

Les Web-Services

Application au domaine de la bioinformatique

<http://lipm-bioinfo.toulouse.inra.fr/formations>

Sébastien Carrere
Sebastien.Carrere@toulouse.inra.fr



Plan

I. Généralités

1. Définitions
2. Motivations
3. Vue globale

II. Technologies

1. XML
2. SOAP
3. WSDL
4. UDDI

IV. Limites

IV. Web-services et bioinformatique ?

1. Intérêt
2. BioMOBY
3. PlayMOBY

V. TP

I. Généralités

I.1. Définitions

- Les web services permettent l'invocation de fonctions **distantes**, présentes sur des systèmes **distribués** et **hétérogènes**, grâce au **protocole** HTTP et à XML.
- « les services web sont des applications auto-descriptives, modulaires et **faiblement couplées** qui fournissent un modèle de programmation et de déploiement d'applications, basé sur des **normes**, et s'exécutant au travers de l'infrastructure web » Les Web Services, H.Kadima, Dunod 2003
- Un service web est une application conçue pour assurer une **interopérabilité** entre machines au travers d'un **réseau**.
- Un web service est une **interface** qui décrit un ensemble d'opérations accessibles via un réseau par des messages **XML** standards
- « Un service web est un composant applicatif mis à la disposition sur un réseau et disposant de méthodes que l'on peut invoquer à distance via l'emploi de protocoles **standards**. Les services Web présentent l'avantage d'être **faiblement couplés**, indépendants des plateformes et réutilisables » Livre Blanc : Les Services Web, S.Bardet, 2003

I.2. Motivations

Pourquoi faire ?

- Le Business to Consumer (B2C)
 - Offrir des ressources au(x) client(s) de façon normalisée
- Business to Business (B2B)
 - Communication inter entreprises / inter labo
- Intra entreprise (A2A)
 - Intégration d'applications

I.2. Motivations

Besoins ?

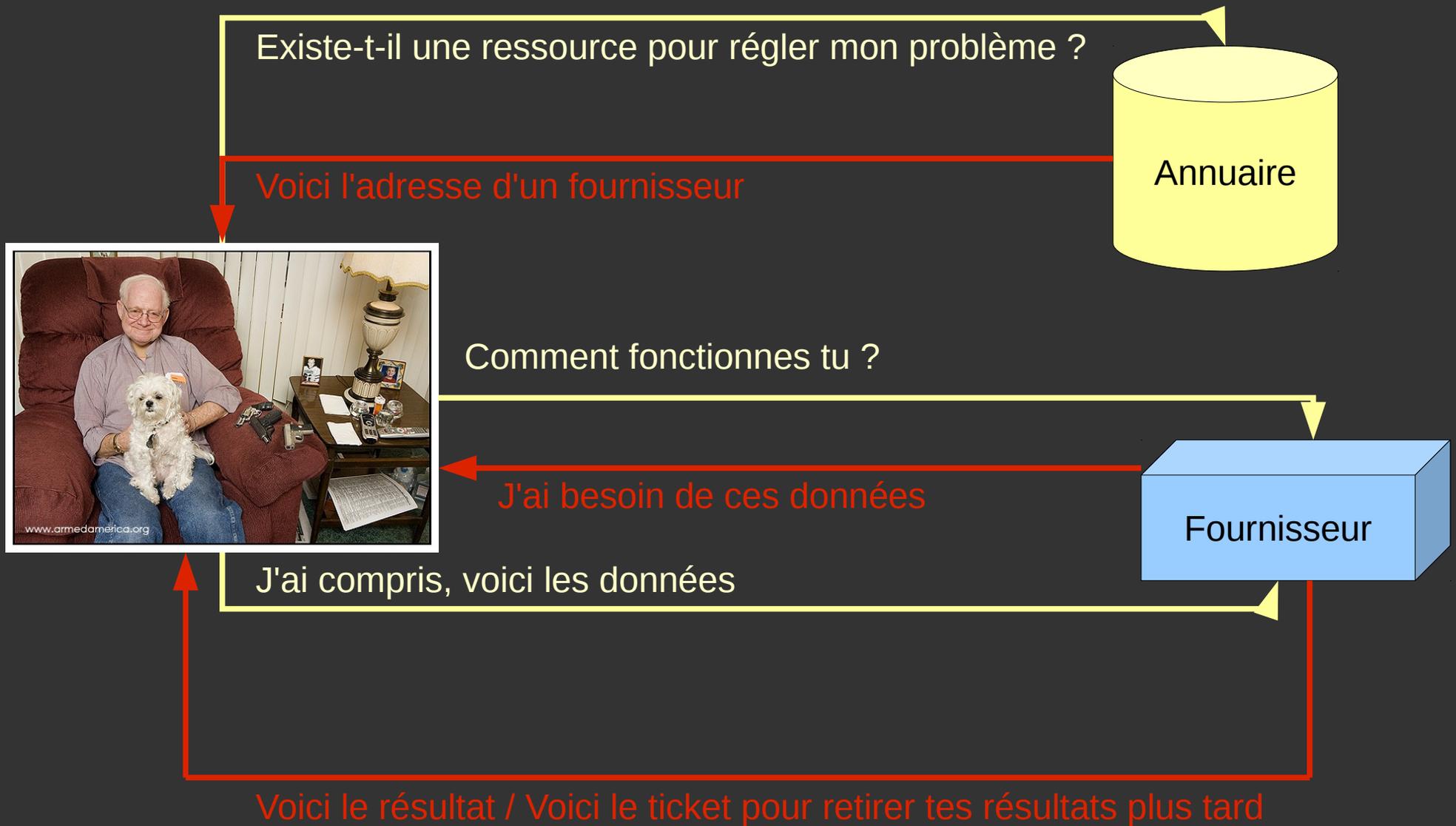
- Flexibilité et indépendance
 - La possibilité de réorganiser le système de façon rapide, efficace et peu chère
- Partage d'informations/d'applications
 - Entre partenaires (communauté bioinfo), clients ou services
- Interopérabilité
 - Tout le monde n'a pas la même organisation, les mêmes systèmes
- Sécurité et confidentialité
 - Comment partager des applications si tout le monde est protégé par un pare-feu ?
 - Garantie de livraison

I.2. Motivations

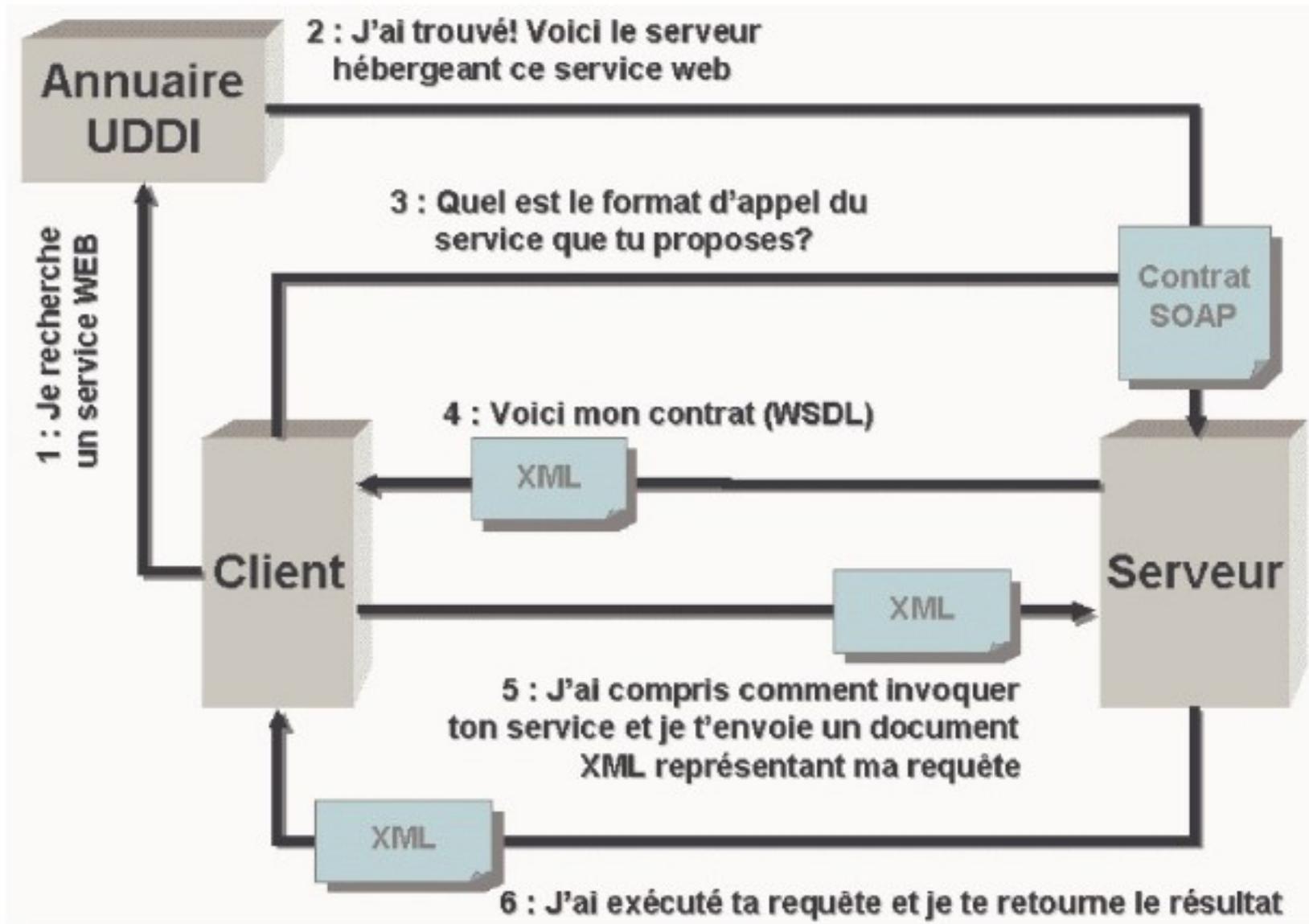
Intérêt des web services

- Technologies acceptées par un grand nombre de fournisseurs de logiciels et d'organismes (Microsoft, OMG, W3C, OASIS)
- XML est utilisé pour l'échange des données et messages
- Ils permettent l'intégration de plate-formes hétérogènes via le protocole HTTP
- Multi-langage : Java, C, C++, Perl, Python, C# ...
- Non intrusif
- Intégration faible (services atomiques)
- Permet la localisation et l'invocation dynamique d'autres services à partir de registres privés ou publics

I.3. Vue Globale Scénario



I.3 Vue Globale Technos



II. Technologies

II.1 XML

Extensible Markup Language

- Langage balisé
- Structure hiérarchique
- Norme W3C
- Description embarquée ou associée (DTD, Schéma)
- Portable (texte)
- Transformable

II.1 XML

Intérêt pour les WS

- Séparation de la structure , des données et de la représentation
 - DTD (pas XML, peu de contraintes)
 - Schema (XML)
 - Namespaces : teinter les balises
 - « Facilement » analysable par une machine
- Favorise l'émergence de standards
 - Réutilisabilité

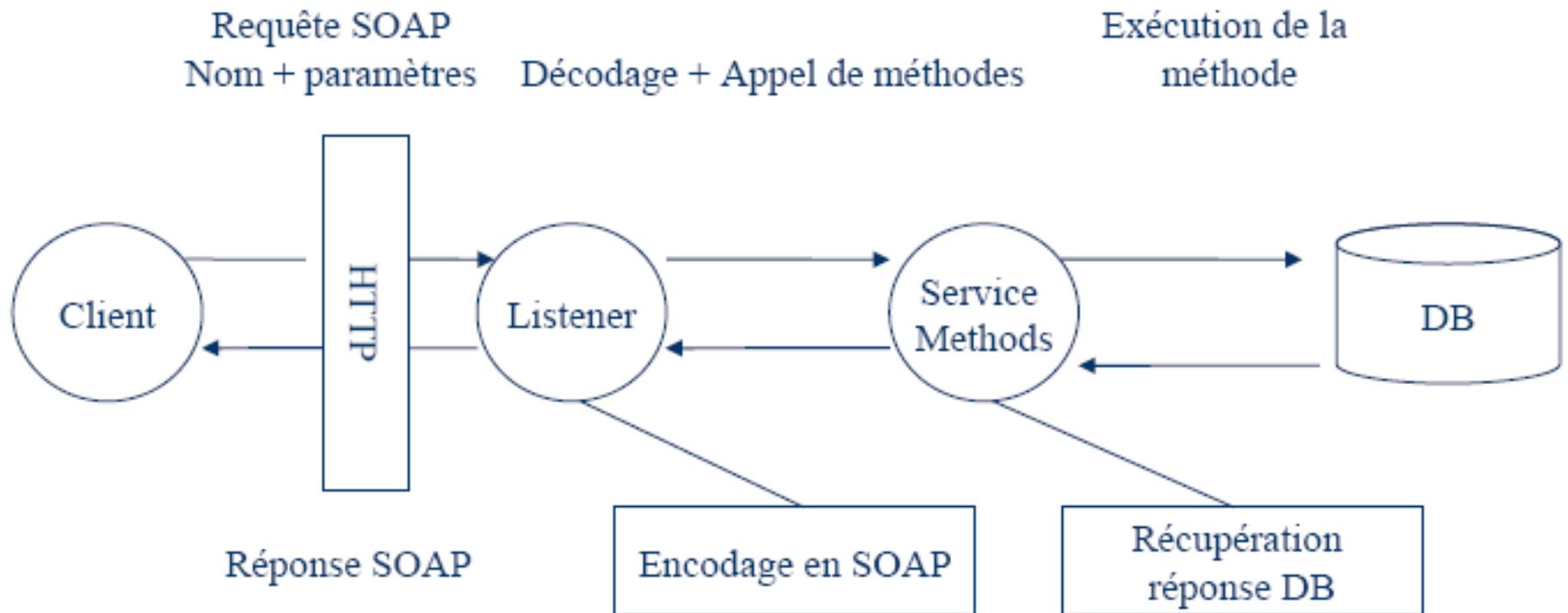
II.2 SOAP

Simple Object Access Protocol

- SOAP permet une normalisation des échanges de données. Les données sont encodées en XML et échangées par des appels de procédures à distance (RPC) en utilisant HTTP/SMTP/POP comme protocole de communication.
- Standard W3C
- Simple, extensible et permet le diagnostic des erreurs
- Fonctionne de manière synchrone et asynchrone.
- Indépendant de la plate-forme et du langage
- N'est pas perturbé par les pare-feu

II.2 SOAP

Fonctionnement



II.2 SOAP

Structure de message

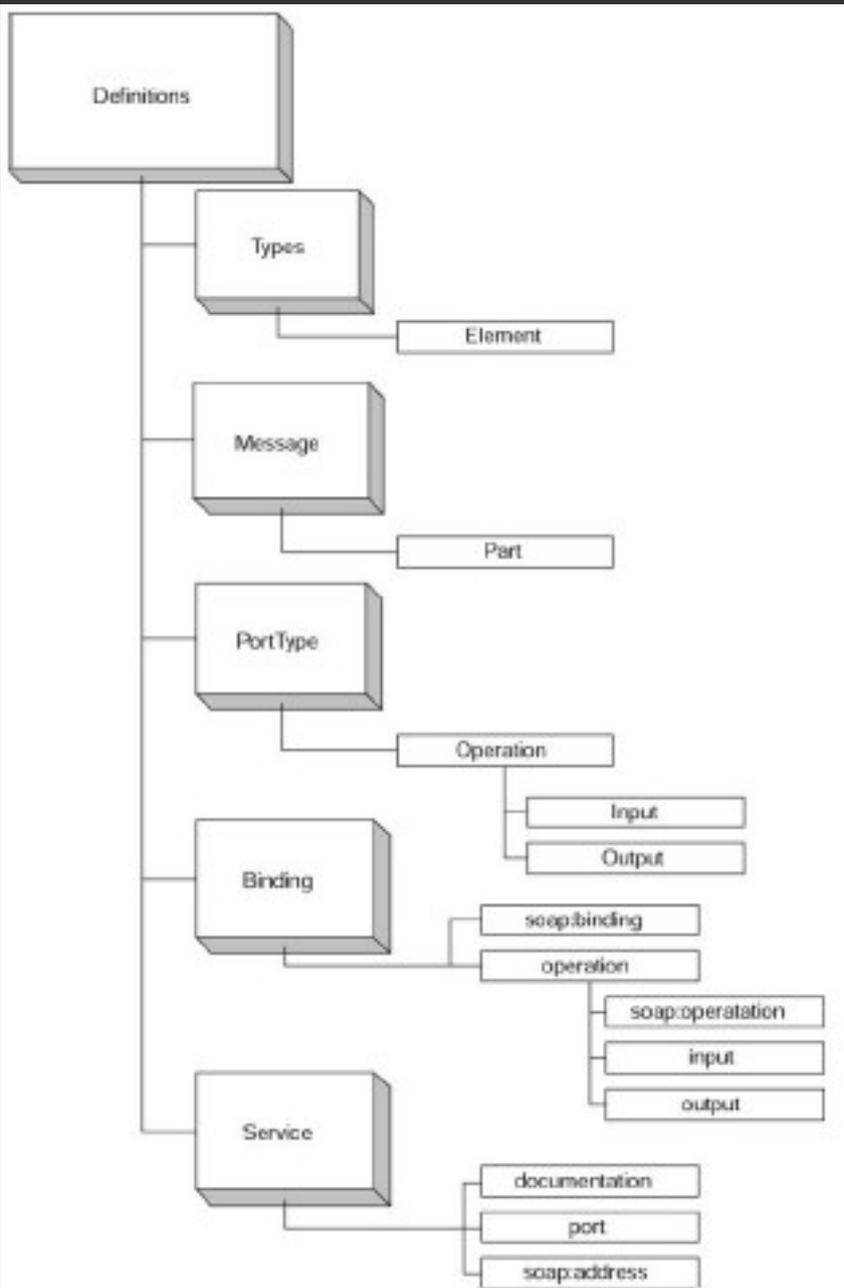
Structure des messages SOAP

- **Enveloppe**
 - Permet de spécifier version de SOAP utilisée
- **Entête**
 - Optionnel, permet de spécifier certaines directives pour le traitement du message
- **Corps**
 - Contient les données
- **Pièces jointes**
 - Permet l'envoi de données non-XML

```
<env:Envelope xmlns:env="http://www.w3.org/2003/05/soap-  
envelope">  
  <env:Header>  
    <n:alertcontrol xmlns:n="http://example.org/alertcontrol">  
      <n:priority>1</n:priority>  
      <n:expires>2005-01-26T10:30:00-11:00</n:expires> </n:alertcontrol>  
    </env:Header>  
    <env:Body>  
      <m:alert xmlns:m="http://example.org/alert">  
        <m:msg>Pas de Pause !!!! </m:msg>  
      </m:alert>  
    </env:Body>  
  </env:Envelope>
```

II.3. WSDL

Web Service Description Language



- Description XML des WS
 - Protocoles
 - Serveurs
 - Port
 - Format
 - E/S
 - Opérations
- Typage des E/S primaire
 - Possibilité d'importer des schémas pour étendre
- Décrit des taches simples
 - 1 service = 1 traitement

II.4. UDDI

Universal Description, Discovery and Integration

- Annuaires
 - Répertoire de services
 - _ Permet aux fournisseurs de publier leur service
 - _ Permet aux clients de rechercher des services
 - Ne contient que les références à ces services
 - _ Le service reste chez le fournisseur
 - Public / Privé
 - _ Accès restreint pour des données sensibles
- Structure
 - Pages blanches : informations sur le fournisseur
 - Pages jaunes : WSDL des services
 - Pages vertes : informations techniques sur les services

III. Limites

III. Limites

Web services vs. Web Sémantique

- Les web services ne traitent que la syntaxe et pas la sémantique
 - L'utilisation d'XML permet de structurer et spécifier les étapes dans la construction d'un document. Ils ne permettent pas de spécifier le sens à donner au document.
- Il est difficile de composer des services complexes avec d'autres services existants.

III. Limites

Sécurité & Confiance

- Sécurité
 - Accès authentifié mal résolu
 - Importance des annuaires privés
 - Accès abusif aux ressources
 - Pas adapté à tous les besoins
 - Coût du transport sur le réseau non négligeable
- Confiance
 - Un service publié peut ne plus être maintenu
 - Un service publié peut avoir été modifié

IV. Web-services et bioinformatique

IV.1. Intérêt

- Utiliser les ressources de gros centres de calcul
- Maintenir les ressources au plus près des compétences
- Publier des ressources originales (données, programmes)
 - Construire des portails intégratifs
 - Mettre en place des architectures orientées services (SOA)

Architecture distribuée	Architecture Orientée Service
Orientée fonctionnalité	Orientée processus
Élaborée pour durer	Élaborée pour gérer le changement
Processus de développement long	Processus de développement court
Centrée sur la réduction des coûts	Centrée sur le support aux métiers
Organisée au sein d'une application	Permet l'orchestration et la composition
Système fortement couplé	Faible couplage, agile
Orienté objet	Orienté message
L'implémentation doit être connue	L'implémentation reste abstraite
Technologie homogène	Technologie hétérogène

IV.2 BioMOBY

Pourquoi une nouvelle plateforme ?

- Absence de sémantique dans WSDL
 - Très important en bioinformatique pour typer les données (ex : séquence protéique au format fasta vs. String)
 - Important pour décrire la ressource
 - Les WS consomment et produisent du XML
 - Absence de standard (depuis il y a BioXSD) pour la bioinformatique
 - Existance de nombreuses ressources basées sur d'autres « standards » (GFF, EMBL, FASTA, etc...)
- Difficulté à chaîner les services sans erreur

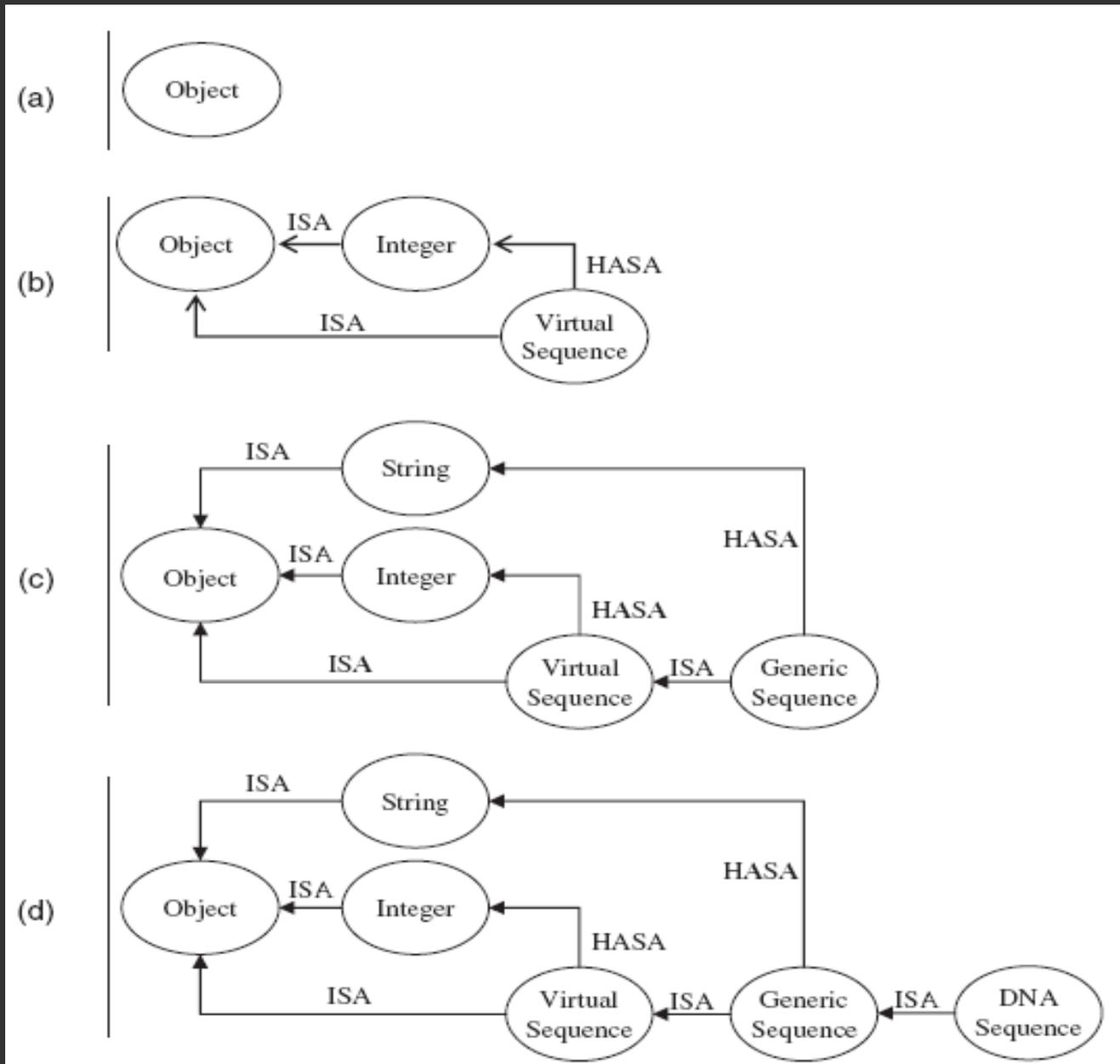
IV.2 BioMOBY

fournir des ressources bioinformatiques (données, programmes) via le web

- Description « XML Métier »
 - Typage métier E/S, Namespaces
- Un/Des « *protocole(s)* »
 - SOAP
 - WSRF (services asynchrones)
- Un annuaire « sémantique »
 - Ontologie des types
 - Ouvert
- Une API multi-langage
 - Perl
 - JAVA
 - *Python*

IV.2 BioMOBY

Data Type Ontology



IV.2 BioMOBY

BioMOBY-XML

(a) `<Object namespace='NCBI_gi' id='111076' articleName=''/>`

(b) `<VirtualSequence namespace='NCBI_gi' id='111076' articleName=''>
 <Integer namespace='' id='' articleName='Length'>38</Integer>
</VirtualSequence>`

(c) `<GenericSequence namespace='NCBI_gi' id='111076'>
 <Integer namespace='' id='' articleName='Length'>38</Integer>
 <String namespace='' id='' articleName='SequenceString'><![CDATA[
 ATGATGATAGATAGAGGGCCCGGCGCGCGCGCGCGC
]]></String>
</GenericSequence>`

(d) `<DNASequence namespace='NCBI_gi' id='111076'>
 <Integer namespace='' id='' articleName='length'>38</Integer>
 <String namespace='' id='' articleName='SequenceString'><![CDATA[
 ATGATGATAGATAGAGGGCCCGGCGCGCGCGCGCGC
]]></String>
</DNASequence>`

b. `<b64_encoded_GIF namespace='DragonDB_Image' id='cho'>
 <String namespace='' id='' articleName='content'><![CDATA[
 TWFuIGlzIGRpc3Rpbmd1aXNoZWQsIG
 5vdCBvbm95IGJ5IGhpcyByZWZzb24sIG
 JldCBieSB0aGlzIHNoYm91bnRlbnQsIG
 Npb24gZnJvbSBvdGhlcilBhbmltYWxzLC
 ...
]]></String>
</b64_encoded_GIF >`

IV.2 BioMOBY

BioMOBY-XML

```
<moby:MOBY xmlns:moby="http://biomoby.org" xmlns="http://biomoby.org">
<moby:mobyContent>
  <ProvisionInformation>
    <serviceSoftware software_name="" software_version="" software_comment=""/>
    <serviceDatabase database_name="" database_version="" database_comment=""/>
    <serviceComment>comment here</serviceComment>
  </ProvisionInformation>
  <moby:mobyData queryID="Q1">
    <moby:Simple>
      <moby:Object namespace="X" id="1"/>
    </moby:Simple>
  </moby:mobyData>
  <moby:mobyData queryID="Q2">
    <moby:Simple>
      <moby:Object namespace="X" id="2"/>
    </moby:Simple>
  </moby:mobyData>
</moby:mobyContent>
</moby:MOBY>
```

IV.2 BioMOBY

Scénario



Spécification des entrées → spécialisation du service

- Un service pouvant manipuler un fichier FASTA (ex: squizz), peut manipuler un fichier FASTA proteique
- Un service analysant une proteine (ex: blastp) ne peut pas traiter une sequence nucleique

IV.3 PlayMOBY

WS @ LIPM

- **Un web-service est l'encapsulation d'un programme déjà existant**
- **Ce programme manipule des fichiers**
 - on peut toujours utiliser ces programmes en ligne de commande
 - on peut les encapsuler via d'autres technologies (CGI)

• Pourquoi PlayMOBY ?

- Déployer de nouveaux web-services
- Automatiquement
- pour des programmes existants (et maintenus !)
- pour les futurs programmes



• Pourquoi Moby ?

- Utiliser un format pivot pour la description
- Experience .acd EMBOSS
- Profiter des programmes déjà décrits

IV.3 PlayMOBY

Mobyle

- **problème**: faciliter l'accès aux outils bioinformatiques pour les biologistes.
 - outils bioinformatiques ~ = outils en ligne de commande le plus souvent.
 - outils en ligne de commande → difficultés d'apprentissage.
 - Fournir une interface web, plus familière aux biologistes.
 - Le développement d'interfaces spécifiques est répétitif, coûteux, et source d'erreurs.

→ **PISE** → **Mobyle**

Qu'est-ce que Mobyle?

- Un portail d'analyses de bioinformatique
- une plate-forme de **publication** et d'**intégration**
- **Un répertoire de description de programmes**

Mobyle: a new full web bioinformatics framework.

Néron B, Ménager H, Maufrais C, Joly N, Maupetit J, Letort S, Carrere S, Tuffery P, Letondal C.

Bioinformatics. 2009 Nov 15;25(22):3005-11. Epub 2009 Aug 17.

IV.3 PlayMOBY

Mobyle

Mobyle@pasteur

hmenager@pasteur.fr (guest)
[set email](#) | [sign-in](#) | [sign-out](#)

Programs

List: by server by category

- ▶ Analysis
- ▶ Annotations
- ▼ Multiple_Sequence_Compare
 - iANTMultalinNucleic@LIP
 - iANTMultalinProteic@LIP
- ▶ Retrieval
- ▼ Sequence
 - ▼ Alignment
 - ▼ multiple
 - ProbCons@RPBS
- ▶ SequenceAnalysis
- ▶ Structure
- ▶ alignment
- ▶ assembly
- ▶ database
- ▶ display
- ▶ hmm
- ▶ nucleic
- ▶ phylogeny
- ▶ protein
- ▶ sequence
- ▶ structure

Data Bookmarks

Protein,Nucleic Sequence:
golden.out
Text: infile.data

Jobs

golden - 11/05/09 14:03:31
 extractfeat - 11/05/09 14:03:52
 squizz_checker - 11/05/09 14:33:04

Welcome Programs Data Bookmarks Jobs Tutorials

ProbCons@RPBS x Clustalw: Multiple alignment x IANTMultalinNucleic@LIPM x

ProbCons@RPBS

ProbCons: Probabilistic Consistency-based Multiple Alignment of Amino Acid Sequences

Input Data

* Input alignment (Sequence)

Paste DB File Result

Options

Output format: ?

fasta
 clustalw

Consistency: ?

Iterations: ?

Pre-training iterations: ?

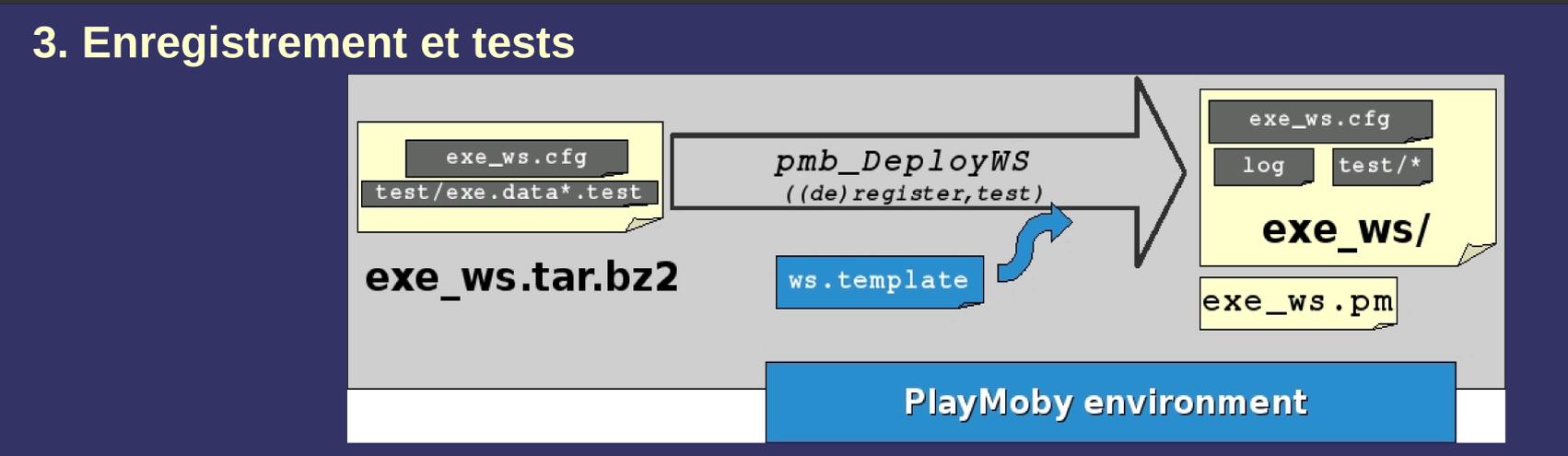
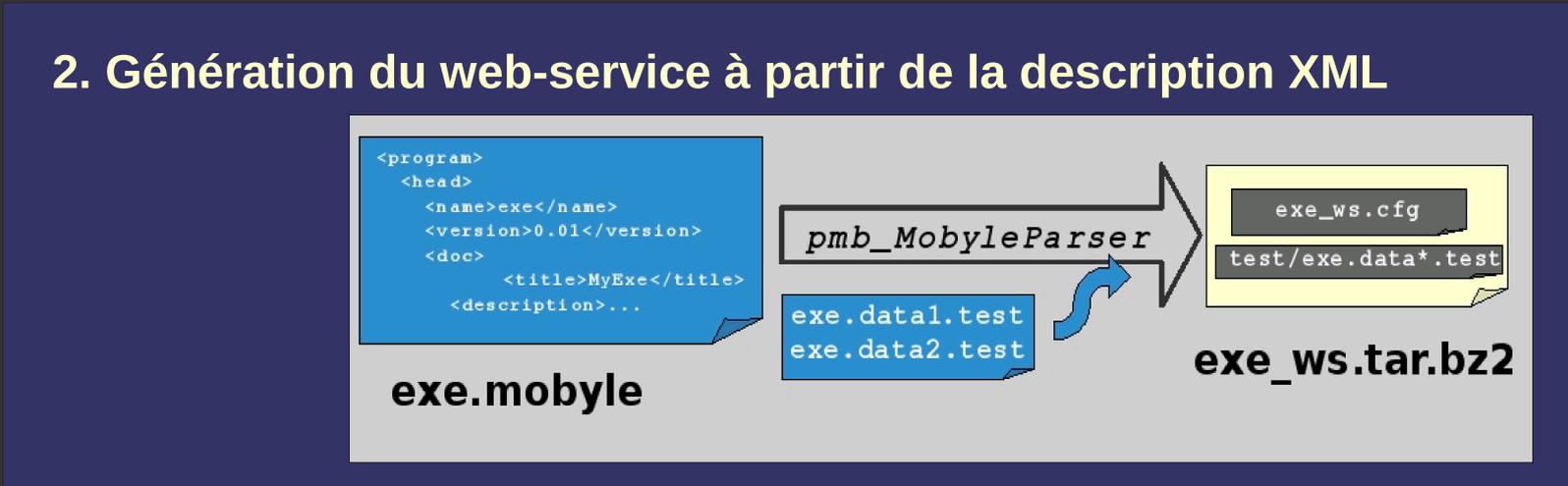
* : mandatory parameter

IV.3 PlayMOBY

déployer un WS BioMOBY en 3 étapes

1. Génération d'un fichier de description Mobyle XML

Appli.pm: un module pour générer ces fichiers XML



IV.3 PlayMOBY

Appli.pm



- **Pourquoi ?**
 - Avoir une description du programme embarquée
 - Restituer la description sous différents formats
 - Mobyly
 - Usage "UNIX-like »
- **Comment ?**
 - Structures Perl de description embarquées
 - Dictionnaire BioMOBY ← → Mobyly

BioMOBY	Biotype	Class	SuperClass	Dataformat	File Extension
text-formatted		Text			
+					
text-html		Html	AbstractText		html
+					
iANT_entry_features-xml		iANTReport	Report	iANTEntryFeaturesReport	xml
+					
iANT_entry-xml		iANTReport	Report	iANTEntryReport	xml
+					
iANT_entry_db-xml		iANTReport	Report	iANTEntryDBReport	xml
+					
iANT_organism-xml		iANTReport	Report	iANTOrganismReport	xml
+					
string		String			
+					
Float		Float			
+					
integer		Integer			
+					
DateTime		String			
+					
Boolean		Boolean			
+					
List_Text		ListText	AbstractText		
+					
FASTA		Sequence		FASTA	
-					
<pre> >Sequence this is a sequence in FASTA format MKFLILLFNILCLFPVLAADNHGUGPQGASGVDPITPDINENQGTGPAFLTAVEMAGVKYL QQQHGSMWN IHR LVEGNV VINENASTPLYTGAIVTNDGSPYMAVVEVLGDFNLQFFIKSG DAWVTLSEHEVLAQLQEIRQAVHIESVFSILMAFQLENNKYEVETHAPKGANMVTFIIPRN GHICMUYHKNVRIYKATGNDIVTSVWGFPRGLRLLINVFSIDDNGMMENRYFQHVDDK YVPISQNYETGIVKLDYGHAYHFVDDIKIDVTMFHLADATYHEPCFKIIPNTGFCI TKLPDGDQVLYESFNPLIHCINEVHIYDRNNGSITCLHLNYSPPSYKAYLVLDKDTGWEAT THPLEEKIEELQDQACELDWNFISDKOLYVAALTNADLNMTVTPRPHRDVIRVSDGS EVLWVYEGLDNFLVCAWIYVSDGVALVHLRIKRIIPANNDIYVLGDLVWTRITKIQT QEIHGLVHLSWGLLAPITEEDSDKHDEPPEGPGASGLPPKAPGDKEGSEGHKGPSWGSDS SHEGKMPGSGHKPGPAREHKPSKIPTLSWMPGPGHDPKHPDPKPEKSKSPRTASPTRR PSPKLPQLSKLPKSTSPFASPPPPTRPSSPERPEGTKIINTSKPPSPKPPFPDPSFKEMFYD </pre>					

IV.3 PlayMOBY

Surveillance fonctionnelle

- **Données de test**
 - Test fonctionnels → Surveillance fonctionnelle → Réseau de confiance

<input type="checkbox"/>		HeliageneGetEntry	2009.02.16 02:06:39	Execution time: 1 s.	
<input type="checkbox"/>		HeliageneGetSequence	2009.02.16 02:06:43	Execution time: 4 s.	
BioMOBY <ul style="list-style-type: none"> URI: www.heliagene.org Registry: http://moby.ucalgary.ca/moby/MOBY-Central.pl RDF: http://www.heliagene.org/playmoby/rdf/aa_dna/mobycentral_HeliageneGetSequence.rdf Contact: sebastien.carrere@toulouse.inra.fr 					
<input type="checkbox"/>		Show perl code used to test sequences format ok			
sequences	<input type="checkbox"/>	Show perl code used to test sequences co... <pre>return (1,"ok") if (\$reference_c PlayMoby where we compare (stric http://www.heliagene.org/playmok http://www.heliagene.org/playmok Differences are reported as "[st ok</pre>			
			Narcisse: a mirror view of conserved syntenies <i>collaboration with Thomas Faraut - Laboratoire de Génétique Cellulaire</i>		
			LEGOO: a bioinformatics gateway towards integrative legume biology		
			HeliaGene: a bioinformatics portal for Helianthus sp. genomics		
			iANT: tools for sequence annotation		
	<input type="checkbox"/>	Show raw output data for sequences  External tools: tools developed by other teams (almost EMBOSS)			
			BIOS: Service Oriented Architecture in Bioinformatics dedicated to RNA-Seq		

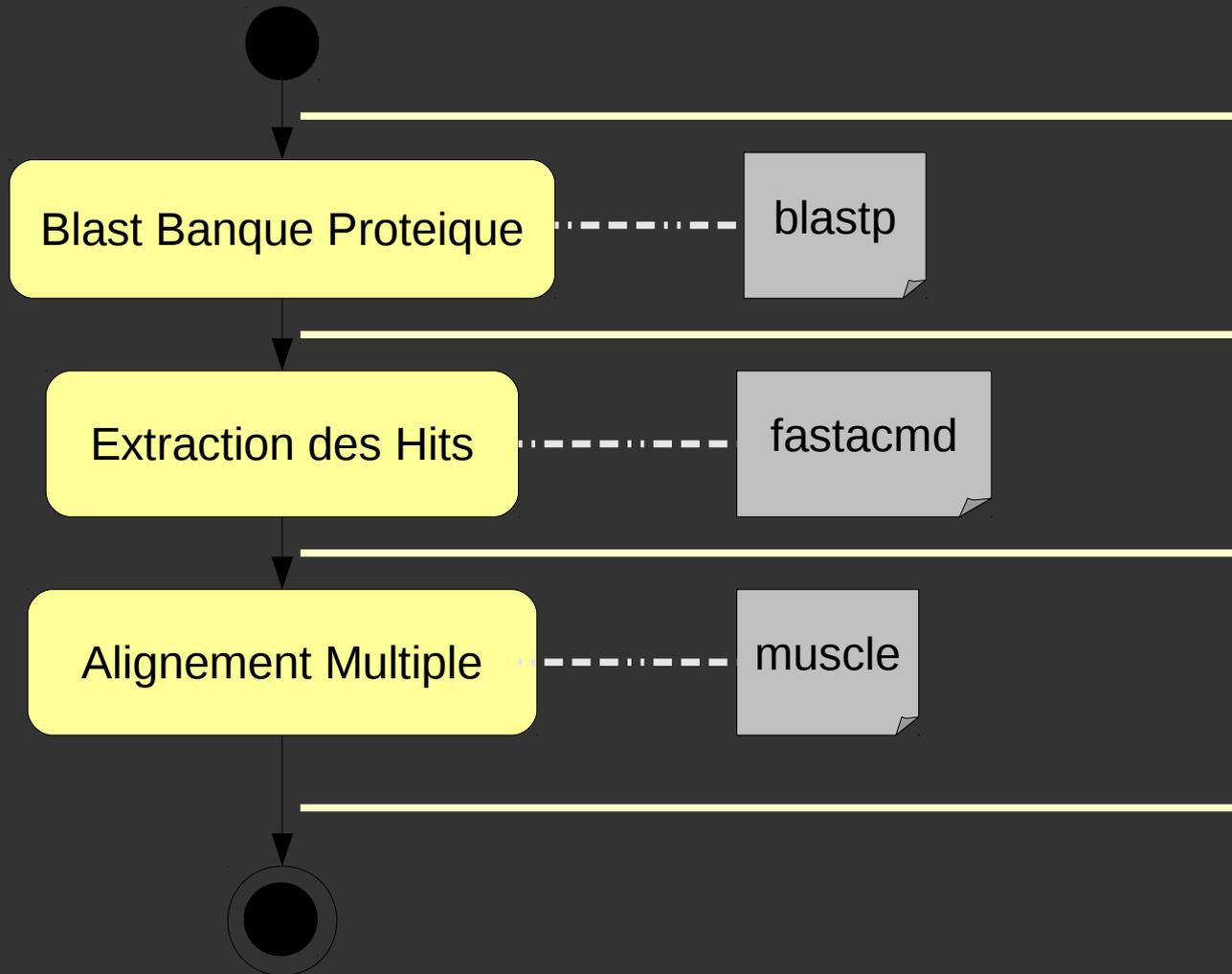
Sources

- Cours Ecole des Mines, Rahee Ghurburn
<http://www.emse.fr/~vercouter/cours/cws/WebService.pdf>
- Interoperability with Moby 1.0--it's better than sharing your toothbrush!
Brief Bioinform. 2008 May;9(3):220-31.
BioMoby Consortium
- Moby: a new full web bioinformatics framework.
Bioinformatics. 2009 Nov 15;25(22):3005-11. Epub 2009 Aug 17.
Néron B, Ménager H, Maufrais C, Joly N, Maupetit J, Letort S, Carrere S, Tuffery P, Letondal C.

V. TP

V. TP

- Mettre en place les Web-services pour accomplir le workflow suivant



Entrées , Sorties ?

V. TP

• Installer l'environnement PlayMOBY

- *Configuration PlayMOBY*
- *Configuration Apache*
- *Test / QuickStart*

• Ecrire le wrapper Perl du programme a déployer

- *ParamParser.pm*

• Décrire le wrapper & Générer la fiche MobyLe

- *Appli.pm*
- *Dictionnaire BioMOBY2MobyLe*
- *Types BioMOBY <http://lipm-bioinfo.toulouse.inra.fr/registry>*
- *Tester la fiche dans le serveur mobyle*

• Déployer le Webservice

- *PlayMOBY::pmb_MobyLeParser.pl*
- *PlayMOBY::pmb_DeployWS.pl*

• Affiner le test fonctionnel

- *PmbTest.pm*

• Mettre en place la surveillance fonctionnelle

- *PlayMOBY::pmb_CheckWS.pl*
- *crontab*

• Chainer les différents services déployés

- *Remora (<http://lipm-bioinfo.toulouse.inra.fr/remora>)*