda			Berriasella		privasensis
22	GEPI	200020	Mollusca	Cephalopoda			Berriasella		
23	GEPI	200021	Mollusca	Cephalopoda			Berriasella		aff. callisto
24	GEPI	200022	Mollusca	Cephalopoda			Berriasella		pontica
25	GEPI	200023	Mollusca	Cephalopoda			Neocomites		occitanicus
26	GEPI	200024	Mollusca	Cephalopoda			Berriasella		carpathica ?
27	GEPI	200025	Mollusca						



Database

OK

Erreurs

Données exportées

Validation,
normalisation

Rapport

* Ne pourra être réalisé ultérieurement qu'en fonction des premières expériences d'intégration des données

Proposition du nœud informatique pour la mise en œuvre du concept de validation automatique

- Vérifications automatiques de règles prédéfinies :
 - Colonne obligatoirement renseignée
 - Valeur appartenant à un thésaurus
 - Unicité
 - ...
- Une liste des enregistrements rejetés est produite pendant le processus
- Des colonnes peuvent être rejetées sans que l'enregistrement dans son entier soit rejeté
- Une liste est produite avec les enregistrements qui possède une ou plusieurs colonnes rejetées
- ...

Proposition du nœud informatique pour la mise en œuvre du concept de normalisation

- Le processus de normalisation vise à attribuer des valeurs normalisées (codes) à certaines colonnes :
 - Stratigraphie
 - BotanicalHigherRankTaxon
 - Pays
 - Systématique
 - TypestatusCode
- La normalisation permet de :
 - Fédérer les données saisies dans des langages différents
 - Étendre les possibilités de corrélation entre les différentes sources de données
 - De renseigner automatiquement les niveaux supérieures lacunaires des valeurs hiérarchiques (Stratigraphie, Systématique, BotanicalHigherRankTaxon)
 - Valider les données

Prochaines étapes

- Finalisation du modèle de « mapping » de la base de données GBIF-CH avec les concepts (août-septembre 2005) : ABCD 2.05, ABCD 1.2 et Darwin core 2.0
- Intégration des données USNEA selon les concepts définis (septembre-décembre 2005)
- Enregistrement de la base de données centralisées GBIF-CH sur GBIF.ORG (dès que des données seront correctement intégrées)
- Développement des outils logiciels pour la validation et la normalisation des données (calendrier à définir)
- Développement d'outils logiciel (interfaces orientées web) pour l'interrogation des données du portail Suisse (2006-2007)

Références

- **Javier Torre**, *BioCAsE Provider Software v. 2.2.0 Installation Guide*, 18.07.2005
- **Javier Torre, Markus Doering**, *BioCAsE Provider Software v. 2.2.0 Configuration Guide*, 25.08.2005
- **François Burri, Mahmoud Bouzelboudjen**, *Architecture du Participant Node suisse - Situation au 1.9.2004*, Berne, Suisse. Présentation disponible sous <http://www.gbif.ch>
- **François Burri, Mahmoud Bouzelboudjen**, *Architecture du Participant Node suisse - Situation au 28.1.2004*, Berne, Suisse. Présentation disponible sous <http://www.gbif.ch>