



**HAL**  
open science

# Démarche centrée utilisateur pour la conception de SIAD basés sur un processus d'Extraction de Connaissances à partir de Données, Application à la lutte contre les infections nosocomiales

Hela Ltifi

► **To cite this version:**

Hela Ltifi. Démarche centrée utilisateur pour la conception de SIAD basés sur un processus d'Extraction de Connaissances à partir de Données, Application à la lutte contre les infections nosocomiales. Informatique [cs]. Université de Valenciennes et du Hainaut-Cambresis; Ecole Nationale d'Ingénieurs de Sfax, 2011. Français. NNT: . tel-00630296

**HAL Id: tel-00630296**

**<https://theses.hal.science/tel-00630296>**

Submitted on 8 Oct 2011

**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.

## THÈSE

pour l'obtention du

**Doctorat en cotutelle**

**de l'Université de Valenciennes et du Hainaut-Cambrésis, France**  
*Spécialité Automatique et Informatique des Systèmes Industriels et Humains*  
*Mention Informatique*

**Et de l'Université de Sfax, Tunisie**  
**École Nationale d'Ingénieurs de Sfax**  
*Discipline Informatique*  
*Ingénierie des Systèmes Informatiques*

**Hela LTIFI**

Mastère informatique

### **Démarche centrée utilisateur pour la conception de SIAD basés sur un processus d'Extraction de Connaissances à partir de Données**

Application à la lutte contre les infections nosocomiales

Soutenue publiquement à Valenciennes le 09 Juillet 2011 devant le jury composé de

<b>M. Bertrand David</b>	Professeur, Ecole Centrale de Lyon	<i>Rapporteur</i>
<b>M. Faouzi Moussa</b>	Maître de Conférences, Université de Tunis, Tunisie	<i>Rapporteur</i>
<b>M. Patrick Girard</b>	Professeur, Université de Poitiers	<i>Examineur</i>
<b>M. Pierre Morizet-Mahoudeaux</b>	Professeur, Université de Technologie de Compiègne	<i>Président</i>
<b>M. Christophe Kolski</b>	Professeur, Université de Valenciennes et du Hainaut-Cambrésis	<i>Co-Directeur</i>
<b>M. Mohamed Adel Alimi</b>	Professeur, Université de Sfax, Tunisie	<i>Co-Directeur</i>
<b>M. Mounir Ben Ayed</b>	Maître-Assistant, Université de Sfax, Tunisie	<i>Invité</i>



# Remerciements

Cette thèse n'aurait pu aboutir sans l'aide et le soutien de nombreuses personnes

*Je tiens d'abord à exprimer ma reconnaissance à mes directeurs :*

*M. Christophe Kolski pour son aide, ses judicieux conseils et ses remarques pertinentes. Cette thèse n'aurait jamais été menée à terme sans le dévouement et l'intérêt constant porté par Christophe Kolski à mes travaux de recherche.*

*M. Adel Alimi et Mounir Ben Ayed dont j'ai pu apprécier tout au long de cette thèse les qualités scientifiques et humaines. Ils ont su me guider avec attention, efficacité et ouverture, tout en m'accordant initiative et autonomie.*

*Je tiens vivement à remercier les membres de jury :*

*M. Bertrand David et M. Faouzi Moussa pour m'avoir fait l'honneur d'accepter d'être les rapporteurs de ce mémoire.*

*M. Patrick Girard et M. Pierre Morizet-Mahoudeaux pour m'avoir fait l'honneur d'accepter d'être les examinateurs de ce mémoire.*

*Merci pour votre disponibilité, pour la lecture de ce mémoire et aussi d'avoir accepté de juger cette thèse.*

*Ce travail a pu voir le jour grâce au soutien du REGIM et du LAMIH, où j'ai pleinement bénéficié d'un environnement humain et matériel stimulant.*

*Je ne peux terminer ces lignes sans remercier profondément ma famille sans qui rien n'aurait été possible.*

*Merci à tous !*



## Table des matières

<b>Table des matières</b> .....	<b>i</b>
<b>Liste des figures et tableaux</b> .....	<b>i</b>
<b>Introduction générale</b> .....	<b>1</b>
<b>Chapitre 1 : Etat de l'art sur les Systèmes interactifs d'aide à la décision</b> .....	<b>5</b>
<b>1.1. Introduction</b> .....	<b>5</b>
<b>1.2. L'aide à la décision</b> .....	<b>5</b>
1.2.1 Processus de décision.....	6
1.2.1.1 L'information ou le renseignement (intelligence en anglais).....	6
1.2.1.2 La conception (design) .....	6
1.2.1.3 Le choix (choice).....	7
1.2.1.4 La révision (review) .....	7
1.2.2 Les systèmes d'aide à la décision (SAD) .....	7
1.2.3 Les Systèmes Interactifs d'Aide à la Décision (SIAD).....	9
1.2.3.1 Caractéristiques .....	9
1.2.3.2 Composition d'un SIAD.....	10
1.2.3.3 Classification des SIAD .....	11
<b>1.3. Les systèmes interactifs d'aide à la décision médicale (SIADM)</b> .....	<b>12</b>
1.3.1 L'approche purement numérique de l'aide à la décision .....	13
1.3.2 L'approche par méthodes d'intelligence artificielle .....	13
1.3.2.1 Les différents modèles de connaissance.....	14
1.3.2.2 Les méthodes de raisonnement.....	15
1.3.2.3 Les formalismes de représentation des connaissances .....	16
1.3.3 Logique et théorie de la décision pour la représentation de l'incertain : le raisonnement sous incertitude.....	16
1.3.3.1 Les logiques non monotones .....	17
1.3.3.2 Processus bayésiens et théorie de la décision .....	17
1.3.3.3 Imprécision et logique floue.....	17
1.3.4 L'Extraction de Connaissances à partir de Données.....	18
1.3.5 Aide à la décision basée sur le Web.....	18
1.3.6 Aide à la décision utilisant des données multimédias .....	19
<b>1.4. Extraction de Connaissances à partir de Données</b> .....	<b>19</b>

---

1.4.1	L'informatique décisionnelle.....	20
1.4.1.1	L'entrepôt de données .....	20
1.4.1.2	De l'OLAP à l'ECD .....	21
1.4.2	Les systèmes d'ECD.....	22
1.4.2.1	Processus d'ECD.....	22
1.4.2.2	Les principaux algorithmes d'ECD.....	24
1.4.3	Manques sous l'angle de l'Interaction Homme-Machine dans les systèmes d'ECD....	26
<b>1.5.</b>	<b>Conclusion.....</b>	<b>27</b>
<b>Chapitre 2 : Conception de système interactif d'aide à la décision .....</b>		<b>31</b>
<b>2.1.</b>	<b>Introduction .....</b>	<b>31</b>
<b>2.2.</b>	<b>Les processus de développement des systèmes, provenant du GL et de l'IHM ....</b>	<b>32</b>
2.2.1	Les processus de développement dans le domaine du GL.....	32
2.2.1.1	Le modèle en cascade.....	32
2.2.1.2	Le modèle en V .....	33
2.2.1.3	Le modèle en spirale.....	34
2.2.1.4	Modèle par incréments (ou incrémental).....	35
2.2.1.5	Modèle transformationnel .....	35
2.2.1.6	Le modèle en Y .....	36
2.2.1.7	Les méthodes agiles.....	37
2.2.1.8	Le processus Unifié basé sur UML .....	39
2.2.1.9	Conclusion sur les modèles étudiés.....	41
2.2.2	Les modèles de développement enrichis sous l'angle de l'IHM.....	42
2.2.2.1	Modèle de Long.....	43
2.2.2.2	Le prototypage.....	43
2.2.2.3	Modèle en couches .....	44
2.2.2.4	Le modèle en étoile .....	45
2.2.2.5	Le modèle en V enrichi .....	45
2.2.2.6	Le modèle de Collins.....	47
2.2.2.7	Le modèle Nabla $\nabla$ .....	47
2.2.2.8	Modèle en U .....	49
2.2.2.9	Conclusion sur les cycles enrichis sous l'angle de l'IHM .....	51
2.2.3	La Norme ISO 13407 et la conception centrée utilisateur .....	52
2.2.3.1	Nature et objectifs de l'ISO 13407.....	52
2.2.3.2	Le processus de conception centrée utilisateur.....	53
<b>2.3.</b>	<b>Méthodes d'analyse et de conception des systèmes .....</b>	<b>54</b>
2.3.1	Les méthodes systémiques .....	54

---

2.3.1.1	La méthode Merise.....	54
2.3.1.2	La méthode Axial.....	55
2.3.2	Les méthodes cartésiennes.....	55
2.3.2.1	La méthode SADT (Structured Analysis and Design Technique).....	56
2.3.2.2	La méthode SA (Structured Analysis).....	56
2.3.2.3	La méthode SA-RT.....	57
2.3.3	Les méthodes orientées objets.....	57
2.3.4	Conclusion sur les méthodes étudiées.....	58
<b>2.4.</b>	<b>Approches existantes de conception de SIAD.....</b>	<b>58</b>
2.4.1	Les processus de développement de SIAD.....	58
2.4.1.1	Démarche itérative.....	58
2.4.1.2	Processus de développement de Nykanen.....	59
2.4.1.3	Le modèle MODESTI.....	60
2.4.1.4	L'approche ADESIAD.....	62
2.4.1.5	Conclusion sur les modèles de développement de SIAD.....	63
<b>2.5.</b>	<b>Conclusion.....</b>	<b>63</b>
<b>Chapitre 3 :</b>	<b>ExUP : Démarche de développement de SIAD/ECD.....</b>	<b>65</b>
<b>3.1.</b>	<b>Introduction.....</b>	<b>65</b>
<b>3.2.</b>	<b>SIAD/ECD : SIAD basé sur un processus d'ECD.....</b>	<b>65</b>
3.2.1	Modules d'un SIAD/ECD.....	67
<b>3.3.</b>	<b>Présentation globale de la démarche proposée.....</b>	<b>70</b>
<b>3.4.</b>	<b>Présentation des acteurs de l'ExUP.....</b>	<b>73</b>
3.4.1	L'utilisateur final.....	73
3.4.2	L'expert ECD.....	73
3.4.3	L'ergonome.....	74
3.4.4	Le concepteur.....	74
3.4.5	Le développeur.....	74
<b>3.5.</b>	<b>Description détaillée des activités de l'ExUP.....</b>	<b>76</b>
3.5.1	Capture des besoins.....	76
3.5.1.1	Analyse des situations décisionnelles normales et anormales.....	77
3.5.1.2	Élaboration des premières maquettes d'IHM.....	79
3.5.1.3	Modélisation de l'utilisateur.....	80
3.5.1.4	Expression des besoins décisionnels.....	82
3.5.1.5	Évaluation des besoins.....	83
3.5.1.6	Définition et répartition des tâches.....	84

3.5.1.7	Evaluation des tâches .....	85
3.5.2	Analyse et conception .....	86
3.5.2.1	Spécification des tâches automatiques .....	88
3.5.2.2	Spécification des tâches interactives et des interfaces homme-machine.....	88
3.5.2.3	Conception des modules automatiques et interactifs d'ECD et des Interfaces Homme-Machine.....	90
3.5.2.4	Evaluation des modèles .....	91
3.5.2.5	Elaboration de maquettes avancées d'IHM.....	92
3.5.2.6	Spécifier l'architecture .....	93
3.5.2.7	Evaluation de l'architecture.....	94
3.5.3	Implémentation .....	94
3.5.3.1	Codage des fonctionnalités et des interfaces Homme-Machine .....	95
3.5.3.2	Construction d'un prototype.....	96
3.5.4	Test.....	97
<b>3.6.</b>	<b>Conclusion.....</b>	<b>99</b>
<b>Chapitre 4 : Application de la démarche proposée pour le développement d'un Système Interactif d'Aide à la Décision Dynamique Médicale basé sur l'ECD 101</b>		
<b>4.1.</b>	<b>Introduction .....</b>	<b>101</b>
<b>4.2.</b>	<b>Contexte de travail .....</b>	<b>101</b>
4.2.1	Les infections nosocomiales .....	101
4.2.2	Travaux antérieurs.....	102
<b>4.3.</b>	<b>SIADDM/ECD : SIAD Dynamique Médicale basé sur l'ECD .....</b>	<b>103</b>
4.3.1	SIADDM pour la lutte contre les IN.....	103
4.3.2	Les modules visés pour le SIADDM/ECD .....	106
<b>4.4.</b>	<b>Mise en application .....</b>	<b>108</b>
4.4.1	Module de fouille de données .....	109
4.4.1.1	Phase d'inception .....	109
4.4.1.2	Phase d'élaboration .....	111
4.4.1.3	Phase de construction .....	114
4.4.1.4	Phase de transition.....	117
4.4.1.5	Les réseaux bayésiens dynamiques exploités pour la réalisation du module de fouille de données.....	119
4.4.2	Module de gestion de connaissances .....	123
4.4.2.1	Phase d'inception .....	123
4.4.2.2	Phase d'élaboration .....	125
4.4.2.3	Phase de construction .....	128

---

4.4.2.4 Phase de transition.....	130
<b>4.5. Conclusion.....</b>	<b>131</b>
<b>Chapitre 5 : Evaluation de l'ExUP et perspectives de recherche.....</b>	<b>133</b>
<b>5.1. Introduction .....</b>	<b>133</b>
<b>5.2. Bilan d'évaluation globale de la démarche .....</b>	<b>133</b>
<b>5.3. Bilan d'évaluation empirique du SIADDM/ECD.....</b>	<b>137</b>
5.3.1 Evaluation de l'utilisabilité du SIADDM/ECD.....	138
5.3.2 Evaluation de l'utilité du SIADDM/ECD.....	140
5.3.2.1 Evaluation des effets du SIADDM/ECD.....	140
5.3.2.2 Expérimentation, résultats et performances.....	141
<b>5.4. Perspectives de recherche .....</b>	<b>144</b>
5.4.1 Perspectives sur le plan méthodologique.....	144
5.4.1.1 Formalisation et modélisation des tâches.....	144
5.4.1.2 Intégration des techniques de visualisation de l'information .....	146
5.4.1.3 Evaluation de la qualité des données.....	149
5.4.2 Perspectives sur le plan expérimental .....	150
<b>5.5. Conclusion.....</b>	<b>151</b>
<b>Conclusion générale .....</b>	<b>153</b>
<b>Bibliographie.....</b>	<b>157</b>
<b>Annexe A. Techniques de visualisation de l'information .....</b>	<b>179</b>
<b>A.1. Introduction .....</b>	<b>179</b>
<b>A.2. Classification des techniques de visualisation.....</b>	<b>179</b>
<b>A.3. Description des techniques .....</b>	<b>180</b>
A.3.1 Approche technique.....	180
A.3.1.1 Données linéaires .....	180
A.3.1.2 Données multidimensionnelles.....	181
A.3.1.3 Hiérarchies .....	182
A.3.1.4 Réseaux .....	190
A.3.1.5 Modèle vectoriel.....	192
A.3.2 Approche interactive .....	194
<b>A.4. Conclusion.....</b>	<b>196</b>



## Liste des figures et tableaux

Figure 1.1 : le processus décisionnel selon le modèle IDCR de Simon [Simon 77] .....	7
Figure 1.2 : composition d'un SIAD [Holtzman 89] [Sage 91] [Turban 93].....	10
Figure 1.3 : les éléments d'un système expert [Degoulet <i>et al.</i> 91].....	16
Figure 1.4 : le processus d'extraction de connaissances à partir de données .....	23
Figure 1.5 : rôle actuel de l'utilisateur lors d'une tâche dans un processus d'ECD [Chevrin <i>et al.</i> 07] .....	27
Figure 2.1 : le modèle en cascade [Royce 70] .....	32
Figure 2.2 : le modèle en V [McDermid <i>et al.</i> 84] .....	33
Figure 2.3 : le modèle en spirale [Boehm 88] .....	34
Figure 2.4 : modèle incrémental [ESA 91] [Laprie <i>et al.</i> 95].....	35
Figure 2.5 : modèle transformationnel .....	36
Figure 2.6 : modèle en Y [Hassine <i>et al.</i> 02].....	36
Figure 2.7 : le cycle de développement RAD [Martin 91] .....	37
Figure 2.8 : cas d'utilisation .....	39
Figure 2.9 : structuration d'un processus unifié autour des cas d'utilisation [Octo 02] .....	40
Figure 2.10 : le processus Unifié [Jacobson <i>et al.</i> 99] .....	41
Figure 2.11 : modèle proposé par Long <i>et al.</i> [Long <i>et al.</i> 90] .....	43
Figure 2.12 : le prototypage [Rettig 94] .....	44
Figure 2.13 : le modèle en couches [Curtis <i>et al.</i> 94].....	44
Figure 2.14 : le modèle en étoile [Hix <i>et al.</i> 93] .....	45
Figure 2.15 : le modèle en V enrichi [Balbo 94] .....	46
Figure 2.16 : le modèle de Collins [Collins 95] .....	47
Figure 2.17 : le modèle $\nabla$ [Kolski 97] [Kolski 98].....	48
Figure 2.18 : le modèle en U [Abed 01] .....	50
Figure 2.19 : processus de conception centré utilisateur [ISO13407 99] .....	53
Figure 2.20 : modèle MODESTI [Duribreux-Cocquebert 95] .....	61
Figure 2.21 : l'approche ADESIAD [Lepreux 05] .....	62
Figure 3.1 : contexte global de recherche.....	66
Figure 3.2 : processus d'un SIAD basé ECD replacé dans le cadre du modèle IDCR de Simon [Ltifi <i>et al.</i> 10] .....	66
Figure 3.3 : modules constitutifs de SIAD/ECD ; différents acteurs sont concernés à chaque étape, exigeant des besoins particuliers vis-à-vis de chaque module concerné.....	69
Figure 3.4 : positionnement de notre contribution par rapport aux modules du SIAD/ECD .....	72

Figure 3.5 : diagramme de contexte SIAD/ECD.....	75
Figure 3.6 : diagramme d'activité de capture des besoins.....	77
Figure 3.7 : diagramme d'activités d'analyse et de conception.....	87
Figure 3.8 : diagramme d'activité d'implémentation .....	95
Figure 3.9 : diagramme d'activité de Test.....	98
Figure 4.1 : les facteurs temporels pour la prévention d'IN .....	106
Figure 4.2 : bilan relativement aux modules visés .....	108
Figure 4.3 : dépendances causales dans un RB statique (modèle statique extrait).....	120
Figure 4.4 : dépendances causales dans un RB dynamique (modèle dynamique extrait).....	121
Figure 4.5 : graphe causal du RBD.....	122
Figure 4.6 : une maquette utilisateur d'IHM du cas d'utilisation "Aide à la décision" .....	124
Figure 5.1 : critères d'évaluation de la démarche.....	134
Figure 5.2 : évaluation globale de l'ExUP .....	137
Figure 5.3 : évaluation de l'utilisabilité du SIADDM/ECD.....	140
Figure 5.4 : les résultats de la prédiction dynamique pour un patient .....	142
Figure 5.5 : résultats de la prédiction pour une base de 58 patients .....	143
Figure 5.6 : raffinement possible des modèles des tâches .....	145
Figure 5.7 : intégration des techniques de visualisation.....	146
Figure 5.8 : évaluation de la qualité des données.....	150
Figure A.1 : table de données « Table lens » [Pirollo et Rao 96].....	180
Figure A.2 : mur fuyant [Robertson et al. 91] .....	181
Figure A.3 : nuages de points [Swayne et al. 98] .....	182
Figure A.4 : système de polyligne, coordonnées parallèles [Inselberg et Dimsdale 90] .....	182
Figure A.5 : représentation matricielle [Abello et Ham 04] .....	183
Figure A.6 : exemple d'outils d'approche par diagrammes : Space Tree [Hascoet et Beaudouin-Lafon 01] .....	185
Figure A.7 : treemaps [Hascoet et Beaudouin-Lafon 01] .....	185
Figure A.8 : arbres circulaires [Andrews et Heidegger 98].....	186
Figure A.9 : arbre hyperbolique [Lamping et al. 95] .....	187
Figure A.10 : arbre conique horizontal [Robertson et al. 91].....	189
Figure A.11 : arbre conique vertical [Robertson et al. 91] .....	189
Figure A.12 : cubes d'informations [Rekimoto et Green 93] .....	190
Figure A.13 : visualisation botanique [Kleinberg <i>et al.</i> 01] .....	190
Figure A.14 : exemple de réseau de liens hypertextes [Chen 05].....	191
Figure A.15 : exemple de réseau social [Heer et Boyd 05] .....	192

---

Figure A.16 : exemple de réseau de co-citation [Chen 05] .....	192
Figure A.17 : approche de similarité [Andrews et al. 04].....	193
Figure A.18 : approche de pertinence [Hascoet et Beaudouin-Lafon 01].....	193
Figure A.19 : filtrage dynamique : application de fisheye [Keim 02] .....	195
Tableau 1.1 : étude comparative entre les approches d'aide à la décision médicale .....	28
Tableau 3.1 : analyse des situations décisionnelles normales et anormales pour les modules du SIAD/ECD .....	78
Tableau 3.2 : élaboration des premières maquettes d'IHM pour les modules du SIAD/ECD .....	80
Tableau 3.3 : modélisation de l'utilisateur pour les modules du SIAD/ECD .....	82
Tableau 3.4 : expression des besoins décisionnels pour les modules du SIAD/ECD .....	83
Tableau 3.5 : évaluation des besoins pour les modules du SIAD/ECD .....	84
Tableau 3.6 : définition et répartition des tâches pour les modules du SIAD/ECD.....	85
Tableau 3.7 : évaluation des tâches pour les modules du SIAD/ECD.....	86
Tableau 3.8 : spécification des tâches automatiques pour les modules du SIAD/ECD .....	88
Tableau 3.9 : spécification des tâches interactives et des interfaces homme-machine pour les modules du SIAD/ECD.....	89
Tableau 3.10 : conception des modules automatiques et interactifs d'ECD et des Interfaces Homme-Machine pour les modules du SIAD/ECD .....	90
Tableau 3.11 : évaluation des modèles pour les modules du SIAD/ECD .....	91
Tableau 3.12 : élaboration des maquettes avancées d'IHM pour les modules du SIAD/ECD.....	92
Tableau 3.13 : spécification de l'architecture des modules du SIAD/ECD .....	93
Tableau 3.14 : évaluation de l'architecture des modules du SIAD/ECD.....	94
Tableau 3.15 : codage des fonctionnalités et des interfaces homme-machine des modules du SIAD/ECD.....	96
Tableau 3.16 : construction de prototype des modules du SIAD/ECD.....	97
Tableau 3.17 : tests des modules du SIAD/ECD .....	99
Tableau 4.1 : déroulement de la phase d'inception pour le module de fouille de données .....	109
Tableau 4.2 : déroulement de la phase d'élaboration pour le module de fouille de données.....	111
Tableau 4.3 : déroulement de la phase de construction pour le module de fouille de données .....	115
Tableau 4.4 : déroulement de la phase de transition pour le module de fouille de données .....	118
Tableau 4.5 : variables pour l'application du RBD .....	120
Tableau 4.6 : déroulement de la phase d'inception pour le module de gestion des connaissances .....	123
Tableau 4.7 : déroulement de la phase d'élaboration pour le module de gestion des connaissances .....	125
Tableau 4.8 : déroulement de la phase de construction pour le module de gestion des connaissances .....	128
Tableau 4.9 : déroulement de la phase de transition pour le module de gestion des connaissances.....	130

Tableau 5.1 : bilan d'évaluation de la démarche .....	135
Tableau 5.2 : grille d'évaluation .....	139
Tableau 5.3 : la matrice de confusion des résultats fournis par notre réseau bayésien dynamique [Ltifi <i>et al.</i> 10b] .....	143
Tableau 5.4 : application possible des techniques de visualisation pour chaque module du SIADDM/ECD .....	148
Tableau A.1 : classification des techniques de visualisation .....	179

## Introduction générale

Le domaine de l'Aide à la décision est particulièrement vaste, source de nombreuses propositions, aussi bien dans le milieu académique qu'industriel. Au milieu des années quatre-vingt sont apparus les outils d'aide à la décision fournissant à un décideur ou une équipe de décideurs des indicateurs et des analyses. Ces outils permettent de faciliter l'accès aux données en ouvrant la possibilité à des analyses plus complètes. La tendance actuelle est d'aller vers les outils permettant le passage de l'information à la connaissance. On parle de l'Extraction de Connaissances à partir des Données (ECD), outil de manipulation et d'exploitation de données. L'ECD est apparu pour explorer la gigantesque quantité de données et d'en extraire des nouvelles connaissances utiles, aidant à prendre des décisions à propos de divers sujets qui touchent le domaine dans lequel on travaille. On s'intéresse donc à des Systèmes Interactifs d'Aide à la Décision (SIAD) basés sur l'ECD (appelés dans la thèse SIAD/ECD).

La modélisation de tels systèmes nécessite une démarche de conception, de développement et d'évaluation intégrant fortement la dimension humaine dans toutes ses étapes. Des modèles, méthodes et outils de conception sont disponibles en Génie Logiciel (GL). Ils permettent d'améliorer la fiabilité, l'évolutivité, la réutilisabilité et la portabilité des logiciels. Ces modèles s'avèrent souvent peu convenables quand le système en question est hautement interactif à cause du manque de prise en compte explicite et systématique de l'utilisateur. Depuis plusieurs années, différents chercheurs dans le domaine de l'Interaction Homme-Machine (IHM) travaillent sur l'enrichissement côté humain de certains processus méthodologiques, mais un long chemin reste encore à parcourir.

Le domaine de l'IHM offre des outils, des techniques, des méthodes et des modèles visant le meilleur usage possible des systèmes interactifs. Il se base sur les résultats issus respectivement, de l'ingénierie pour les aspects techniques (à savoir les méthodes d'analyse et de conception, techniques et systèmes d'acquisition, de représentation et exploitation des connaissances et des outils graphiques) et de l'ergonomie pour les facteurs humains (à savoir les méthodes d'évaluation, d'analyse des tâches et des activités humaines, des connaissances de l'homme au travail). Notre objectif est de tirer partie de ce domaine qu'est l'IHM.

Ce présent travail s'inscrit dans le cadre de la proposition d'une démarche de conception de SIAD/ECD centrée utilisateur, la démarche devant donc favoriser l'intégration des besoins

décisionnels et des besoins utilisateurs dans le processus de conception du SIAD/ECD. Suite à ce constat, nous avons pris en compte deux axes de connaissances constitutives du développement de tels systèmes, ce sont les connaissances décisionnelles (extraites par l'outil ECD) relatives au domaine d'application du système et les connaissances relatives à l'activité d'utilisation du système (modèle de l'utilisateur, analyse de tâches, etc.).

A ce sujet, nous proposons l'exploitation d'une méthode de conception existante issue du domaine du génie logiciel puis son enrichissement sous l'angle des interactions homme-machine afin de mettre en place une démarche contribuant à la conception centrée utilisateur des SIAD/ECD. La méthode que nous avons choisie est le processus unifié, et ce sont ses activités qui seront enrichies sous l'angle de l'IHM plaçant les acteurs humains en position centrale. L'approche proposée est nommée : le processus unifié étendu (ExUP pour **Extended Unified Process**).

La réalisation d'un SIAD/ECD nécessite le développement des modules interactifs qui le composent en se référant aux étapes du processus ECD. Chaque module du SIAD/ECD se comporte comme un système devant être conçu et implémenté. Pour ce faire, nous proposons d'appliquer les phases et les activités de notre approche ExUP.

Cette démarche est mise en application dans le domaine médical en vue du développement d'un SIAD/ECD pour la lutte contre les infections nosocomiales (IN) dans le service de réanimation de l'Hôpital Habib Bourguiba de Sfax, Tunisie. L'objectif de ce système à réaliser est de résoudre le problème de décision relatif à l'apparition d'une IN lors de l'hospitalisation d'un patient pouvant affaiblir ou retarder son traitement. Nous visons de ce fait la prévention de l'occurrence d'une IN chaque jour durant la période d'hospitalisation du patient dans le service de réanimation. Ceci illustre que, dans des situations journalières, des décisions sont incluses dans un plus large contexte et font partie d'un cycle décision-action qui est affecté par la surveillance et la rétroaction plutôt qu'un jugement simple. Notre travail concerne l'analyse des données temporelles rassemblées concernant les actes que subissent les patients, pour l'acquisition de la connaissance utile pour l'aide à la décision. Notre décision médicale est de ce fait dynamique ce qui donne naissance à un **Système Interactif d'Aide à la Décision Dynamique Médicale** basé sur l'ECD (SIADDM/ECD).

Le plan que nous suivrons dans ce mémoire est le suivant.

Le premier chapitre a pour but de présenter et définir les systèmes d'aide à la décision ainsi que les évolutions qu'ils ont subies. La problématique d'aide à la décision médicale fait

l'objet de la seconde partie de ce chapitre. Les différents systèmes et méthodologies de ce domaine seront présentés. Pour un SIAD, une théorie de décision doit être prise en compte. Nous optons pour l'extraction de connaissance à partir des données (ECD) ; c'est l'objet de la troisième partie de ce chapitre. L'ECD permet de découvrir les connaissances pour aider le décideur à prendre la bonne décision.

Le deuxième chapitre propose un état de l'art sur les approches de développement des systèmes présentés auparavant. Nous nous intéressons dans un premier lieu, aux processus de développement de systèmes. Nous distinguons deux grandes familles, la première concerne les processus dits « centrés architecture » (domaine du GL), et la deuxième concerne les processus dits « centrés utilisateur » (domaine de l'IHM). Dans un second lieu, nous passons en revue les méthodes d'analyse et conception. Par la suite nous exposons des exemples d'approches spécifiques pour le développement des SIAD. Ce chapitre se termine avec une évaluation des approches présentées.

Le troisième chapitre traite les fondements de notre principale contribution : le processus unifié étendu (ExUP). Ce chapitre démarre par une étude de caractéristiques essentielles du SIAD/ECD ainsi que sa décomposition possible en modules interactifs. Par la suite, nous nous focalisons sur la description de l'ExUP. Elle structure des activités enrichies sous l'angle de l'interaction homme-machine indispensable pour augmenter l'utilité et l'utilisabilité du système. Chaque activité est supportée par une démarche méthodologique intégrant des aspects itératifs, architecturaux et interactifs et mettant en avant des principes génériques à tous les modules et d'autres spécifiques à chaque module du SIAD/ECD.

La démarche de l'ExUP est mise en application pour la conception et le développement des modules de SIADDM/ECD pour la lutte contre les IN, représentant un des problèmes majeurs dans le service de réanimation de l'hôpital Habib Bourguiba de Sfax, Tunisie. Pour chaque module à développer, nous nous intéressons à l'application des phases de l'ExUP. Dans le cadre de chaque phase, nous présentons le déroulement itératif des activités enrichies ainsi que la prise en compte continue des médecins (futurs utilisateurs du système). Nous discutons aussi dans ce chapitre, l'aspect dynamique du SIADDM/ECD par le choix et l'implémentation de la technique de fouille de données qui répond au besoin de traitement des données temporelles du service de réanimation. L'objectif de ce chapitre est de pouvoir valider *à priori* le SIADDM/ECD réalisé et par conséquent, la démarche étudiée.

Le cinquième chapitre est divisé en deux parties. La première a pour but de proposer un bilan d'évaluation de l'approche proposée et de son application dans le domaine médical (SIADDM/ECD). Les résultats de cette évaluation ouvrent plusieurs perspectives de recherche en vue de l'amélioration et l'extensibilité de l'ExUP. Ces perspectives sont dressées en seconde partie de ce chapitre.

# Chapitre 1 : Etat de l'art sur les Systèmes interactifs d'aide à la décision

## 1.1. Introduction

Ce chapitre a pour but de présenter les systèmes d'aide à la décision. La première partie de ce chapitre est axée sur la définition de l'aide à la décision, son processus ainsi que les différents types de systèmes d'aide à la décision. Seuls les systèmes interactifs d'aide à la décision, tels que nous les définissons, nous intéresserons par la suite. Dans la deuxième partie, nous nous concentrerons sur les systèmes d'aide à la décision médicale qui ont fait l'objet de multiples réalisations. Un ensemble de systèmes faisant appel aux méthodes de l'intelligence artificielle (IA), se multiplie en médecine comme dans d'autres domaines scientifiques et techniques. Avant eux, des approches plus traditionnelles, basées sur des méthodes statistiques ou probabilistes, avaient été largement développées. Ce chapitre illustre l'intérêt et les possibilités des systèmes d'aide à la décision, et plus particulièrement des systèmes experts dans le domaine médical ainsi que le problème de leur intégration au sein des systèmes d'information hospitaliers. Les types de décision et les sortes de processus de prise de décision sont nombreux. Certaines décisions nécessitent l'accès à des connaissances du domaine. **Un lien peut donc être établi entre les systèmes d'aide à la décision et les systèmes de découverte de connaissances.** En conséquence, un état de l'art sur la technologie d'Extraction de Connaissances à partir de données (ECD) fera l'objet de la troisième partie de ce chapitre.

## 1.2. L'aide à la décision

L'aide à la décision est un domaine d'étude ancré dans la recherche académique. En fait, [Eom 04] indique qu'en moyenne 100 articles de recherche sont consacrés annuellement aux systèmes d'aide à la décision. L'aide à la décision est destinée à assister le décideur, l'accompagner dans sa compréhension de la situation décisionnelle en lui proposant une justification des choix retenus et en lui permettant d'évaluer les risques qu'il prend lorsqu'il adopte telle ou telle stratégie (solution) [Akharraz 04].

### **1.2.1 Processus de décision**

"Décider" ne correspond pas à une phase précise, clairement identifiable [Akharraz 04]. La décision est un cheminement : elle se construit, se négocie et suit des voies sinueuses au cours du temps [Doise 92]. L'objectif du processus de décision est de définir un espace de solution répondant à une problématique donnée, un besoin à satisfaire ou un souhait d'amélioration, de changement ou d'adaptation en tenant compte de contraintes diverses [Montmain *et al.* 03].

H.A. Simon [Simon 77] ne parle pas de la décision mais de "*decision making process*". Il propose le processus de décision IDCR (**I**ntelligence-**D**esign-**C**hoice-**R**eview) (cf. Figure 1.1). Ce processus se déroule selon quatre phases dans le processus de décision [Lévine *et al.* 89] [Turban 93] [Power 02] : la recherche d'information, la conception, le choix et la révision.

En raison de la faible capacité cognitive des hommes et la complexité de certains problèmes de décision de nos jours, les différentes phases de la décision ne se présentent pas de façon linéaire, mais en boucles. En fait, de nombreuses itérations sont nécessaires (cf. Figure 1.1) [Montmain *et al.* 03] [Akharraz 04] [Turban *et al.* 08].

#### **1.2.1.1 L'information ou le renseignement (intelligence en anglais)**

Il s'agit de définir le problème à résoudre, particulièrement les objectifs ou buts du décideur. Pour cela, il est nécessaire de rechercher les informations pertinentes en fonction des questions que se pose le décideur. L'acquisition d'informations pertinentes pendant cette phase peut se poser elle-même en termes de décision. En effet, ces informations pertinentes sont à l'origine du processus de décision et leur choix est crucial. Elles influencent fortement les autres phases puisque tous les choix suivants en découlent.

#### **1.2.1.2 La conception (design)**

Il s'agit de la génération, du développement et de l'analyse des différentes actions possibles. Le décideur construit des solutions, imagine des scénarios, ce qui peut le mener à rechercher des informations supplémentaires. De ce fait, il sera nécessaire de choisir un ou plusieurs modèles de décision en fonction de la complexité du problème à traiter. Pour le ou les modèles choisis, il faut déterminer les variables de décision, la sélection des critères d'évaluation, ainsi que les relations mathématiques ou symboliques ou qualitatives entre ces variables et construire les différentes alternatives.

### 1.2.1.3 Le choix (choice)

Durant cette phase, le décideur choisit entre les différentes solutions qu'il a été capable de construire pendant la phase de conception. Il faut donner les critères d'évaluation des différentes solutions possibles et étudier les conséquences de chaque alternative. L'évaluation des alternatives et le choix final dépendent du type de critères utilisés (trouver la meilleure alternative, c'est-à-dire assez bonne ou satisfaisante, prendre des risques ou pas, etc.). Cette phase comprend la recherche, l'évaluation et la recommandation de la solution la plus appropriée au modèle.

### 1.2.1.4 La révision (review)

Cette phase est souvent négligée, bien qu'elle nous semble extrêmement importante, en particulier dans le cas où la décision s'intégrerait dans un processus dynamique. De nouvelles informations pertinentes peuvent influencer tel ou tel choix, voir le modifier complètement.

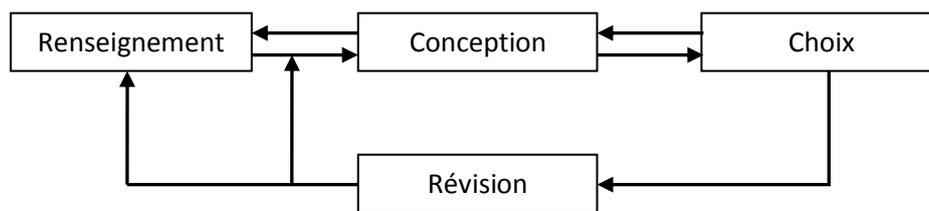


Figure 1.1 : le processus décisionnel selon le modèle IDCRC de Simon [Simon 77]

Le processus d'aide à la décision consiste à construire une décision satisfaisante et non pas une décision optimale objective. Si une décision optimale existe, il est la plupart du temps presque impossible de pouvoir prouver le caractère optimal de celle-ci. L'objectif en aide à la décision va donc être de construire une décision satisfaisante que le décideur va être capable d'argumenter [Akharraz 04].

Pour pouvoir faciliter le processus d'aide à la prise de décision, un système informatique est nécessaire. Différentes classes de systèmes informatiques d'aide à la décision existent dans la littérature. Elles font l'objet de la section suivante.

## 1.2.2 Les systèmes d'aide à la décision (SAD)

Depuis les travaux de Scott Morton [Morton 71], le domaine des SAD n'a cessé d'évoluer. Le concept de *système d'aide à la décision* est extrêmement vaste et ses définitions dépendent du point de vue de chaque auteur. Un tel système peut prendre de nombreuses formes et peut

s'utiliser de diverses manières. D'une manière générale, on peut le définir comme étant "*un système informatique qui facilite le processus de prise de décision*".

On peut opérer différentes classifications des systèmes d'aide à la décision. Mais en axant sur l'aspect opérationnel de tels systèmes, on peut adopter la typologie proposée par Turban [Turban 93] qui en distingue trois catégories :

**Les *Executive information system (EIS)*** : ce sont des outils fournissant au décideur l'information utile. Ces applications ne permettent que de recenser les informations pertinentes par l'amélioration des accès et des manipulations des informations existantes pour confronter objectifs et observations [Watson *et al.* 93] [Turban *et al.* 95] [Arnott *et al.* 07].

**Les *Executive Support System (ESS)*** : l'ESS va au delà de l'EIS car il inclut des communications, des outils d'analyse et de l'intelligence. L'ESS doit permettre l'analyse des données et permettre d'avoir une appréciation du futur de l'organisation. Ainsi pour passer de l'EIS à l'ESS ou de l'information à l'intelligence, il faut pouvoir affecter le futur (axe temporel et visualisation des liens) [Kuo 98].

**Les *Decision support System (DSS)*** : un DSS (SIAD : **S**ystème **I**nteractif d'**A**ide à la **D**écision) est un système informatique qui aide les décideurs de façon *interactive* à utiliser des données et des modèles pour résoudre des problèmes de décision. Un SIAD doit offrir un langage sophistiqué et des possibilités d'analyser les effets d'éventuels changements de l'environnement sur l'organisation [Turban *et al.* 98] [Dargam *et al.* 06].

Les deux premières catégories privilégient surtout l'étape d'Intelligence. Celle-ci est formée des activités d'exploration nécessaires pour identifier les situations où il existe un problème, ou encore les opportunités de décider et d'agir. Précisément, il permet :

- La compréhension du problème : en donnant un modèle précis de la situation avec les diverses hypothèses,
- La génération des solutions possibles à partir de la manipulation des modèles implémentés, en procurant des éclaircissements et des idées sur les solutions possibles du problème,
- La faisabilité des solutions : elles sont analysées en termes d'impact sur l'environnement.

Les SIAD sont des systèmes mettent en œuvre l'importance de l'Interaction Homme-Machine (IHM) dans la prise de décision. Les SIAD devraient être adaptés aux styles cognitifs des décideurs [Lévine *et al.* 89] [Power 02].

### 1.2.3 Les Systèmes Interactifs d'Aide à la Décision (SIAD)

De nombreuses définitions des SIAD existent dans la littérature, ces diverses définitions portent soit sur le type de problème de décision, soit sur les fonctions du système, soit sur sa constitution ou encore sur le processus de développement. Nous reprenons ici la définition de Turban [Turban 93], qui porte à la fois sur les fonctions et les composants du système.

*"Un SIAD est un système d'information interactif, flexible, adaptable et spécifiquement développé pour aider à la résolution d'un problème de décision en améliorant la prise de décision. Il utilise des données, fournit une interface utilisateur simple et autorise l'utilisateur à développer ses propres idées ou points de vue. Il peut utiliser des modèles – soit standards, soit spécifiques -, supporter les différentes phases de la prise de décision et inclure une base de connaissances".*

#### 1.2.3.1 Caractéristiques

Parmi les caractéristiques que l'on associe aux SIAD, on peut citer [Akharraz 04] :

- Ils apportent principalement une aide pour les problèmes en connectant ensemble jugements humains et données calculées.
- Ils fournissent une aide pour différentes catégories d'utilisateurs ou des groupes d'utilisateurs.
- Ils supportent des processus séquentiels ou interdépendants.
- Ils s'adaptent dans le temps.
- Ils sont suffisamment flexibles pour que le décideur soit capable de les adapter pour faire face à de nouvelles conditions (ajouter, combiner, changer et réarranger les variables du processus de décision, ainsi que les différents calculs fournissant ainsi une réponse rapide à des situations inattendues).
- Le décideur a le contrôle de toutes les étapes du processus de décision et peut à tout moment remettre en cause les recommandations faites par le SIAD. *Ce dernier doit aider le décideur et non se substituer à lui.*
- Ils utilisent des modèles. La modélisation permet d'expérimenter différentes stratégies sous différentes conditions.

- Les SIAD les plus avancés utilisent un système d'acquisition de connaissances qui apporte notamment une aide efficace et effective dans des problèmes nécessitant une expertise.

### 1.2.3.2 Composition d'un SIAD

En vue de satisfaire les critères cités ci-dessus, un SIAD se compose généralement d'une interface Homme-Machine, d'une base d'informations, d'une base de connaissances et d'une base de modèles [Holtzman 89] [Sage 91] [Turban 93] (cf. figure 1.2).

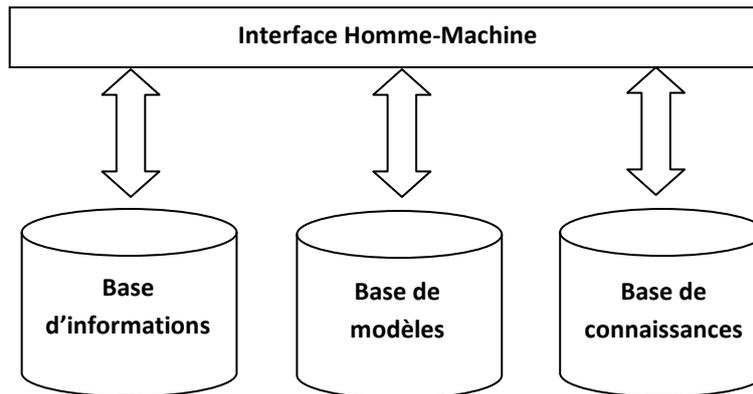


Figure 1.2 : composition d'un SIAD [Holtzman 89] [Sage 91] [Turban 93]

#### (a) L'interface Homme-Machine

L'interface Homme-Machine permet d'établir une collaboration entre le décideur et la machine. Elle est au centre du SIAD et sa réalisation est primordiale. En fait, une étude [Myers 95] a montré qu'au moins 50% du code des applications interactives correspond à l'IHM et 50 % du temps du développement est dépensé pour son implantation. C'est par l'intermédiaire des interfaces que le décideur accède aux données et aux fonctions de calcul du SIAD. Une fois les manipulations demandées par le décideur effectuées, le système lui renvoie le résultat via les interfaces. Les échanges sont d'autant plus favorisés que les représentations des résultats, tout comme le mode de questionnement du système, correspondent aux représentations mentales du décideur. Ainsi, le décideur peut exercer son contrôle et effectuer sa recherche heuristique dans de bonnes conditions. Une bonne interaction homme-machine doit permettre d'afficher les informations sous différentes formes (graphiques 2D ou 3D, textes, vidéo ou autres). Il doit aussi fournir une aide à l'utilisateur pour que ce dernier mène à bien sa tâche et il doit le guider à l'aide d'exemples précis tout en étant suffisamment flexible pour s'adapter aux besoins des différents usagers.

**(b) La base d'informations**

La base d'informations assure la fonction de mémoire, elle stocke non seulement les données, de façon permanente ou passagère, mais elle gère aussi l'enregistrement de données volatiles ainsi que l'effacement de ces mêmes données selon le souhait de l'utilisateur. Ces données volatiles correspondent aux résultats obtenus lors de traitements de données. Les données que nous avons qualifiées de permanentes sont les informations statistiques ou autres données qui décrivent les situations courantes et passées. Parmi ces données, il peut aussi y avoir des estimations concernant l'évolution de certains paramètres environnementaux.

**(c) La base de modèles**

La base de modèle se compose d'un ensemble de modèles et d'un système de gestion de ceux-ci. Les modèles peuvent être : des outils de recherche opérationnelle, des modèles statistiques ou autres. Pour avoir davantage de flexibilité, un SIAD doit posséder plusieurs modèles. Dans cette optique, le SIAD organise les liens et le passage de paramètres entre les différents modèles, de même qu'il gère le module de dialogue [Lévine *et al.* 90].

**(d) La base de connaissances**

La base de connaissance regroupe pour sa part un ensemble de connaissances sur le domaine du problème, sur les modèles et sur les stratégies de constructions des modèles. Elle permet d'apporter une aide active à la résolution du problème de décision pendant toutes les phases du processus. Elle introduit la notion d'apprentissage dans le SIAD. La base de connaissance peut aussi jouer dans certains cas le rôle de base de modèles.

### **1.2.3.3 Classification des SIAD**

Il existe plusieurs classifications des SIAD. On peut citer celle de [Hackathorn et Keen 81] qui distingue trois catégories de systèmes de ce type, mais qui sont en corrélation : (1) aide personnelle, (2) aide de groupe, et (3) aide d'organisation. Pour leur part, [Holsapple et Whinston 96] classifient les SIAD en six catégories : (1) les SIAD orientés textes (Text-oriented), (2) les SIAD orientés bases de données (Database-oriented DSS), (3) les SIAD orientés Bilan (Spreadsheet-oriented DSS), (4) les SAD orientés résolution (Solver-oriented DSS), (5) les SIAD orientés Règle (Rule-oriented DSS), et (6) les SIAD composés (Compound DSS).

Utilisant le mode d'assistance comme critère, [Power 02] différencie entre :

1. **les SIAD dirigés par les modèles** (model-driven DSS) soulignant l'accès à et la manipulation d'un modèle statistique, financier, d'optimisation, ou de simulation. Ce système utilise des données et des paramètres fournis par des utilisateurs pour aider les décideurs dans l'analyse d'une situation ;
2. **les SIAD dirigés par la communication** (communication-driven DSS) supportant plus d'une seule personne travaillant sur une tâche partagée ;
3. **les SIAD dirigés par les données (dits aussi orientés données)** (data-driven DSS) soulignant l'accès et la manipulation d'une série chronologique de données internes d'entreprise et, parfois, de données externes ;
4. **les SIAD dirigés par les documents** (document-driven DSS) qui contrôlent, recherchent, et manipulent l'information non structurée dans une variété de formats électroniques ;
5. **les SIAD dirigés par la connaissance** (knowledge-driven DSS) fournissant l'expertise de résolution des problèmes stockée comme faits, règles, procédures, ou en structures semblables. Un concept lié est l'Extraction de Connaissances à partir de Données (ECD) qui sera abordée en profondeur dans la partie I.4, et qui sera importante dans le cadre de nos recherches.

Dans notre contexte, nous nous concentrons sur les SIAD dirigés par la connaissance. De tels systèmes sont conçus pour tamiser de grands volumes de données et identifier des modèles cachés dans ces données.

Le système visé dans la cadre de cette recherche est appliqué dans le domaine médical. En conséquence, nous nous focalisons sur l'aide à la décision médicale.

### **1.3. Les systèmes interactifs d'aide à la décision médicale (SIADM)**

Les systèmes interactifs d'aide à la décision médicale (SIADM) jouent un rôle de plus en plus important dans la pratique médicale. En aidant des médecins ou d'autres professionnels du domaine médical à prendre des décisions cliniques, les SIADM exercent une influence grandissante sur le processus de soins ; on s'attend à ce qu'ils améliorent la qualité du soin médical [Lussier *et al.* 07]. Leur impact devrait s'intensifier en raison de notre capacité croissante à traiter plus de données plus efficacement [Leroy *et al.* 07].

Un SIADM combine le plus souvent plusieurs composants dans un système se voulant simple. Leroy et ses collègues [Leroy *et al.* 07] proposent pour les SIADM, la définition suivante: "*it is a software system that targets a specific group of decision-makers, uses algorithms to process and prepare the relevant data in a useful way for the decision-makers, and optionally combine these with additional knowledge sources*".

Les SIADM peuvent aider les médecins à organiser, stocker, extraire, et exploiter des connaissances médicales. On s'attend à ce qu'ils améliorent la qualité du soin en fournissant des diagnostics et des traitements plus précis, plus efficaces, et fiables, et en évitant d'éventuelles erreurs de décision [Marckmann 01] [Dreiseitl et Binder 05]. Le SIADM peut avoir un impact positif sur les performances cliniques. En plus, un SIADM peut diminuer des coûts médicaux en fournissant un diagnostic plus spécifique et plus rapide, par la prescription des médicaments les plus adéquates, et en réduisant le besoin de consultations de spécialistes [Marckmann 01]. Dans ce cadre, nous nous intéressons aux SIADM permettant de contrôler les infections nosocomiales qui constituent un défi important de la médecine moderne et qui sont considérées comme un des indicateurs les plus précis de la qualité du soin des patients [Gaynes *et al.* 99]. Dans cette section nous présentons un bref état de l'art sur les approches de SIADM en vue de positionner notre contribution, en passant en revue les différentes approches.

### **1.3.1 L'approche purement numérique de l'aide à la décision**

Cette approche qui est basée principalement sur la mise en œuvre d'algorithmes numériques (comme l'analyse discriminante ou le théorème de Bayes) a été historiquement la première utilisée dans les systèmes d'aide à la décision. Elle a donné lieu à de très nombreuses réalisations qui ne s'opposent pas mais qui complètent l'approche symbolique de l'intelligence artificielle [Degoulet *et al.* 91].

Dans ce qui suit, on s'intéresse aux méthodologies donnant naissance aux **systèmes experts**. Ces derniers ont fait appel aux méthodes de *l'intelligence artificielle (IA)*.

### **1.3.2 L'approche par méthodes d'intelligence artificielle**

L'intelligence artificielle médicale est principalement concernée par la réalisation des programmes d'IA qui font des diagnostics et des recommandations de thérapie [Rialle 95]. Ces programmes sont basés sur des modèles symboliques des entités de maladies en rapport

avec les facteurs patients et manifestations cliniques. Les systèmes experts constituent l'un des secteurs primaires dans le vaste domaine de l'intelligence artificielle. Un système expert est un système informatique basé sur la connaissance qui émule le pouvoir de prise de décision d'un expert humain [Ritchie 90] [Darmoni *et al.* 94]. Des directions principales de ces systèmes sont connues comme les SIADM basés sur le raisonnement, les SIADM basés sur l'apprentissage automatique, les SIADM basés sur les réseaux neuronaux artificiels, les SIADM basés sur les agents, etc.

### 1.3.2.1 Les différents modèles de connaissance

A la différence des types de connaissances qui sont variés, les modèles de connaissances utilisables ne sont pas très nombreux. Peter Szolovits propose de regrouper les modèles de connaissances utilisées en trois catégories [Szolovits 85] :

- **Les modèles empiriques** : les connaissances *empiriques* sont relatives aux associations entre les maladies et les signes. Elles peuvent être fournies par un expert ou dérivées de l'analyse d'une base de données. Ce type de connaissance est très souvent utilisé dans les systèmes experts.
- **Les modèles empiriques quantitatifs** : ce type de connaissance est plus complexe. Il s'agit de combiner une connaissance de type empirique et une connaissance modélisée de manière quantitative.

C'est le cas, par exemple, du système *Digitalis Therapy Advisor* développé par W. Swartout [Swartout 83]. Ces deux connaissances sont nécessaires pour calculer la dose de digitaline à prescrire en fonction de l'état du patient et des doses prescrites antérieurement. Peter Szolovits souligne l'intérêt de ce modèle et justifie l'usage d'un modèle empirique pour traduire une connaissance difficile à apprécier [Szolovits 85]. Par exemple : comment ajuster les doses devant une insuffisance rénale qui s'aggrave ?

- **Les modèles physiologiques et physio-pathologiques** : cette connaissance, lorsqu'elle existe, permet de mettre en œuvre des raisonnements plus profonds décrivant les mécanismes qui sous-tendent des processus morbides [Degoulet *et al.* 91].

*CASNET* est un exemple de système basé sur une connaissance des mécanismes physiologiques intervenant dans le glaucome [Kulikowski 82]. Les explications fournies à partir de connaissances causales sont plus faciles à comprendre pour un utilisateur n'ayant pas participé au développement du système. Dans *CASNET* deux types différents

de connaissances sont utilisés. Les connaissances causales servent à identifier des états physiopathologiques. D'autres connaissances servent à classer le cas en fonction des états précédemment confirmés ou infirmés. Ce dernier type de connaissances, appliqué à des états identifiés et non à des données de départ, est utilisé à un niveau d'abstraction plus élevé que les règles de causalité. Il a pour but de produire des conclusions diagnostiques et pronostiques.

### 1.3.2.2 Les méthodes de raisonnement

Les méthodes de résolution de problème et de raisonnement sont très variées et peuvent être mises en œuvre sur les différents modèles de connaissance. On peut avoir recours aux types de raisonnement suivants [Paolaggi *et al.* 01] [Nendaz 05] :

- Le *raisonnement déductif*, qui porte sur des données catégoriques ou des données incertaines et/ou imprécises et mettant en œuvre le principe de l'implication logique ou l'une de ses généralisations.
- Le *raisonnement hypothético-déductif*, qui est une généralisation du raisonnement par l'absurde, permet de concentrer la recherche d'une solution à un problème déterminé.
- Le *raisonnement par défaut* suppose une valeur possible pour des données incomplètes puis met en cause cette valeur lors de l'acquisition de nouvelles informations.
- Le *raisonnement qualitatif* permet de décrire des connaissances de bon sens courant. Il s'agit de prédire et expliquer le comportement d'un mécanisme physique en termes qualitatifs.
- Le *raisonnement inductif* et le *raisonnement par analogie* sont aussi utilisés.

Ces raisonnements peuvent être mis en œuvre aussi bien pour résoudre des problèmes où toutes les situations possibles sont énumérables (par exemple, les problèmes de classification dont la solution consiste à affecter un individu à une ou plusieurs classes prédéfinies<sup>1</sup>) que pour résoudre des problèmes où cette énumération n'est pas possible (il s'agit dans ce cas de construire la solution).

---

<sup>1</sup> Ces classes peuvent définir des diagnostics ou des attitudes thérapeutiques.

### 1.3.2.3 Les formalismes de représentation des connaissances

Les formalismes de représentation des connaissances les plus utilisés sont les règles de production et les objets structurés (*frames*) [Sabah 90]. Dans ce type de formalisme, on utilise souvent des règles de production pour représenter des connaissances qui s'expriment par de simples phrases conditionnelles (Si A et B Alors C).

L'architecture des systèmes qui utilisent ce type de représentation des connaissances, peut être schématisée par les trois éléments de base (cf. Figure 1.3) : une *base de connaissances* (ensemble de règles), une *base de faits* (base de données) et un *moteur d'inférence* (interpréteur de règles).

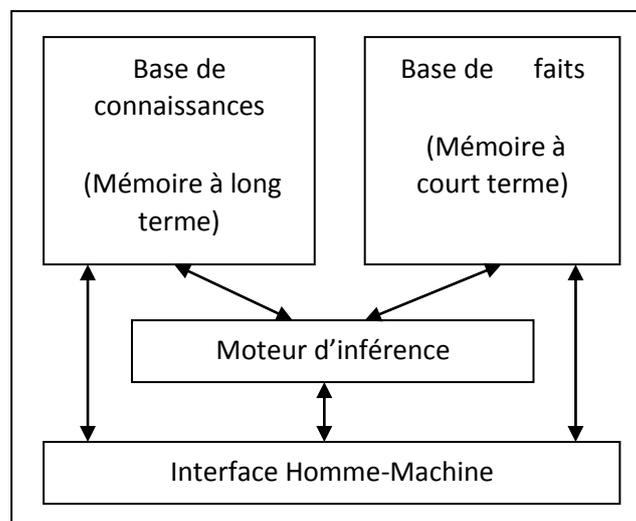


Figure 1.3 : les éléments d'un système expert [Degoulet *et al.* 91]

### 1.3.3 Logique et théorie de la décision pour la représentation de l'incertain : le raisonnement sous incertitude

La représentation de l'incertitude est l'un de très nombreux problèmes de décision médicale. Les aspects de décision les plus courants concernent la fréquence des maladies, des associations entre présence d'un signe et présence d'une maladie, l'indisponibilité d'une information utile au moment de la décision. Les formalismes les plus souvent employés pour représenter ces connaissances et les utiliser dans des raisonnements sont successivement présentés [Degoulet *et al.* 91] [Brachman *et al.* 04].

### 1.3.3.1 Les logiques non monotones

Bien que la logique classique soit un formalisme intéressant pour représenter les connaissances puisqu'il ne présuppose pas de la façon dont la connaissance sera utilisée et son mécanisme d'inférence mis en œuvre assure la validité des conclusions atteintes, elle ne permet pas d'exprimer des connaissances incertaines. Pour remédier à ce problème, d'autres systèmes logiques ont été proposés permettant d'attribuer à une proposition une *valeur de vérité* (par défaut=vrai). Celle-ci signifie que la proposition n'est pas en contradiction avec les données du problème [Reiter 80]. Une nouvelle information fournie au système pourra infirmer cette valeur par défaut et amener le retrait de la proposition. Ces systèmes logiques sont dits *non monotones*. Ils permettent d'exprimer l'incertitude par des règles d'inférence par défaut [Lavoie 07].

### 1.3.3.2 Processus bayésiens et théorie de la décision

La théorie de la décision permet aussi de prendre en compte l'incertitude. Une fois dénombré l'ensemble des actions possibles permettant de résoudre le problème, on dénombre l'ensemble des éventualités, liées au hasard, qui peuvent se produire. Au niveau de la théorie de décision, les bases axiomatiques sont celles des probabilités permettant de fournir des mesures de *crédibilité* pour les différentes propositions incertaines.

Afin de choisir les solutions du problème de décision à résoudre, on introduit des quantités appelées *utilités* pour accorder des degrés de préférences aux différentes solutions et pour choisir celle qui présente le score d'utilité maximum (*Maximum Expected Utility*) [Itzhak *et al.* 89].

L'arbre de décision est la plus commune des représentations adoptées. Elle est appropriée pour résoudre des problèmes de petite taille mais d'une utilisation difficile dans les autres cas. De ce fait, l'utilisation des *diagrammes d'influence* est intéressante dans ce cas, car ils grandissent linéairement avec la complexité du problème [Holtzman 89].

### 1.3.3.3 Imprécision et logique floue

L'imprécision des données et du vocabulaire médical sont caractéristiques des situations de décisions médicales. La logique floue fut développée par Lofti A. Zadeh en 1965 [Zadeh 65]. Elle a été proposée pour représenter les concepts médicaux et les raisonnements que l'on peut développer à partir de ces concepts imprécis [Degoulet *et al.* 91]. PROTIS est un système expert en diabétologie développé à partir de cette théorie [Soula *et al.* 87].

### **1.3.4 L'Extraction de Connaissances à partir de Données**

L'extraction de connaissance à partir de données (ECD) médicales, est un nouveau paradigme qui surmonte l'approche statistique classique [Paetz *et al.* 04]. Il permet de trouver des modèles, des tendances, et des régularités dans des grandes quantités de données médicales (stockées dans des entrepôts de données médicales). Un but des systèmes interactifs d'aide à la décision médicaux basés sur l'ECD est de capturer les données médicales qui peuvent fournir l'aide à la décision au moment du soin [Krzysztof et Moore 02] [Chen *et al.* 05] [Jinyan *et al.* 06]. En particulier l'ECD :

1. Évalue le passé (à travers les informations historisées dans l'entrepôt de données) pour déterminer si un besoin existe.
2. Projette les avantages potentiels d'une intervention réussie d'un système d'aide à la décision médicale.

On considère que la découverte de connaissance facilite la conversion des données médicales en évidence pour une future aide à la prise de décision. L'ECD combine des analyses statistiques, de l'apprentissage automatique et la technologie des bases des données pour extraire des modèles et des rapports cachés à partir de grandes bases de données [Thuraisingham 00]. Dans certains cas cette connaissance est exprimée en tant que rapports probabilistes entre les dispositifs cliniques et les solutions proposées. Dans d'autres cas une règle basée représentation est choisie afin de fournir au décideur une explication de la décision [Lavrac 99] [Villmann *et al.* 00] [Ordonez *et al.* 01]. Nous reviendrons sur l'ECD dans la partie 1.4.

### **1.3.5 Aide à la décision basée sur le Web**

L'Internet a évolué comme outil potentiellement utile pour l'étude, la diffusion, et l'exécution de directives en raison de ses normes ouvertes et de sa capacité à fournir l'information clinique concise et appropriée. Ce potentiel contingent sur les outils en développement, ouverts et librement disponibles, peut être adapté aux conditions locales, et incorporé au déroulement des opérations de médecin [Meredith 03] [Atkinson *et al.* 09].

### **1.3.6 Aide à la décision utilisant des données multimédias**

L'évolution de l'informatique, la nature et la complexité croissante des problèmes médicaux abordés amènent à l'utilisation du multimédia. Cette dernière technologie permet de fournir à l'utilisateur des informations de natures différentes qui mises ensemble fournissent une interaction plus naturelle. Le multimédia peut être exploité dans plusieurs applications pour faciliter plus efficacement l'accomplissement de différentes tâches des utilisateurs ; notons par exemple le cas de la radiologie pour faire du diagnostic avec des images IRM, des angiographiques<sup>2</sup>, 2D ou encore 3D, ... [Garlatti 96].

La tendance actuelle est à la fouille de données multimédia ; celle-ci est à la confluence des technologies des bases de données, des statistiques, de l'intelligence artificielle et de la visualisation. Elle offre des outils et des méthodologies pour la gestion, l'analyse, l'interprétation des données et la prédiction [Detyniecki 06].

Notre travail s'inscrit dans une approche interactive d'aide à la décision médicale. L'objectif de tels SIAD est d'apporter une aide interactive à des utilisateurs se trouvant face à des problèmes de décision médicale. Le développement de tels systèmes exige une connaissance réelle du domaine d'application, ce qui donne naissance à une approche d'Extraction de Connaissances à partir des Données, objet de la section suivante.

## **1.4. Extraction de Connaissances à partir de Données**

La complexité des systèmes dans lesquels l'homme est impliqué aujourd'hui conduit à l'émergence de systèmes dynamiques de traitement de l'information (STI) de plus en plus sophistiqués et incontournables où la prise de décision est de plus en plus difficile [Montmain *et al.* 03]. De ce fait, trouver la donnée utile et pertinente au bon moment devient un enjeu crucial de la décision. Les systèmes d'information dynamiques et intelligents deviennent ainsi les outils privilégiés de la décision [Akharraz 04].

---

<sup>2</sup> L'angiographie est un examen radiologique qui permet de visualiser certaines artères et certaines veines, ainsi que leurs branches de division, après utilisation d'un produit de contraste que l'on injecte dans le courant sanguin.

### 1.4.1 L'informatique décisionnelle

L'informatique décisionnelle (en anglais BI pour *Business Intelligence*) [Victor 97] désigne les moyens, les outils et les méthodes qui permettent de collecter, consolider, modéliser et restituer les données d'une entreprise en vue d'offrir une aide à la décision et de permettre aux responsables de la stratégie d'une entreprise d'avoir une vue d'ensemble de l'activité traitée.

Ce type d'application utilise en règle générale un **entrepôt de données** pour stocker des données transverses provenant de plusieurs sources hétérogènes et fait appel à des traitements lourds de type batch pour la collecte de ces informations.

La technologie d'entrepôt de données permet aux responsables et aux analystes de prendre connaissance des données à un niveau global et ainsi prendre des décisions plus pertinentes, d'où le nom d'informatique décisionnelle.

#### 1.4.1.1 L'entrepôt de données

Les données applicatives métier sont stockées dans une ou plusieurs bases de données relationnelles ou non relationnelles. Ces données sont extraites, transformées et chargées dans un entrepôt de données généralement par un outil de type ETC (*Extraction-Transformation-Chargement*).

D'un entrepôt de données (**datawarehouse**), on peut extraire un ou plusieurs magasins de données (**datamarts**). En règle générale, le datawarehouse globalise toutes les données applicatives de l'entreprise, tandis que les datamarts (généralement alimentés depuis les données du datawarehouse) sont des sous-ensembles d'informations concernant un métier particulier de l'entreprise (marketing, risque, contrôle de gestion, ...).

L'entrepôt de données constitue une grande étape des SIAD. A ce sujet, la définition proposée par Inmon, est la suivante : « *Un système datawarehouse organise et conserve les données nécessaires aux processus informationnels et analytiques dans une perspective de long terme. Ce système correspond à un ensemble de données orientées selon un sujet, intégrées, évoluant dans le temps et non volatiles, qui a pour but l'aide aux processus de prise de décision de gestion* » [Inmon 96].

Il est utile d'expliquer les différents éléments de cette définition :

- *orienté-sujet* : signifie que les données sont structurées en fonction des centres d'intérêt des managers. Afin de réaliser une telle structure de données, de nouveaux

modèles de données ont été mis en œuvre (modèle multidimensionnel en étoile ou en flocon).

- *intégré* : les données proviennent de différentes sources (par exemple, les bases transactionnelles de l'organisation) et sont réunies au sein de l'entrepôt de données.
- *évoluant avec le temps* : les données de chaque champ sont liées à une date.
- *non volatile* renvoie au fait que les données ne sont pas remplacées dans un processus de mise à jour, mais, les données sont conservées.

De ce constat, la structure de l'entrepôt de données est orientée vers la prise de décision et se positionne entre les applications transactionnelles et les SIAD.

Les potentialités offertes par ce nouveau type d'architecture ont permis aux SIAD de faire partie intégrante du système d'information et donc de fonctionner de manière nettement plus satisfaisante pour l'organisation [Lebraty 98]. Dès lors, de nouveaux outils d'aide à la décision sont apparus et notamment les outils OLAP (**O**n-**L**ine **A**nalYTical **P**rocessing - Traitement Analytique en Ligne) et ceux de l'ECD (communément appelé data mining) [Watson *et al.* 98]. La principale caractéristique de ces outils est de générer des connaissances, que ce soit via les manipulations du décideur dans le cas des cubes OLAP [Akinde *et al.* 03] ou de manière automatique via les outils d'ECD [Zhou 03].

#### **1.4.1.2 De l'OLAP à l'ECD**

Dans un entrepôt de données, les données sont organisées en fonction des besoins analytiques des utilisateurs. Contrairement aux données opérationnelles qui ont une durée de vie limitée, les entrepôts de données doivent permettre une analyse historique. Les requêtes des bases de données opérationnelles s'effectuent sous forme de transactions qui lisent et écrivent un nombre réduit de lignes dans différentes tables liées par des liens référentiels. Ce type de requêtes est dit "On line Analytical Transactional Processing" (OLTP). Au contraire, le type de requêtes effectuées sur un entrepôt de données est appelé "On line Analytical Processing" (OLAP). Les requêtes OLAP nécessitent la lecture d'une énorme quantité de données pour produire un ensemble de valeurs numériques. C'est une technologie permettant une analyse multidimensionnelle, évolutive et interactive des données.

Néanmoins, si cette technologie permet de répondre à des questions et de vérifier des hypothèses beaucoup plus complexes que les systèmes d'interrogation de Bases de Données, elle n'a pas d'imagination par elle-même : elle ne permet pas de découvrir des associations

nouvelles par exemple, sans intervention de l'utilisateur ou de l'expert. Pour cette raison, on se tourne alors vers l'extraction de connaissances à partir de données (ECD), collection de techniques et d'outils qui détectent automatiquement des tendances, des modèles, des corrélations cachées dans les données et qui, utilisant des techniques statistiques et d'Intelligence Artificielle, avec aussi la capacité d'identifier ce qui vaut la peine d'être examiné plus avant et ce qui peut être laissé de côté.

## 1.4.2 Les systèmes d'ECD

L'Extraction de Connaissance à partir de Données (ECD en français, KDD<sup>3</sup> en Anglais) est aussi connue sous les noms suivants : fouille de données, data mining (forage de données) et exploration de données [Zhu *et al.* 07]. C'est un domaine de recherche qui se situe à l'intersection de nombreuses disciplines comme l'apprentissage automatique, la reconnaissance de formes, les bases de données, les statistiques, la représentation de connaissances, l'intelligence artificielle ou encore les systèmes experts.

Grâce aux techniques en ECD [Kantardzic 03], les bases de données volumineuses deviennent potentiellement des sources riches et fiables pour la génération et la validation de connaissances. La fouille de données (Data Mining) constitue la phase centrale du processus, et consiste à appliquer des algorithmes d'apprentissage sur les données afin d'en extraire des modèles (patterns).

### 1.4.2.1 Processus d'ECD

L'Extraction de Connaissances à partir de Données est un processus interactif et itératif d'analyse d'une grande quantité de données brutes afin d'en extraire des connaissances pouvant être exploitables par un utilisateur-analyste [Fayyad *et al.* 96] [Fayyad *et al.* 01] [Kodratoff *et al.* 01]. L'interactivité est liée aux différents choix que l'utilisateur est amené à effectuer. L'itérativité est liée au fait que l'ECD est composée de plusieurs étapes et que l'utilisateur peut décider de revenir en arrière à tout moment si les résultats ne lui conviennent pas. La Figure 1.4 présente le processus d'ECD tel qu'il a été présenté par [Lefébure *et al.* 01]. Ce processus est décomposé en huit tâches distinctes qui sont décrites ci-après :

---

<sup>3</sup> Knowledge Discovery from Databases

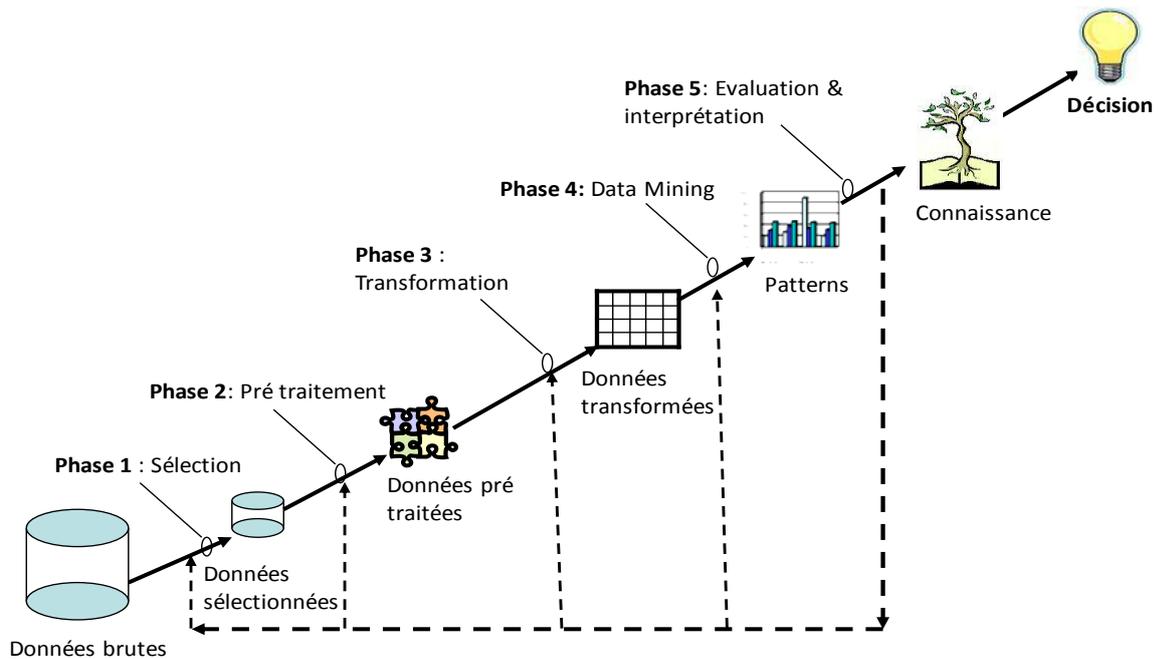


Figure 1.4 : le processus d'extraction de connaissances à partir de données

- 1. poser le problème** : avant d'engager une action, il faut d'abord bien définir le problème ainsi que son contexte.
- 2. recherche des données** : elle consiste à identifier les informations et les sources, vérifier leur qualité ainsi que leur facilité d'accès (documents papier, supports électroniques, fichiers internes ou externes, Data Warehouse ou Data Mart, etc.).
- 3. sélection des données** : les données ayant un rapport avec l'analyse demandée sont retrouvées dans la base.
- 4. nettoyage des données** : le nettoyage des données a pour objectif de corriger des inexactitudes ou erreurs de données.
- 5. transformation des données** : les données sont regroupées, normalisées, et transformées dans un format qui les prépare à la fouille.
- 6. fouille de données** : elle est définie comme l'art d'extraire des connaissances à partir des données. Il s'agit d'appliquer des méthodes intelligentes dans le but d'extraire des motifs.

- 7. évaluation du résultat :** elle consiste à mesurer l'intérêt des patterns générés. L'évaluation est généralement qualitative (facilité de compréhension du résultat) et quantitative (tel que l'intervalle de confiance<sup>4</sup>, etc.).
- 8. intégration de la connaissance :** elle consiste à implanter le modèle dans un système informatique. Des techniques de visualisation sont utilisées pour présenter clairement à l'utilisateur le savoir extrait des données.

**Les systèmes d'ECD sont des systèmes décisionnels basés sur un processus d'ECD.** Il s'agit de développer des systèmes d'information orientés vers la décision. Tout décideur ayant à prendre des décisions doit pouvoir accéder aux données de l'entreprise, doit pouvoir traiter ces données, extraire la connaissance pertinente de ces données pour prendre les "meilleures" décisions. Les techniques de fouille de données (les algorithmes de segmentation, les règles d'association, les arbres de décision, les réseaux de neurones, etc. [Wesphal *et al.* 98]), sont proposées selon le problème à résoudre. Il peut s'agir de méthodes de structuration ou de classification [Berkhin 02], d'explication ou de prédiction [Zighed *et al.* 02].

#### **1.4.2.2 Les principaux algorithmes d'ECD**

Il existe un grand nombre d'algorithmes et techniques d'ECD dans la littérature, nous présentons dans cette section ceux qui nous semblent les plus couramment employées [Lefébure *et al.* 01] :

- 1. Apprentissage fondé sur l'explication (EBL pour Explanation Based Learning) :** l'apprentissage est fondé sur des explications dérivées d'une théorie (généralement incomplète) fournie en entrée. Cette forme d'apprentissage repose sur des déductions pour expliquer les données à partir de la théorie et sur des arbres de décision pour générer de la nouvelle connaissance.
- 2. Apprentissage statistique (STL pour Statistical Learning) :** c'est un apprentissage qui repose sur des opérations statistiques telles que la classification bayésienne ou la régression pour apprendre à partir de données.

---

<sup>4</sup> L'intervalle de confiance traduit la zone d'incertitude relative au résultat d'une question d'enquête utilisant une méthode d'échantillonnage probabiliste (échantillon aléatoire).

3. **Apprentissage par réseaux neuronaux** (NNL pour Neuronal Network Learning) : un réseau de neurones est défini par un ensemble d'unités de traitement qui peuvent être des unités soit d'entrée, soit de sortie, soit cachées. L'apprentissage s'effectue par l'injection de cas en entrée et par la mesure des conclusions en sortie.
4. **Apprentissage par algorithmes génétiques** (GAL pour Genetic Algorithm Learning) : les algorithmes génétiques sont des procédures de recherche basées sur la dynamique de la génétique biologique. Ils comprennent trois opérateurs : la sélection, la combinaison et la mutation, qui sont appliqués à des générations successives d'ensembles de données.
5. **Apprentissage par similarité** (SBL pour Similarity Based Learning) : ces techniques utilisent des indicateurs de similarité pour regrouper des données ou des observations et pour définir des règles (le raisonnement à partir de cas).
6. **Apprentissage symbolique empirique** (SEL pour Symbolic Empirical Learning) : cette forme d'apprentissage extrait des règles symboliques compréhensibles par l'utilisateur à partir de données. On retrouve dans cette catégorie les algorithmes ID3/C4.5 [Quinlan 86] [Quinlan 93] et CN2 [Clark *et al.* 89].
7. **Apprentissage par analogie** (ANL pour Analogy Learning) : cet apprentissage s'appuie sur l'analogie entre un nouveau cas et des cas ressemblants soumis préalablement.

Les principales natures de problèmes qui sont résolues par ces techniques sont les suivantes :

1. **Classification** : c'est la capacité de classer des objets ou des événements comme membres de classes prédéfinies.
2. **Prédiction** : liée à la classification, cette tâche vise à prédire une ou plusieurs caractéristiques inconnues à partir d'un ensemble de caractéristiques connues.
3. **Optimisation** : il s'agit d'optimiser un ou plusieurs paramètres d'un système, compte tenu d'un ensemble de contraintes.
4. **Planning** : cette tâche consiste à déterminer un ensemble d'actions ordonnées qui satisfont un ensemble donné de buts.
5. **Ordonnancement** : l'ordonnancement suit le planning et consiste à positionner des actions dans le temps et à leur affecter des ressources.

6. **Acquisition de connaissance** : l'acquisition de connaissances consiste à créer une représentation efficace et fidèle de la connaissance d'experts.
7. **Résolution de conflits** : la résolution de conflits peut, par exemple, aider à départager des experts qui sont en désaccord ou s'appliquer dans le cadre de processus de négociation.

### 1.4.3 Manques sous l'angle de l'Interaction Homme-Machine dans les systèmes d'ECD

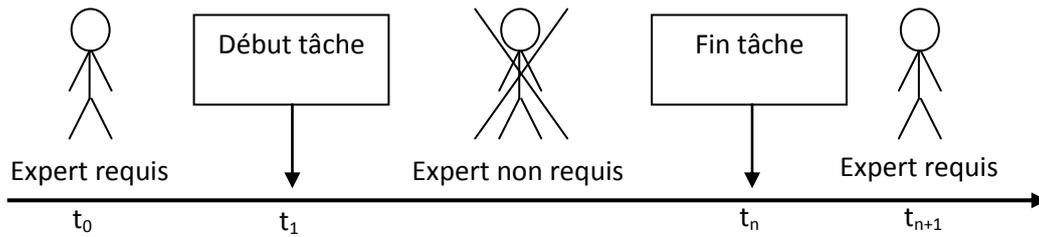
L'une des constatations faites par les chercheurs du domaine lors de la conférence KDD en 2003, est que l'interaction avec les outils existants reste limitée [Fayyad *et al.* 03]. Lors de son essor au milieu des années 90, l'ECD était censée respecter les trois « I » : Intégration, Itérativité et Interaction. Or, on observe dans la littérature que relativement peu de travaux, anciens ou récents, intègrent ces trois notions.

En fait, la notion d'interactivité n'est pas respectée en pratique tout au long d'un processus d'ECD car certaines phases restent exclusivement *système* (ne demandent aucune intervention de l'utilisateur). Nous pensons que les utilisateurs devraient être invités à intervenir au sein du processus d'ECD. Ils devraient interagir avec le processus pendant son exécution mais aussi pendant le déroulement des différentes phases qui le composent [Chevrin *et al.* 07].

Différents travaux de recherche ont pour objectif l'augmentation de la place de l'IHM dans le déroulement du processus d'ECD. Par exemple [Chevrin *et al.* 07] s'intéressent à la recherche interactive de règles d'association. En effet, les systèmes d'ECD doivent s'inspirer des notions de systèmes orientés-tâches tels qu'ils sont définis en IHM [Diaper 04] [Greenberg 04].

Dans les systèmes orientés-tâches, l'activité de l'utilisateur est décomposée en tâches, de manière à construire un modèle de tâches de l'application finale. Ces tâches constituent les briques de base de la conception de systèmes à base de connaissances. L'exécution réussie d'une tâche permet d'atteindre le but qui lui est associé.

Une tâche regroupe un ensemble de traitements et peut alors être décomposée en sous-tâches interactives, humaines ou système. [Chevrin *et al.* 07] parlent de rôle de l'utilisateur dans un processus d'ECD (cf. Figure 1.5), dont son intervention n'est sollicitée qu'entre les tâches et non pas pendant le déroulement desdites tâches.



**Figure 1.5** : rôle actuel de l'utilisateur lors d'une tâche dans un processus d'ECD [Chevrin *et al.* 07]

L'utilisateur doit pouvoir interagir facilement avec un environnement de fouille de données de manière à comprendre plus aisément les résultats qui lui sont proposés, ce qui est essentiel à la performance globale du système [Couturier *et al.* 06]

Il nous semble donc nécessaire de redéfinir le rôle de l'utilisateur, de manière à ce qu'il puisse intervenir tout au long du processus afin de le guider.

## 1.5. Conclusion

Dans ce chapitre, nous avons décrit notre problématique de recherche qui concerne l'aide à la prise de décision. Il s'agit d'un processus itératif qui se déroule en quatre phases (renseignement, conception, choix et révision) selon le modèle IDCR de Simon [Simon 77]. Les systèmes informatiques facilitant ce processus d'aide à la décision peuvent prendre de nombreuses formes allant des Executive Information Systems (EIS) fournissant au décideur l'information utile, aux Executive Support Systems (ESS) permettant d'analyser les données et d'avoir une appréciation du futur de l'organisation, en passant par les Decision Support Systems (DSS, en français **S**ystème **I**nteractif d'**A**ide à la **D**écision) aidant le décideur à utiliser des données et des modèles pour résoudre des problèmes de décision. Pour répondre à notre problématique, les DSS soulignant l'importance de l'interaction homme-machine dans la prise de décision nous semblent les plus adaptés.

Les décisions qui nous intéressent sont des décisions médicales. L'aide à la décision médicale permet d'assister un médecin dans les tâches de décisions diagnostiques ou thérapeutiques. En fait, nous nous intéressons à l'analyse de l'apport et des limites des systèmes d'aide à la décision médicale (cf. tableau 1.1).

**Tableau 1.1** : étude comparative entre les approches d'aide à la décision médicale

Approche de SIADM	Principe	Intérêt à nos travaux
Approche purement numérique d'aide à la décision	Etant donnée une hypothèse particulière, des méthodes et tests statistiques et probabilistes sont appliqués aux données pour étudier si des rapports peuvent être trouvés entre différents paramètres.	Cette approche statistique peut représenter des problèmes d'une manière simple. Mais le domaine de l'intelligence artificielle peut aller beaucoup plus loin en ajoutant <i>l'intelligence</i> aux systèmes médicaux.
Approche par méthode(s) d'Intelligence Artificielle (IA)	L'intelligence artificielle médicale est principalement concernée par la réalisation de programmes d'IA qui viennent en support des diagnostics et des recommandations de thérapie.	Ces méthodes d'IA permettent d'améliorer la qualité des soins et des prestations fournis aux patients mais pas à gérer l'hôpital ou à analyser son activité.
Raisonnement sous incertitude	Ce type d'approche permet d'exprimer l'incertitude caractérisant les problèmes médicaux.	<i>Même conclusion que l'approche par intelligence artificielle.</i>
Extraction de connaissance à partir de données	Les systèmes d'ECD permettent d'analyser et fouiller les données médicales en vue de rechercher une nouvelle et intéressante connaissance permettant de guider la résolution des problèmes médicaux. Ces systèmes combinent des méthodes statistiques, d'IA, de raisonnement sous incertitude et la technologie des bases de données.	L'extraction ou la découverte de connaissance, facilite la conversion des données médicales en évidences pour une future aide à la prise de décision. C'est ce qui nous intéresse dans le cadre de ce travail de recherche.
Aide à la décision basée sur le Web	Cette approche d'aide à la décision est conçue pour les environnements Internet/Intranet.	Notre contexte n'étant pas orientée web, cette approche n'apporte pas d'intérêt à nos travaux.
Aide à la décision utilisant des données multimédias	Elle offre des outils et des méthodologies pour la gestion, l'analyse, l'interprétation des données multimédia et la prédiction.	<i>Même conclusion que l'approche d'aide à la décision basée sur le web.</i>

En se basant sur les constatations présentées par le tableau 1.1, nous pouvons conclure que notre contexte concerne les systèmes d'aide à la décision médicale basés sur un processus d'extraction de connaissances à partir de données.

Dans ce chapitre nous avons présenté aussi les principaux concepts relatifs à notre outil décisionnel qu'est l'extraction de connaissances à partir de données (ECD). Contrairement à l'OLAP qui permet de répondre à des questions et de vérifier des hypothèses beaucoup plus complexes que les systèmes d'interrogation de bases de données, l'ECD détecte automatiquement de nouvelles connaissances (sous forme de tendances, modèles et corrélations cachées dans les données). Une connaissance extraite (rôle de l'ECD) permet de prendre des décisions plus pertinentes (rôle du SIAD). On propose donc de travailler sur des SIAD basés sur un processus d'ECD (SIAD/ECD), sujet du chapitre 3.

Le processus d'ECD est un processus itératif et interactif qui se déroule en phases allant de la sélection de données (à partir de bases de données ou d'entrepôt de données) jusqu'à l'intégration des connaissances extraites. Une phase centrale de ce processus est la fouille de données qui consiste à appliquer des algorithmes d'apprentissage pour extraire des patterns (segmentation, règles d'association, arbres de décisions, réseaux de neurones, etc.) utiles pour la résolution d'un problème de décision. C'est sur de telles étapes que repose l'acceptation ou le refus par l'utilisateur final de l'outil d'aide à la décision visé [Lefébure *et al.* 01]. Les interactions homme-machine au niveau de ce système devraient permettre de guider les utilisateurs tout au long des étapes d'ECD ; il est important aussi d'adapter au mieux l'IHM à chaque classe de décideur et/ou à chaque décideur [Lajnef *et al.* 05].

Les premiers travaux de recherche dans le domaine de l'ECD ont principalement porté sur le développement d'algorithmes automatiques performants. Actuellement, le rôle de l'utilisateur est devenu peu à peu un sujet de préoccupation majeur [Chevrin *et al.* 07]. Ce besoin d'intégration a conduit à l'émergence de nombreux outils de visualisation et d'interaction [Fayyad *et al.* 01]. Le rôle de l'utilisateur dans un processus d'ECD avait été mis en avant dès le début des années 90 dans l'article fondateur de W.J. Frawley et al. [Frawley *et al.* 92]. Ces derniers insistaient sur la nécessité d'intégrer l'utilisateur dans la "boucle de découverte" afin de combiner les potentialités humaines de jugement avec les capacités de calcul de la machine. Dans ce fait, pour la modélisation et la réalisation d'un SIAD basé sur un processus interactif d'ECD, il faut fortement intégrer le facteur humain ; c'est-à-dire suivre une démarche méthodologique centrée utilisateur.

Il nous semble donc utile d'étudier, dans le chapitre suivant, les méthodes et les modèles de conception des systèmes en général et des systèmes d'aide à la décision en particulier.

## Chapitre 2 : Conception de système interactif d'aide à la décision

### 2.1. Introduction

Depuis quelques temps, l'importance accordée aux facteurs humains dans la conception et l'utilisation des systèmes informatiques en général, des SIAD en particulier, s'accroît. En effet, le facteur humain n'est plus seulement vu comme un paramètre dans un système d'aide à la décision, mais il fait bel et bien partie du système. Son degré d'implication dans le système va faire fluctuer les performances qui vont en résulter. Sa prise en compte s'avère donc plus que jamais incontournable. En effet, l'utilisateur (le décideur) doit jouer un rôle central dans les systèmes qualifiés de décisionnels ; en pratique son potentiel n'est souvent pas exploité et ceci au détriment d'approches automatiques [Chevrin *et al.* 07]. Nous tentons ici de conceptualiser les SIAD en fonction de leurs spécificités et de leur architecture, en accordant une part importante à l'humain.

Il existe plusieurs méthodes et techniques de conception dans la littérature, destinées à l'analyse et la modélisation des systèmes. Nous passons en revue un ensemble de méthodes d'analyse et de spécification, et ceci sans souci d'exhaustivité, mais plutôt de représentativité. En fait, « *La difficulté réside dans le choix et la combinaison des méthodes et techniques, selon les caractéristiques et les objectifs du projet concerné* » [Moussa *et al.* 06].

Ce chapitre est composé de quatre sections. La première comprend une description de plusieurs processus de développement provenant respectivement du domaine du Génie Logiciel et de celui de l'Interaction Homme-Machine. Dans la seconde, on recense des méthodes de conception issues de chacun de ces domaines. La troisième section sera consacrée à l'étude d'approches existantes pour la conception des SIAD. A la fin de ce chapitre, une évaluation de ces approches par domaine sera réalisée afin d'analyser les besoins en vue de la mise au point d'une approche de développement de SIAD basé sur un processus d'Extraction de Connaissances à partir de Données (ECD).

## 2.2. Les processus de développement des systèmes, provenant du GL et de l'IHM

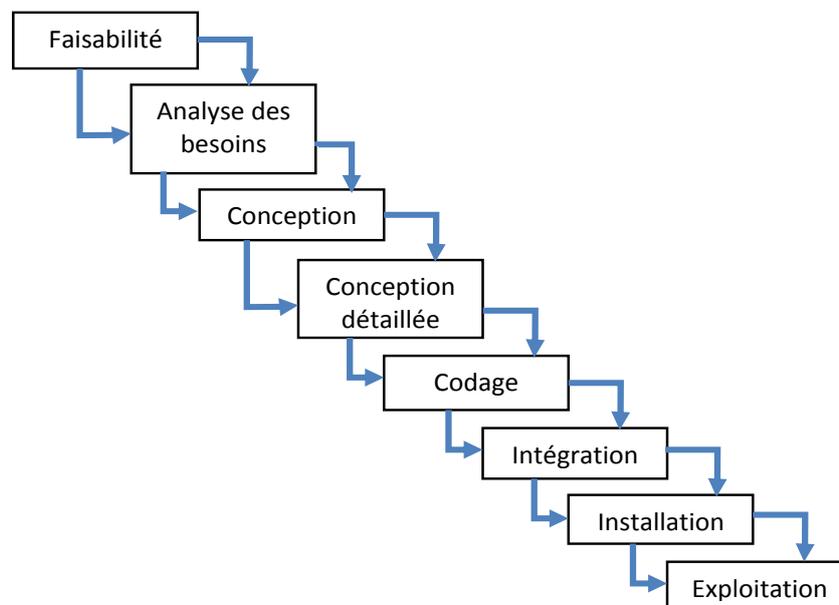
Les processus ou cycles de développement de logiciels sont souvent généraux. Ils peuvent être vus comme une façon de séparer la production d'un logiciel en un ensemble de phases plus ou moins complexes, dont l'interrelation permet de décrire le procédé, ainsi que d'indiquer la logique ou l'ordre temporel dans lequel ces phases se produisent.

### 2.2.1 Les processus de développement dans le domaine du GL

#### 2.2.1.1 Le modèle en cascade

Le modèle en cascade [Royce 70] est l'un des premiers modèles qui ont été proposés afin de satisfaire les besoins industriels en termes de qualité logicielle et de productivité dans ce domaine (cf. Figure 2.1.).

Il permet d'identifier les principales étapes par lesquelles il est nécessaire de passer pour produire et maintenir une application. Les retours ne sont possibles que vers l'étape précédente afin de prendre toutes les insuffisances identifiées en considération et de les corriger (vérification et validation). Avec ce modèle, les principaux acteurs sont l'analyste du système, le concepteur de logiciels, les testeurs et l'équipe de maintenance.



**Figure 2.1** : le modèle en cascade [Royce 70]

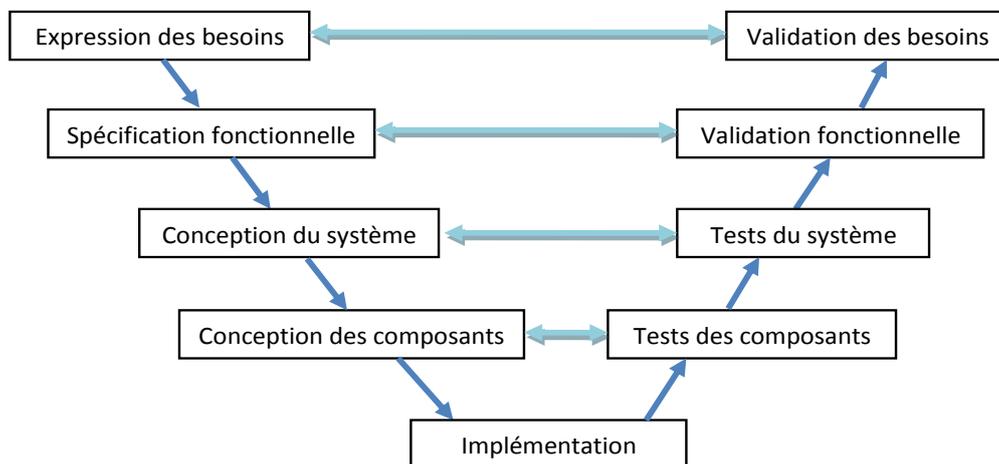
C'est un modèle générique qui n'incite pas à la prise en compte de l'IHM même lorsque le système est hautement interactif [Kolski 97].

### 2.2.1.2 Le modèle en V

Le modèle en cascade est séduisant de par sa simplicité, mais, souvent, il lui est préféré celui en V [McDermid *et al.* 84] (cf. Figure 2.2), préconisé par les organismes de qualité pour son souci d'articulation entre les activités de spécification et de conception, avec celles de validation et vérification. En effet, contrairement au modèle en cascade, ce modèle fait apparaître le fait que le début du processus de développement conditionne ses dernières étapes. Il est adapté pour les projets complexes et il structure les étapes du cycle, qui demeurent identiques globalement à ceux du modèle en cascade, dans deux processus :

1. Un sous processus descendant visant la spécification et la conception,
2. Un sous processus ascendant ayant pour objet les validations et les tests.

Le plan, les moyens et les méthodes d'évaluation et de validation des résultats de la phase doivent être inclus dans chaque phase de l'approche allant de haut en bas. Ce souci, avec celui de l'évaluation du système aussi loin en amont que possible, et avec précision pour chaque phase, est un point indéniablement fort du modèle en V. Cependant, il prévoit seulement des retours très limités, qui peuvent être insuffisants pour une conception itérative.



**Figure 2.2** : le modèle en V [McDermid *et al.* 84]

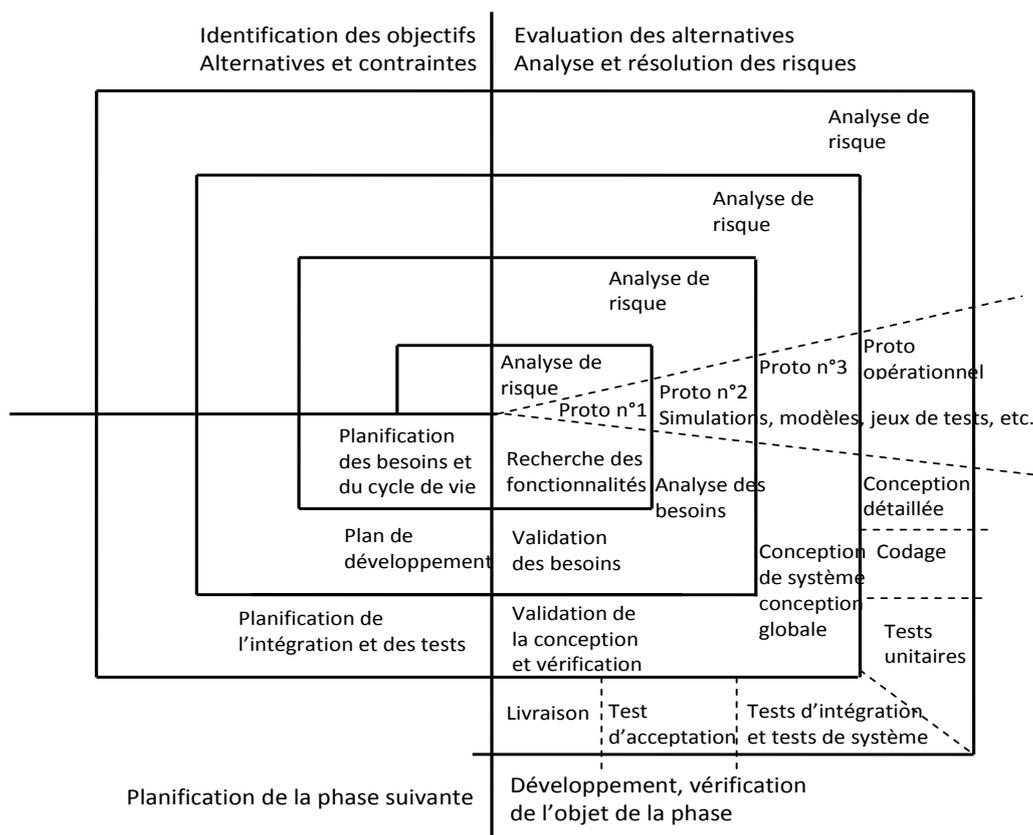
Ce modèle est intéressant pour son incitation à bien préparer l'évaluation du système, mais il n'introduit pas de considérations d'IHM si bien qu'il reste dirigé souvent principalement vers l'angle technique du système visé.

### 2.2.1.3 Le modèle en spirale

Le modèle en spirale [Boehm 88] s'appuie sur une succession de cycles dont chacun se déroule en quatre phases (cf. Figure 2.3). Ce modèle de cycle de vie tient compte de la possibilité de réévaluation des risques en cours de développement. Il couvre l'ensemble du cycle de développement d'un produit. Il met l'accent sur l'activité d'analyse des risques.

La démarche poursuivie par ce modèle est la suivante :

- Identifier les risques et leur affecter une priorité,
- Développer une série de prototypes pour identifier les risques en commençant par le plus grand,
- Si un cycle concernant un risque a été terminé avec succès :
  - Evaluer le résultat du cycle en cours et planifier le cycle suivant,
  - Si un risque n'a pu être résolu, terminer le projet immédiatement.



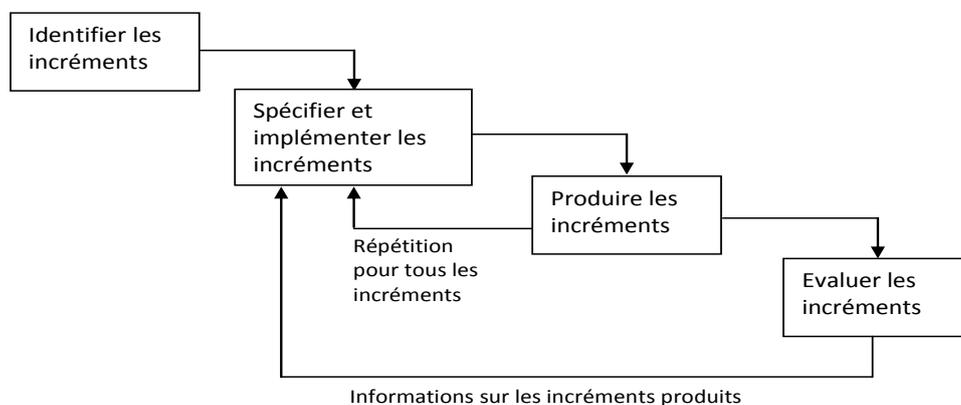
**Figure 2.3 :** le modèle en spirale [Boehm 88]

À chaque cycle, les prototypes créés permettent de réduire les risques et de guider la conception pour obtenir un système qui répond au besoin de l'utilisateur. Chaque cycle fait en sorte que le système soit de plus en plus complet.

Ce cycle est itératif et introduit la notion de prototypage et de raffinement, ce qui le rend très intéressant. Néanmoins, comme les autres processus, il n'invite pas explicitement à la prise en compte de l'utilisateur durant le déroulement du projet.

#### 2.2.1.4 Modèle par incréments (ou incrémental)

Le modèle incrémental suit le modèle en cascade mais l'implantation se fait par incréments : à partir d'une phase donnée (généralement la spécification ou la conception architecturale), le processus est itéré plusieurs fois, et donne lieu à chaque fois, à la production d'un incrément (cf. Figure 2.4). Chaque incrément correspond à un logiciel opérationnel s'approchant à chaque fois davantage du produit final par adjonction de fonctionnalités ; les évolutions entre les incréments sont guidées par l'expérience opérationnelle [ESA 91] [Laprie *et al.* 95].



**Figure 2.4** : modèle incrémental [ESA 91] [Laprie *et al.* 95]

Les inconvénients de ce modèle sont relatifs aux cas de remise en cause du noyau ou d'incrément et à la difficulté d'intégration.

#### 2.2.1.5 Modèle transformationnel

Ce modèle indique une séquence d'étapes qui transforment graduellement une spécification en une implantation (cf. Figure 2.5). De cette façon, chaque transformation est prouvée formellement. Ces preuves sont stockées dans un document. Ce processus exige une certaine expertise en méthodes formelles pour le développement de systèmes complexes d'une façon correcte et documentée formellement. Ce processus est basé sur des outils d'aide à la preuve ; il doit aussi permettre de développer des composants réutilisables. Cette approche est surtout destinée aux systèmes à risques comme les systèmes embarqués et ne concerne pas directement nos travaux sur les systèmes interactifs d'aide à la décision. Les méthodes

supports à ce processus sont le plus souvent les méthodes/notations Z [Spivey 89], VDM (Vienna Development Method) [Jones 90] ou B [Abrial 96].

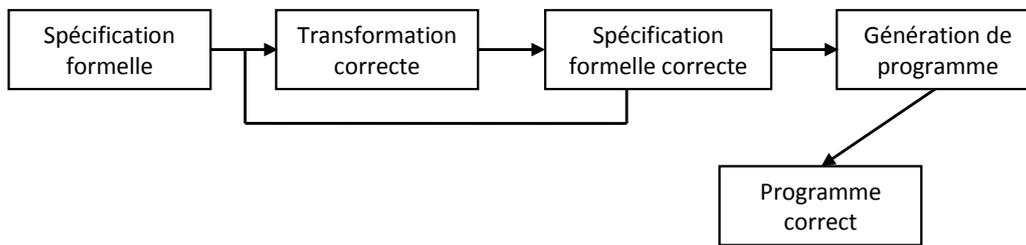


Figure 2.5 : modèle transformationnel

### 2.2.1.6 Le modèle en Y

Le modèle en Y, proposé par [André 93] et cité et adapté dans [Larvet 94], a inspiré [Hassine *et al.* 02] pour aboutir au modèle présenté dans la figure suivante. C'est un modèle composé de trois branches :

- La branche fonctionnelle permettant d'identifier et analyser les composants métier (CM)
- La branche technique permettant d'identifier et analyser les composants techniques (CT)
- Et la branche centrale du "Y" (conception, codage, etc.) : dans laquelle les composants conceptuels (CM et CT) sont progressivement transformés en composants logiciels.

Ce modèle apporte une aide à la maîtrise du développement et à l'optimisation des coûts et des délais en parallélisant les tâches d'analyse et de conception [Lepreux 05].

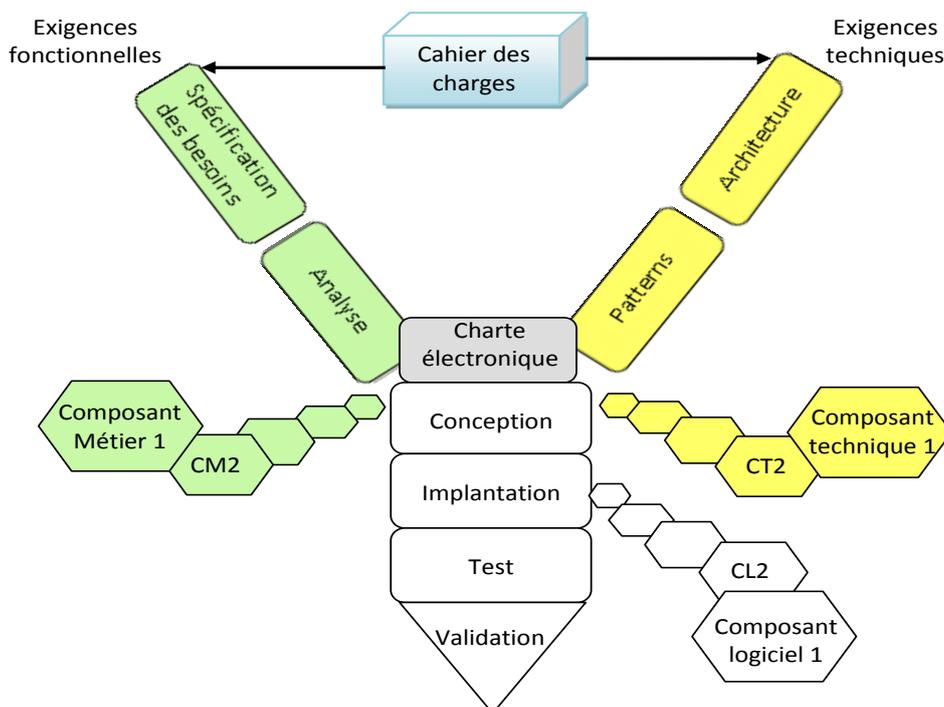


Figure 2.6 : modèle en Y [Hassine *et al.* 02]

Le modèle en Y est un processus de développement par réutilisation ; il prône en effet l'utilisation des composants réutilisables pour concevoir et réaliser un système. Néanmoins, il ne prend pas en compte l'aspect IHM.

### 2.2.1.7 Les méthodes agiles

La démarche de conception en spirale a donné naissance à un courant méthodologique autour des méthodes dites « agiles » [Vickoff 09] telles que RAD [Martin 91], DSDM [Stapleton 97] et XP [Médina 00]. L'agilité des méthodes fait référence à leur capacité à s'adapter aux particularités de chaque projet et aux obstacles de chaque phase du cycle de vie.

**RAD (Rapid Application Development)** : RAD est une méthode de développement rapide d'applications [Martin 91] qui s'appuie sur une implication forte de l'utilisateur (cf. Figure 2.7). Il permet de construire le système avec l'utilisateur, en étant totalement axé sur des prototypages successifs. La question de l'identification des besoins devient alors primordiale. RAD préconise d'organiser la participation des utilisateurs tout au long du processus de conception en fournissant des supports à leur prise de décision, tout en leur faisant intégrer l'objectif délai pour focaliser l'expression des besoins vers l'essentiel. Les utilisateurs effectuent la collecte des informations nécessaires à l'expression des besoins. Ils se réunissent ensuite en session avec les informaticiens à qu'ils présentent leurs résultats qui pourront être alors confrontés, amendés et acceptés [Morley 06].

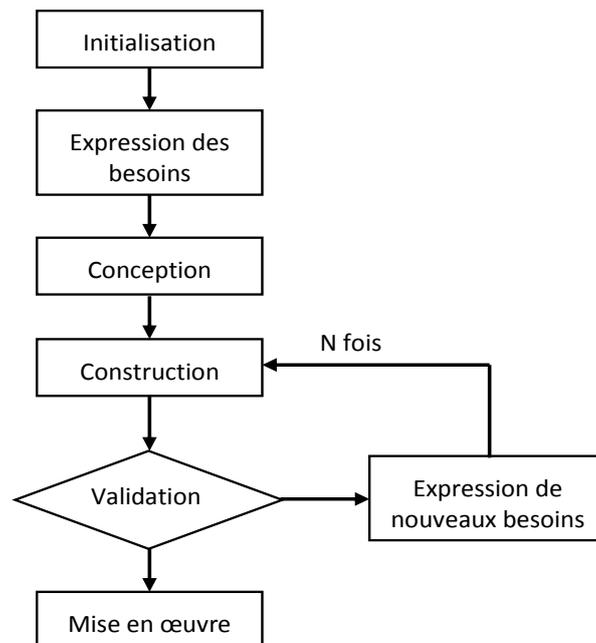


Figure 2.7 : le cycle de développement RAD [Martin 91]

Cependant, l'approche itérative RAD pose trois exigences majeures et difficiles à satisfaire :

- L'exigence d'une grande implication des acteurs qu'il est difficile d'obtenir au même degré pour tous les projets ;
- La nécessité d'une équipe experte capable de proposer dans un temps réduit des solutions aux demandes des utilisateurs ;
- La nécessité de s'appuyer sur un solide atelier de génie logiciel qui garantisse un passage rapide du concept au prototype.

**DSDM** (Dynamic Software Development Method) : la méthode DSDM [Stapleton 97] a été mise au point en s'appuyant sur la méthode RAD afin de combler certaines de ses lacunes, notamment en offrant un canevas prenant en compte l'ensemble du cycle de développement.

La méthode DSDM se base sur des principes d'implication des utilisateurs, de développement itératif et incrémental, de fréquence de livraison élevée, d'intégration des tests au sein de chaque étape ainsi que d'acceptation des produits livrés dépendant directement de la satisfaction des besoins.

**XP** (eXtreme Programming) : la méthode XP [Médina 00] définit un certain nombre de bonnes pratiques permettant de développer un logiciel dans des conditions optimales en plaçant le client au cœur du processus de développement, en relation étroite avec lui. XP est basé sur les concepts suivants :

- Les équipes de développement travaillent directement avec l'utilisateur sur des cycles très courts d'une à deux semaines maximum.
- Les livraisons de différentes versions du logiciel interviennent très tôt et à une fréquence élevée pour maximiser l'impact des retours utilisateurs.
- L'équipe de développement travaille en collaboration totale sur la base de binômes.
- Le code est testé et nettoyé tout au long du processus de développement.
- Des indicateurs permettent de mesurer l'avancement du projet afin de permettre de mettre à jour le plan de développement.

Dans les méthodes agiles les utilisateurs sont certes fortement impliqués dans la conception du système d'information mais toutefois encouragés à établir des priorités dans leurs attentes en fonction des délais du projet. Cette stratégie contraint l'expression des besoins car l'utilisateur doit jouer le double rôle d'utilisateur/décideur [Mouloudi 07].

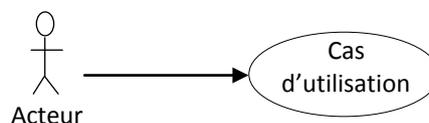
### 2.2.1.8 Le processus Unifié basé sur UML

Le processus de développement orienté objets UP (Unified Process) offre un cadre méthodologique générique qui repose sur la notation objet UML. Ce processus est utilisé dans l'industrie du développement logiciel à une telle échelle qu'il y est perçu comme une norme de fait [Lemieux *et al.* 06].

Le Processus Unifié (PU) est un ensemble de principes génériques adaptables en fonction des spécificités des projets. Il s'agit donc en quelque sorte d'un patron de processus pouvant être adapté à une large classe de systèmes logiciels, à différents domaines d'application, types d'entreprises, niveaux de compétences et à différentes tailles de projets.

Le PU est piloté par les cas d'utilisation, centré sur l'architecture, itératif et incrémental [Jacobson *et al.* 99] :

Piloté par les cas d'utilisation : l'objectif principal d'un système logiciel est de rendre service à ses utilisateurs ; il faut donc bien comprendre les désirs et les besoins des futurs utilisateurs. Le terme utilisateur désigne les utilisateurs humains mais également les autres systèmes, l'utilisateur représente donc une personne ou un système dialoguant avec le système en cours de développement. Ce type d'interaction est appelé cas d'utilisation (cf. Figure 2.8).



**Figure 2.8** : cas d'utilisation

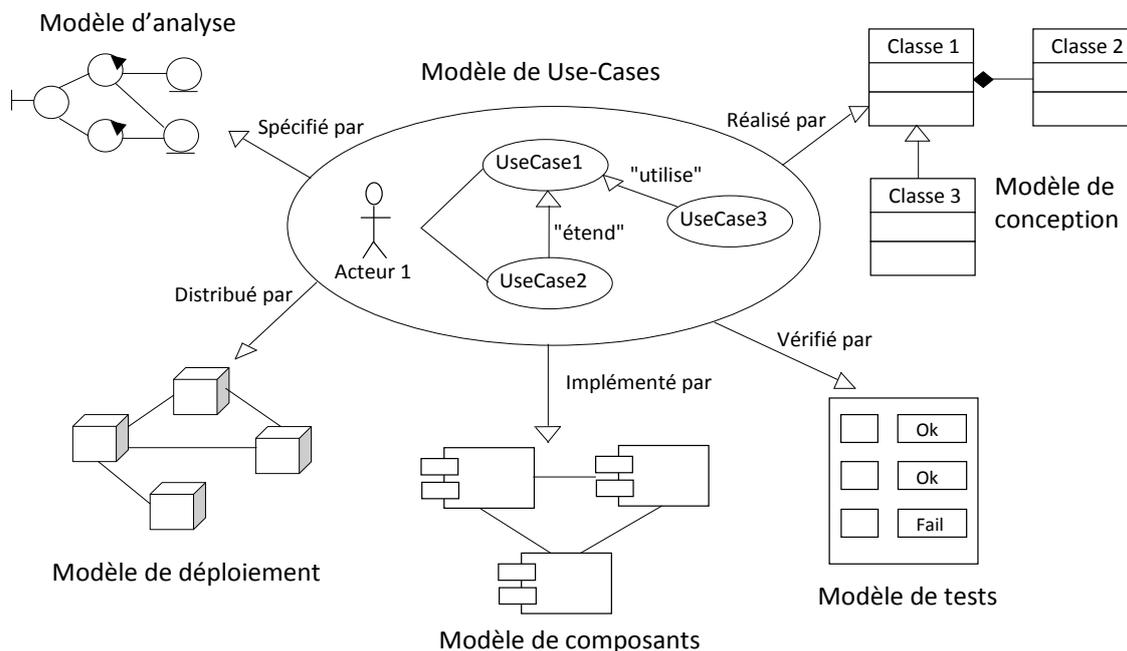
Un cas d'utilisation est une séquence de transactions effectuées par le système et qui procure un résultat significatif et mesurable pour un acteur donné [Roques 05] (cf. Figure 2.9).

Les cas d'utilisation font apparaître les besoins fonctionnels et leur ensemble constitue le modèle des cas d'utilisation qui décrit les fonctionnalités complètes du système.

Centré sur l'architecture : l'architecture d'un système logiciel peut être décrite comme étant les différentes vues du système. À partir d'une vision de haut niveau des besoins, l'architecte se focalise sur une partie en l'affinant et en créant un système. Après avoir affiné successivement les modèles produits, l'architecture se précise.

L'architecture fournit la structure qui servira de cadre au travail effectué au cours des itérations, tandis que les cas d'utilisation définissent les objectifs et orientent le travail de

chaque itération. La décomposition suivant un mode itératif et incrémental permet de mieux gérer les risques et de limiter la complexité des applications.



**Figure 2.9** : structuration d'un processus unifié autour des cas d'utilisation [Octo 02]

**Itératif et incrémental** : dans le but de réduire la complexité en la maîtrisant, un projet informatique se trouve décomposé en sous projets qui représentent chacun une itération. Les itérations montrent des étapes de l'enchaînement d'activités, tandis que les incréments conviennent à des stades de développement du produit.

Le choix de ce qui doit être implémenté au cours d'une itération repose sur deux facteurs :

- Une itération prend en compte un certain nombre de cas d'utilisation qui ensemble, améliorent l'utilisabilité du produit à un certain stade de développement.
- L'itération traite en priorité les risques majeurs.

Un incrément constitue souvent un additif.

Les aspects de modélisation en analyse et conception se basent sur le langage UML (Unified Modelling Language) [Rumbaugh *et al.* 99] [Larman 05]. Ce processus comprend quatre phases (cf. Figure 2.10) :

- L'initialisation (ou l'Inception) pour définir l'étendue du projet par l'utilisation de cas d'utilisation et l'étude de faisabilité,
- L'élaboration pour définir les besoins et spécifier l'architecture,

- La construction au cours de laquelle le logiciel est bâti au moyen de plusieurs itérations et de nombreuses versions du système, et
- La transition pour remettre le système aux utilisateurs finaux avec la mise en service et pour les former et les soutenir à l'utilisation [Jacobson *et al.* 99] [Larman 05].

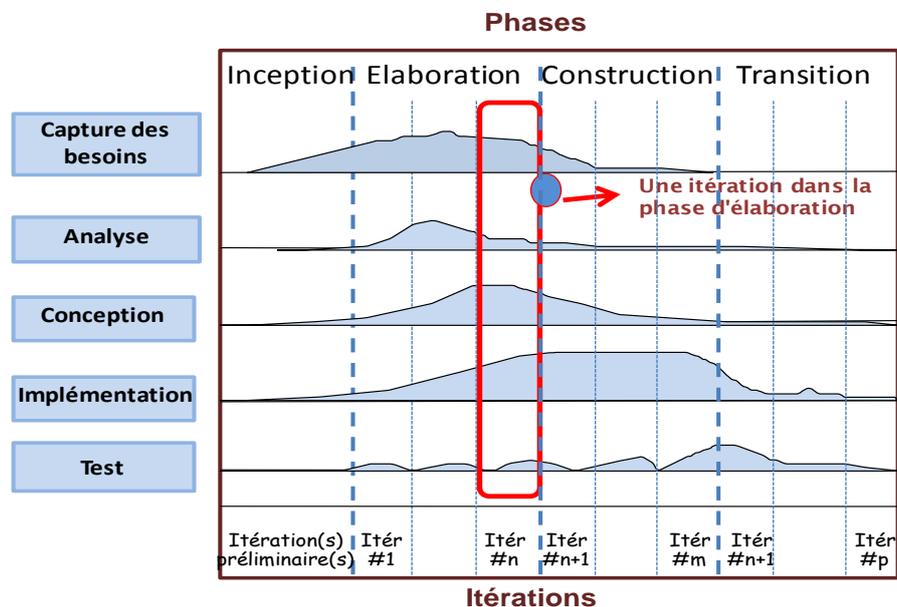


Figure 2.10 : le processus Unifié [Jacobson *et al.* 99]

Parmi les principes du Processus Unifié (PU), on peut parler de son développement itératif, la spécification du dialogue entre les différents intervenants du projet : les livrables, les plannings, les prototypes... En plus, il propose des modèles de documents, et des canevas pour des projets types. L'intégration de phases de prototypage et de validation successives favorise en principe l'implication des utilisateurs au cours de la conception.

Toutefois le processus unifié est coûteux à personnaliser, très axé processus, au détriment du développement. Il donne peu de place pour le code et la technologie. On notera aussi un article très intéressant de Lemieux et Desmarais montrant les nombreuses limites d'UP vis-à-vis de la conception centrée utilisateur [Lemieux *et al.* 06]. La principale observation est que l'intérêt théorique du pilotage par les cas d'utilisation peut être altéré par le manque de formalisation de leur définition.

### 2.2.1.9 Conclusion sur les modèles étudiés

Les modèles cités auparavant, ont un objectif commun qui est la production de logiciels de qualité. Mais on s'aperçoit que les modèles les plus classiques sont bien trop souvent orientés vers la partie technique (le code) et non vers l'utilisateur, excepté jusqu'à une certaine mesure

seulement le modèle PU (par l'utilisation surtout des cas d'utilisation). En effet, ils sont orientés système et n'intègrent pas explicitement par exemple l'analyse et la modélisation des tâches potentielles des utilisateurs. Ces aspects sont laissés à l'appréciation du concepteur en particulier, la prise en compte explicite des facteurs humains en général.

En plus, nous pouvons remarquer que la tendance va aux processus itératifs (spirale, méthodes agiles, PU, etc.). Et même si les utilisateurs sont mentionnés pour les étapes d'analyse et de validation de prototype, les modèles et processus s'accompagnent le plus souvent de peu d'explications relatives à la prise en compte des utilisateurs. La conception et l'évaluation des IHM sont rarement spécifiées dans ces processus.

Sous l'angle du développement d'un système interactif, il est important de prendre en compte finement les aspects liés à l'utilisateur. Dans la mesure où les processus de développement issus de GL sont jugés insuffisants pour la prise en compte de l'utilisateur dans la conception d'un système interactif, il y a eu, depuis plus d'une vingtaine d'années, un enrichissement de modèles classiques en essayant d'intégrer la dimension humaine, tout en favorisant le prototypage. On parle de modèles enrichis sous l'angle de l'IHM.

### **2.2.2 Les modèles de développement enrichis sous l'angle de l'IHM**

Les modèles précédents ne sont pas directement applicables au domaine de l'IHM. En effet, loin de prendre en compte l'utilisateur, ces modèles génériques restent souvent axés sur la qualité du code produit. Ils ne sont pas à écarter, mais bien au contraire, il est souhaitable de les adapter pour favoriser une démarche centrée sur l'utilisateur [Mouloudi et 07] [Mouloudi *et al.* 11]. C'est ainsi qu'ont été proposés les modèles enrichis sous l'angle de l'IHM positionnant des notions le plus souvent inexistantes dans les modèles de développement issus du GL et rendant nécessaire la reconsidération de leur structure et organisation [Kolski *et al.* 01].

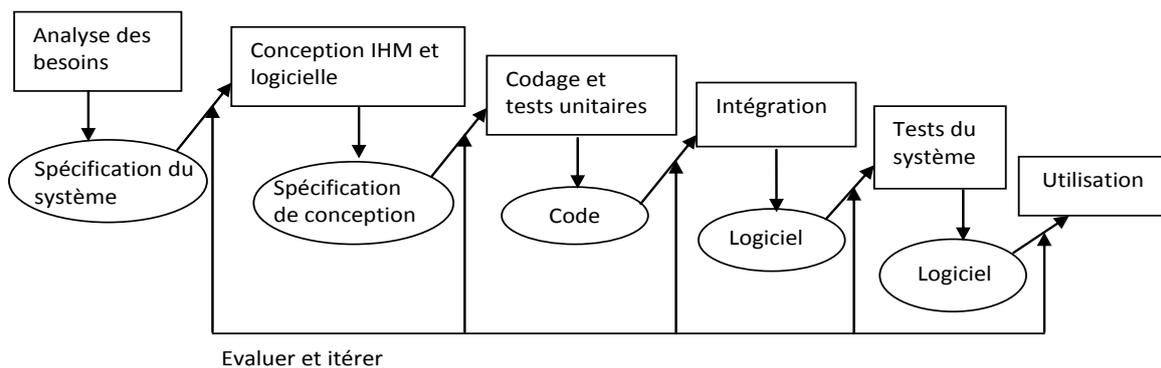
Cependant, les modèles suggérés ne prétendent pas nécessairement fournir toute l'assurance qu'un projet visant la conception et le développement d'un système interactif soit réalisé avec un succès total. Le souci principal de ces modèles doit surtout souligner, du point de vue méthodologique, des aspects fondamentaux tels que la modélisation des tâches humaines, le développement itératif des prototypes et l'évaluation du système interactif.

Quoique ces modèles aient des limitations possibles, il est intéressant d'examiner certains d'entre eux pour identifier la spécificité de chacun en ce qui concerne l'aspect interactif. Ces

modèles seront considérés en tant que cadres théoriques et méthodologiques pour un processus interactif de développement de système interactif ; aucun d'entre n'est parfait. Ils sont composés de phases basées sur celles proposées par les modèles issus du GL.

### 2.2.2.1 Modèle de Long

Ce modèle [Long *et al.* 90] est proche du modèle en cascade du génie logiciel classique (cf. Figure 2.11), tout en positionnant l'IHM dans l'étape de conception, et en insistant sur l'importance de l'évaluation et des itérations lors du projet.



**Figure 2.11** : modèle proposé par Long *et al.* [Long *et al.* 90]

Loin d'être parfait, ce modèle s'avère toutefois incitateur sous l'angle des interactions homme-machine, tout en restant très proche d'un modèle classique.

### 2.2.2.2 Le prototypage

La Figure 2.12 présente le processus de développement fondé sur le prototypage, présenté dans [Rettig 94]. Dans cette approche, des prototypes sont construits jusqu'à ce qu'une bonne compréhension des besoins du logiciel soit mise en évidence. Comme dans les processus classiques, on commence ici par l'étape de collecte des besoins, puis on procède à une conception rapide afin de construire un prototype, qui est évalué par l'utilisateur, pour décider s'il convient, auquel cas on peut obtenir deux types de sorties :

- Soit on explicite les besoins finaux, voir la sortie gris clair du test « adéquat ? » dans la Figure 2.12, (on parle dans ce cas-là de prototypage basse-fidélité) ;
- Soit on produit le système interactif final ; cf. en sortie noire du test « adéquat ? » dans la Figure 2.12, (on parle dans ce cas-là de prototypage haute-fidélité).

Si tel n'est pas le cas, on reprend à l'étape de prototypage.

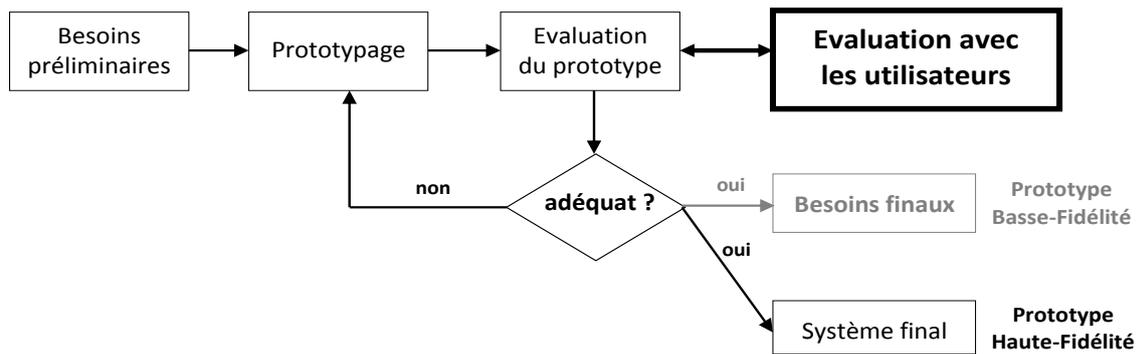


Figure 2.12 : le prototypage [Rettig 94]

Ce processus permet de construire un système interactif en permettant de faire plusieurs tentatives jusqu'à ce que les besoins soient correctement identifiés. Il met ainsi en avant le bien fondé de l'itération des différentes phases d'un processus de développement, et surtout l'intérêt de l'intervention de l'utilisateur au cours du processus de conception.

Par contre, il faut noter que la qualité du logiciel produit, dans le cas du prototypage haute-fidélité, est inversement proportionnelle au nombre d'itérations y conduisant, grâce à l'agrégation de défauts (ou *bugs*) à chaque itération.

### 2.2.2.3 Modèle en couches

Le principe général du modèle en couches, issu de [Curtis *et al.* 94], est de proposer une séparation en deux parties de chaque phase du développement, une partie concernant le développement classique de logiciel (partie de droite), et une partie concernant le problème de l'interaction homme-machine (partie de gauche) (cf. Figure 2.13).

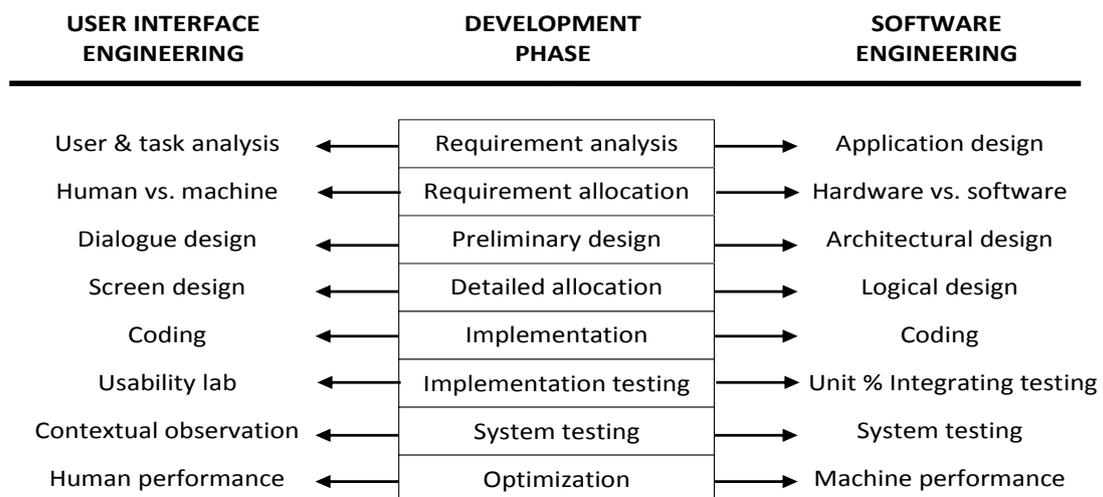


Figure 2.13 : le modèle en couches [Curtis *et al.* 94]

Un des avantages de ce processus est qu'il précise, pour chaque phase du développement, ce qui est à faire pour la partie fonctionnelle de l'application et sur sa partie interactive, en conservant une séparation claire entre ces deux parties d'un système interactif.

Cette approche présente deux inconvénients :

- aucune mise en relation explicite des deux parties n'est faite ;
- le modèle est trop séquentiel, perdant les avantages d'un processus itératif comme le prototypage.

#### 2.2.2.4 Le modèle en étoile

Le modèle en étoile [Hix *et al.* 93] situe l'évaluation au centre même du cycle complet. Ceci permet des interactions/itérations entre chacune des étapes. L'étape d'évaluation est vue comme une étape intermédiaire ; elle permet de protéger l'équipe de développement contre un rejet final (cf. Figure 2.14). Particulièrement ouvert, ce modèle est très flexible, voire un peu trop. Il n'impose pas un ordre dans lequel les étapes du processus doivent être exécutées, bien que dans la pratique les activités de développement soient placées à la fin du cycle. Il est à noter qu'il implique une conception participative qui vise la détection des problèmes d'utilisabilité et qu'il exige un degré élevé d'implication de l'utilisateur [Lepreux 05].

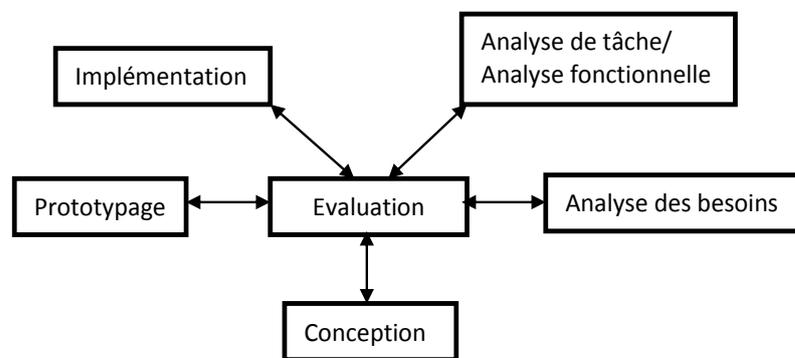


Figure 2.14 : le modèle en étoile [Hix *et al.* 93]

#### 2.2.2.5 Le modèle en V enrichi

Le GL propose un cadre structurant du processus de développement des systèmes informatiques. Dans ce cadre, l'ergonomie s'inscrit en des étapes bien précises avec ses méthodes et techniques. La figure suivante illustre cette relation pour le cas simplificateur, mais réaliste, du modèle en V [McDermid *et al.* 84].

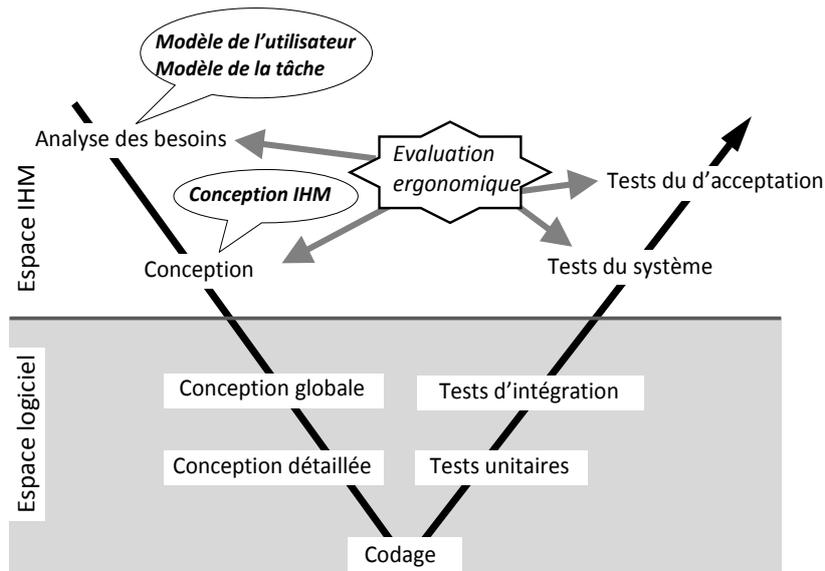


Figure 2.15 : le modèle en V enrichi [Balbo 94]

La Figure 2.15 montre la répartition des étapes du processus de développement en deux espaces : l'espace IHM et l'espace logiciel. Le premier se caractérise par la priorité qu'il faudrait accorder aux aspects ergonomiques, le second par l'accent mis sur les techniques d'implémentation logicielle.

En effet, l'espace IHM inclut l'analyse des besoins, la conception et les tests du système et d'acceptation. Dans l'analyse des besoins, il faut inscrire l'apport de l'ergonomie et de la psychologie pour l'analyse de tâche et la modélisation de l'utilisateur. Dans l'étape de conception intervient l'ergonomie pour la définition et la spécification de l'interface utilisateur. Les tests système et d'acceptation ainsi que l'évaluation des interfaces utilisateur devraient nécessairement faire appel à des techniques d'évaluation.

L'espace logiciel laisse la place aux compétences informatiques avec les conceptions globales et détaillées, le codage et les tests unitaires et d'intégration. Par souci de clarification, on appelle **conception logicielle**, les activités de conception globale et détaillée. Alors qu'on appelle **conception d'IHM**, l'activité amont qui implique une analyse centrée sur l'utilisateur [Balbo 94].

Néanmoins, le modèle en V enrichi autorise des retours arrière limités, ce qui peut nuire à l'aspect souvent nécessairement itératif d'une conception centrée utilisateur.

### 2.2.2.6 Le modèle de Collins

Le modèle présenté en Figure 2.16 est un modèle itératif proposé par [Collins 95]. Ce modèle accorde une grande importance à l'analyse des tâches, reconnaissant ainsi l'importance de cette activité dans une conception centrée sur l'utilisateur. Il en résulte une tentative d'intégration des facteurs humains et du GL.

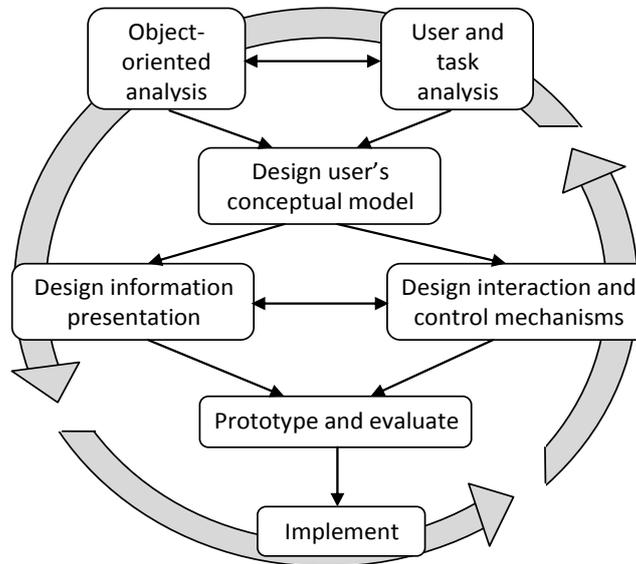


Figure 2.16 : le modèle de Collins [Collins 95]

Malheureusement, le processus reste confus et fournit plus une description de cette relation qu'une façon de la respecter.

### 2.2.2.7 Le modèle Nabla ∇

Le modèle Nabla [Kolski 97] [Kolski 98] est dédié à la conception de systèmes interactifs et il intègre de façon importante le futur utilisateur dans le cycle de vie du projet.

Ce modèle ∇ construit selon un double cycle en V, situe les différentes étapes du GL nécessaires pour développer un système interactif, tout en différenciant l'interface proprement dite (partie gauche du modèle) des modules d'aide (ou applicatifs) éventuellement accessibles à partir de ceux-ci (partie droite). Nabla se base sur une confrontation progressive entre un modèle réel et un modèle de référence, où le modèle de référence correspond à celui d'un système Homme-Machine dit idéal, en considérant les points de vue et besoins des différents intervenants concernés par le système Homme-Machine visé. Le résultat de cette confrontation conduit à identifier les données pertinentes pour spécifier un système interactif adapté aux besoins informationnels des utilisateurs et aux besoins en mode de coopération utilisateur-modules d'aide. L'ensemble des spécifications est ensuite évalué et validé d'un

point de vue socio-ergonomique, afin de vérifier la pertinence de l'intégration des solutions nouvelles dans le système Homme-Machine visé.

L'évaluation se situe au centre du projet suggérant une démarche itérative aussi bien dans les parties gauche que droite. Elle se termine par l'étape de recette qui se différencie symboliquement en une recette orientée interface et une recette orientée modules d'aide.

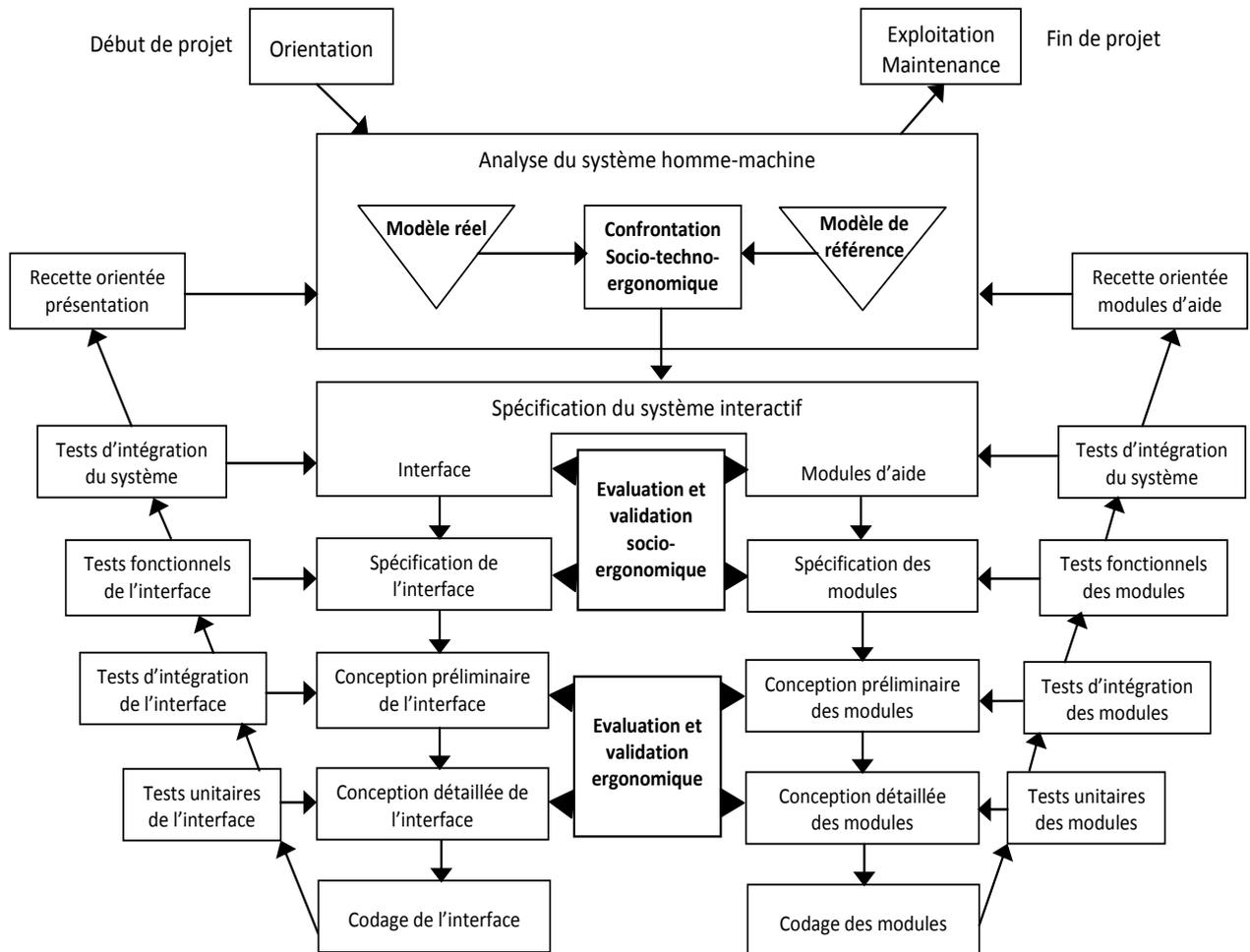


Figure 2.17 : le modèle V [Kolski 97] [Kolski 98]

Le modèle Nabla n'explique pas clairement la modélisation de l'utilisateur et les tâches humaines en montrant leurs relations avec les spécifications d'interface (elles sont en fait intégrées dans la boîte d'analyse du système Homme-Machine). D'ailleurs, comme le modèle en V, Nabla s'exprime dans une série de retours très limités qui constituent un handicap pour la conception itérative. Le modèle n'indique rien au sujet du prototypage. Il présente une tentative intéressante de la part du GL pour se relier à l'ergonomie cognitive en tenant compte des facteurs humains, et également de l'évaluation ergonomique [Lepreux 05].

### 2.2.2.8 Modèle en U

Le modèle en U [Abed 90] [Milot 90] [Abed 01] [Lepreux *et al.* 03] situe les étapes qui n'existent pas dans les modèles classiques du GL, lesquels demeurent très généraux, tout en partant sur l'hypothèse que les facteurs humains doivent y être considérés par l'équipe de développement.

Le modèle en U est structuré en deux phases :

1. Une phase descendante avec la spécification et la conception du système homme-machine (où on prend de la hauteur par rapport au logiciel interactif, vu par rapport à une organisation, à un environnement socio-technique), qui mène à son implémentation.
2. Une phase ascendante composée de l'évaluation du système global, selon des critères d'efficacité du système et également des critères strictement humains. La validation consiste à confronter le modèle des tâches théoriques (prescrites) avec le modèle des activités réelles défini en phase ascendante; et ceci en se basant sur les principes originaux proposés par [Abed *et al.* 91]. Le résultat de cette confrontation permet soit de valider le système Homme-Machine soit de dégager ses carences et à affiner progressivement celui-ci, particulièrement au niveau des interfaces Homme-Machine et des outils d'aide.

Le modèle final résultant de la confrontation permet ainsi de généraliser des comportements spécifiques de l'homme dans des conditions particulières de travail, réutilisable dans des situations liées à des systèmes similaires [Lepreux 05].

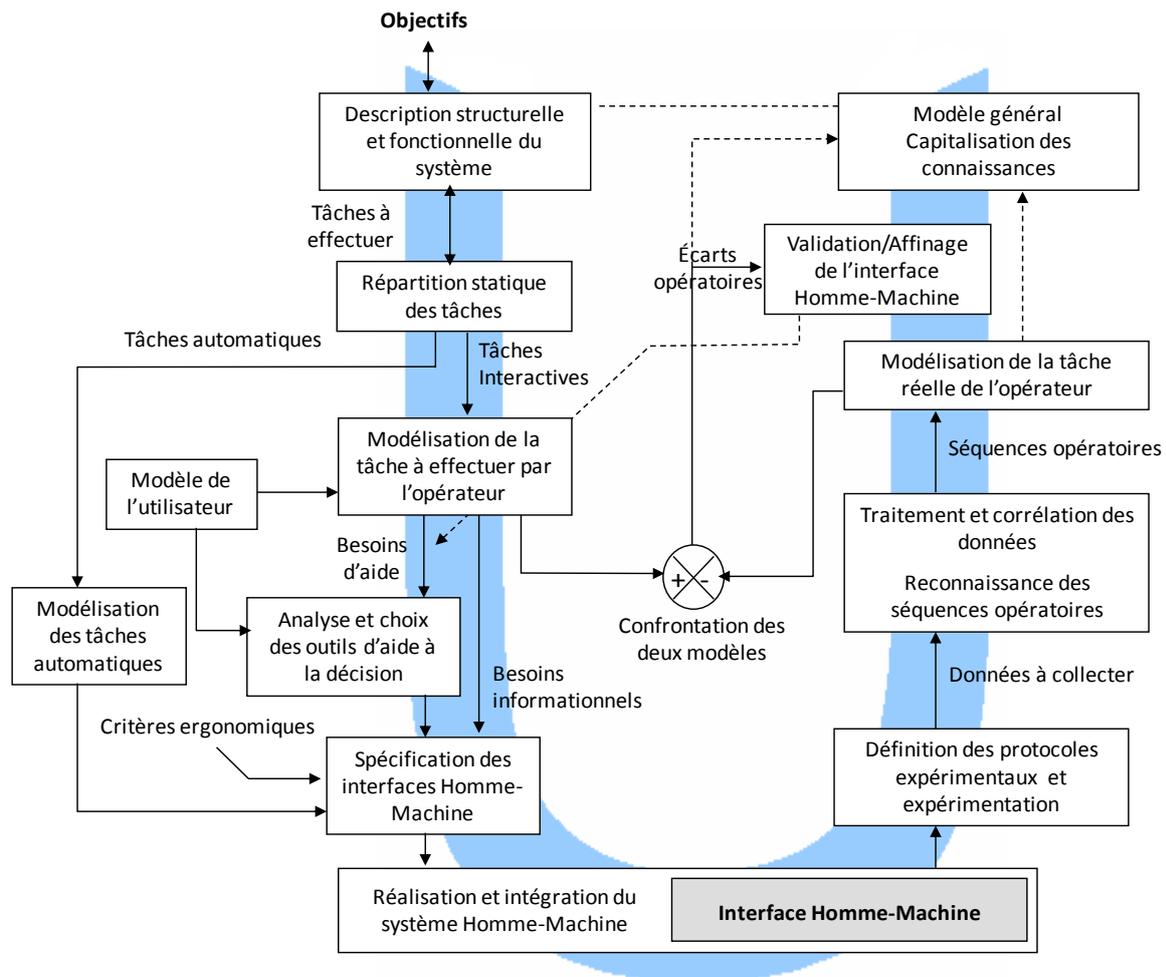


Figure 2.18 : le modèle en U [Abed 01]

Dans sa version originale, ce modèle a permis de commencer à placer les premières étapes qui ont semblé être fondamentales en ce qui concerne la **conception et l'évaluation de systèmes interactifs**. Au cours des vingt dernières années, il a été progressivement enrichi en ajoutant plusieurs étapes, qui sont le résultat de la recherche effectuée dans le cadre de plusieurs projets industriels, par exemple :

- L'étape appelée "Analyse de l'existant" et/ou "Analyse de la situation de référence" n'existait pas dans le modèle initial.
- La comparaison ou la confrontation du modèle d'activité à un modèle de référence est également devenue de plus en plus détaillée selon l'expérience acquise.
- Actuellement, d'autres étapes sont proposées au sujet de l'intégration et de l'utilisation de la connaissance ergonomique et experte dans les phases initiales du modèle de développement.

C'est un modèle qui peut être adapté selon les caractéristiques de l'application. Par exemple, dans [Lepreux *et al.* 03], il a été adapté pour des types particuliers de SIAD dans des projets où l'extraction de l'expertise des experts dans un but de conception de composants logiciels correspondant chacun à des modules d'aide aux activités expertes est centrale.

Le modèle en U est axé sur l'IHM, Néanmoins, il ne présente pas clairement le développement itératif et incrémental du système interactif à réaliser. En effet, c'est après la confrontation des deux modèles de la tâche prescrite et de la tâche réelle qu'on aura soit un retour arrière, soit une validation. Par la suite, le modèle en U sera appliqué pour la tâche suivante. De ce fait, le développement itératif se fait tâche par tâche, on revient donc au parcours séquentiel des étapes des modèles classiques de GL.

### **2.2.2.9 Conclusion sur les cycles enrichis sous l'angle de l'IHM**

Les insuffisances des modèles du GL vis-à-vis des systèmes interactifs (analyse de tâches, facteurs humains, etc.) ont amené à définir des modèles dits enrichis sous l'angle de l'IHM. Ces modèles centrés utilisateur montrent des évolutions apportées par le domaine des IHM au GL en se focalisant sur des idées essentielles pour le développement de systèmes interactifs tels que :

- Fixer les activités pour les différents intervenants,
- Placer l'évaluation au centre du processus,
- Modéliser les activités humaines, les interfaces Homme-Machine et le système et
- Suite à des évaluations, confronter les tâches théoriques prévues par les concepteurs à effectuer du début du cycle avec les activités réellement effectuées par les utilisateurs.

Néanmoins, ces cycles enrichis sont difficilement utilisables car ils ne sont pas suffisamment complets et ils montrent de leur part des insuffisances tel que le développement itératif qui reste limité (par exemple dans le modèle Nabla).

Aussi ces modèles enrichis ne tiennent pas compte des environnements d'évolution des utilisateurs et des technologies supportant ces environnements caractérisant l'hétérogénéité de ces derniers. Concernant l'évaluation, même si elle est généralement présente dans ces modèles, elle est souvent effectuée à la fin de chaque étape ou à la fin de la construction du système interactif d'où le risque de rejet en cas d'inadéquation entre le système construit et les attentes et besoins des utilisateurs. Par conséquent, ils ne permettent pas toujours une évaluation précoce.

Une manière de pallier cet écueil est de s'appuyer sur la norme de conception centrée utilisateur ISO 13407. Cette norme, sans être un modèle de conception, propose des recommandations pour considérer les utilisateurs et leurs besoins tout au long du processus de conception du système interactif. Elle offre un panel d'outils et une succession d'étapes visant à planifier le processus, à spécifier les besoins et le contexte d'utilisation et propose des orientations en termes de production de solution et d'évaluation du système.

## **2.2.3 La Norme ISO 13407 et la conception centrée utilisateur**

### **2.2.3.1 Nature et objectifs de l'ISO 13407**

L'approche de conception centrée utilisateur (en anglais User Centered Design) préconise une participation active des utilisateurs tout au long du processus de développement de produit ainsi qu'une compréhension claire des exigences liées à l'utilisateur et à la tâche. Cette démarche est normalisée par la norme ISO (International Standard Organization) 13407 [ISO13407 99].

Selon cette norme, l'utilisateur final est le mieux placé pour évaluer et influencer le développement d'un produit. Si le produit final correspond à ses besoins, il aura toutes les chances d'être adopté. La conception centrée utilisateur impose que le développement soit guidé par les besoins des utilisateurs plutôt que par les possibilités technologiques.

La norme ISO 13407 a été complétée une première fois en 2000 par le rapport technique 18529 [ISO/TR 18529 00] qui décrit sept ensembles de pratiques de base pour mettre en œuvre le processus de conception centrée sur l'humain. Un second rapport technique a été rédigé en 2002 sur proposition du groupe d'experts AFNOR (<http://www.afnor.fr>) en ergonomie des logiciels. Ce rapport [ISO/TR 16982 02] permet d'identifier quelles méthodes d'utilisabilité sont recommandées à chacun des stades du cycle de conception. Douze méthodes génériques ont été identifiées et décrites avec leurs principaux avantages et inconvénients selon les caractéristiques du projet.

Concevoir une application performante selon la conception centrée utilisateur nécessite de se demander à chaque étape critique de la conception si le produit correspond aux besoins des utilisateurs finaux. Cette approche a été traduite en une norme internationale l'ISO 13407 sur le processus de conception des systèmes interactifs centrés sur l'humain [Mouloudi 07].

### 2.2.3.2 Le processus de conception centrée utilisateur

La norme ISO 13407 [ISO13407 99] définit les conditions de mise en œuvre d'un processus centré sur l'humain à travers cinq principes nécessaires :

- Une préoccupation précoce des utilisateurs, de leurs tâches et de leur environnement ;
- Une participation active de ces utilisateurs, ainsi qu'une compréhension claire de leurs besoins et des exigences liées à leurs tâches ;
- Une répartition appropriée des fonctionnalités entre les utilisateurs et la technologie ;
- Une itérativité des solutions de conception ;
- Une intervention d'une équipe de conception multidisciplinaire.

Cette norme concerne le cycle de conception d'applications informatiques et détermine les exigences auxquelles un projet doit répondre pour être considéré comme "centré utilisateur". Elle concerne la méthodologie de conception et l'intégration de la démarche ergonomique dans le cycle de développement.

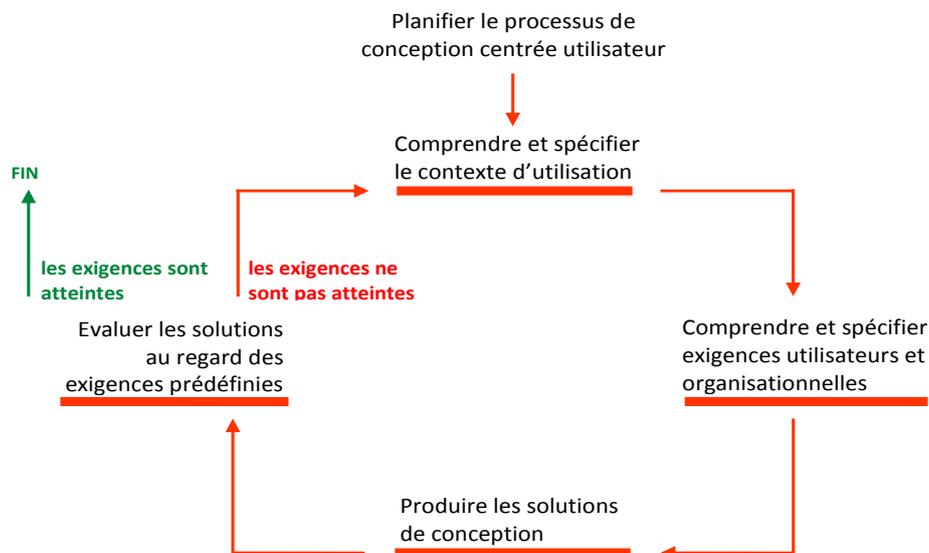


Figure 2.19 : processus de conception centrée utilisateur [ISO13407 99]

1. **Planification du processus de conception** : cette étape concerne la planification des activités de développement dans le cadre d'une approche de conception centrée utilisateur. L'adaptation des outils et méthodes se base sur la consultation de documentation et des discussions autour des pratiques et des contraintes de la conception.

2. **Spécification du contexte d'utilisation** : cette étape vise à comprendre et spécifier le contexte d'utilisation. Il s'agit donc de saisir les caractéristiques, buts et tâches des utilisateurs ainsi que leur environnement d'utilisation. A cet effet, il s'agit d'abord de décrire l'environnement des points de vue technique, matériel, social, organisationnel et législatif.
3. **Spécification des besoins utilisateurs** : cette étape consiste à spécifier les exigences liées à l'utilisateur et à l'organisation. Il s'agit de prendre en compte les besoins, compétences et l'environnement de travail de tous les intervenants sur le système.
4. **Production des solutions** : il s'agit d'utiliser les connaissances acquises lors des étapes précédentes pour matérialiser les solutions sous forme de prototypes.
5. **Evaluation des solutions** : les prototypes créés au stade précédent sont utilisés pour évaluer les solutions conçues en fonction des exigences. Le pilotage de tests utilisateurs selon un protocole d'évaluation précis doit permettre de détecter les défauts.

L'étude que nous effectuons, au niveau de ce mémoire, sur les différentes approches de développement en vue d'aboutir à celle adaptée au contexte de notre SIAD nécessite un passage en revue de quelques méthodes d'analyse et de conception des systèmes interactifs : c'est l'objet de la section suivante de ce chapitre.

## **2.3. Méthodes d'analyse et de conception des systèmes**

Pour pouvoir modéliser et formaliser les différentes étapes d'un processus de développement d'un système, une méthode d'analyse et de conception doit être utilisée. C'est un procédé qui consiste à créer une représentation virtuelle d'une réalité de façon à faire ressortir les points auxquels on s'intéresse. C'est dans ce cadre que s'inscrit cette section.

### **2.3.1 Les méthodes systémiques**

Elles permettant des modélisations centrées sur les données et les traitements.

#### **2.3.1.1 La méthode Merise**

La méthode Merise [Tardieu *et al.* 83] [Tardieu *et al.* 91] était la méthode systémique la plus connue et la plus employée en France à la fin des années 50 [Silvestre *et al.* 96]. Un des

points forts de Merise est de proposer trois niveaux d'abstractions avec des préoccupations complémentaires :

- Le niveau conceptuel comprend le modèle conceptuel de données et le modèle conceptuel des traitements,
- Le niveau logique (ou organisationnel) est représenté par le modèle logique des données et le modèle organisationnel des traitements,
- Le niveau physique (ou opérationnel) comprend le modèle physique des données et le modèle opérationnel des traitements.

Suivant cette méthode tout projet informatique suit trois cycles : abstraction, décision et vie. Le cycle de vie permet de décrire la vie du système. Il distingue les périodes qui vont de la conception à la maintenance. Le cycle de décision concerne les différentes décisions et les choix qui sont effectués tout au long du cycle de vie. Le cycle d'abstraction offre les concepts pour pouvoir décrire les différents éléments du monde réel qui seront représentés dans le système d'information. Merise a évolué au fur et à mesure des nouvelles technologies, par exemple en intégrant les notions objets dans une de ses versions récentes [Gabay 01].

Pour ce qui est de l'étude de l'existant et de l'analyse des besoins, la méthode Merise propose d'intéressants modèles centrés d'une part sur les données et d'autre part sur les traitements [Nanci *et al.* 01].

### **2.3.1.2 La méthode Axial**

La méthode Axial [Pelleaumail 86] C'est une méthode d'analyse, de conception et de gestion de projet. C'est une méthode systémique qui a été concurrente de la méthode Merise. Les différents modèles de cette méthode mettent l'accent sur le rapprochement des données et des traitements.

### **2.3.2 Les méthodes cartésiennes**

Elles permettent de procéder à un découpage fonctionnel et hiérarchique du système ; elles sont très pratiques pour représenter des systèmes complexes, et pour faciliter la communication dans les équipes projet.

### **2.3.2.1 La méthode SADT (Structured Analysis and Design Technique)**

C'est une méthode d'analyse bien adaptée à la phase de spécification fonctionnelle d'un système. Conçue au début des années 70 par Doug Ross aux Etats-Unis [Ross 77] [Dickover *et al.* 77], elle a été mise au point par la société Softech, en collaboration avec ITT [Marca 88]. Elle est diffusée en France par IGL Technologie [IGL 89].

Cette méthode permet de décrire un système complexe existant sur les deux plans : *le plan fonctionnel* (la mission du système, les différentes fonctions, leurs niveaux et leurs interactions) et *le plan structurel* (les frontières du système, les entrées-sorties, les principaux sous-systèmes et leurs liaisons). L'analyse d'un système par la méthode SADT suit une démarche descendante, modulaire, hiérarchique et structurée.

En effet, la technique consiste à découper hiérarchiquement l'objectif global en n niveaux (décrivant complètement les fonctionnalités du système) dépendant de la complexité du système, sous forme d'actigrammes. Chaque niveau est composé d'un ensemble de tâches composées ou terminales ; la tâche composée présente un niveau opérationnel qui fait appel à un enchaînement structuré de sous-tâches. Les tâches terminales, non décomposables, correspondent à la description des séquences composées d'actions élémentaires de l'opérateur ou de la machine pour atteindre le but fixé. Son niveau opérationnel est défini par une structure décrivant le corps de la tâche.

### **2.3.2.2 La méthode SA (Structured Analysis)**

Elle a été développée par Demarco [DeMarco 79] ; elle a pour base le modèle diagramme de flot de données (DFD). Son objectif est d'identifier et représenter les fonctions d'un système par niveaux de détails successifs et par niveaux d'abstraction successifs. A tout niveau, chaque fonction peut être décomposée en sous-fonctions de niveau inférieur représentant le détail des traitements de la fonction mère. Les fonctions de SA sont appelées processus et ont pour but de transformer des données portées par des flots de données.

L'apport de la méthode SA est qu'elle représente les fonctions d'un système par niveaux de détails successifs et par niveaux d'abstraction successives. Sa limite est qu'elle n'exprime ni le contrôle ni le séquençement temporel des actions.

### 2.3.2.3 La méthode SA-RT

Elle a été proposée par Ward et Mellor en 85 [Ward et Mellor 85], puis étendue en 87 aux aspects architecture système par Hatley et Pirbhai [Hatley *et al.* 87]. Elle se présente comme une extension sous l'angle du temps réel de la méthode fonctionnelle SA avec une bonne couverture de l'axe dynamique par la prise en compte des informations de contrôle (événements, exceptions, activations, synchronisation, etc.). La modélisation du comportement du système se fait dans un modèle dynamique séparé mais couplé au modèle fonctionnel. Un troisième modèle est proposé par Hatley et Pirbhai [Hatley *et al.* 91], c'est le modèle d'architecture. Il est proposé dans le but de couvrir le troisième axe d'analyse à savoir l'axe de l'aspect structurel. Sous l'angle du diagramme de flots de données, il est comme SA très proche du modèle SADT de décomposition en actigrammes ; néanmoins, il reste moins convaincant et moins approfondi.

### 2.3.3 Les méthodes orientées objets

Ces méthodes sont toutes basées sur UML<sup>5</sup> actuellement suite à l'unification des trop nombreuses méthodes existantes, durant les années 90 [Jacobson *et al.* 99]. UML est définie comme un langage d'unification des différentes méthodes objet. Ce n'est pas véritablement une méthode puisqu'elle doit être associée à un processus méthodologique. UML provient d'une fusion des principales modélisations objets. En cela, elle reprend et améliore leurs formalismes.

Parmi les diagrammes que propose UML, on retrouve le diagramme de classes, le diagramme d'objets, le diagramme des cas d'utilisation, le diagramme de séquences, le diagramme de collaboration, le diagramme d'activité, le diagramme de temps, le diagramme d'états / transitions, le diagramme de composants, le diagramme de déploiement, etc. [Larman 05].

L'emploi des cas d'utilisation de Jacobson [Jacobson *et al.* 93] incite à intégrer ou tout au moins à prendre en compte l'utilisateur au moins à un niveau d'abstraction élevé, ce qui constitue un progrès important par rapport à la plupart des méthodes objets qui ont procédé

---

<sup>5</sup> UML est au centre de premières tentatives pour aller vers des versions enrichies sous l'angle de l'IHM [Ruault 02] [Larman 05].

UML. Les méthodes orientées objets du domaine de GL nécessitent des extensions adaptées aux systèmes complexes [Moussa *et al.* 06].

### **2.3.4 Conclusion sur les méthodes étudiées**

Les méthodes systémiques sont mieux adaptées au domaine de l'informatique de gestion et des systèmes d'information qu'à celui des applications industrielles critiques [Moussa *et al.* 06]. La méthode cartésienne SADT permet de formaliser clairement l'état fonctionnel du système et d'en faire ressortir les points critiques. Néanmoins, pour un système complexe mal décomposé, elle aboutit à des schémas difficiles à exploiter. Par ailleurs, cette méthode ne fournit pas d'éléments de réponse concernant l'aspect dynamique. De même, la formalisation des notions de contraintes et moyens de réalisation des fonctions reste assez limitée. De ce fait et malgré les propositions de SA-RT [Hatley *et al.* 91], les méthodes cartésiennes présentent des carences au niveau de la description fine de la composante dynamique du système [Moussa *et al.* 06].

Les méthodes systémiques et cartésiennes présentées succinctement sont deux des plus anciennes. Elles ont l'inconvénient de ne pas faire intervenir les utilisateurs explicitement. Par contre, avec UML, on peut suggérer grâce aux cas d'utilisation une certaine implication de l'utilisateur dans le processus de développement du système en question. Quoique, d'une manière générale, l'utilisateur reste trop peu impliqué.

Dans la section suivante, on présente des exemples d'approches existantes dans la littérature pour la conception et le développement de SIAD.

## **2.4. Approches existantes de conception de SIAD**

Cette section présente des exemples de processus de développement et des méthodes de conception spécifiques aux SIAD.

### **2.4.1 Les processus de développement de SIAD**

#### **2.4.1.1 Démarche itérative**

Selon Checroun [Checroun 92], Courbon et ses collègues [Courbon *et al.* 81] ont été les premiers à proposer une démarche itérative basée sur la conception progressive d'un SIAD,

en utilisant des cycles aussi courts que possible, où les versions successives du système en construction sont utilisées par l'utilisateur final. Les étapes de cette démarche peuvent être résumées en quatre étapes :

1. identification d'un sous problème important,
2. développement d'un petit système utilisable par le décideur pour évaluer son adéquation au problème posé (maquette)
3. reprise, développement et modification progressive de ce « noyau » pour aboutir au SIAD spécifique à ce problème,
4. évaluation continue du système.

Ce cycle ayant abouti à la réalisation d'un sous-système satisfaisant, on le reproduit au niveau d'autres sous problèmes, puis au niveau du problème global, de manière à obtenir progressivement l'ensemble des applications et l'outil de développement et de maintenance de l'ensemble.

#### **2.4.1.2 Processus de développement de Nykanen**

Nykanen [Nykanen 00] a cité un processus de développement en le considérant comme "classique". Ce cycle de vie comprend les principales phases suivantes :

- La phase de planification intègre l'évaluation des besoins, le diagnostic du problème, la définition des objectifs et des buts du système ;
- La phase de recherche consiste à identifier l'approche utilisée pour adresser les besoins de l'utilisateur et les ressources disponibles, et à définir l'environnement du SIAD ;
- La phase de conception comprend la sélection de la meilleure approche de conception, la définition des ressources requises, l'étude de faisabilité ;
- La phase de conception détaillée cible la conception des composants, de la structure, des interfaces, du dialogue, des bases de données et de la gestion des connaissances du système ;
- La phase de construction correspond à l'implémentation technique et à l'intégration ;
- La phase de développement intègre les tests et évaluations, les démonstrations, les entraînements, et le déploiement ;

- La phase de maintenance, de documentation et d'adaptation interviennent pour la gestion de la documentation, pour adapter le système lors d'un changement éventuel d'environnement et permet de le faire évoluer selon les besoins des utilisateurs.

Ce cycle est classique par le contenu des phases. Il parait peu spécifique à un système particulier, tel qu'un SIAD. Les seuls traits caractéristiques concernent :

- 1) les notions de problème qui n'apparaissent pas dans d'autres modèles du GL,
- 2) la gestion des connaissances mais seulement dans la partie conception détaillée.

### **2.4.1.3 Le modèle MODESTI**

Le modèle MODESTI (Méthodologie de développement de systèmes à base de connaissances) [Duribreux-Cocquebert 95] fournit le cadre d'une méthodologie de développement de Système à Base de Connaissances interactive et centrée sur l'utilisateur, cf. Figure 2.20. Elle propose une approche multi-modèles qui fait cohabiter des acteurs, des techniques et des méthodes issues de disciplines différentes mais complémentaires telles que l'informatique, l'ingénierie des connaissances, la psychologie et l'ergonomie cognitives.

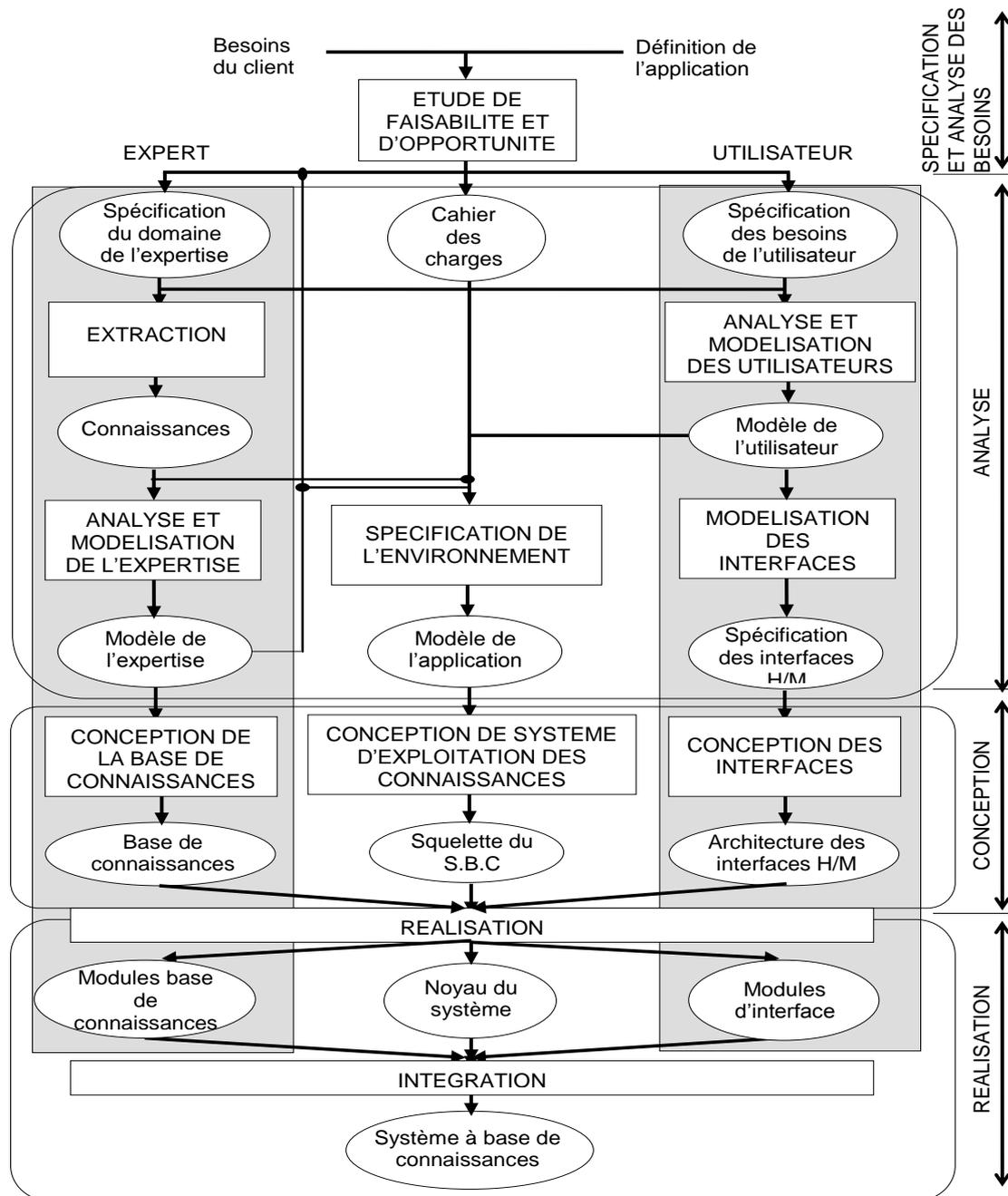


Figure 2.20 : modèle MODESTI [Duribreux-Cocquebert 95]

Le modèle MODESTI couple les principes des IHM et l'intégration des experts. Il propose trois phases descendantes en parallèle de manière à distinguer les tâches de l'expert, de l'utilisateur, et des autres intervenants de la conception comportant les étapes classiques du génie logiciel. Il permet de cette manière d'intégrer les besoins de l'utilisateur, les connaissances de l'expert dans le cycle « classique » de développement.

### 2.4.1.4 L'approche ADESIAD

Cette approche, intitulée *Approche de DEveloppement de SIAD* (ADESIAD) [Lepreux 05], est spécifique aux SIAD, au sens où elle sera centrée sur les décideurs, et par voie de conséquence sur les experts du domaine et le processus de décision. Elle a pour but d'intégrer les spécificités des approches de développement des différents domaines que nous avons décrits (GL et IHM) et qui sont couverts par les SIAD.

L'approche proposée se compose d'un modèle, d'une méthode et d'une architecture. Le modèle est une combinaison des cycles en V, en U et en étoile, en X et MODESTI. Ce cycle se veut également itératif. Le cycle proposé, vu en Figure 2.21, permet d'impliquer les utilisateurs et experts lors des phases (1) descendante dite de conception, (2) ascendante qui comprend l'intégration et les évaluations techniques et terminales (ou sommatives) et (3) centrale dite d'évaluations formatives.

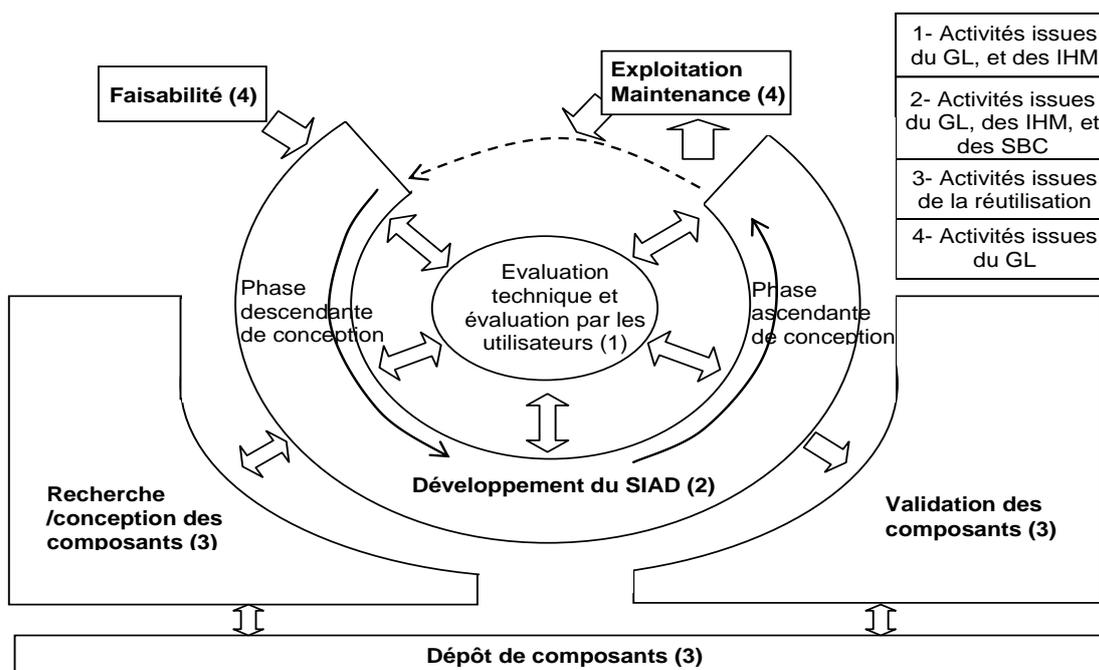


Figure 2.21 : l'approche ADESIAD [Lepreux 05]

La méthode permet de détailler l'ensemble des étapes du modèle. Elle introduit la notion de patron dans chacune des étapes de la phase descendante de conception. Ces patrons sont utilisés pour la réutilisation des connaissances, à la fois expertes (métier) et informatiques avant de devenir des composants réutilisables. De sa part, l'architecture d'ADESIAD permet de situer les éléments par rapport aux couches interface, spécifique au système et plateforme logicielle.

#### **2.4.1.5 Conclusion sur les modèles de développement de SIAD**

L'ensemble des modèles présentés ci-dessus offre des principes intéressants tels que la prise en compte de l'utilisateur et de l'expert, la réutilisabilité et l'intégration des connaissances. Cependant ils ne sont pas individuellement suffisants pour le développement d'un SIAD. Ils montrent qu'il n'y a pas eu de grandes modifications des cycles du Génie logiciel pour la conception de SIAD, excepté le dernier modèle. Les quelques modifications concernent la terminologie ; on retrouve les notions de problème, sous-problèmes, la notion de décision, de la gestion des connaissances.

### **2.5. Conclusion**

Nos travaux de recherche visent à élaborer une démarche de conception et de réalisation d'un SIAD basé sur un processus d'ECD (appelés dans la suite du mémoire SIAD/ECD). Pour cette raison, nous avons fait un survol des processus de développement et méthodes d'analyse et conception les plus répandus aussi bien dans le domaine de GL, que dans celui de l'IHM, pour en tirer les enseignements nécessaires à la mise en place d'un processus plus complet supportant la conception et le développement des SIAD/ECD.

Les modèles de développement existants ont montré des avantages mais aussi souvent des insuffisances pour le développement des SIAD. Du domaine de GL, nous avons remarqué que la tendance va aux processus itératifs (spirale, PU). Nous avons constaté que, même si les utilisateurs sont mentionnés pour les étapes d'analyse et de validation de prototype, les modèles et processus s'accompagnent le plus souvent de peu d'explications relatives à la prise en compte des utilisateurs ; la conception et l'évaluation des IHM sont rarement spécifiées dans ces processus. Pour cette raison, nous avons étudié des modèles enrichis sous l'angle des IHM. Ils montrent les évolutions apportées par le domaine des IHM au GL. Parfois, ils sont difficilement utilisables car ils ne sont pas suffisamment complets mais ils montrent le manque dans les modèles du GL. Ces modèles mettent l'accent sur des idées essentielles pour le développement de systèmes interactifs tels que la modélisation des tâches, l'évaluation, etc. De leur part, les méthodes d'analyse et de conception étudiées dans ce chapitre, mis à part UML, ne font vraiment intervenir l'utilisateur du début à la fin du projet. D'une manière générale, l'utilisateur reste trop peu impliqué.

Par la suite nous avons présenté aussi quelques démarches de développement représentatives orientées vers les SIAD. On peut remarquer que les démarches rapportées sont générales.

Elles traitent la conception de SIAD sans réellement prendre en compte les spécificités de ce type de systèmes. Les quelques spécifications qui peuvent être citées concernent la terminologie qui diffère par rapport aux méthodes classiques du GL. D'autres démarches beaucoup plus spécifiques existent mais elles sont tellement spécifiques à un type de SIAD particulier qu'elles n'apportent souvent pas beaucoup d'intérêt à ces travaux.

Sachant que nous travaillons sur des systèmes se situant à la confluence de trois domaines : l'aide à la décision, l'Extraction de connaissances à partir de données et l'Interaction Homme-Machine, la proposition d'une démarche de conception adéquate doit se baser sur des aspects potentiellement importants, que nous les représentons sous forme de trois *points clés* :

1. Point #1 : les principes de base de Génie Logiciel ; y compris le développement itératif, l'organisation des cycles, stratégie de tests, etc.
2. Point #2 : l'intégration des modules relatifs au processus d'ECD.
3. Point #3 : la prise en compte de l'Interaction Homme-Machine (IHM).

On se rend compte qu'il s'agit de concilier et d'intégrer des méthodes provenant d'une part du Génie Logiciel (Point #1), d'autre part de l'IHM (Point #3) dans un but de modélisation fine des SIAD basés sur l'ECD tout en tenant compte des spécificités des différents modules d'ECD (Point #2). Ces modules sont interactifs ; de ce fait, cette modélisation doit nécessairement s'appuyer sur une démarche centrée utilisateur. C'est pourquoi le chapitre suivant présente la démarche que nous proposons pour la conception centrée utilisateur de SIAD/ECD.

## **Chapitre 3 : ExUP : Démarche de développement de SIAD/ECD**

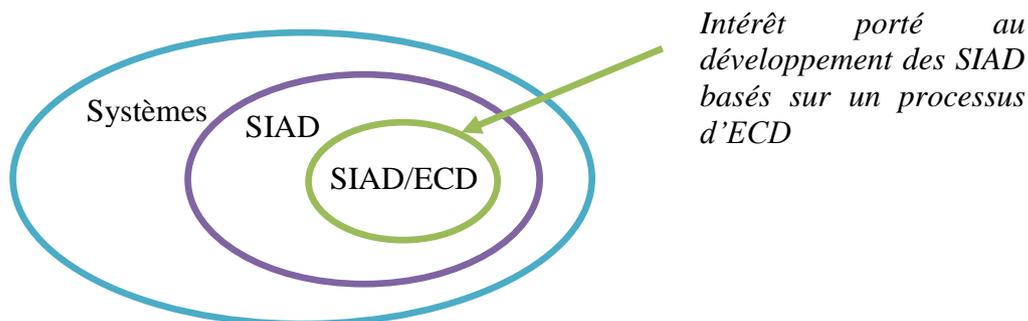
### **3.1. Introduction**

Des nombreux progrès sont réalisés régulièrement pour améliorer les méthodes et algorithmes en **Extraction de Connaissances à partir de Données (ECD)**. En revanche, peu d'efforts ont été déployés sur les aspects méthodologiques. Ainsi, dans le chapitre précédent, un état de l'art sur les méthodes et modèles de développement de systèmes, aussi bien classiques qu'enrichis sous l'angle de l'IHM, a permis de mettre en relief l'absence d'une démarche de conception précise et appropriée de systèmes décisionnels interactifs exploitant un processus d'Extraction de Connaissances à partir de Données. Rappelons que de tels systèmes facilitent l'accès aux données en offrant la possibilité d'effectuer des analyses en profondeur.

Dans ce chapitre, nous commençons par la présentation des caractéristiques essentielles de tels systèmes. Par la suite, nous proposons une démarche de conception et de développement qui nous semble appropriée pour répondre aux attentes considérées dans un tel contexte.

### **3.2. SIAD/ECD : SIAD basé sur un processus d'ECD**

La recherche dans le domaine des SIAD a eu comme conséquence l'apparence de nouvelles technologies et concepts concernant le stockage, le traitement et l'analyse des données et informations nécessaires au processus d'aide à la prise de décision. En conséquence dans le domaine des SIAD, la technologie d'Extraction de Connaissances à partir des Données [Fayyad *et al.* 96] occupe une place de plus en plus importante. Dans ce cadre, notre travail s'intéresse aux SIAD qui se basent sur la technologie la plus répandue de découverte de connaissances, qu'est l'ECD (figure 3.1).

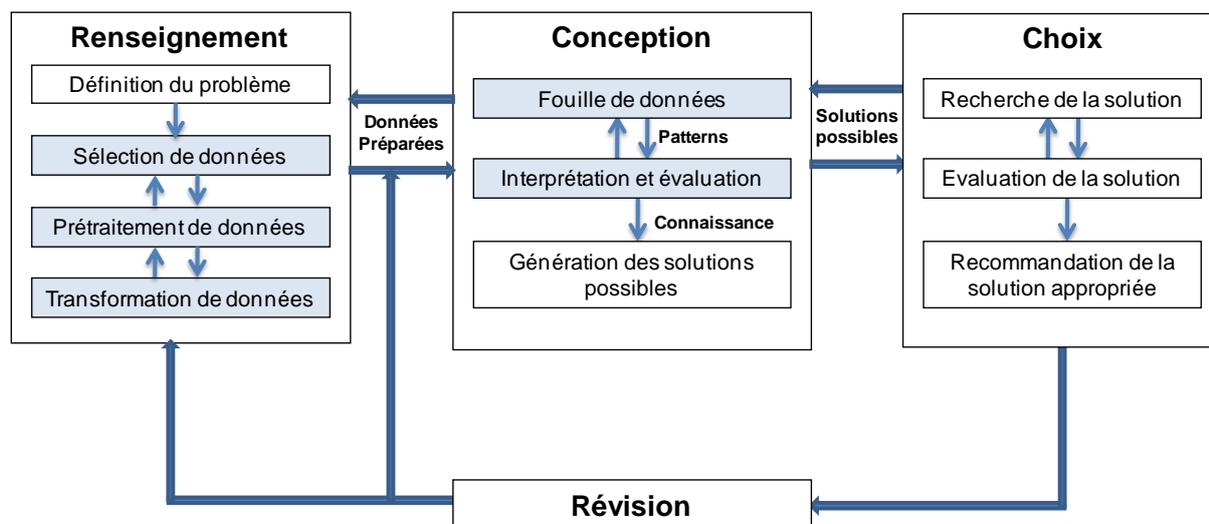


*Intérêt porté au développement des SIAD basés sur un processus d'ECD*

**Figure 3.1** : contexte global de recherche

Un SIAD/ECD est un système qui permet de détecter les stratégies de résolution d'un problème de décision par le biais d'une fouille de données. Dans ce processus, l'analyse des besoins des décideurs, les différentes activités réalisées en rapport avec la préparation et la manipulation des données pertinentes, de même que l'intégration de connaissances pour l'aide à la prise de décision constituent des étapes très importantes.

Le schéma visible en figure 3.2 propose une première représentation du déroulement d'un SIAD basé sur un processus ECD.



**Figure 3.2** : processus d'un SIAD basé ECD replacé dans le cadre du modèle IDCRC<sup>6</sup> de Simon [Ltfi *et al.* 10]

La figure 3.2 nous montre l'intégration des étapes de processus d'ECD dans le processus d'aide à la prise de décision. En effet, l'identification du problème, première phase du processus d'ECD, permet de cerner et définir les différents objectifs principaux du futur

<sup>6</sup> Le modèle IDCRC (Intelligence-Design-Choice-Review) a été décrit dans le chapitre 1, §1.2.1

système. Les étapes de prétraitements consistent à construire des corpus de données spécifiques ainsi qu'à procéder au nettoyage des données, au traitement des données manquantes, à la sélection d'attributs ou à la sélection d'instances puis à la transformation de ces données. Ces étapes sont considérées comme cruciales pour la recherche des informations pertinentes du processus de prise de décision [Turban 93].

La fouille de données peut alors être effectuée pour obtenir des connaissances mises sous la forme de modèles qui doivent être validés. Des post-traitements sont nécessaires pour rendre ces modèles intelligibles soit par un humain soit par une machine [Turban 93]. D'où la génération, l'analyse et le développement des solutions possibles au problème posé, basées sur les connaissances découvertes par le processus d'ECD.

Un SIAD/ECD est hautement interactif : comme la prise de décision implique l'acteur humain, qui est le décideur, le processus d'ECD nécessite de sa part une interactivité homme-machine pour qu'il soit réalisé. Dans ce cadre, nous proposons une démarche centrée utilisateur permettant de guider le concepteur du futur SIAD/ECD à réaliser la conception du système. Un tel système doit fournir de façon synthétique et simple les éléments nécessaires aux décideurs pour extraire des connaissances. Il nous semble important de construire des interfaces homme-machine adaptées aux besoins de chaque étape du processus d'ECD en tenant compte de chaque type d'utilisateur impliqué dans celles-ci.

A notre connaissance, aucune approche n'a été proposée dans la littérature pour le développement des SIAD centrés utilisateur et basés sur un processus d'ECD. Par conséquent, le paradigme de développement traditionnel de tels systèmes doit être changé ou adapté en s'appuyant entre autres sur des aspects relatifs aux processus interactifs, ceux-ci pouvant servir de point de départ pour aller vers une nouvelle démarche voyant ces SIAD comme des systèmes interactifs.

### **3.2.1 Modules d'un SIAD/ECD**

La modélisation et la réalisation d'un SIAD basé sur un processus interactif d'ECD, nécessite de développer en principe six modules successivement décrits ci-après (voir aussi la figure 3.2) en se référant aux différentes étapes du processus d'ECD [Ben Ayed *et al.* 10].

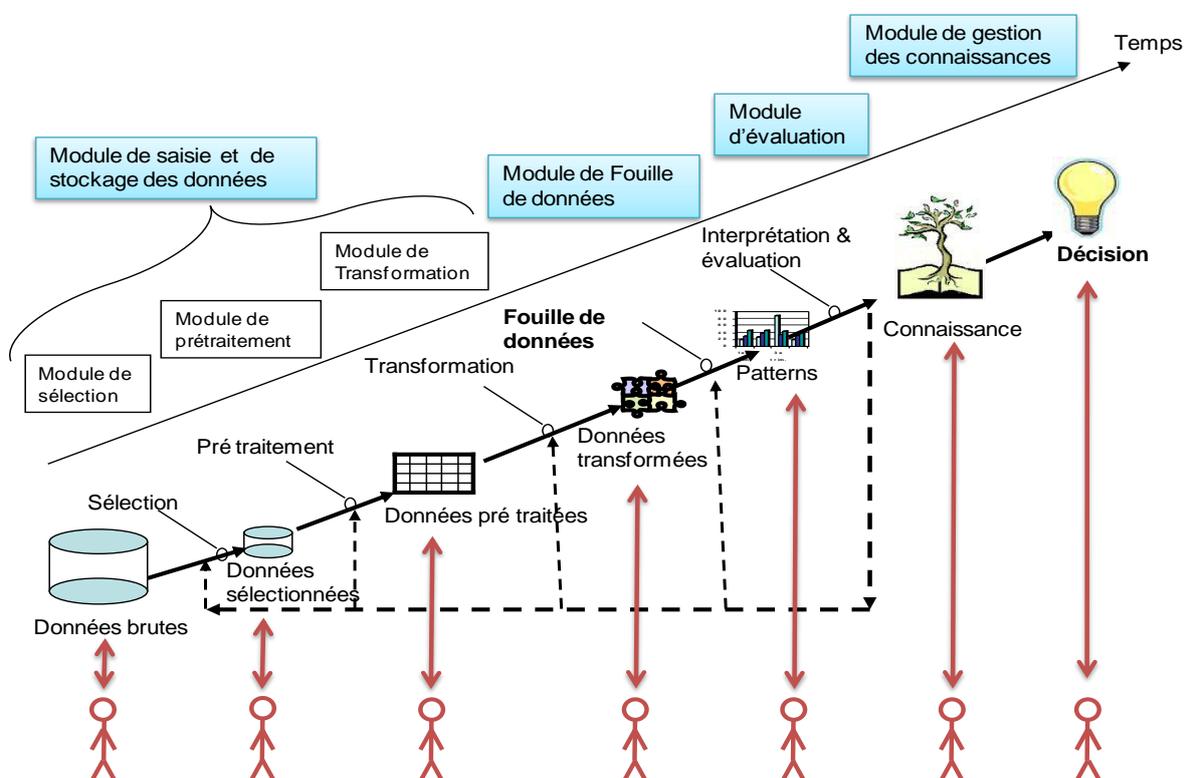
Dans une première phase, on expose le problème décisionnel et on définit les objectifs. Pour ce faire, on recueille les avis et les connaissances existantes des experts du domaine concerné, et on formule le problème à résoudre. La formulation du problème doit tenir compte des

techniques et outils qui vont être utilisés par la fouille de données. Ayant défini le type de problème, on doit bien savoir ce qu'on attend comme résultat et l'exploitation qu'on en fera. Ces connaissances faciliteront les choix à effectuer dans les différents modules constitutifs du SIAD/ECD. Une fois que les objectifs sont définis, il s'agit de déterminer la structure générale des données ainsi que les règles utilisées pour les constituer. Les modules concernés sont les suivants :

- (1) **Module pour la sélection des données** : une application pour la sélection des données doit être mise en œuvre en collaboration avec un ou des experts du domaine, afin de déterminer les attributs nécessaires pour décrire au mieux la problématique. Ce module peut fortement conditionner la qualité des résultats du processus d'ECD. Dans ce module, on effectue une collecte et une sélection de données. Il faut constituer une base d'informations qui permet de construire l'apprentissage, c'est-à-dire la construction de modèles en recherchant dans le passé des événements similaires. La sélection des données peut aboutir à un échantillon ou un ensemble exhaustif de données qui seront ensuite nettoyées.
- (2) **Module pour le prétraitement de données** : celui-ci permet de définir la taille de la base de données, et choisir la manière de la constituer ; il faut effectuer un diagnostic de qualité potentielle des données. Ce module permet d'améliorer la qualité des données afin de minimiser l'effet d'anomalies telles que des erreurs de saisie, des champs nuls, des valeurs aberrantes.
- (3) **Module pour la transformation des données** : une application de transformation des données permet de transformer des données fiables avec une visée de préparation du travail d'analyse. En se référant aux objectifs du SIAD/ECD, l'expert ECD peut déterminer les fonctions de transformation nécessaires. Ces fonctions interviennent sur des variables pour qu'elles soient mieux exploitables par des outils de fouille.
- (4) **Modules pour la fouille de données (Data mining)** : une application pour la fouille de données consiste à extraire un ou plusieurs modèles (patterns ou motifs) à partir du volume de données prétraitées et à la présenter sous une forme synthétique. La construction de ce modèle peut se faire de manière aussi bien automatique qu'interactive. Sa performance dépend du choix d'algorithmes de fouille de données. La précision du modèle extrait devra être vérifiée par une évaluation.

- (5) **Module pour l'évaluation et l'interprétation des patterns (motifs)** : l'évaluation du résultat permet d'estimer la qualité du modèle découvert, c'est-à-dire sa capacité à déterminer correctement les valeurs qu'il est censé avoir apprises sur des cas nouveaux. Le module d'interprétation du modèle évalué permet d'extraire la ou les connaissances.
- (6) **Module pour la gestion de connaissances** : la connaissance n'est rien tant qu'elle n'est pas convertie en décision, puis en action. Il est essentiel d'implanter le modèle et ses résultats dans le SIAD/ECD. Cette intégration peut se faire soit sous la forme de données (résultat du modèle) ou sous la forme d'un traitement (algorithme du modèle).

Dans de nombreux cas de figure, ces présents modules peuvent être rassemblés. Dans le cas où il n'y a pas de sources de données initiales, il sera nécessaire d'en créer une dédiée au projet. En fait, les modules de sélection, de prétraitement et de transformation des données ayant un seul objectif global qui est la préparation des données pour la fouille de données. Pour cette raison, ils peuvent être fusionnés en un seul module "Saisie et stockage des données". Le SIAD/ECD visé à terme comprend 4 modules : "Saisie et stockage des données", "Fouille de données", "Evaluation" et "Gestion de connaissances" (Figure 3.3).



**Figure 3.3** : modules constitutifs de SIAD/ECD ; différents acteurs sont concernés à chaque étape, exigeant des besoins particuliers vis-à-vis de chaque module concerné.

Notre contribution au domaine est d'intégrer ou considérer un développement des différents modules du SIAD/ECD sous les angles du GL et de l'IHM. C'est dans cette optique que nous proposons un cadre méthodologique global. Il prend la forme d'une démarche s'appuyant sur des méthodes, et des modèles issus du GL d'une part, et de l'IHM d'autre part. Il s'agit aussi de mettre en avant dans cette démarche les aspects architecturaux (étapes ECD) sur lesquels repose notre proposition.

Une démarche hybride utilisant le principe du processus unifié est proposée. La section suivante de ce chapitre dresse une description globale de cette démarche en présentant ses quatre phases principales ainsi que les articulations entre ses différentes étapes. Ensuite, chacune des ces étapes est successivement détaillée.

### **3.3. Présentation globale de la démarche proposée**

L'étude bibliographique, ayant fait l'objet du chapitre 2, a montré que de nombreuses méthodes intégrant des dimensions IHM étendent des processus ou méthodes existantes en GL. Nous optons pour ce point de vue, qui permet de profiter d'une méthode existante en GL, dont on connaît les forces et les faiblesses. Sur ce noyau orienté GL doivent se greffer harmonieusement les préoccupations et pratiques de l'IHM.

De ce fait, nous définissons un cadre méthodologique global relatif à la conception d'un SIAD/ECD. La démarche préconisée par ce cadre méthodologique s'appuie sur les principes d'une conception centrée utilisateur comme définie par [Gould et Lewis 85], [ISO1347 99] et [Norman 86] où plusieurs phases peuvent être mises en avant, en cohérence avec différents auteurs ayant proposé des démarches méthodologiques centrées utilisateur [Hackos et Redish 98] [Robert 03] [Lepreux *et al.* 03] [Shneiderman 09].

Ce cadre général de développement s'inscrit dans un processus se basant sur les trois principes du processus générique unifié déjà évoqué dans le chapitre 2 (démarche itérative, pilotée par les cas d'utilisation, et centrée sur l'architecture) et proposant un découpage du projet en quatre grandes phases (Inception, Élaboration, Construction, Transition) où on retrouve les principales activités de développement conduites de façon itérative et incrémentale. En fait, ce choix est basé sur les critères suivants :

- La tendance actuelle est vers les processus itératifs.
- Le processus unifié présente des principes génériques.

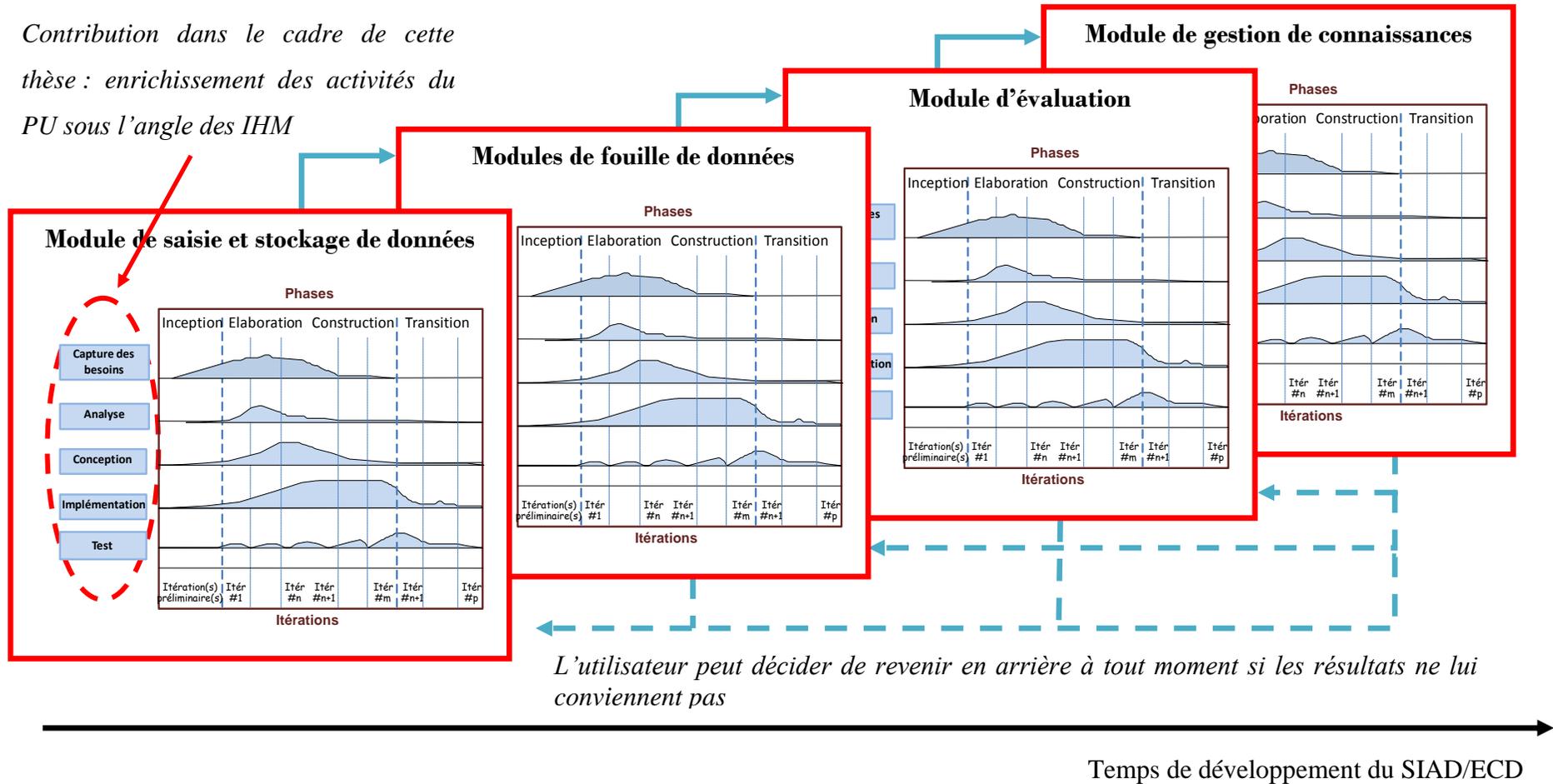
- Le processus unifié s'adapte aux spécificités des projets.

Dans le cadre du processus unifié issu du domaine de GL, nous nous intéressons particulièrement aux activités comme illustré par la Figure 3.4 où nous situons notre contribution.

L'idée de notre proposition est d'étendre les activités du processus unifié (qui est à l'origine une méthodologie générique) sous l'angle des interactions Homme-Machine pour le développement de chacun des modules de SIAD/ECD (figure 3.4). Pour cette raison, nous appelons notre approche le processus unifié étendu (en anglais **EX**tended **U**nified **P**rocess). On utilise l'abréviation : **ExUP**. Cette approche doit être appliquée au SIAD/ECD, qui est composé de quatre modules. Ces modules sont liés entre eux puisqu'ils constituent ensemble le processus d'ECD. En effet, pour le module de saisie et de stockage des données, les tâches de prétraitement de données dépend de la technique à utiliser dans le module de fouille de données. De même, cette technique conditionne aussi l'évaluation et l'interprétation des motifs découverts ainsi que l'intégration des connaissances extraites.

Cependant, chaque module possède individuellement ses propres objectifs et besoins décisionnels. Par conséquent, la conception et la réalisation de chaque module peut se faire en parallèle ou en chevauchement avec les autres modules du SIAD/ECD (figure 3.4).

*Contribution dans le cadre de cette thèse : enrichissement des activités du PU sous l'angle des IHM*



**Figure 3.4 :** positionnement de notre contribution par rapport aux modules du SIAD/ECD

L'étude de la participation des différents acteurs dans le déroulement de l'approche proposée pour le développement d'un SIAD/ECD, nous semble une étape importante. Pour cette raison, la section suivante sera consacrée à la présentation des intérêts des acteurs et la manière avec laquelle ils peuvent contribuer au développement du SIAD/ECD.

### **3.4. Présentation des acteurs de l'ExUP**

Le succès de la mise en œuvre d'une démarche de conception centrée utilisateur repose sur l'implication d'une équipe projet en interne, rassemblant des acteurs de différentes spécialités. Les types d'acteurs principalement concernés sont successivement décrits.

#### **3.4.1 L'utilisateur final**

Il s'agit du futur utilisateur du système à concevoir. C'est le décideur qui doit être considéré dans sa variabilité afin d'assurer la performance du système. L'utilisateur/décideur est l'acteur qui exploite les connaissances relatives à l'activité d'utilisation du système.

#### **3.4.2 L'expert ECD**

Les efforts de recherche en ECD depuis son émergence ont été essentiellement consacrés au développement d'algorithmes automatiques. Nous nous sommes intéressés aux systèmes ECD orientés utilisateurs. Notre approche propose d'introduire l'utilisateur final, le plus souvent expert de domaine, mais aussi un expert en ECD. C'est un acteur spécialisé en ECD et qui a acquis des connaissances et des compétences grâce à son expérience dans ce domaine.

Il s'agit de prendre en compte cet expert dès l'analyse et la conception du module de saisie et stockage de données. Il va décider à chaque phase des modules, les types d'informations qui vont être utilisées pour la suite de son exécution en sélectionnant les plus pertinentes. Cet expert pose les apports technologiques et attendus du système basé sur l'ECD. Il permet aussi de définir les contraintes et limitations techniques et organisationnelles inhérentes à la mise en place du nouveau SIAD/ECD.

### **3.4.3 L'ergonome**

La compétence ergonomique est mobilisée pour sa capacité à analyser l'activité et à en déduire les possibilités d'amélioration de l'existant. L'ergonome est donc le plus compétent pour avoir un dialogue structuré avec l'utilisateur. Ce dialogue lui permet d'émettre des recommandations fonctionnelles du système sous l'angle de son interaction avec l'utilisateur. Il peut aussi contribuer aux évaluations du système tout au long du projet.

Cet acteur appelé ergonome dans ce chapitre peut en fait avoir un profil et des compétences variables selon les projets et sa disponibilité dans ceux-ci (ergonome, psychologue ergonome, spécialiste de l'IHM, etc.)

### **3.4.4 Le concepteur**

C'est le créateur du SIAD/ECD. Son objectif est de traduire les besoins de l'utilisateur en spécifications fonctionnelles prêtes pour l'implémentation. Pour réaliser cette traduction, il utilise des modèles et des formalismes de représentation permettant l'abstraction de concepts réels vers des modules manipulant des données.

### **3.4.5 Le développeur**

La réalisation d'un SIAD découlant d'un processus d'ECD est à la charge de l'informaticien. Il définit les orientations pour le développement et concrétise le système en implémentant les IHMs et les modules d'ECD définis et en les intégrant.

Après avoir présenté les différents acteurs, il s'agit de définir leurs interactions avec le système. On utilise pour ce faire, le diagramme de contexte<sup>7</sup> UML [Roques 04] (cf. Figure 3.5).

---

<sup>7</sup> Un diagramme de contexte est un diagramme de classe qui contient une classe système et tous les acteurs autour.

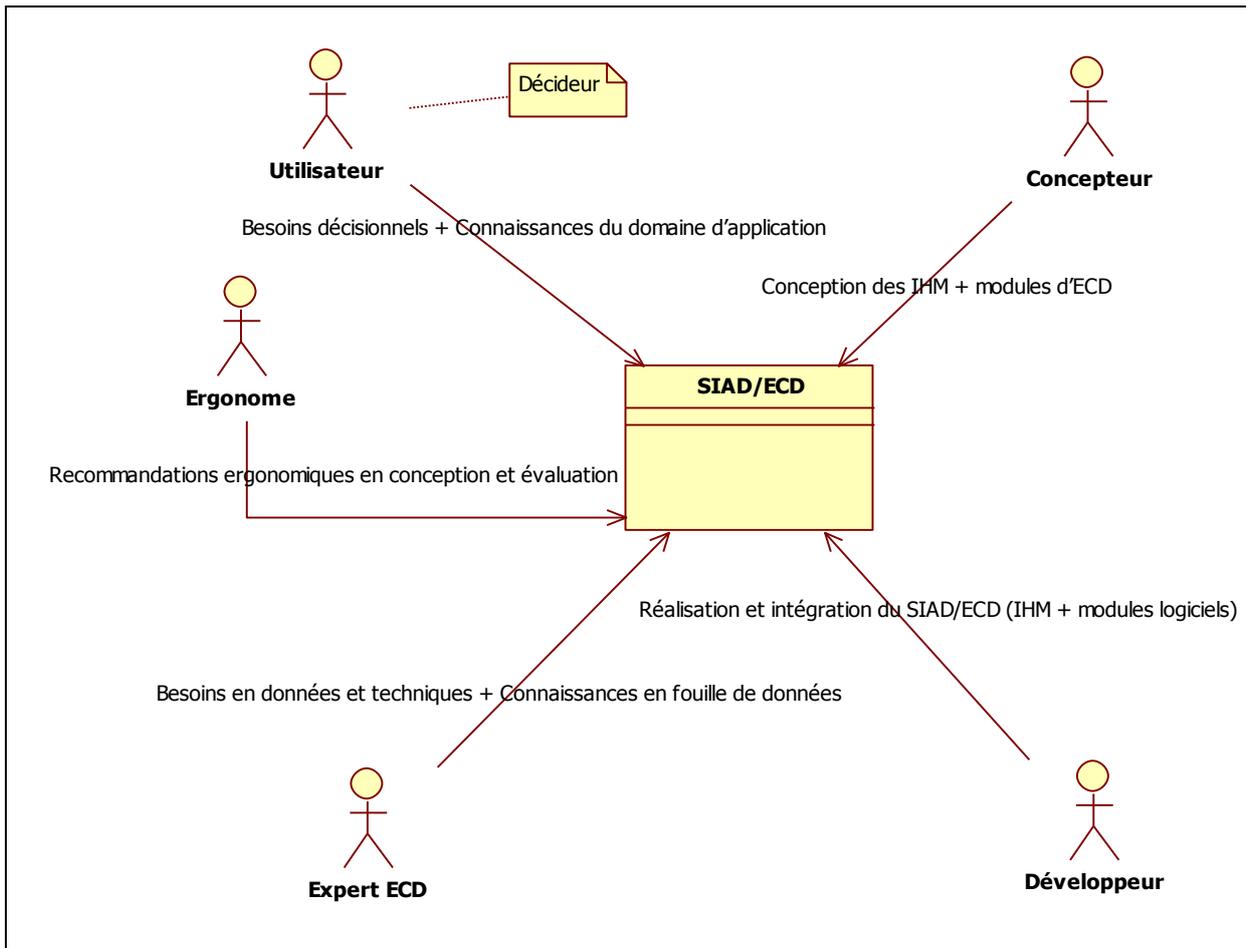


Figure 3.5 : diagramme de contexte SIAD/ECD

L'utilisateur, l'expert ECD ainsi que l'ergonome produisent des connaissances en fonction de leur spécialité. Ces connaissances doivent être transmises au concepteur qui doit construire des modèles de conception. A l'issue de cette phase, les premières spécifications du système sont rédigées et peuvent être validées. Ce travail collaboratif entre les acteurs est itératif, aboutissant à la construction des spécifications fonctionnelles du SIAD/ECD visé.

Dans la phase de conception, le concepteur dispose donc des corpus de données nécessaires à l'élaboration des spécifications du système SIAD/ECD interactif : les données concernant les tâches automatiques et celles relatives à l'activité des utilisateurs (avec IHM).

L'extension proposée, dans le cadre de l'ExUP, ne concerne pas les phases de processus unifié mais plutôt ses activités. La section suivante présente l'extension faite au niveau des activités qui composent notre approche. Ces activités se déroulent d'une manière itérative et incrémentale permettant d'enrichir le retour de connaissances sur le SIAD basé sur l'ECD.

Dans ce qui suit, nous présentons une description détaillée du processus unifié étendu (ExUP).

### **3.5. Description détaillée des activités de l'ExUP**

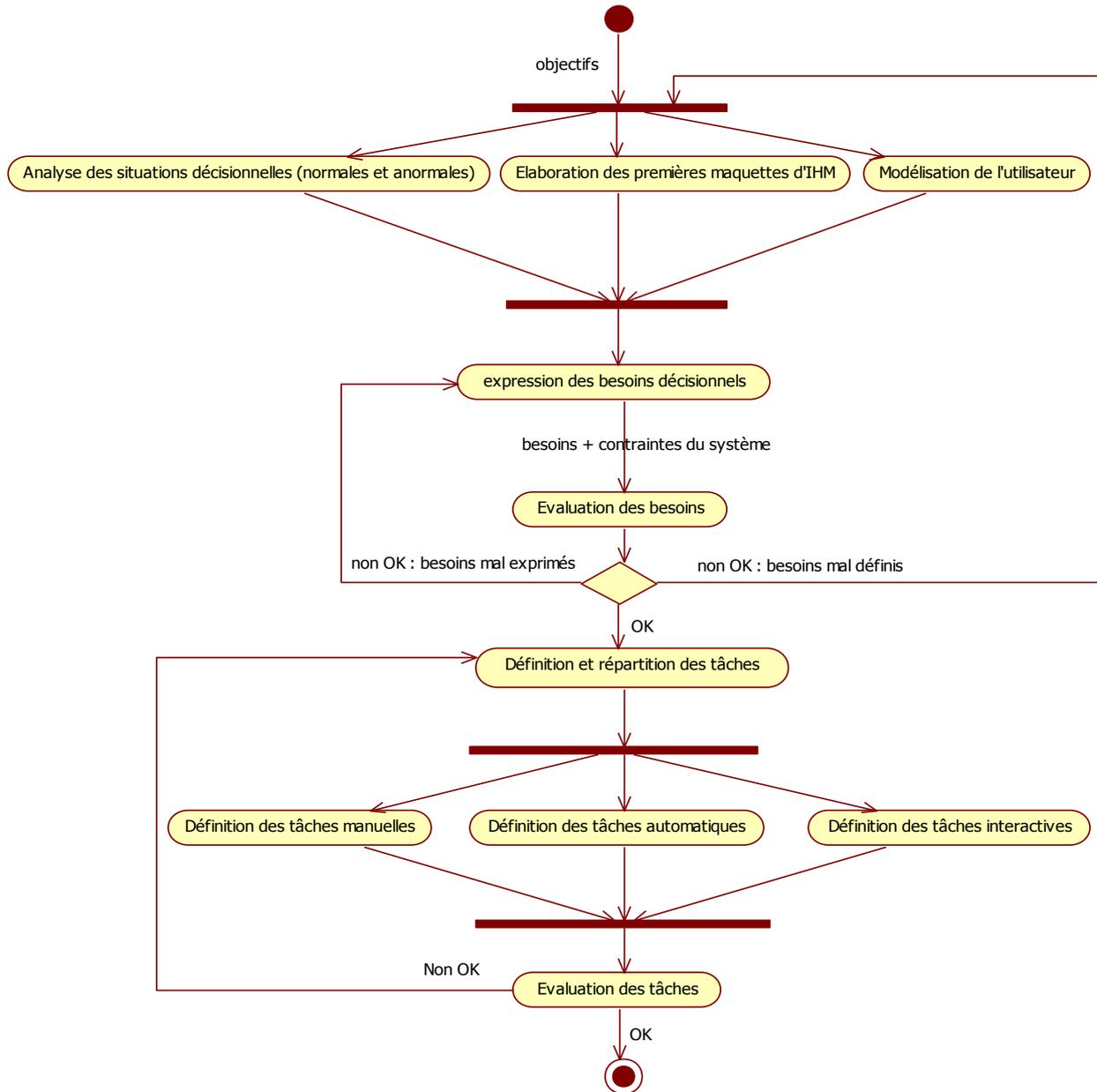
Rappelons que notre objectif est d'étendre les activités du processus unifié sous l'angle des interactions homme-machine pour une conception anthropocentrée des différents modules du SIAD/ECD. Cette section est donc consacrée à la présentation des activités enrichies qui composent l'approche proposée. Dans chaque itération de l'ExUP, il s'agit d'exécuter ces activités qui sont soit d'état initial (inception) soit d'état intermédiaire (élaboration + construction) soit d'état final (transition vers l'utilisateur). Les catégories d'activités divisent la réalisation du produit en une succession temporelle (raffinage successif et itération) et proposent une organisation matricielle du volume d'activités total fourni.

Pour chaque module, nous allons prendre les activités une par une (capture des besoins, analyse, conception, implémentation et test) et les détailler en intégrant les aspects IHM. Ces aspects concernent notamment l'élaboration des maquettes, la modélisation de l'utilisateur, etc. [Ltifi *et al.* 10] [Ltifi 07] [Kolski *et al.* 01]. Dans ce qui suit, nous allons donner les différentes activités et sous-activités pour chacun des quatre modules du SIAD/ECD.

#### **3.5.1 Capture des besoins**

La première activité de notre démarche centrée utilisateur est celle de capture et d'expression des besoins (cf. figure 3.6). Elle consiste en l'analyse des objectifs du module SIAD/ECD en termes des besoins fonctionnels et des besoins relatifs à l'IHM.

Dans les premières itérations de l'ExUP, une grande partie du temps sera consacrée à cette activité ; ainsi dans les dernières itérations elle est presque inexistante.



**Figure 3.6** : diagramme d'activité de capture des besoins

Cette activité marque le commencement du projet. Elle s'appuie sur un ensemble d'étapes successivement décrites.

### 3.5.1.1 Analyse des situations décisionnelles normales et anormales

Elle consiste en une étude préalable du problème décisionnel du module SIAD/ECD concerné par le développement. Cette étude vise l'identification des situations décisionnelles normales et anormales. L'analyse de ces situations doit permettre d'identifier, pour chaque classe de

situation (normale ou anormale), les objectifs, les critères de décision et d'intervention considérés en priorité ainsi que les stratégies à ce sujet.

Dans le cadre de cette analyse, il s'agit d'intégrer des connaissances expertes des décideurs et des experts ECD. L'acquisition de ces connaissances se fait lors de réunions de travail régulières en s'appuyant principalement sur des techniques d'entretien, des analyses de traces écrites et de documents.

Dans le cadre de cette étape, on vise également l'identification des acteurs qui répondent aux types de besoins exprimés par l'environnement de décision tels que les acteurs impliqués dans le processus de prise de décision, leur coopération et collaboration dans le cadre de leur missions respectives, leurs besoins en informations, etc.

La définition des situations décisionnelles normales et anormales du module SIAD/ECD permet de cerner les objectifs, définir les cibles et vérifier les besoins, mais aussi d'identifier les données et leurs sources, vérifier leur qualité ainsi que leur facilité d'accès (documents papier, supports électroniques, fichiers internes ou externes, Data Warehouse, etc.).

Le tableau 3.1 présente l'application de cette étape dans le cadre de chaque module du SIAD/ECD.

**Tableau 3.1** : analyse des situations décisionnelles normales et anormales pour les modules du SIAD/ECD

Module	Spécificités
<b>Module de saisie et stockage de données</b>	<p>A travers des réunions avec les futurs utilisateurs et l'expert ECD, le concepteur analyse les situations décisionnelles normales et anormales des différents sous-modules de :</p> <ul style="list-style-type: none"> <li>- Saisie des données relatives au problème décisionnel du SIAD/ECD, et</li> <li>- Prétraitement de données (sélection, nettoyage et transformation des données) dont le degré de complexité dépend principalement de la quantité et de la nature des données en question ainsi que de la technique de fouille de données qui sera utilisée (réseaux de Neurones, algorithmes génétiques, réseaux Bayésiens, règles d'associations, etc.).</li> </ul>
<b>Module de fouille de données</b>	<p>Il s'agit d'analyser les situations décisionnelles relatives à l'application des différentes techniques de fouille de données pouvant être utilisées dans le contexte de SIAD/ECD, en se référant aux connaissances et à l'expérience de l'expert ECD.</p> <p>En effet, le choix d'un ou plusieurs algorithmes reflète les objectifs de fouille qui ont été fixés précédemment. Est-ce que le processus de fouille a pour but la classification, la régression, le clustering, l'extraction de règles... ? Une fois la (ou les) technique(s) choisie(s), il faut décider quels paramètres sont les plus appropriés par rapport aux données étudiées.</p>
<b>Module d'évaluation</b>	<p>Cette analyse concerne l'étude des critères d'évaluation des modèles découverts et les situations décisionnelles résultantes. En fonction de ces situations, seront dégagées les méthodes d'interprétation de ces modèles pour l'extraction des connaissances.</p>

**Tableau 3.1** : analyse des situations décisionnelles normales et anormales pour les modules du SIAD/ECD (suite)

Module	Spécificités
<b>Module de gestion de connaissances</b>	<p>Il s'agit d'analyser chaque situation décisionnelle représentée par une valeur possible de la connaissance extraite par la fouille de données (intégration de connaissances).</p> <p>A chaque valeur de la connaissance, une liste des solutions ou recommandations sont formulées pour aider le décideur à choisir une alternative appropriée au problème décisionnel.</p>

Cette étape est donc prévue pour fournir un cadre structurant pour chaque module SIAD/ECD, en ce qui concerne aussi bien les futures activités que les solutions techniques. A l'issue de cette étape, on obtient une définition d'un modèle de processus de travail.

### 3.5.1.2 Élaboration des premières maquettes d'IHM

L'utilisateur élabore ces premières maquettes en vue d'aider à identifier les fonctions principales de chacun des futurs modules du SIAD/ECD, et à définir la structure globale des données ainsi que les actions que chaque utilisateur est censé exécuter en utilisant l'interface Homme-Machine appropriée.

L'élaboration de premières maquettes d'interfaces des modules visés permettent d'impliquer le plus rapidement possible les futurs utilisateurs en leur donnant un aperçu de solutions possibles (ou alternatives de solutions). Ce type de maquettage permet de définir toutes les contraintes ergonomiques liées à la mission de l'utilisateur et à son contexte de travail en termes de besoins. Ce maquettage peut être fait sur un support papier inspiré de croquis des utilisateurs ; il peut aussi exploiter des outils logiciels (éventuellement très basiques). En s'appuyant sur les maquettes préparées, il sera possible d'identifier la répartition des tâches entre l'utilisateur et chaque module SIAD/ECD concerné.

La plupart du temps, c'est l'ergonome, en collaboration avec l'utilisateur et l'expert ECD, qui réalise les premières maquettes liées à chaque module du SIAD/ECD, car la maquette dérive souvent d'une analyse de l'activité ou d'un recueil des besoins utilisateurs. Le concepteur peut également réaliser des maquettes, il peut demander le concours de l'ergonome.

Le tableau 3.2 discute de l'élaboration de ces premières maquettes pour chacun des modules du SIAD/ECD.

**Tableau 3.2** : élaboration des premières maquettes d'IHM pour les modules du SIAD/ECD

Module	Spécificités
<b>Module de saisie et stockage de données</b>	<p>Des réunions avec les futurs utilisateurs (ou décideurs) et élaboration des premières maquettes d'IHM permettent de déterminer l'architecture générale du sous-module de saisie des données. Les utilisateurs peuvent proposer des fenêtres (avec leur contenu, même vu globalement) qu'ils trouvent adaptées à leurs besoins.</p> <p>Concernant les sous-modules de prétraitement, c'est l'expert ECD qui peut élaborer des maquettes d'IHM pour la sélection, le nettoyage et la transformation des données et ceci en s'appuyant d'une part sur ses connaissances et d'autre part sur les situations décisionnelles définies dans l'étape précédente. Ces maquettes peuvent être révisées avec le concours de l'ergonome.</p>
<b>Module de fouille de données</b>	<p>Les premières maquettes d'IHM réalisées par ou avec l'utilisateur dans ce module permettent de définir la manière avec laquelle l'utilisateur souhaiterait visualiser les connaissances extraites. D'autres maquettes peuvent être préparées par l'expert ECD pour donner une idée sur les variables devant être renseignées par l'utilisateur.</p> <p>L'ensemble des maquettes permettent de prévoir la (les) interface(s) de l'application de la technique de fouille.</p>
<b>Module d'évaluation</b>	<p>Une élaboration des maquettes d'IHM peut être faite par l'expert ECD en collaboration avec l'ergonome pour donner une idée sur l'architecture globale des interfaces concernant :</p> <ul style="list-style-type: none"> <li>- l'évaluation des patterns (modèles),</li> <li>- l'interprétation de ces modèles pour l'extraction des connaissances.</li> </ul>
<b>Module de gestion de connaissances</b>	<p>Les maquettes d'IHM dans le cadre de ce module permettent de proposer l'architecture globale des interfaces homme-machine de :</p> <ul style="list-style-type: none"> <li>- Présentation des valeurs prédites par la fouille de données (par l'intégration des connaissances extraites)</li> <li>- Présentation des solutions possibles générées automatiquement par le système.</li> </ul>

### 3.5.1.3 Modélisation de l'utilisateur

Dans cette étape venant en parallèle des deux précédentes, il s'agit de recueillir des données concernant les futurs utilisateurs du SIAD/ECD qui seront utiles vis-à-vis des futures interfaces homme-machine relatives au module du système à développer (pour cela de nombreuses caractéristiques sont possibles, comme rappelé par [Amalberti *et al.* 91] et [Robert 03]).

Il s'agit de définir et identifier les profils des utilisateurs potentiels des futurs modules du SIAD/ECD, considéré comme hautement interactif, en recueillant des données et informations pertinentes les concernant en rapport avec l'utilisation potentielle des IHM et des outils d'ECD : classification par fonction, typologie par rapport à chaque classe de tâche (novice, expérimenté,

expert, occasionnel), niveau de formation, connaissances, expérience par rapport aux tâches à effectuer et aux outils informatiques, etc.

La modélisation des utilisateurs permet de considérer l'ensemble des acteurs humains de profils différents qui ont en conséquence des besoins différents vis-à-vis de l'IHM envisagée (et ceci pour l'ensemble des modules), qu'il s'agit de considérer dans la suite du processus de conception. Un objectif sous-jacent peut être de pouvoir personnaliser les réponses du système. La modélisation de l'utilisateur est donc la façon de représenter un utilisateur et ses comportements. Cela concerne également la façon d'exploiter les connaissances disponibles à son sujet.

Au sein d'une organisation, on peut envisager l'utilisateur de façon individuelle. Mais on peut également considérer qu'il appartient à une communauté, celle des utilisateurs exploitant un ou plusieurs modules du SIAD/ECD. Ainsi, dans ce contexte, la notion de collaboration<sup>8</sup> peut émerger. Ces utilisateurs sont liés entre eux par des relations, le plus souvent de hiérarchie et/ou de responsabilité ; ces relations influent sur leurs coopérations<sup>9</sup> et leur rendement. Par conséquent, ceci doit être bien compris et modélisé.

Le tableau 3.3 est relatif à la modélisation de l'utilisateur pour chacun des modules du SIAD/ECD<sup>10</sup>.

---

<sup>8</sup> Action de collaborer, de participer à une œuvre avec d'autres.

<sup>9</sup> Action de coopérer, de participer à une œuvre commune.

<sup>10</sup> Chaque module peut être exploité par une ou plusieurs personnes et chaque personne peut utiliser un ou plusieurs modules.

**Tableau 3.3** : modélisation de l'utilisateur pour les modules du SIAD/ECD

Module	Spécificités
<b>Module de saisie et stockage de données</b>	Cette étape consiste à modéliser l'utilisateur qui est expert dans son domaine ; il peut être considéré comme expert en ce qui concerne l'utilisation des outils informatiques pour la saisie et le prétraitement de données.
<b>Module de fouille de données</b>	Cette étape consiste à modéliser l'utilisateur qui doit pouvoir interagir avec ce module pour la découverte de patterns intéressants relativement à la problématique. Il n'est pas aussi censé connaître les détails de la (ou des) technique(s) de fouille de données utilisée(s) (au sens des mécanismes logiciels sous-jacents).
<b>Module d'évaluation</b>	Dans ce module, l'utilisateur est appelé à pouvoir interpréter les résultats de l'extraction des connaissances. Il faut prendre en compte les capacités de l'utilisateur et les modéliser pour pouvoir réaliser un module d'évaluation utile et utilisable [Nielsen 93] [Bahloul 10].
<b>Module de gestion de connaissances</b>	L'utilisateur est à ce niveau un décideur. Il faut analyser son comportement en tenant compte de son expérience et de ses capacités cognitives.

#### 3.5.1.4 Expression des besoins décisionnels

Notre démarche est centrée utilisateur, nous faisons donc appel à des méthodes spécifiques (que l'on pourrait qualifier de classiques en IHM) destinées à recueillir et collecter les besoins utilisateurs, comme par exemple l'étude de divers documents (décrivant l'organisation, exploités par les membres de celle-ci), l'interview des différents acteurs humains concernés, l'exploitation de questionnaire ou encore l'observation d'acteurs en situation de travail. Le plus souvent, l'utilisation combinée de telles techniques s'avère nécessaire.

Il s'agit de souligner que durant cette activité, les trois étapes présentées précédemment (analyse des situations décisionnelles, élaboration des maquettes et modélisation de l'utilisateur) aident à faire ressortir les besoins fonctionnels, les besoins en IHM et les besoins en données du futur système décisionnel. Le recueil et la collecte des besoins se font d'une façon itérative jusqu'à leur validation. Les besoins ainsi captés et collectés sont recensés, triés, classés et priorisés. Ces besoins expriment des scénarios pouvant prendre la forme de cas d'utilisation UML.

Les cas d'utilisation expriment essentiellement les fonctionnalités voulues et recherchées. Les utilisateurs ont souvent beaucoup de difficultés à exprimer clairement et précisément ce qu'ils attendent du système. L'objectif de cette étape est de les aider à formuler et formaliser ces besoins. A l'issue de cette étape, nous obtenons un modèle de cas d'utilisation exprimant les différents besoins captés, ces besoins devant par la suite être évalués.

Le tableau 3.4 est centré sur l'expression des besoins décisionnels pour chacun des modules du SIAD/ECD.

**Tableau 3.4** : expression des besoins décisionnels pour les modules du SIAD/ECD

Module	Spécificités
<b>Module de saisie et stockage de données</b>	Les besoins relatifs à la saisie et au prétraitement de données et qui sont collectés lors des étapes précédentes permettent au concepteur de réaliser le modèle des cas d'utilisation qui s'enrichit au fur et à mesure de l'avancement dans les itérations. Ce modèle concerne les besoins nécessaires pour la création de l'ensemble de données sur lequel va être appliquée la fouille de données.
<b>Module de fouille de données</b>	Les cas d'utilisation de ce module doivent montrer les fonctionnalités relatives à la fouille de données. Les descriptions textuelles de ces cas permettent de détailler davantage les spécificités techniques de ce module (classification, apprentissage, etc.).
<b>Module d'évaluation</b>	Le diagramme des cas d'utilisation de ce module concerne les fonctionnalités d'évaluation des patterns découverts ainsi que leurs interprétations qui seront évaluées pour extraire la (ou les) connaissance(s).
<b>Module de gestion de connaissances</b>	Le diagramme des cas d'utilisation décrit les fonctionnalités d'intégration de connaissances, de génération des solutions possibles et, éventuellement, une proposition de décision (en se basant sur les connaissances extraites par le (ou les) module(s) de la fouille de données).

### 3.5.1.5 Evaluation des besoins

A l'issue de l'étape d'expression et d'identification de besoins, une évaluation précoce de ces derniers est effectuée par rapport aux contraintes du système définies par les utilisateurs dès le début (contraintes d'organisation, contraintes logiques et/ou temporelles, etc.). Ce type d'évaluation s'inscrit dans une démarche participative où les utilisateurs doivent réellement s'impliquer et valider les différents besoins dégagés.

L'évaluation des besoins conduira à deux résultats possibles :

- a. En cas de non validation des besoins, une vérification des étapes décrites ci-dessus (avec re-bouclage préalable sur celles-ci) doit être effectuée dans l'itération suivante.
- b. Autrement, une liste des tâches du module SIAD/ECD à développer doit être dressée (cf. § 3.4.1.6).

Le tableau 3.5 positionne l'évaluation des besoins pour chacun des modules du SIAD/ECD.

**Tableau 3.5** : évaluation des besoins pour les modules du SIAD/ECD

Module	Spécificités
<b>Module de saisie et stockage de données</b>	Dans le cadre de ce module, une démonstration du diagramme des cas d'utilisation (raffiné et enrichi par le développement itératif de l'ExUP) aux utilisateurs doit être faite pour une évaluation précoce des fonctionnalités des sous-modules de saisie, sélection, nettoyage et transformation des données.
<b>Module de fouille de données</b>	Des réunions avec l'expert ECD doivent offrir l'opportunité d'évaluer les fonctionnalités représentées par le diagramme de cas d'utilisation de ce module. Une telle évaluation permet de vérifier la possibilité de réalisation des fonctionnalités de la ou des techniques de fouille de données choisies.
<b>Module d'évaluation</b>	L'utilisateur et l'expert ECD sont appelés à valider les fonctionnalités d'évaluation et d'interprétation des patterns découverts ; et faire en sorte que ces fonctionnalités répondent aux contraintes du système.
<b>Module de gestion de connaissances</b>	Les fonctionnalités du processus d'intégration de connaissances et de prise de décision (schématisées sous forme des cas d'utilisation) doivent être validées par le futur décideur avant de passer à l'étape suivante.

### 3.5.1.6 Définition et répartition des tâches

Par la suite, l'objectif est d'identifier et d'organiser l'ensemble de tâches devant être remplies par le couple Homme-Machine, en fonction des buts fixés par la logique de fonctionnement liée aux différents futurs modules du SIAD/ECD.

Chaque tâche dans le module interactif a un degré d'interactivité. Trois catégories principales de tâches peuvent être identifiées (ces catégories sont classiques en Génie Logiciel et IHM depuis les années 80) [Barthet 88] [Breedvelt *et al.* 97] [Paternò 99] :

- 1) les tâches dans lesquelles seul l'utilisateur est impliqué, appelées tâches manuelles,
- 2) les tâches dans lesquelles seul l'aspect applicatif est représenté, appelées tâches système ou tâches automatiques,
- 3) les tâches dans lesquelles l'utilisateur interagit avec le système, appelées tâches interactives.

Concernant les tâches automatiques, il s'agit de les extraire à partir des besoins fonctionnels déjà exprimées (à partir des cas d'utilisation). Ces tâches se manifestent généralement lors du fonctionnement du système par une succession chronologique d'actions relatives à des informations. La définition d'une tâche consiste à montrer comment elle peut être réalisée dans le projet décisionnel, quelles décisions sont prises, quels problèmes sont rencontrés, quels résultats sont obtenus, etc.

En plus des tâches induites des besoins fonctionnelles, il peut exister aussi de nombreuses autres tâches pouvant être accomplies d'une façon interactive. Par conséquent cette étape concerne l'allocation et l'expression des tâches et des fonctions interactives aux différents acteurs humains.

Pour chacun des modules du SIAD/ECD, la définition et la répartition des tâches se font tel qu'il est montré par le tableau 3.6 :

**Tableau 3.6** : définition et répartition des tâches pour les modules du SIAD/ECD

Module	Spécificités
<b>Module de saisie et stockage de données</b>	<p>A partir du modèle des cas d'utilisation de ce module, sont définies et décrites :</p> <ul style="list-style-type: none"> <li>- des tâches manuelles,</li> <li>- des tâches automatiques (calculs, enlèvement du bruit et des valeurs aberrantes -si nécessaire) et ;</li> <li>- des tâches interactives (choix des libellés des zones de saisie, définition du type de données à sélectionner, décisions sur les stratégies qui vont être utilisées pour traiter les valeurs manquantes, les variables à transformer, etc.).</li> </ul>
<b>Module de fouille de données</b>	<p>Les tâches relatives à ce module peuvent être réparties en tâches :</p> <ul style="list-style-type: none"> <li>- automatiques : l'exécution de la ou des techniques de fouille de données choisie(s) lors des étapes précédentes.</li> <li>- interactives : l'insertion des valeurs nécessaires pour les tâches automatiques.</li> </ul>
<b>Module d'évaluation</b>	<p>Les tâches d'évaluation des patterns consistent à faire correspondre les patterns obtenus avec les objectifs initialement fixés par l'utilisateur du système. Elles se distinguent en :</p> <ul style="list-style-type: none"> <li>- tâches pour une évaluation qualitative : tâches interactives pour lesquelles l'utilisateur répond à un questionnaire posé par le système d'évaluation.</li> <li>- tâches pour une évaluation quantitative : tâches automatiques d'évaluation par des outils statistiques, moucharid électronique, etc.</li> </ul> <p>L'interprétation des patterns, par rapport à la problématique, quant à elle se fait d'une manière automatique pour extraire une nouvelle connaissance en vue de l'intégrer dans le module de gestion de connaissances.</p>
<b>Module de gestion de connaissances</b>	<p>Les tâches de ce module se distinguent principalement en :</p> <ul style="list-style-type: none"> <li>- tâches automatiques pour l'intégration des connaissances et la génération des solutions possibles.</li> <li>- tâches interactives pour que le décideur puisse valider ou annuler la décision générée par le système.</li> </ul>

### 3.5.1.7 Evaluation des tâches

Une fois les tâches définies et réparties entre humain et machine, il est question de les évaluer par rapport aux besoins des utilisateurs et des experts ECD. Cette évaluation a pour résultat

soit la validation des tâches soit la planification d'un feedback sur la définition et la répartition des tâches dans la prochaine itération.

Le tableau 3.7 fournit des éléments d'évaluation des différents modules du SIAD/ECD :

**Tableau 3.7** : évaluation des tâches pour les modules du SIAD/ECD

Module	Spécificités
<b>Module de saisie et stockage de données</b>	L'évaluation consiste à vérifier si les tâches de saisie, de recherche d'informations et de prétraitement de données peuvent répondre aux besoins des futurs utilisateurs/décideurs. Le développement itératif de l'ExUP permet d'offrir un support pour le raffinement des tâches de manière incrémentale.
<b>Module de fouille de données</b>	Cette étape concerne l'évaluation des possibilités offertes par la ou les techniques de fouille de données choisie(s).
<b>Module d'évaluation</b>	Il s'agit de s'assurer que : <ol style="list-style-type: none"> <li>1. les patterns (motifs) découverts par l'outil de fouille de données peuvent être bien évalués quantitativement et qualitativement, et que le système arrive à interpréter ces patterns.</li> <li>2. La (ou les) techniques de visualisation des résultats et des données à utiliser est (sont) bonne(s) [Ltifi <i>et al.</i> 09]. C'est un support pour l'exploration interactive du pattern découvert, support pour l'interchangement des patterns, etc.</li> </ol>
<b>Module de gestion de connaissances</b>	Il s'agit de s'assurer que les tâches définies permettent au décideur d'arriver à prendre des décisions satisfaisantes. Sinon, un feed-back permettra de clarifier, d'améliorer et de faire évoluer la définition des besoins lors des itérations suivantes.

Après la capture des besoins, il faut procéder à l'analyse et la conception centrée tâche du SIAD aussi bien pour la partie IHM que pour les outils d'aide basés sur l'ECD. Les différentes tâches définies et évaluées doivent, par la suite, être traduites en spécifications fonctionnelles ou fonctionnalités d'IHM.

### 3.5.2 Analyse et conception

Dans la section précédente, nous avons décrit en détails la première activité de notre démarche, les résultats de cette dernière sont principalement les différentes tâches ainsi que le modèle de l'utilisateur<sup>11</sup>.

<sup>11</sup> Rappelons que la formulation "modèle de l'utilisateur" est vue au sens large (un ensemble de types ou profils d'utilisateurs pouvant être concernés).

Une fois que les tâches sont définies, elles doivent être analysées en profondeur et contribuer à la conception du module concerné. L'activité d'analyse de notre démarche doit déboucher à la spécification des tâches automatiques et interactives ainsi que des interfaces homme-machines ; alors que l'activité de conception concerne celle des modules automatiques et interactifs d'ECD et des interfaces homme-machine en rapport (cf. figure 3.7). Les éléments méthodologiques visibles dans la figure sont successivement décrits.

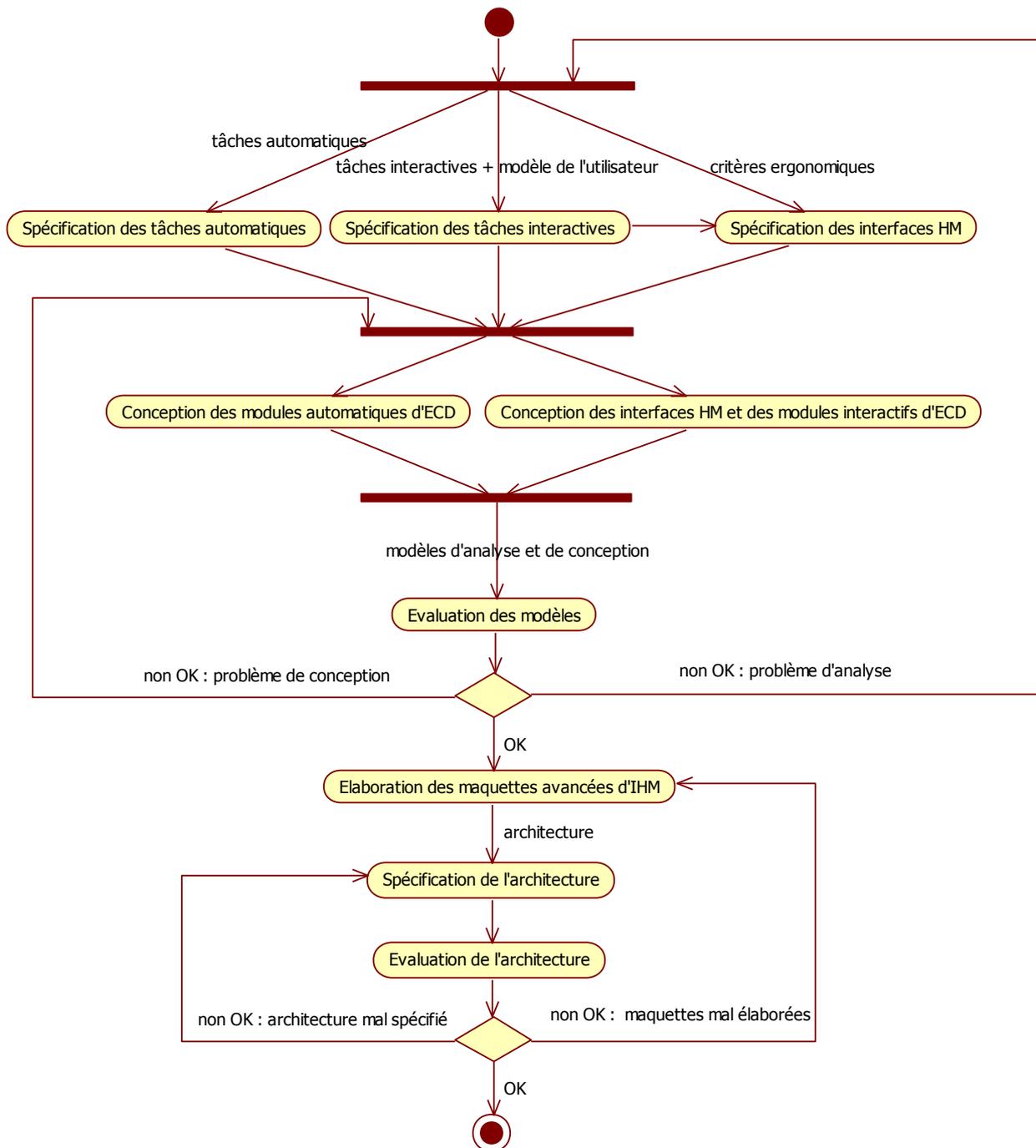


Figure 3.7 : diagramme d'activités d'analyse et de conception

### 3.5.2.1 Spécification des tâches automatiques

La spécification consiste à analyser les tâches. Elle consiste à relier objectifs, tâches, sous-tâches et actions. Il s'agit de comprendre la manière avec laquelle passent les objectifs aux tâches puis aux actions que ce soit en situation normale ou en situation anormale. On s'intéresse particulièrement aux aspects de fonctionnement normal et anormal selon différents contextes opérationnels.

L'analyse des tâches automatiques concerne l'analyse des tâches fonctionnelles pouvant être accomplies au sein d'un processus d'ECD. Par ailleurs, elle permet de recueillir des données et informations pertinentes sur de nombreux aspects de celles-ci.

Pour chacun des modules du SIAD/ECD, la spécification des tâches automatiques a des spécificités montrées par le tableau 3.8.

**Tableau 3.8** : spécification des tâches automatiques pour les modules du SIAD/ECD

Module	Spécificités
<b>Module de saisie et stockage de données</b>	Cette étape doit permettre de modéliser (par des diagrammes de communication et d'activités) les tâches automatiques de gestion des données ainsi que les méthodes de représentation des données. Ces méthodes incluent le nettoyage des données, la réduction des dimensions et la transformation des attributs en se basant sur la (ou les) technique(s) de fouille de données utilisée(s).
<b>Module de fouille de données</b>	L'analyse des tâches automatiques relatives à l'application de la (ou les) technique(s) de la fouille de données permet de spécifier son fonctionnement. Ce dernier dépend du type de fouille de données à utiliser et donc du but du modèle à extraire (par exemple : classification, régression, regroupement, etc.).
<b>Module d'évaluation</b>	Cette étape concerne l'analyse des tâches automatiques : <ul style="list-style-type: none"> <li>- de l'évaluation des patterns tout en tenant compte des critères d'évaluation de l'utilité et de la fiabilité des résultats de la fouille de données</li> <li>- de l'interprétation de ces résultats</li> </ul>
<b>Module de gestion de connaissances</b>	Il s'agit dans ce cadre d'analyser les tâches automatiques permettant : <ul style="list-style-type: none"> <li>- d'intégrer la ou (les) connaissance(s) découverte(s)</li> <li>- de générer les solutions possibles au problème décisionnel en se basant sur les modèles de la prédiction.</li> </ul>

### 3.5.2.2 Spécification des tâches interactives et des interfaces homme-machine

Pour l'analyse des tâches interactives, il est important de souligner que cette étape concerne aussi bien les situations normales qu'anormales afin de mieux cerner les différentes tâches interactives. Cette analyse est en étroite liaison avec l'identification des caractéristiques, ressources et limites cognitives de l'utilisateur venant du modèle de l'utilisateur. Elle se réfère

aux procédures composées d'opérations élémentaires que l'utilisateur est supposé effectuer pour réaliser la tâche. Ces procédures formalisent les séquences de dialogue homme-machine définissant les stratégies et les requêtes de l'utilisateur nécessaires à l'accomplissement du but fixé.

Les interfaces homme-machine peuvent ensuite être analysées en vue de définir leur comportement. Cette analyse se focalise sur les relations entre l'utilisateur et le module SIAD/ECD en cours de développement. En se basant sur les critères ergonomiques, il s'agit de recenser rigoureusement les besoins ergonomiques et techniques, puis de définir le nombre d'écrans à utiliser, l'enchaînement des vues, les modes de présentation des informations, les modalités de dialogue homme-machine, etc. Cette combinaison possible d'options d'affichage, d'interaction et de navigation aboutie à des interfaces de plus en plus riches et adaptées.

Le tableau 3.9 présente l'application de cette étape dans le cadre de chaque module du SIAD/ECD.

**Tableau 3.9 :** spécification des tâches interactives et des interfaces homme-machine pour les modules du SIAD/ECD

Module	Spécificités
<b>Module de saisie et stockage de données</b>	<p>L'analyse des tâches interactives nécessite la prise en compte du modèle utilisateur, résultat de l'étape précédente. Il s'agit d'analyser les interactions nécessaires pour la saisie des données, la sélection des données, le choix des valeurs à nettoyer ainsi que les variables à transformer.</p> <p>Dans cette étape, il s'agit aussi de décrire l'enchaînement des interfaces homme-machine en définissant les options d'affichages tels que les boutons, les onglets, etc. (cette remarque est valable pour les autres modules).</p>
<b>Module de fouille de données</b>	<p>Il s'agit d'analyser les interactions de l'utilisateur avec la machine pour la fouille de données. Il s'agit de modéliser la manière avec laquelle l'utilisateur peut guider la recherche des motifs et décider les paramètres appropriés pour l'application de l'algorithme de fouille.</p>
<b>Module d'évaluation</b>	<p>Cette analyse concerne les tâches d'évaluation qualitative qui nécessitent la prise en compte du comportement du décideur (modèle de l'utilisateur) lors de la mise en place du questionnaire d'évaluation.</p> <p>L'analyse concerne aussi les enchaînements des interfaces homme-machine de l'application de la (ou des) technique(s) de visualisation à utiliser.</p>
<b>Module de gestion de connaissances</b>	<p>Dans ce module, il faut analyser les interactions du décideur avec le système pour prendre les décisions (valider ou non la décision générée par le système).</p> <p>Une modélisation de la logique d'enchaînement des interfaces homme-machine, en se référant aux critères ergonomiques, doit permettre de tolérer les changements de décision des utilisateurs.</p>

### 3.5.2.3 Conception des modules automatiques et interactifs d'ECD et des Interfaces Homme-Machine

L'étape de spécification est suivie par l'étape de conception qui permet de réaliser les modèles de conception des tâches automatiques et des interfaces homme-machine analysées. La conception consiste à traiter les analyses qui ont été réalisées en regardant la manière dont elles doivent être traitées. Cette conception permet de générer des modèles plus détaillés que ceux d'analyse. Cette activité permet de formaliser les spécifications des besoins pour arriver à la définition des algorithmes qui seront nécessaires. Il s'agit aussi de rechercher les composants de code fonctionnels ainsi que les composants de code spécifiques aux interfaces homme-machine. Les algorithmes définis à cette étape seront "transmis" à l'étape d'implémentation en vue d'être développés.

Le tableau suivant présente la conception des modules automatiques et interactifs ainsi que des interfaces homme-machine pour chacun des modules du SIAD/ECD (cf. tableau 3.10).

**Tableau 3.10** : conception des modules automatiques et interactifs d'ECD et des Interfaces Homme-Machine pour les modules du SIAD/ECD

Module	Spécificités
<b>Module de saisie et stockage de données</b>	<p>La conception des tâches automatiques et des IHM du module de saisie et de stockage de données permet de définir les différents scénarios possibles<sup>12</sup> de chaque analyse faite dans l'étape précédente.</p> <p>La conception permet de formaliser les interfaces homme-machine et les fonctionnalités associées pour la saisie des données ainsi que le prétraitement de ces données.</p> <p>Il faut savoir concevoir la manière avec laquelle sera filtré un nouvel enregistrement dans la base de données, nettoyé et transformé si nécessaire.</p>
<b>Module de fouille de données</b>	<p>La conception dans ce module permet de formaliser les différents modules de fouille de données (c'est-à-dire la ou les techniques utilisées).</p>
<b>Module d'évaluation</b>	<p>La conception du module d'évaluation concerne la définition des formalismes (et donc des algorithmes) d'évaluation quantitative et qualitative ainsi que d'interprétation et extraction de connaissances.</p>
<b>Module de gestion de connaissances</b>	<p>Il est question maintenant de concevoir les différentes tâches d'intégration de connaissances, de génération des solutions possibles pour l'aide à la décision dans leurs différents scénarios possibles.</p>

<sup>12</sup> Pour expliquer l'idée des scénarios, on prend l'exemple de la tâche d'ajout d'un patient. Deux scénarios sont possibles :

1. Ajout avec succès
2. Echec d'ajout (patient existant, etc.)

### 3.5.2.4 Evaluation des modèles

Une évaluation des modèles d'analyse et de conception peut s'effectuer à ce niveau. Cette évaluation vise à vérifier leur correspondance avec les tâches, le modèle d'utilisateur (définies dans l'activité de capture des besoins) et les critères ergonomiques. Tant que le résultat est jugé non satisfaisant (éventuellement pour des raisons d'incomplétude), une modification doit être introduite dans la prochaine itération au niveau des modèles d'analyse et de conception de manière à produire des modèles en accord avec les besoins.

Le tableau 3.11 présente l'évaluation des modèles pour chacun des modules du SIAD/ECD.

**Tableau 3.11** : évaluation des modèles pour les modules du SIAD/ECD

Module	Spécificités
<b>Module de saisie et stockage de données</b>	<p>Avant de passer au codage (de l'itération en cours), il est important d'évaluer les modèles d'analyse et de conception de saisie et de prétraitement de données. Ces modèles (diagrammes + descriptions textuelles) doivent être lisibles par les utilisateurs, qui sont pour la plupart des cas non experts. Ces derniers peuvent proposer des adaptations, modifier des tâches, etc.</p> <p>L'aspect interactif dans les modèles doit être évalué par rapport aux critères ergonomiques.</p>
<b>Module de fouille de données</b>	<p>Les modèles d'analyse et de conception de ce module doivent aussi être évalués d'une part par l'utilisateur et l'expert ECD pour vérifier si la ou les techniques de fouille de données sont bien formalisées. Sinon, il peut être très utile de les corriger avant l'étape d'implémentation. D'une autre part ces modèles sont évalués par rapport aux critères ergonomiques.</p>
<b>Module d'évaluation</b>	<p>Les méthodes d'évaluation et d'interprétation des patterns présentées par des modèles de conception et d'analyse. C'est dans le cadre de cette étape, que l'utilisateur et l'expert ECD vont pouvoir les valider. Ces modèles doivent être aussi évalués par rapport aux critères ergonomiques</p>
<b>Module de gestion de connaissances</b>	<p>Le décideur est appelé à valider les formalismes, les interfaces homme-machine (par rapport aux critères ergonomiques) et les algorithmes (présentés graphiquement par les modèles) de processus de gestion de connaissances et de prise de décision.</p>

### 3.5.2.5 Elaboration de maquettes avancées d'IHM

Une fois la spécification et la conception des interfaces validées, un maquettage avancé<sup>13</sup> des interfaces Homme-Machine peut être réalisé avec la collaboration des experts et des utilisateurs. Les maquettes d'IHM permettent aux utilisateurs d'avoir une vue concrète mais non définitive de la future interface de l'application. La maquette peut comprendre un ensemble de dessins produits par un logiciel de présentation ou de dessin. Par la suite, la maquette pourra intégrer des fonctionnalités de navigation permettant à l'utilisateur de tester l'enchaînement des écrans ou des menus, même si les fonctionnalités restent fictives. La maquette d'IHM doit être développée rapidement afin de provoquer des retours de la part des utilisateurs.

L'élaboration des maquettes avancées d'IHM pour chacun des modules du SIAD/ECD se présente dans le tableau 3.12.

**Tableau 3.12** : élaboration des maquettes avancées d'IHM pour les modules du SIAD/ECD

Module	Spécificités
<b>Module de saisie et stockage de données</b>	A ce niveau, le concepteur connaît bien ce que le développeur est censé implémenter. Et pour bien satisfaire les besoins des utilisateurs, l'expert ECD prépare des maquettes avancées d'IHM (peuvent être réalisées à l'aide d'un outil informatique) surtout pour la sélection, le nettoyage et la transformation des données qui semblent des tâches non "habituelles". Ces maquettes peuvent être considérées comme un premier pas vers le développement du module en question.
<b>Module de fouille de données</b>	Des maquettes avancées d'IHM permettent aussi de donner une idée claire sur les interactions possibles de l'utilisateur avec le module de fouille de données en présentant par exemple les zones de sélection de plage de données, de choix des tests et d'affichage de résultat.
<b>Module d'évaluation</b>	Dans ce module, les maquettes avancées d'IHM permettent d'aider à définir la manière avec laquelle l'utilisateur pourra interagir avec le module pour évaluer les patterns.
<b>Module de gestion de connaissances</b>	Pour ce module, l'élaboration des maquettes avancées d'IHM montre comment l'utilisateur pourra valider ou annuler une solution possible générée par le système lors de la phase de prise de décision.

<sup>13</sup> Au niveau de la phase de capture des besoins, le maquettage effectué par les utilisateurs est de "haut niveau" (généralement des maquettes papier) alors qu'en phase de conception, le maquettage est avancé (réalisé avec un éditeur de présentation ou d'IHM, ce type de maquette est plus détaillé).

### 3.5.2.6 Spécifier l'architecture

Toutes les informations fournies dans le cadre de ces activités permettent de spécifier l'architecture de chaque futur module du SIAD/ECD à réaliser.

L'architecture est un ensemble de structures comprenant chacune des composants, les propriétés extérieurement visibles de ces composants et les relations qu'ils entretiennent [Bass *et al.* 98]. Cette architecture se développe et s'enrichie au fur et à mesure de l'évolution des itérations de notre démarche de développement (diagramme UML de composants).

L'architecture se compose des éléments suivants :

- IHM permettant des interactions avec l'utilisateur
- La base de données contenant les données à fouiller, les procédures, les solutions générées ainsi que les décisions prises.
- Les modules fonctionnels.

La spécification de l'architecture de chacun des modules du SIAD/ECD fait l'objet du tableau 3.13.

**Tableau 3.13** : spécification de l'architecture des modules du SIAD/ECD

Module	Spécificités
<b>Module de saisie et stockage de données</b>	Dans ce module, l'architecture du système doit considérer les composants suivants : <ul style="list-style-type: none"> <li>- les interfaces homme-machine de stockage et de prétraitement de données,</li> <li>- les modules automatiques de saisie et de prétraitement de données,</li> <li>- la base de données.</li> </ul>
<b>Module de fouille de données</b>	Dans ce module, l'architecture de système doit considérer les composants suivants : <ul style="list-style-type: none"> <li>- les interfaces relatives aux techniques de la fouille de données,</li> <li>- les interfaces relatives à la prédiction automatique,</li> <li>- l'algorithme de fouille de données,</li> <li>- la base de données "préparée" enrichie par les cas traités.</li> </ul>
<b>Module d'évaluation</b>	Dans ce module, l'architecture du système doit considérer les composants suivants : <ul style="list-style-type: none"> <li>- les interfaces relatives à l'évaluation,</li> <li>- les interfaces relatives à l'interprétation,</li> <li>- les outils d'évaluation et d'interprétation,</li> <li>- la base de données enrichie par les connaissances.</li> </ul>
<b>Module de gestion de connaissances</b>	Dans ce module, l'architecture du système doit considérer les composants suivants : <ul style="list-style-type: none"> <li>- les interfaces relatives à la génération automatique des solutions possibles au problème décisionnel,</li> <li>- la génération automatique des solutions possibles, en se basant sur la connaissance extraite par la fouille de données,</li> <li>- la base de données et de connaissances.</li> </ul>

### 3.5.2.7 Evaluation de l'architecture

L'évaluation de l'architecture de chacun des modules du SIAD/ECD est importante car elle conditionne l'évolutivité, les performances et le développement de ces modules.

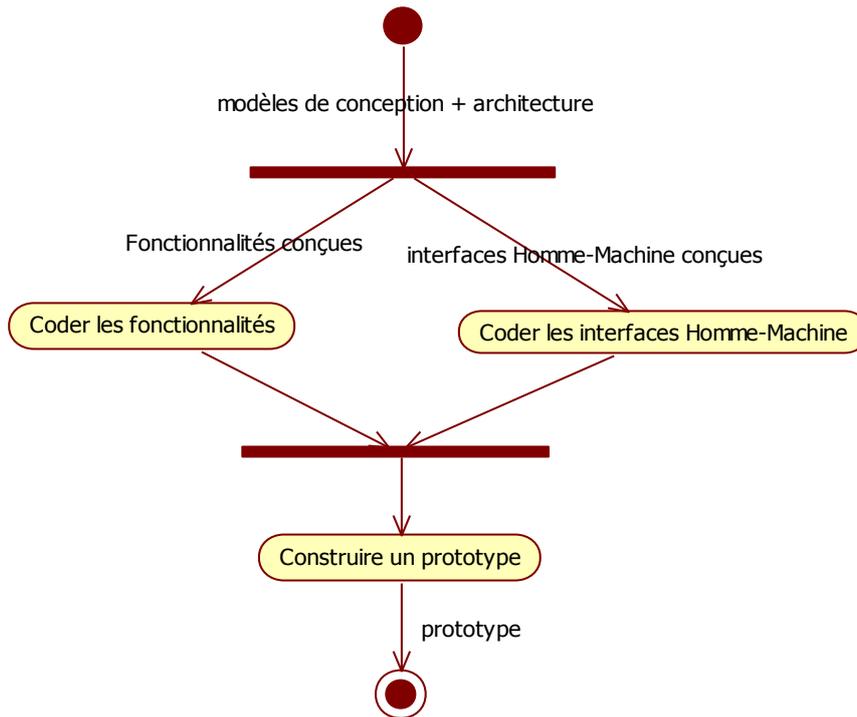
Selon le résultat de cette évaluation, il s'agit de revenir en arrière (pour résoudre les anomalies dans l'itération suivante) ou passer à l'étape suivante qui aura lieu dans l'itération en cours ou la suivante (tout dépend du plan de l'itération). Pour chacun des modules du SIAD/ECD, le tableau 3.14 présente des éléments d'évaluation de leur architecture.

**Tableau 3.14** : évaluation de l'architecture des modules du SIAD/ECD

Module	Spécificités
<b>Module de saisie et stockage de données</b>	L'expert ECD est appelé à évaluer la structure des sous-modules de saisie (transactions avec la base de données) et de prétraitement de données (base de données flexible : possibilité de réduction des dimensions et transformation des attributs).
<b>Module de fouille de données</b>	L'architecture de ce module doit être évaluée par l'expert ECD en se basant sur ses connaissances et les spécificités de la (des) technique(s) utilisée(s). Il faut s'assurer aussi que les cas à traiter peuvent être intégrés dans la base de données.
<b>Module d'évaluation</b>	L'utilisateur doit évaluer la structure du sous-module d'évaluation qualitative et quantitative. De sa part, l'expert ECD s'intéresse à l'évaluation de la structure du sous-module d'interprétation et d'intégration de la (ou des) technique(s) de visualisation à utiliser dans ce module.
<b>Module de gestion de connaissances</b>	L'architecture de ce module doit être évaluée par le décideur selon les objectifs décisionnels de départ. Elle devait permettre : <ul style="list-style-type: none"> <li>- l'intégration des modèles de connaissances pour la prise de décision,</li> <li>- l'incrémentation continue des connaissances dans la base de données du module.</li> </ul>

### 3.5.3 Implémentation

L'activité d'implémentation se décompose également en plusieurs sous-activités (cf. figure 3.8).



**Figure 3.8** : diagramme d'activité d'implémentation

### 3.5.3.1 Codage des fonctionnalités et des interfaces Homme-Machine

L'implémentation consiste à coder les parties fonctionnelles et IHM pour chaque module du SIAD/ECD à réaliser. Le codage des fonctionnalités doit respecter les algorithmes définis à l'étape précédente (conception).

Cette étape d'implémentation transforme les caractéristiques d'interfaces concrètes (définies dans les diagrammes de séquences UML associés à d'éventuelles maquettes plus ou moins avancées) en représentations qui peuvent être implémentées directement par une boîte à outils graphiques ou par un générateur d'interface homme-machine.

Le tableau 3.15 précise des éléments de codage des fonctionnalités et des interfaces homme-machine pour chacun des modules de SIAD/ECD.

**Tableau 3.15** : codage des fonctionnalités et des interfaces homme-machine des modules du SIAD/ECD

Module	Spécificités
<b>Module de saisie et stockage de données</b>	<ul style="list-style-type: none"> <li>- Implémentation des interfaces Homme-Machine et des fonctionnalités de saisie et de gestion des transactions avec la base de données.</li> <li>- Implémentation des interfaces Homme-Machine et des fonctionnalités de sélection, nettoyage et transformation des données en tenant compte des techniques de fouille choisies.</li> <li>- Implémentation de la base de données contenant toutes les tables nécessaires à l'application de la (ou des) technique(s) de fouille de données.</li> </ul>
<b>Modules de fouille de données</b>	<ul style="list-style-type: none"> <li>- Implémentation des interfaces homme-machine de la (ou des) technique(s) de fouille de données choisi(s) dans les étapes antérieures au cours des itérations précédentes.</li> <li>- Implémentation du (ou des) algorithme(s) de fouille de données. Il est important d'appliquer le (ou les) algorithme(s) développé(s) sur les données préparées pour avoir le résultat attendu.</li> </ul>
<b>Module d'évaluation</b>	<ul style="list-style-type: none"> <li>- Implémentation des interfaces homme-machine pour : (1) la représentation visuelle des motifs découverts et (2) l'application des techniques interactives (questionnaire, etc.) et des résultats de l'évaluation et de l'interprétation de ces motifs (visualisation de la connaissance extraite).</li> <li>- Implémentation des techniques d'évaluation et d'interprétation des motifs découverts. Ce code doit donner la possibilité d'enlever les motifs redondants ou non-représentatifs et de les transformer dans des termes compréhensibles pour l'utilisateur.</li> </ul>
<b>Module de gestion de connaissances</b>	<ul style="list-style-type: none"> <li>- Implémentation des interfaces homme-machine et des fonctionnalités de l'intégration des connaissances et de la génération des solutions possibles associées. Ces fonctionnalités dépendent de la nature de la connaissance extraite par les outils de fouille de données.</li> <li>- Implémentation des interfaces homme-machine et des modules fonctionnels de la prise de décision.</li> </ul>

### 3.5.3.2 Construction d'un prototype

L'ensemble des composants de code sont ensuite assemblés et intégrés [Meinadier 02] en vue de construire un prototype en fin d'itération.

En fait, le système se développe par prototypage en incrémentant ses fonctionnalités graduellement. Les prototypes de chaque module du SIAD/ECD tiennent compte de plus en plus des détails de spécifications. En particulier, les considérations de l'efficacité, de la robustesse et de la maintenance sont prises en compte dès même le début du développement (selon une démarche classique en Génie Logiciel).

Le tableau 3.16 positionne la construction de prototype pour chaque module du SIAD/ECD.

**Tableau 3.16** : construction de prototype des modules du SIAD/ECD

Module	Spécificités
<b>Module de saisie et stockage de données</b>	<p>Cette étape consiste à assembler les composants de code des interfaces homme-machine et des fonctionnalités de saisie et de prétraitement de données. Un prototype de saisie et de stockage des données est construit.</p> <p>Chaque itération fournit un prototype de plus en plus élaboré et de plus en plus proche du module de saisie et de stockage des données à produire.</p>
<b>Modules de fouille de données</b>	<p>Il s'agit d'assembler les composants de code des interfaces homme-machine et de (ou des) algorithmes de fouille de données pour l'obtention d'un prototype.</p>
<b>Module d'évaluation</b>	<p>Les composants de code à assembler concernent les interfaces homme-machine et les modules fonctionnels d'évaluation et d'interprétation des patterns pour fournir un prototype du module d'évaluation.</p>
<b>Module de gestion de connaissances</b>	<p>Les composants de code à assembler concernent les interfaces homme-machine et les modules fonctionnels de génération des solutions possibles ainsi que la prise de décision. Un prototype du module de gestion de connaissances est construit.</p>

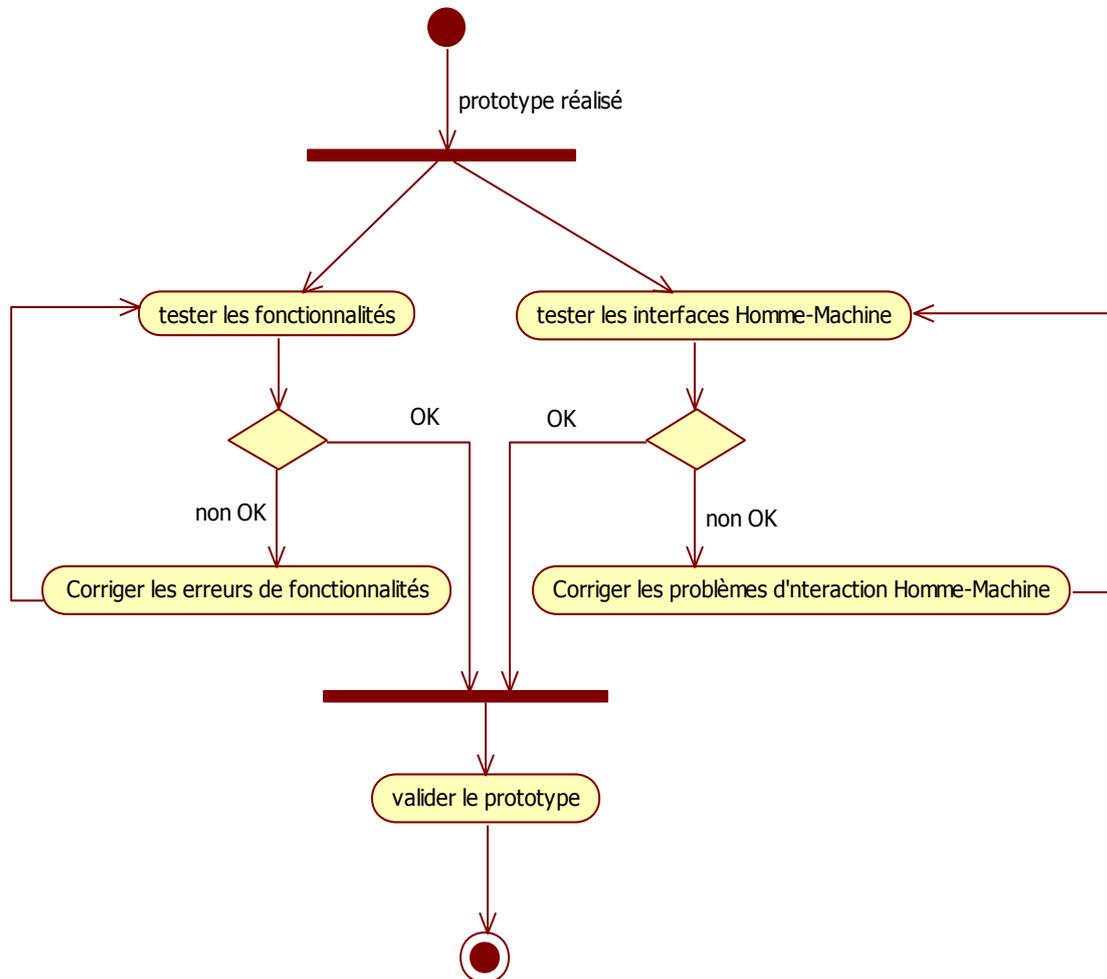
### 3.5.4 Test

L'activité de test consiste à évaluer le prototype réalisé dans l'étape d'implémentation. Il s'agit de s'assurer en particulier que l'utilisateur est capable de réaliser sa tâche au moyen de l'interface proposée [Scapin 86] [Nielsen 93] [Bahloul 10]. Les tests d'implémentation mettront en évidence les défauts éventuels qui peuvent être de type fonctionnel, en rapport avec la performance ou de type interactif.

Cette étape commence par le test des fonctions, qui ont été développées durant l'itération. Elle peut débiter dès que la fonction a été codée. Lorsqu'une non-conformité est détectée, il faut alors corriger les bugs et les anomalies. Sinon, les composants de code vont être progressivement assemblés au code testé lors des itérations précédentes.

Lors de cette étape, il s'agit de tester aussi les interfaces Homme-Machine du prototype réalisé, la façon dont chacune d'elles communique et se comporte dans le nouvel environnement. Dans cette évaluation, on s'intéresse généralement aux performances du système global, selon les comportements des utilisateurs lors de leur interaction avec le système (le temps requis pour l'exécution d'une tâche, l'exactitude du résultat, le nombre et le type des erreurs, les difficultés rencontrées, le jugement de l'utilisateur concernant notamment l'interface de dialogue, etc.).

L'itération aboutit à une validation d'un prototype, si ses fonctionnalités et ses interfaces homme-machine sont validées. Sinon, les corrections seront planifiées pour l'itération suivante.



**Figure 3.9** : diagramme d'activité de Test

Le tableau 3.17 présente des exemples de tests pour chaque module du SIAD/ECD :

**Tableau 3.17** : tests des modules du SIAD/ECD

Module	Spécificités
<b>Module de saisie et stockage de données</b>	Mise en œuvre de tests utilisateurs pour vérifier : <ul style="list-style-type: none"> <li>- l'utilisabilité au niveau des saisies, le choix des données à sélectionner, à nettoyer et à transformer.</li> <li>- La facilité d'interaction avec les interfaces homme-machine du module surtout lors des opérations de prétraitement.</li> </ul>
<b>Modules de fouille de données</b>	En vue de la validation et l'affinage du prototype réalisé de fouille de données des tests utilisateurs doivent être planifiés pour : <ul style="list-style-type: none"> <li>- Vérifier si les interfaces homme-machine sont utiles c'est-à-dire qu'elles fournissent les fonctions nécessaires à l'utilisateur. Ce type de test permet de vérifier si l'algorithme de fouille développé répond bien aux besoins de l'utilisateur (validité du résultat, temps d'exécution, etc.).</li> <li>- Vérifier la facilité de manipulation et rendre compte de la qualité de l'interface en termes de facilité d'apprentissage et d'utilisation.</li> </ul> Ce cycle de production-évaluation d'un prototype de fouille de données est répété jusqu'à atteindre des patterns satisfaisants pour les utilisateurs.
<b>Module d'évaluation</b>	Mise en œuvre des tests pour vérifier si l'utilisateur arrive à bien utiliser le questionnaire d'évaluation. D'autres tests peuvent être effectués par l'expert ECD pour vérifier si les modèles sont bien visualisés et interprétés et que les utilisateurs arrivent à extraire des connaissances pour l'aide à la prise de décision.  D'autres tests peuvent être réalisés pour valider les interfaces Homme-Machine du module.
<b>Module de gestion de connaissances</b>	Mise en œuvre des tests pour vérifier si les utilisateurs peuvent interagir facilement avec les solutions possibles générées et prendre des décisions.  Les interfaces homme-machine de ce module devant être aussi testées par les utilisateurs en vue de les valider.

### 3.6. Conclusion

Dans ce chapitre, nous avons présenté l'ensemble des phases et des activités de notre démarche que nous avons appelée le processus unifié étendu (**ExUP** : **Extended Unified Process**) et qui permet un développement centré utilisateur d'un système interactif d'aide à la décision basé sur l'extraction de connaissances à partir des données (SIAD/ECD).

Nous avons envisagé une décomposition du SIAD/ECD en quatre modules (en se référant aux étapes du processus d'ECD) : saisie et stockage de données, fouille de données, évaluation des données et gestion de connaissances. Chacun de ces modules est considéré lui-même comme un système interactif.

L'ExUP traite chaque module comme étant un système à développer. Le développement de chaque module SIAD/ECD se déroule suivant les phases du processus unifié. Dans le cadre de chaque phase, il s'agit de procéder en itérations (principe de développement itératif). Les itérations désignent des étapes de l'enchaînement d'activités. Notre travail consiste à enrichir les activités du processus unifié sous l'angle des interactions homme-machine (d'où l'appellation de processus unifié étendu). Elle préconise une évaluation continue (démarrant dès la phase de capture des besoins) pour une meilleure maîtrise des risques et une prise en compte permanente de l'utilisateur.

Dans le chapitre suivant, nous nous focalisons sur l'application de la démarche proposée à un cas concret pour le développement d'un SIAD/ECD dans le domaine médical. Ce système doit apporter un support à la lutte contre les infections nosocomiales dans les services de réanimation. L'objectif de ce cas d'application est d'évaluer et de valider la démarche proposée dans le cadre de cette thèse.

## **Chapitre 4 : Application de la démarche proposée pour le développement d'un Système Interactif d'Aide à la Décision Dynamique Médicale basé sur l'ECD**

### **4.1. Introduction**

Les chapitres précédents ont abouti à la présentation de notre démarche de développement d'un SIAD/ECD intitulée ExUP (**Extended Unified Process**). Cette approche va maintenant être appliquée à un cas concret dans le domaine médical pour le développement d'un SIAD visant la lutte contre les infections nosocomiales. Ce système a été développé en collaboration avec l'équipe du service de réanimation du Centre Hospitalo-universitaire Habib Bourguiba à Sfax, Tunisie. Il a pour but d'aider les médecins à prévoir l'apparence des infections nosocomiales pour les patients en réanimation.

Ce chapitre est organisé en quatre sections. Dans la deuxième section, nous allons présenter le contexte de travail. La troisième section sera consacrée à la définition de l'aspect dynamique de notre SIAD/ECD pour la lutte contre les IN, on parle de SIADDM/ECD. En finalement, nous allons appliquer l'ExUP au développement de SIADDM/ECD.

### **4.2. Contexte de travail**

#### **4.2.1 Les infections nosocomiales**

Les infections nosocomiales (IN) représentent un des problèmes majeurs de la santé publique. Ce sont des infections contractées dans un établissement de soins. Une infection est considérée comme telle lorsqu'elle était absente au moment de l'admission du patient [Garner *et al.* 88]. Lorsque l'état infectieux du patient à l'admission est inconnu, l'infection est classiquement considérée comme nosocomiale si elle apparaît après un délai de 48 heures d'hospitalisation. Ces infections peuvent être directement liées aux soins ou simplement survenir lors de l'hospitalisation indépendamment de tout acte médical. Dans les services de

réanimation le problème des IN est d'autant plus préoccupant, que les patients qui y sont hospitalisés sont plus fragiles. La lutte contre les IN est un problème complexe.

La plupart des patients admis en réanimation du CHU Habib Bourguiba de Sfax sont dits "lourds" car ils nécessitent beaucoup de soins et une surveillance continue 24h/24h. Ils sont souvent branchés à des machines (respirateur artificiel, électrocardiogramme, seringue électrique, etc.) et/ou connectés à des cathéters (cathéters veineux, sonde urinaire, drain thoracique...). Ces patients sont souvent très fragiles. Ils sont sensibles à tout nouveau germe qui entre dans leurs corps. A chaque apparition d'infection (nosocomiale ou non), un échantillon est envoyé au laboratoire pour y réaliser un antibiogramme<sup>14</sup>. Selon le résultat de l'antibiogramme, une antibiothérapie est prescrite. Le problème de l'antibiothérapie est qu'un germe peut être sensible à un antibiotique à une période et résistant quelques semaines ou mois plus tard. Cette sensibilité peut être différente d'un individu à un autre.

Plusieurs travaux ont été effectués pour lutter contre ces infections nosocomiales. Une brève description de ces travaux fait l'objet de la section suivante.

#### **4.2.2 Travaux antérieurs**

Une étude sur la prévalence de la survenue d'infections nosocomiales dans le CHU Habib Bourguiba à Sfax, Tunisie, a montré que 17,9 % des 280 patients hospitalisés dans l'ensemble de l'hôpital ont été victimes d'une IN entre le 17 avril 2002 (minuit) et le 18 avril 2002 (minuit) [Kallel *et al.* 05]. Notre projet est en lien avec les médecins du service de réanimation de cet hôpital. Ces médecins sont les futurs utilisateurs (experts du domaine) du SIAD/ECD que nous étudions.

Plusieurs thèses et travaux ont été publiés sur le domaine par l'équipe des médecins du service [Gafsi-Moalla 05] [Hergafi 06]. Pour la réalisation de ces études, il a été demandé, jusqu'à maintenant, de saisir des informations sur des formulaires. Ces informations sont ensuite saisies et stockées dans un fichier Excel et analysées par un logiciel de statistiques tel que SPSS. Ces outils ne permettent d'obtenir que des statistiques "classiques" (pourcentages, moyennes, comparaison de moyennes, etc.). Les médecins voudraient aller plus loin dans leur

---

<sup>14</sup> Un antibiogramme est un examen qui permet de déterminer les antibiotiques auxquels un germe est sensible ou résistant.

analyse afin d'extraire des connaissances et des informations leurs permettant de mieux comprendre, prédire et prévenir l'apparition des IN.

Dans ce contexte, quelques travaux ont proposé des systèmes informatiques de surveillance des IN basés sur les techniques de fouille de données [Brossette *et al.* 98] [Brossette *et al.* 00]. Les travaux montrent leur efficacité et leur capacité à produire des règles utiles. Mais, tels qu'elles sont décrites dans les articles, leur utilisation directe par des médecins nous paraît difficile. De ce fait, nous proposons d'appliquer notre démarche de conception centrée utilisateur de SIAD/ECD, proposée dans le chapitre 3, pour répondre à ces besoins.

### **4.3. SIADDM/ECD : SIAD Dynamique Médicale basé sur l'ECD**

Le risque d'une IN peut affaiblir le patient ou retarder sa guérison. Ce risque est principalement conditionné par la fragilité du patient et les techniques de réanimation utilisées pour sa survie. Notre objectif est de prévoir l'occurrence d'une IN **chaque jour** durant la période d'hospitalisation. C'est dû au fait qu'un patient peut voir son état empirer ou s'améliorer du point de vue du risque d'attraper une IN pendant son séjour dans un service de soins intensifs. Ceci illustre que dans des situations journalières, des décisions sont incluses dans un plus large contexte et font partie d'un cycle décision-action affecté par la surveillance et la rétroaction. Les éléments intervenant dans ce cycle sont des données extraites à partir des rapports microbiologiques, d'interviews médicales avec des infirmières et des médecins responsables du patient, etc. Ces données sont généralement temporelles. Notre travail concerne ainsi l'exploitation des données temporelles rassemblées, pour l'acquisition **dynamique** de la connaissance utile pour la décision.

#### **4.3.1 SIADDM pour la lutte contre les IN**

L'aspect dynamique du système est observé à divers niveaux de prise de décision. Pour l'analyser, il faut prendre en considération un ensemble de facteurs critiques de décision qui sont identifiés à l'aide d'un ensemble d'interviews avec les médecins responsables dans le service de réanimation. L'identification des facteurs supportant l'apparence des infections est une étape très significative qui influence les résultats de la prise de décision. Ces facteurs sont classifiés en deux catégories :

**1) Les données caractéristiques collectées à l'admission du patient :** ce sont les informations démographiques et administratives (âge, sexe, dates d'admission et de sortie du service de réanimation, localisation avant l'admission, etc.), les antécédents, les diagnostics d'admission, et le type d'admission. On peut noter aussi dans cette catégorie, les deux scores suivants :

- **Le score SAPS II** (Simplified Acute Physiology Scores II) : ce score est utilisé pour évaluer et comparer la gravité des patients aux soins intensifs. Il s'agit d'un modèle prédictif de mortalité des patients. Plusieurs études de recherche [Bueno-Cavanillas *et al.* 91] [Lortholary *et al.* 95] ont prouvé que ce score mesuré dans les premières 24 heures du soin intensif est un indicateur de risque pour l'IN.
- **La catégorisation APACHE** développée par William Knaus [Knaus *et al.* 81] : ce n'est pas un score pour prédire l'IN, mais plutôt une classification de l'état antérieur du malade qui est statistiquement liée à la survenue d'IN. On distingue **A** : bon état de santé antérieur ; **B** : limitation modérée de l'activité ; **C** : limitation majeure de l'activité ; et **D** : grabataire (ne bouge pas de son lit).

Statistiquement, plus l'état antérieur est altéré, plus le risque d'acquérir une infection en réanimation est grand. SAPS II et APACHE sont des valeurs mesurées à l'admission du patient pour prévoir le risque de mortalité. Plus le patient est fragile, plus il risque de contracter une IN.

**2) Les facteurs quotidiens de contrôle :** celles-ci concernent l'exposition à divers facteurs de risque d'infection (dispositifs invasifs, antibiotiques, traitements immunosuppresseurs, etc.). Plus spécifiquement, on parle de :

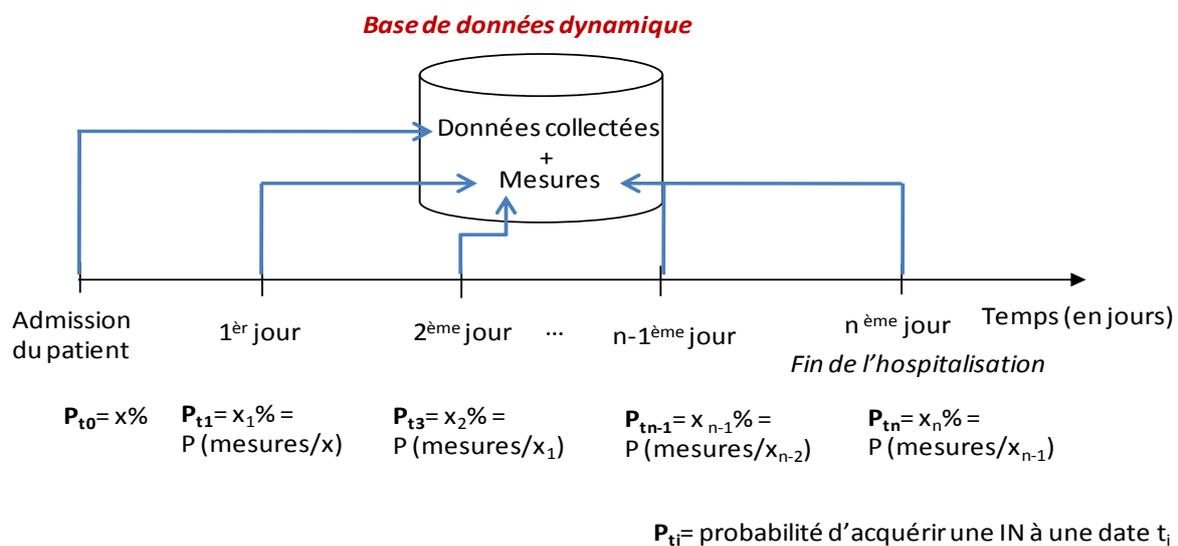
- *L'intubation*: pénétration d'une sonde, à l'intérieur de la trachée, pour assurer la liberté des voies aériennes.
- *Le Cathéter Veineux Central (C.V.C.)* : un cathéter est un tube creux de plastique souple que le médecin place dans une veine du cou, de l'épaule ou de la cuisse. La présence de micro-organismes à la surface interne et/ou externe du C.V.C. peut causer une infection locale et/ou générale. Merrer *et al.* [Merrer *et al.* 01] ont montré que plus de 15% des patients qui reçoivent ces cathéters ont des complications infectieuses.

- *La sonde urinaire* : le sondage urinaire est un geste invasif couramment réalisé dans les services hospitaliers. La durée du sondage est un facteur de risque d'IN. En fait, le risque d'infection nosocomiale urinaire est multiplié par plus de 10 en cas de sondage à demeure et augmente avec la durée du sondage. L'utilisation des dispositifs urinaires peut causer l'infection urinaire qui est l'infection nosocomiale la plus fréquente (40 % des cas) [Haley *et al.* 85].
- *Les examens infectieux* : s'il y a une infection (pneumonie, bactériémie, infection liée au C.V.C., etc.), il s'agit de définir la localisation et le germe responsable ainsi que les antibiotiques à prescrire. En fait, les germes responsables de l'apparition d'infections hospitalières sont variables selon la pathologie présentée par le patient. Plusieurs études dans la littérature montrent des fréquences élevées d'apparition d'IN [Kallel *et al.* 05]. On peut citer par exemple [Pittet *et al.* 93] pour les bactériémies, et [Veenstra *et al.* 99] pour les infections liées au C.V.C.
- *La prise des antibiotiques* : les antibiotiques utilisés peuvent avoir un effet néfaste à long terme. Les micro-organismes sont de plus en plus résistants aux antibiotiques mêmes les plus récents. Donc, il s'agit de procéder tout d'abord à une recherche de sensibilité de façon à ne pas utiliser n'importe quel antibiotique (utilité de l'antibiogramme). Sinon on voit apparaître une sélection de souches hospitalières multi-résistantes pouvant se transmettre de patient à patient.

Notre étude vise l'estimation quotidienne de la probabilité d'acquérir une IN pendant l'hospitalisation de patient dans le service de réanimation. Cette probabilité est calculée en fonction des facteurs décrits ci-dessus afin de prévoir le risque d'infection sur la base des patients enregistrés. Chaque jour  $i$  ( $1 \leq i \leq n$  = durée d'hospitalisation), la décision sur l'état du patient dépend de la probabilité  $P_i$  d'IN et donc des valeurs des facteurs décrits ci-dessus, mais également des connaissances obtenues par l'étude temporelle d'anciens événements (cf. Figure 4.1). Dans ce cadre, il s'agit de mettre l'accent sur des décisions liées ou interdépendantes, ou encore des décisions qui s'auto-influencent.

En fait, "une décision de base" est prise à l'admission du patient ( $t_0$ ). La "future décision" se rapporte à une décision à prendre après que les conséquences d'une décision de base deviennent (partiellement) connues. Une future décision est liée "à la décision de base" parce que les solutions qui seront disponibles à l'avenir dépendent du choix fait dans "la décision de base". Pendant que le temps passe, la "future décision" à l'étape courante ( $t$ ) devient "la

décision de base" à la prochaine étape de décision (t+1), quand de nouvelles connaissances (probabilité d'acquérir une IN) et "future décision" devraient être adressées. Ce lien se répète jusqu'à la fin d'hospitalisation du patient. Le modèle "**décider-apprendre**" [Ge *et al.* 07] décrit comment le décideur répond à de nouvelles connaissances acquises pendant le processus décisionnel. Les éléments décrits ci-dessus, particulièrement l'existence des décisions liées, prouvent clairement que la prise de décision pour la surveillance des IN est un processus **dynamique**. Ce modèle dynamique de prise de décision est une chaîne consistant à apprendre puis décider, puis apprendre plus puis décider ; et ainsi de suite<sup>15</sup>. On parle donc de Système Interactif d'Aide à la **D**écision **D**ynamique Médicale (SIADDM)



**Figure 4.1** : les facteurs temporels pour la prévention d'IN

Le SIADDM visé est basé sur l'ECD. Comme expliqué dans le chapitre 3, ceci nécessite donc la conception et le développement de quatre modules constitutifs de SIADDM/ECD.

### 4.3.2 Les modules visés pour le SIADDM/ECD

**1. Saisie et stockage des données** : ce module permet de saisir les données personnelles des patients ainsi que les données relatives à leur hospitalisation. Les étapes de prétraitement consistent à construire un corpus des données spécifiques, nettoyant les données, traitant les

<sup>15</sup> On peut souligner que l'aspect dynamique de la décision a été traité par des tentatives de modélisation humaine, issues des sciences cognitives. On peut citer par exemple la modélisation de la démarche de résolution d'un problème de décision par Rasmussen [Rasmussen 86] (qui vise la modélisation des différents types de comportement humains identifiés), ce modèle ayant été révisé par Hoc et Amalberti sous l'angle de la dynamique [Hoc et Amalberti 95].

données absentes, sélectionnant des attributs et puis transformant ces données pour être utilisables par un algorithme d'extraction de connaissances. Ces étapes sont cruciales pour la recherche de l'information pertinente nécessaire au processus décisionnel [Turban 93].

Le sous-module de saisie des données a été développé dans le cadre de mon travail de maîtrise [Ltifi 07]. Alors que les sous-modules de sélection, nettoyage et transformation des données ont été développés dans le cadre d'un autre travail de maîtrise [Trabelsi 09] [Trabelsi *et al.* 10] (cf. Figure 4.2).

**2. Fouille des données :** dans ce module, la (ou les) technique(s) d'ECD utilisées dans le cadre de la prévention des IN, permet la découverte de la connaissance nécessaire pour interpréter les données recueillies des patients. Dans certains cas cette connaissance est exprimée en tant que rapports probabilistes entre les dispositifs cliniques et les solutions proposées. Dans d'autres cas, une représentation à base de règles peut être choisie afin de fournir au décideur une explication de la décision. La fouille de données permet donc de mener à la connaissance mise sous forme de modèles qui doivent être validés.

Une première application a été développée dans le cadre de mon travail de maîtrise [Ltifi 07]. La technique de fouille de données utilisée est les k-plus proches voisins (KPPV), technique de raisonnement à partir de cas [Riesbeck 89]. Pour mieux traiter l'aspect dynamique de décision, il nous semble important de développer une autre application de fouille de données qui sera présenté dans la section suivante. De ce fait, le SIADDM/ECD va contenir deux applications, utilisant chacune une technique différente pour des objectifs différents (cf. Figure 4.2).

**3. Evaluation des modèles :** ce module permet d'évaluer les modèles découverts. Une fois les modèles découverts sont évalués, ils seront ensuite interprétés pour les rendre compréhensibles par le décideur.

Ce module a été modélisé et développé dans le cadre d'un travail de maîtrise de [Bahloul 07] [Bahloul *et al.* 10] (cf. Figure 4.2).

**4. Gestion des connaissances :** ce module a pour objectif l'intégration des connaissances extraites pour la génération des solutions possibles en vue de guider la résolution du problème de décision médicale [Haug *et al.* 03].

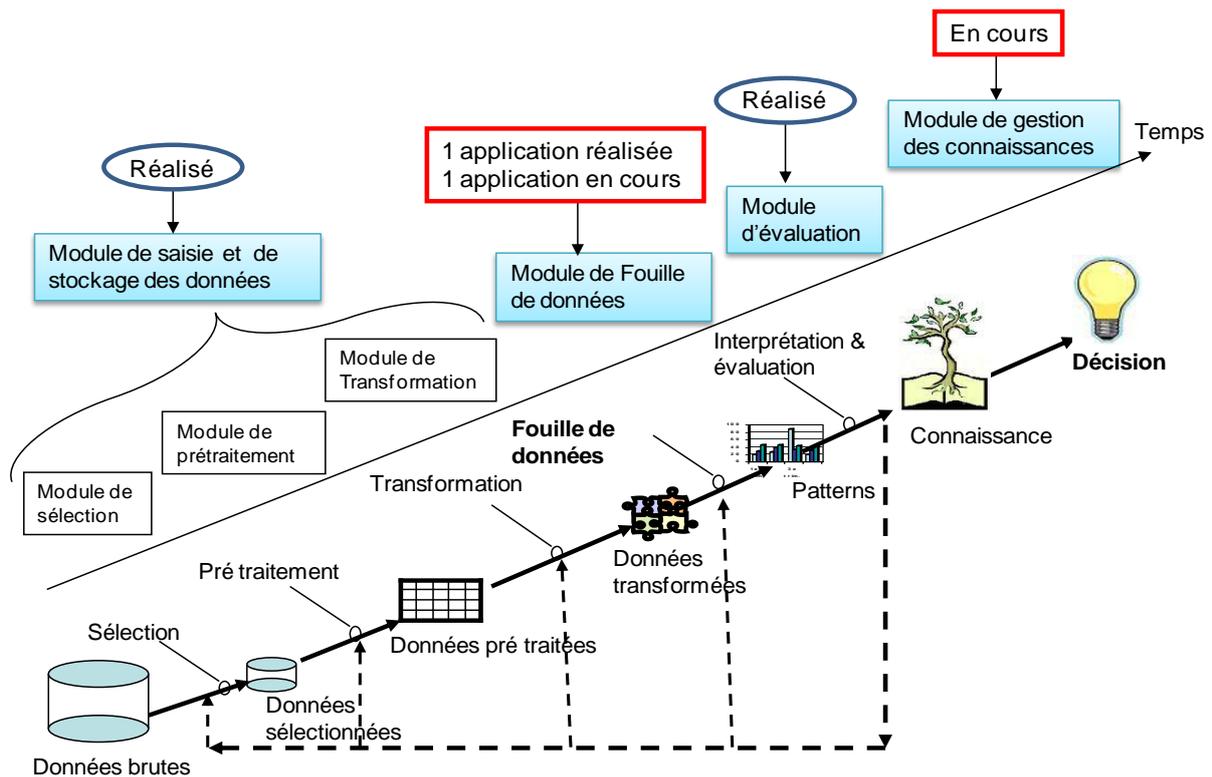


Figure 4.2 : bilan relativement aux modules visés

Dans le cadre de ce travail, nous envisageons d'appliquer l'ExUP pour développer les modules restants du SIADDM/ECD (cf. Figure 4.2) : technique dynamique de fouille de données et module de gestion des connaissances. C'est l'objet de la section suivante.

#### 4.4. Mise en application

Nous abordons dans cette section l'aspect applicatif de notre projet décisionnel visant la lutte contre les infections nosocomiales. Il s'agit de concevoir et réaliser les deux modules SIADDM/ECD non encore développés (cf. Figure 4.2). Tout au long de ce chapitre, nous utilisons l'ExUP comme démarche de conception. De ce fait, Pour chaque module, le travail abordé sera divisé en quatre phases :

- *La phase d'inception* : au cours de laquelle les principaux cas d'utilisation sont identifiés.
- *La phase d'élaboration* : où l'analyse et la conception de la plupart des fonctionnalités et des IHM du module sont abordées.
- *La phase de construction* : au cours de laquelle la conception et la réalisation du module sont achevées.
- *La phase de transition* : celle-ci est dédiée au test des fonctionnalités et des IHM du module. La correction et l'amélioration de ces dernières sont aussi abordées.

#### 4.4.1 Module de fouille de données

Nous nous consacrerons dans cette section à la conception et le développement du module de fouille de données.

##### 4.4.1.1 Phase d'inception

Durant cette première phase de l'ExUP, nous avons essayé d'établir une vision initiale du module de fouille de données, de comprendre le principe de la technique de fouille à utiliser, de définir les besoins fonctionnels et non fonctionnels, d'identifier et de décrire les principaux cas d'utilisation ainsi que de modéliser l'utilisateur. Cette phase s'est déroulée en une seule itération.

**Tableau 4.1** : déroulement de la phase d'inception pour le module de fouille de données

Activités	Itération 1
<b>Capture des besoins</b>	
Analyse des situations décisionnelles	<p>Cette itération démarre par une étude des différentes techniques de fouille de données pouvant être utilisées pour un système dynamique de d'aide à la prise de décision.</p> <p>La situation décisionnelle dans le cadre de ce module est dynamique. La technique de fouille à utiliser doit respecter l'aspect dynamique de la décision. La plupart des techniques utilisent des algorithmes qui fonctionnent sur des données fixes non temporelles (chaque individu est représenté par un vecteur de caractéristiques). Dans plusieurs domaines, tels que celui de la santé, les données concernant un individu sont saisies à des moments différents d'une manière plus ou moins périodique.</p> <p>Les Réseaux Bayésiens Dynamiques (RBD) [Darwich 01] [Murphy 02] sont des modèles représentant des connaissances incertaines sur des phénomènes complexes dans le cadre d'un processus dynamique. Il s'agit d'obtenir des modèles de connaissances qui évoluent avec le temps. De ce fait nous avons choisi les réseaux bayésiens dynamiques qui nous semblent la technique la plus adéquate à notre contexte.</p> <p>La description de la technique des <i>Réseaux bayésiens dynamiques</i> est fournie en § 4.4.1.5</p>
Élaboration des premières maquettes d'IHM	Des réunions avec des futurs utilisateurs (médecins) et l'expert ECD permettent d'élaborer des prototypes d'IHM de l'application de la technique de fouille des données. Ces prototypes déterminent l'architecture générale de ce module.
Modélisation de l'utilisateur	Une modélisation préliminaire de l'utilisateur (type unique pour le moment) est réalisée : il est expert dans son domaine (nombreuses publications et encadrements de thèses en lien avec la lutte contre les IN), expert aussi concernant l'utilisation des outils informatiques (plus

	de 3 heures par jour devant son PC).
Expression des besoins décisionnels	<p>Les premiers contacts avec les médecins (futurs utilisateurs du module de fouille de données) et l'expert ECD, ont permis une étude de l'existant. Les fonctionnalités que nous envisageons de réaliser dans le cadre de ce module peuvent être décrites comme suit :</p> <ul style="list-style-type: none"> <li>- le système doit permettre de classifier quotidiennement l'état de tout nouveau patient. Ce dernier sera considéré comme un nouveau cas de prédiction à étudier. On parle donc d'une classification dynamique.</li> <li>- une fois la classification du cas de prédiction est faite, le système doit permettre de l'ajouter en tant que cas d'apprentissage dans la base de données. Cet apprentissage sert pour une future classification d'un nouveau cas.</li> <li>- le système doit permettre aussi de calculer la probabilité quotidienne d'acquisition d'une IN. C'est une prédiction dynamique.</li> </ul> <p>Dans cette itération, un diagramme de cas d'utilisation a été réalisé. Les diagrammes de cas d'utilisation sont faciles à représenter et constituent des outils essentiels et pédagogiques pour expliquer aux médecins, généralement non-experts en fouille de données, les caractéristiques externes du module concerné.</p>
Evaluation des besoins	<p>Dans le cadre d'une réunion avec les médecins et l'expert ECD, une démonstration du modèle de cas d'utilisation (le diagramme de cas d'utilisation et les descriptions textuelles associées) a été effectuée pour une évaluation précoce des besoins fonctionnels définis.</p> <p>Le concepteur et l'expert ECD ont expliqué aux médecins comment les fonctionnalités définies permettent de prédire la probabilité d'acquérir une IN. Les médecins ont validé par la suite le modèle de cas d'utilisation.</p>
Répartition et définition des tâches	<p>Une première description et répartition des tâches du module de fouille de données est faite dans le cadre de cette itération.</p> <p>Les tâches de ce module sont les suivants : la classification dynamique d'un cas de prédiction, l'ajout d'un cas d'apprentissage et le calcul de la probabilité quotidienne d'acquisition d'une IN.</p>
Définition des tâches manuelles	Dans le cadre de ce module, il n'y a pas des tâches manuelles.
Définition des tâches automatiques	Les tâches automatiques de ce module sont : l'ajout d'un cas d'apprentissage et le calcul de la probabilité quotidienne d'acquisition d'une IN.
Définition des tâches interactives	La tâche interactive de ce module est la classification des cas de prédiction au sujet desquels l'utilisateur doit introduire des informations (tel que le choix du jour de prédiction).
Evaluation des tâches	Une documentation détaillée des tâches dégagées dans les étapes précédentes a été proposée aux médecins et à l'expert ECD pour l'évaluation des ces tâches automatiques et interactives.
<b>Analyse et conception</b>	
<i>Sans objet dans cette phase</i>	
<b>Implémentation</b>	
<i>Sans objet dans cette phase</i>	
<b>Test</b>	
<i>Sans objet dans cette phase</i>	

Cette phase nous a permis d'avoir une compréhension assez claire de la vision et de la faisabilité du module, nous pourrions passer à la phase d'élaboration où il y a lieu d'entamer l'analyse et la conception des cas d'utilisation.

#### 4.4.1.2 Phase d'élaboration

La phase d'élaboration permet de spécifier, analyser et concevoir la plupart des cas d'utilisation. C'est une phase qui s'est déroulée en trois itérations. Dans chaque itération, nous avons présenté les nouveaux cas d'utilisation, l'analyse de ces cas ainsi que la conception de ceux qui sont déjà analysés dans l'itération en cours ou dans les itérations précédentes.

**Tableau 4.2 :** déroulement de la phase d'élaboration pour le module de fouille de données

Activités	Itération 1	Itération 2	Itération 3
<b>Capture des besoins</b>			
Analyse des situations décisionnelles	Nous avons commencé par une étude des travaux internationaux et nationaux réalisés sur la technique des RBD, surtout en ce qui concerne la classification.  En fait, pour pouvoir appliquer une classification dynamique d'un nouveau cas de prédiction, il faut tout d'abord l'ajouter dans la base de données préparée.	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
Élaboration des premières maquettes d'IHM	Une maquette papier a été réalisée par l'expert ECD lors d'une réunion de capture des besoins pour expliquer le déroulement de la classification dynamique.	L'utilisateur a proposé une maquette papier présentant des options de navigation dans l'IHM de classification dynamique.	Aucun changement au niveau de l'architecture initiale des IHM proposées.
Modélisation de l'utilisateur	L'utilisateur a bien compris le principe de l'ExUP. Il trouve avantageux de pouvoir exprimer d'autres besoins et proposer des modifications au cours du déroulement du projet (acteur d'une démarche participative).	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>

Expression des besoins décisionnels	Après la compréhension du principe du cas d'utilisation de classification dynamique, on a pu le raffiner en ajoutant un nouveau cas d'utilisation : <i>ajout d'un cas de prédiction dynamique</i>	Aucun changement au niveau du diagramme de cas d'utilisation.	<i>Sans objet dans cette itération</i>
Evaluation des besoins	Les utilisateurs ont évalué le diagramme de cas d'utilisation raffiné. Ils ont trouvé que les fonctionnalités du module répondent mieux aux besoins de la fouille des données.	<i>Sans objet dans cette itération</i>	Les nouveaux besoins exprimés par le diagramme de cas d'utilisation et la documentation associée sont présentés à l'utilisateur dans un but d'évaluation. Le résultat d'évaluation est satisfaisant.
Répartition et définition des tâches	Une nouvelle tâche est définie dans ce niveau de conception. C'est la tâche d'ajout de cas de prédiction dynamique.	En se basant sur la maquette papier proposée, on dégage deux nouvelles tâches interactives (cf. ci-dessous).	<i>Sans objet dans cette itération</i>
Définition des tâches manuelles	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
Définition des tâches automatiques	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
Définition des tâches interactives	La tâche d'ajout de cas de prédiction dynamique est une tâche interactive nécessitant l'intervention de l'utilisateur lors de son exécution.	Ajout des deux tâches suivantes : <ul style="list-style-type: none"> <li>• Ajout de "Jour Suivant" et "Jour Précédent". La saisie de la date au clavier reste toujours possible.</li> <li>• Lorsque l'utilisateur saisit un nom ou un index (code administratif), le ou les patients concernés(s) sont affichés(s).</li> </ul>	<i>Sans objet dans cette itération</i>
Evaluation des tâches	L'utilisateur a détaillé certains points pour une meilleure analyse et conception de ces tâches.	L'utilisateur a trouvé que le concepteur a bien compris sa maquette papier.	
<b>Analyse et conception</b>			
Spécification des tâches automatiques	Analyse du cas d'utilisation d'ajout d'un nouveau cas (patient) pour une future classification dynamique. On a utilisé les diagrammes de communication et d'activités pour l'étape de spécification.	Analyse des tâches automatiques de classification dynamique d'un cas de prédiction.	Analyse du cas d'utilisation d'ajout d'un cas d'apprentissage.

Spécification des tâches interactives	Analyse des tâches de choix des informations relatives au patient à classifier.	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
Spécification des IHM	Spécification des IHM d'ajout d'un cas de classification dynamique.	C'est la même IHM que pour la tâche d'ajout d'un cas de classification dynamique.	<i>Sans objet dans cette itération</i>
Conception des modules automatiques d'ECD	Conception du cas d'utilisation d'ajout d'un cas de classification dynamique dans les différents scénarios. On a utilisé les diagrammes de séquence pour l'étape de conception.	Conception des tâches automatiques de classification dynamique d'un cas de prédiction.	Conception du cas d'utilisation d'ajout d'un cas d'apprentissage dans les différents scénarios (succès et échec).
Conception des IHM et des modules interactifs d'ECD	Conception de l'IHM d'ajout d'un cas de classification dynamique.	C'est la même IHM que pour la tâche d'ajout d'un cas de classification dynamique.	<i>Sans objet dans cette itération</i>
Evaluation des modèles	Les diagrammes de communication et de séquence ont été évalués par les utilisateurs pour vérifier si les concepteurs ont bien compris les besoins. Résultat satisfaisant.	Les diagrammes UML de communication et de séquence ont été évalués par l'équipe de développement pour passer à l'implémentation des tâches de classification dynamique.	<i>Sans objet dans cette itération</i>
Elaboration des maquettes avancées d'IHM	Une maquette électronique de classification dynamique a été préparée pour traduire les besoins utilisateurs.	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
Spécification de l'architecture	En se basant sur les modèles et les maquettes, on a pu définir l'architecture initiale de module de fouille de données.	L'architecture spécifiée dans l'itération précédente est légèrement modifiée.	<i>Sans objet dans cette itération</i>
Evaluation de l'architecture	L'architecture est composé des IHM et modules fonctionnels de classification dynamique. L'utilisateur a évalué cette architecture et a montré sa satisfaction de cette première vision du module.	Une re-évaluation de l'architecture de module en cours de réalisation. Résultat : satisfaction du module.	Avec les nouvelles informations, l'architecture du futur système commence à s'enrichir.
<b>Implémentation</b>			
Codage des fonctionnalités	<i>Sans objet dans cette itération</i>	Notre SIADDM/ECD fonctionne sur les données temporelles. La base des données	Codage de la fonctionnalité d'ajout d'un cas de prédiction dynamique

		temporelles a déjà été prétraitée lors des travaux antérieurs et implémentée sous SQL server. Donc il nous reste à implémenter la technique des Réseaux Bayésiens Dynamiques. Pour se faire nous avons utilisé le langage de programmation C#. net. Le codage des fonctionnalités se fait dans un fichier cs (fichier écrit en C SHARP)	
Codage des Interfaces Homme-Machine	<i>Sans objet dans cette itération</i>	Le codage des IHM se fait dans un fichier <b>resx</b> (RESource file eXtension)	Préparation de l'IHM d'ajout d'un cas de prédiction dynamique.
Construction du prototype	<i>Sans objet dans cette itération</i>	Pour exécuter une fonctionnalité sous C#, il faut assembler le fichier cs et celui resx.	Assembler le fichier cs et celui resx d'ajout d'un cas de prédiction dynamique. Le premier prototype de ce module permet un ajout d'un nouveau cas (patient) pour une future classification dynamique
<b>Test</b>			
<i>Sans objet dans cette phase</i>			

Dans cette phase, notre attention a été portée principalement sur la formulation d'une architecture stable du module et sur l'enrichissement de l'environnement de développement. Nous avons, à ce stade du projet, suffisamment d'informations pour entamer la phase de construction qui fait l'objet de la section suivante.

#### 4.4.1.3 Phase de construction

Durant cette phase qui s'est déroulé en deux itérations, nous avons achevé l'analyse, la conception et l'implémentation du module de gestion de connaissances dans l'optique de livrer un produit exploitable.

**Tableau 4.3** : déroulement de la phase de construction pour le module de fouille de données

Activités	Itération 1	Itération 2
<b>Capture des besoins</b>		
Analyse des situations décisionnelles	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
Élaboration des premières maquettes d'IHM	L'utilisateur a préparé une maquette montrant la manière avec laquelle il souhaite visualiser la probabilité d'acquisition d'IN.	<i>Sans objet dans cette itération</i>
Modélisation de l'utilisateur	L'étude de la maquette préparée dans l'étape précédente permet de modéliser les préférences de l'utilisateur en matière d'affichage des patterns découverts.	<i>Sans objet dans cette itération</i>
Expression des besoins décisionnels	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
<b>Evaluation des besoins</b>	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
Répartition et définition des tâches	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
Définition des tâches manuelles	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
Définition des tâches automatiques	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
Définition des tâches interactives	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
<b>Evaluation des tâches</b>	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
<b>Analyse et conception</b>		
Spécification des tâches automatiques	Analyse du cas d'utilisation de prédiction automatique. Le diagramme de communication de ce cas doit présenter les tâches et les sous-tâches pour le calcul de la probabilité de l'acquisition des IN.	<i>Sans objet dans cette itération</i>
Spécification des tâches interactives	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
Spécification des IHM	Spécification d'une IHM qui présente la probabilité calculée d'acquisition d'IN.	<i>Sans objet dans cette itération</i>
Conception des modules automatiques d'ECD	Conception du cas d'utilisation de calcul de probabilité quotidienne d'acquisition d'IN.	<i>Sans objet dans cette itération</i>
Conception des IHM et des modules	Conception de l'IHM présentant la probabilité d'occurrence d'IN, en se	<i>Sans objet dans cette itération</i>

interactifs d'ECD	référant à la maquette préparée par l'utilisateur dans ce contexte.	
<b>Evaluation des modèles</b>	Les diagrammes de communication et de séquences de prédiction quotidienne de la probabilité d'occurrence d'IN ont été évalués par les utilisateurs.	<i>Sans objet dans cette itération</i>
Elaboration des maquettes avancées d'IHM	A ce stade, une maquette électronique de l'IHM d'affichage de la probabilité calculée a été élaborée.	<i>Sans objet dans cette itération</i>
Spécification de l'architecture	En se basant sur les modèles et les maquettes, on a pu spécifier l'architecture finale de module de fouille de données.	<i>Sans objet dans cette itération</i>
<b>Evaluation de l'architecture</b>	L'utilisateur et l'expert ECD ont évalué cette architecture et procédé à sa validation.	<i>Sans objet dans cette itération</i>
<b>Implémentation</b>		
Codage des fonctionnalités	La fonctionnalité de classification dynamique est implémentée dans le cadre de cette itération.	Codage de la fonctionnalité de calcul quotidien de la probabilité d'occurrence d'IN. L'implémentation se base sur les scénarios du diagramme de séquence de cette fonctionnalité. Correction des bugs de la fonctionnalité d'ajout d'un cas de prédiction dans la base de données.
Codage des Interfaces Homme-Machine	L'IHM de la classification dynamique et celle d'ajout d'un cas de prédiction dynamique sont codées.	Préparation de l'IHM relative à la prédiction de la probabilité de contracter une infection nosocomiale.
Construction du prototype	Un assemblage du fichier cs de classification dynamique avec le fichier resx d'ajout d'un cas de prédiction dynamique permet d'enrichir le prototype de fouille des données. Il permet une classification dynamique d'un nouveau patient. Le prototype construit doit être évalué par les utilisateurs pour vérifier sa facilité d'utilisation et par l'expert ECD pour vérifier la validité de la classification obtenue.	Assemblage du fichier cs et du fichier resx de prédiction dynamique. Avec toutes les fonctionnalités implémentées, le prototype de fouille des données est maintenant complet.
<b>Test</b>		
Test des fonctionnalités	L'évaluation des fonctionnalités de classification permet de mettre en évidence une erreur dans l'insertion d'un cas de prédiction dans la base de données.	Deux problèmes sont détectés : 1) Problème d'exécution de l'algorithme des RBD surtout lorsque le nombre de séries de temps est important 2) Problème d'utilisabilité relatif à la date de prédiction.

Test des IHM	L'utilisateur n'a pas accepté l'IHM de classification à cause de l'illisibilité des données affichées (petites polices).	Le test montre un affichage erroné de la valeur de la probabilité d'acquisition d'une IN lorsque cette valeur contient plusieurs chiffres après la virgule.
Correction des erreurs de fonctionnalités	Les anomalies détectées seront corrigées dans la deuxième itération de cette phase de construction.	On procède à la correction du format des dates de prédiction.
Correction des problèmes d'interaction HM	Modification de l'IHM de classification en ajoutant les boutons : "Jour suivant" et "Jour précédent".	Le sous-système de fouille de données corrigé, au niveau de cette itération, doit être réévalué jusqu'à sa validation finale.
Validation du prototype	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>

Cette phase nous a permis d'atteindre les objectifs fixés et plus particulièrement, la livraison d'un produit logiciel sous forme de version opérationnelle et conforme à la conception élaborée. Une question fondamentale qui se pose à ce stade est : « le module est-il assez valide et favorable pour être livré à partir de ce niveau de développement ? ».

En effet, il n'est pas exclu que le module puisse encore être perfectionné pour atteindre un niveau de qualité permettant sa mise en exploitation. C'est l'objet de la phase de transition qui fera l'objet de la section suivante.

#### 4.4.1.4 Phase de transition

La phase de transition représente la dernière activité de l'ExUP dans laquelle nous allons essayer de stabiliser le module de fouille de données réalisé à travers les tests. Cette phase est réalisée en une seule itération, lors de laquelle les principaux objectifs qui doivent être atteints sont : la correction des bugs, le raffinement des performances et de l'utilisabilité ainsi que les tests de validation de la conformité du générateur par rapport aux besoins capturés dans les phases précédentes.

**Tableau 4.4** : déroulement de la phase de transition pour le module de fouille de données

Activités	Itération 1
<b>Capture des besoins</b>	
<i>Sans objet dans cette itération</i>	
<b>Analyse et conception</b>	
<i>Sans objet dans cette itération</i>	
<b>Implémentation</b>	
Codage des fonctionnalités	Correction des bugs détectés lors de l'application de l'algorithme de fouille de données lorsque le nombre de séries de temps est important.
Codage des Interfaces Homme-Machine	Correction de l'affichage de la valeur de la probabilité d'acquisition d'une IN.
Construction du prototype	L'exécution du prototype satisfait l'utilisateur.
<b>Test</b>	
Test des fonctionnalités	Le test de plusieurs cas de patients à la file a généré des erreurs d'implémentation. Ces erreurs sont détectées surtout lors du passage d'un cas à un autre.
Test des IHM	Quelques améliorations de détail sont apportées aux IHM (présentation de textes).
Correction des erreurs de fonctionnalités	Ajout d'un script de code d'actualisation
Correction des problèmes d'interaction HM	Amélioration de l'IHM d'affichage de probabilité d'occurrence d'IN.
Validation du prototype	Expression de la satisfaction de l'utilisateur et acceptation finale du module de fouille des données.

Cette phase nous a permis d'obtenir une version fonctionnelle et stable du module de fouille de données prête à être exploitée.

#### 4.4.1.5 Les réseaux bayésiens dynamiques exploités pour la réalisation du module de fouille de données

##### Généralités :

Un RB est un graphe causal orienté et acyclique (DAG) permettant de représenter des variables aléatoires avec leurs dépendances. Il présente la distribution des probabilités conditionnelles d'un ensemble de variables. Ses nœuds représentent les variables aléatoires et ses arcs représentent les dépendances entre ces variables [Jensen 96] [Lauritzen *et al.* 99].

La distribution de probabilité des variables aléatoires  $S = \{X_1, \dots, X_N\}$  dans un RB est calculé par la multiplication des probabilités conditionnelles locales de tous les nœuds. Soit un nœud  $X_i$  dans  $S$  dénotant la variable aléatoire  $X_i$ , et soit  $Pa(X_i)$  dénotant les nœuds parents de  $X_i$ , alors, la distribution de probabilités de  $S = \{X_1, \dots, X_N\}$  est exprimée comme suit :

$$P(X_1, X_2, \dots, X_N) = \prod_{i=1}^N p(X_i | Pa(X_i))$$

Les RB ne permettent pas de représenter les relations temporelles entre les variables aléatoires. Pour cette raison, on utilise les RB Dynamiques (RBD) [Darwich 01] [Murphy 02].

Les RBD formalisent la distribution de probabilité d'un ensemble de variables temporelles  $X[t] = \{X_1[t], \dots, X_N[t]\}$ . Si on considère  $T$  tranches de temps, le RBD peut être considéré comme un RB statique avec  $T \times N$  variables. En utilisant la propriété de factorisation des RB [Darwich 01] [Murphy 02], la densité de probabilité de  $X^T = \{X[1], \dots, X[T]\}$  peut être exprimée comme suit :

$$P(X[1], \dots, X[N]) = \prod_{t=1}^T \prod_{i=1}^N p(X_i[t] | Pa(X_i[t])) \quad \text{avec } Pa(X_i[t]) \text{ dénotent les parents de } X_i[t]$$

En raison de leur capacité à représenter les connaissances incertaines, les réseaux bayésiens (RB) jouent un rôle de plus en plus important dans beaucoup d'applications médicales. Ils ont été proposés dans les années 80 comme formalisme de représentation et de raisonnement pour des problèmes impliquant l'incertitude ; ils adoptent pour cela la théorie des probabilités. L'exploitation de ce formalisme dans le contexte de la prise de décision médicale a commencé à la fin des années 80 [Pearl 88].

La littérature médicale contient des exemples d'utilisation des RB. On peut citer un modèle de RB développé et mis à la disposition de cliniciens pour l'aide au traitement des pneumonies

en soins intensifs [Lucas *et al.* 00]. De leur part, Burnside et ses collègues [Burnside *et al.* 06] ont proposé l'utilisation des RB pour la prédiction du risque de cancer du sein.

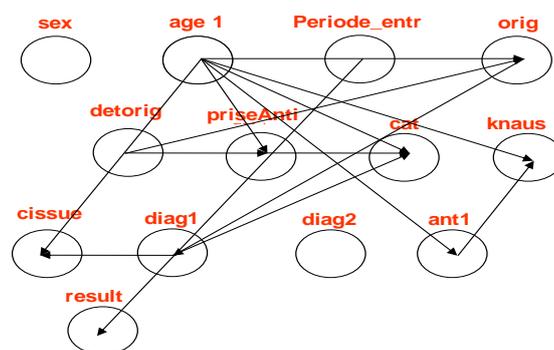
**L'application des RBD pour la lutte contre les IN** : la structure d'un RBD est obtenue par un algorithme itératif et récursif d'apprentissage dynamique. Les variables fixes et temporelles utilisées sont listées dans le tableau 4.5.

**Tableau 4.5** : variables pour l'application du RBD

Variables fixes		Variables temporelles	
Acronyme	Description	Acronyme	description
Sex	Le sexe du patient	dsj	La durée du séjour
age1	L'âge du patient	act <sub>i</sub>	Acte effectué le jour i
Periode_entr	Indique la saison d'entrée en réanimation	cissue <sub>i</sub>	Issue pour le jour i
Orig	Le service d'origine	examinf <sub>i</sub>	Examen infectieux effectué le jour i
Detorig	Détails sur le service d'origine	sens <sub>i</sub>	Sensibilité du germe (de l'infection acquise le jour i) à l'antibiotique prescrit
priseAnti	Prise d'antibiotique	result <sub>i</sub>	Probabilité de prédiction dynamique d'IN pour le jour i
Knaus	La catégorisation Apache		
Cissue	patient décédé ou ayant survécu		
Diag	Diagnostic		
Ant	Antécédent		
Result	Probabilité de prédiction statique d'IN		

La théorie des RB permet de représenter des rapports probabilistes entre les variables observées ce qui est bien adapté à l'incertitude inhérente aux questions médicales.

**Construction des modèles de connaissances sur des données fixes (RB statique)** : les dépendances causales entre les variables fixes sont représentées par la figure 4.3. Cependant les observations faites sur le RB statique ne sont pas suffisantes pour estimer la probabilité quotidienne d'occurrence d'IN.



**Figure 4.3** : dépendances causales dans un RB statique (modèle statique extrait)

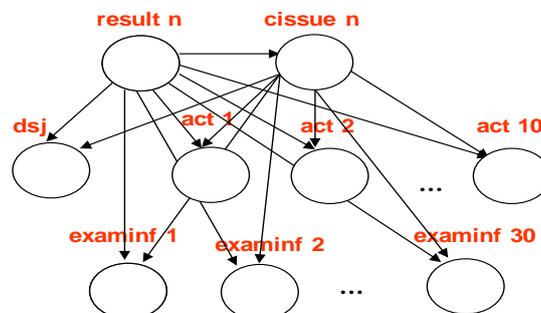
Le modèle extrait pourrait détecter des relations entre les variables logiques comme la relation entre l'âge et l'antécédent, entre l'âge et l'issue. Toutefois le graphe causal obtenu, contient des liens "illogiques" entre les nœuds (par exemple, l'âge agit sur la prise d'antibiotique). On peut également noter l'absence de liens présentant des relations intéressantes (par exemple, la relation entre le résultat et l'issue).

Les probabilités sont calculées en utilisant  $P(V_i|C)$  avec :

- $V_i$  : le nœud (sexe, age1, periode\_entr... diag1) ayant des valeurs discrètes, et
- $C$  : la classe à prévoir (issue et résultat) ayant les valeurs booléennes.

L'utilisation des probabilités et du graphe causal fournit des modèles de connaissances qui ne donnent une grande valeur à ajouter. Afin de représenter l'influence des événements passés sur l'état actuel du patient, il est nécessaire de prolonger ce modèle avec un RB dynamique.

**Construction des modèles de connaissances sur des données temporelles (RB dynamique)** : la figure 4.4 montre un modèle dynamique basé sur des variables temporelles. Le graphe causal représente l'interdépendance entre les variables temporelles pour une tranche de temps  $T$ . Nous avons employé pour cette structure dynamique les valeurs de chaque série de temps ( $act_1 \dots act_{10}$ ,  $exinfn_1 \dots exinfn_{30}$ ) connectées directement à deux nœuds prédictifs qui sont le résultat et l'issue.



**Figure 4.4** : dépendances causales dans un RB dynamique (modèle dynamique extrait)

Le principe de notre RBD peut être défini par :

- À  $t=0$ , on utilise le modèle statique extrait (cf. figure 4.3).
- Pour  $1 \leq t \leq T$  (durée d'hospitalisation) ; il s'agit de "dérouler" le modèle dynamique pour chaque jour de la durée d'hospitalisation.

On obtient un RBD final dont le graphe causal est visible en figure 4.5.

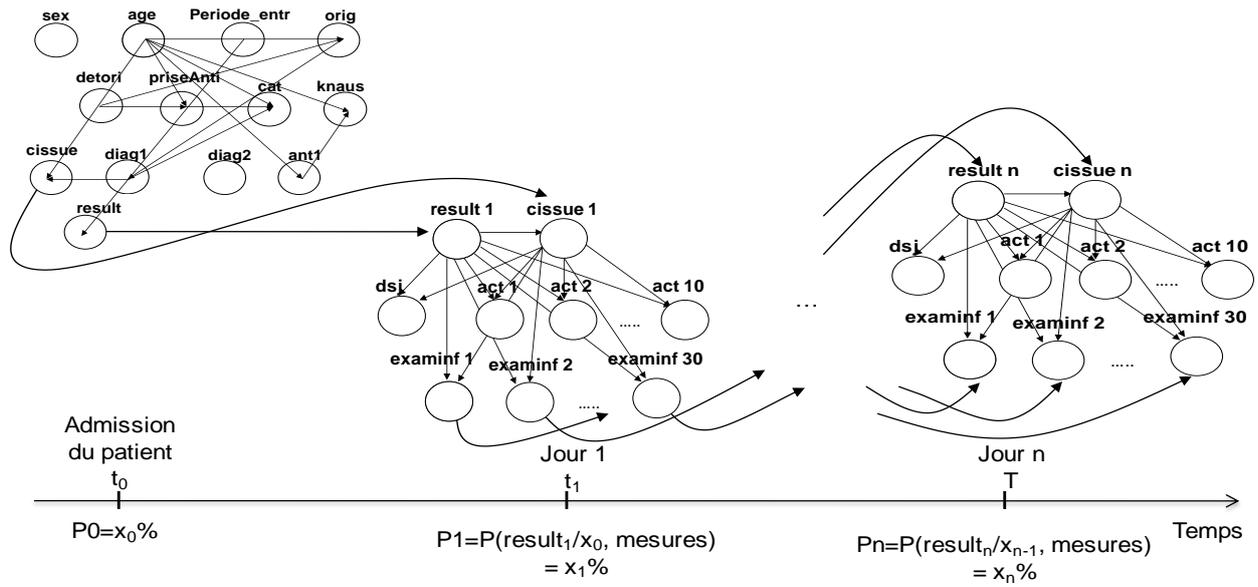


Figure 4.5 : graphe causal du RBD

Le résultat de distribution des probabilités est donné par :

$$P(\text{result}_{1,T}) = \prod_{t=1}^T \prod_{i=1}^N P(\text{result}_t^i | Pa(\text{result}_t^i))$$

Avec :

- T : la durée de temps d'hospitalisation,
- N : le nombre de variables pour chaque modèle extrait.

L'application des RBD pour la lutte contre les IN donne des bons résultats de prédiction qui seront présentés dans le chapitre suivant.

## 4.4.2 Module de gestion de connaissances

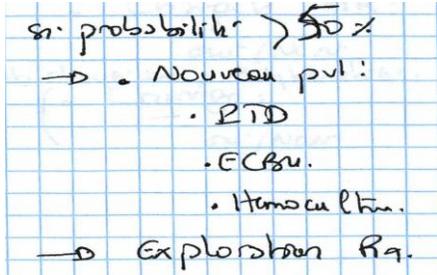
Après avoir finir avec le module de fouille de données, nous allons entamer la conception et le développement du module de gestion des connaissances en suivant les phases et les activités de l'ExUP.

### 4.4.2.1 Phase d'inception

La première phase de l'ExUP s'est déroulée en une seule itération. Elle nous a permis d'établir une vision initiale des fonctionnalités et des IHM du module de gestion de connaissances.

**Tableau 4.6** : déroulement de la phase d'inception pour le module de gestion des connaissances

Activités	Itération 1
<b>Capture des besoins</b>	
Analyse des situations décisionnelles	<p>En fonction de la valeur de probabilité découverte par une analyse avec les RBD ; une liste des solutions possibles doit être générée.</p> <p>Si (probabilité &gt; 50% et probabilité &lt; 70%) alors</p> <ul style="list-style-type: none"> <li>Solution 1 : Nouveau Prélèvement PTD</li> <li>Solution 2 : Nouveau Prélèvement HC</li> <li>Solution 3 : Nouveau Prélèvement ECBU</li> <li>Solution 4 : Exploration Rx</li> </ul> <p>Sinon</p> <p>Si (Probabilité &gt; 70%) alors</p> <ul style="list-style-type: none"> <li>Solution 1 : Nouveau Prélèvement PTD</li> <li>Solution 2 : Nouveau Prélèvement HC</li> <li>Solution 3 : Nouveau Prélèvement ECBU</li> <li>Solution 4 : Changement antibiotique</li> <li>Solution 5 : Si patient chirurgical Alors chirurgie Fin Si</li> </ul>

	<p>Sinon : la situation est non urgente</p> <p>Fin Si</p> <p>Fin Si</p>
Élaboration des premières maquettes d'IHM	<p>Une maquette papier a été réalisée par l'utilisateur lors d'une réunion de capture des besoins (cf. Figure 4.5). Cette maquette donne un aperçu des solutions possibles.</p>  <p><b>Figure 4.6</b> : une maquette utilisateur d'IHM du cas d'utilisation "Aide à la décision"</p>
Modélisation de l'utilisateur	<p>Le médecin (futur décideur) préfère avoir des interfaces HM sans détails. Il veut aussi cliquer sur la solution, proposée par le système, pour qu'elle soit validée. Ces préférences sont prises en compte dans cette étape.</p>
Expression des besoins décisionnels	<p>Les deux étapes précédentes permettent, dès la première itération, de fournir un cadre structurant en ce qui concerne les futures activités aussi bien que les solutions techniques.</p> <p>Les fonctionnalités du module se représentent sous forme de cas d'utilisation :</p> <ol style="list-style-type: none"> <li>1. Intégration de la connaissance : le pattern découvert par le module de fouille des données est une probabilité. Elle est évaluée puis interprétée par le module d'évaluation. La connaissance extraite se représente sous forme d'un <i>pourcentage</i>.</li> <li>2. Génération des solutions possibles : en se basant sur le pourcentage d'acquisition d'IN, il s'agit de générer les différentes solutions possibles pour minimiser le risque d'occurrence d'une IN.</li> <li>3. Prise de décision : le décideur visualisant les différentes solutions générées, choisie la solution qui lui semble la meilleure.</li> </ol>
<b>Evaluation des besoins</b>	<p>La description des besoins du module de gestion des connaissances a été montrée aux utilisateurs (décideurs) pour vérification.</p> <p>Résultat : validation des besoins.</p>
Répartition et définition des tâches	<p>Le modèle des cas d'utilisation permet de cerner la liste des tâches à effectuer pour le fonctionnement du futur module de gestion des connaissances à réaliser. Les tâches sont : l'affichage de la connaissance extraite, la génération des solutions possibles et la prise de décision.</p>
Définition des tâches manuelles	<p><i>Sans objet dans cette itération</i></p>
Définition des tâches automatiques	<p>Les tâches automatiques consistent à afficher la connaissance et à générer les différentes solutions possibles.</p>

Définition des tâches interactives	La tâche de prise de décision est la seule tâche interactive de ce module qui nécessite l'intervention du décideur pour valider la solution qui lui semble la meilleure.
<b>Evaluation des tâches</b>	Une évaluation des tâches citées ci-dessus a été réalisée dans le cadre de cette itération. Résultats : satisfaction de l'utilisateur.
<b>Analyse et conception</b>	
<i>Sans objet dans cette itération</i>	
<b>Implémentation</b>	
<i>Sans objet dans cette itération</i>	
<b>Test</b>	
<i>Sans objet dans cette itération</i>	

Au cours de cette phase, nous avons pu comprendre le contexte du module, définir ses besoins, identifier les principaux cas d'utilisation, ce qui nous a permis de délimiter la portée du module envisagé et de cerner le champ d'investigation du projet de développement.

#### 4.4.2.2 Phase d'élaboration

La phase d'élaboration s'est déroulée en deux itérations. Elle a permis d'analyser la plupart des cas d'utilisation ainsi que de concevoir l'architecture du module de gestion de connaissances. Cette phase a abouti à une spécification détaillée du module concerné.

**Tableau 4.7 :** déroulement de la phase d'élaboration pour le module de gestion des connaissances

Activités	Itération 1	Itération 2
<b>Capture des besoins</b>		
Analyse des situations décisionnelles	Lors d'une autre réunion avec les médecins décideurs, ceux-ci ont proposé d'associer au pourcentage d'occurrence d'une IN un texte et une image d'alerte si le risque d'avoir une IN est grand.	<i>Sans objet dans cette itération</i>
Élaboration des premières maquettes d'IHM	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>

Modélisation de l'utilisateur	Les préférences du décideur doivent être modélisées pour fournir des IHM plus adaptées à leurs besoins sous l'angle de l'ergonomie d'utilisation.	<i>Sans objet dans cette itération</i>
Expression des besoins décisionnels	Le cas d'utilisation d'intégration de la connaissance doit permettre de présenter le pourcentage d'occurrence d'une IN.	<i>Sans objet dans cette itération</i>
<b>Evaluation des besoins</b>	Cette version du modèle de cas d'utilisation a été évaluée par les utilisateurs. Elle converge bien vers les besoins exprimés.	<i>Sans objet dans cette itération</i>
Répartition et définition des tâches	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
Définition des tâches manuelles	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
Définition des tâches automatiques	La tâche d'affichage de la connaissance consiste à afficher le pourcentage avec un texte et une image d'alerte.	<i>Sans objet dans cette itération</i>
Définition des tâches interactives	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
<b>Evaluation des tâches</b>	Les tâches sont claires et répondent aux besoins des utilisateurs.	<i>Sans objet dans cette itération</i>
<b>Analyse et conception</b>		
Spécification des tâches automatiques	Analyse de la tâche d'affichage du pourcentage de prédiction quotidienne de l'acquisition d'une IN.	Modification du diagramme de communication de l'affichage de pourcentage de prédiction quotidienne de l'acquisition d'une IN pour ajouter les autres options d'affichage (texte et image d'alerte). Analyse de la tâche de génération des solutions possibles en fonction de la connaissance découverte par la technique de fouille de données.
Spécification des tâches interactives	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
Spécification des interfaces HM	Analyse de l'IHM qui va afficher le pourcentage de prédiction quotidienne de l'acquisition d'une IN.	Analyse de l'IHM qui va afficher les solutions possibles au problème d'occurrence d'IN.
Conception des modules automatiques d'ECD	Conception de la tâche d'affichage de pourcentage de prédiction quotidienne de l'acquisition d'une IN.	Révision de la conception de la prédiction quotidienne de l'acquisition d'une IN. Conception de la tâche de génération des solutions possibles au problème de décision ainsi que des tâches de validation et d'annulation du choix de décideur.
Conception des interfaces HM	Conception de l'IHM d'affichage du pourcentage de prédiction	Conception de l'IHM d'affichage des solutions possibles au problème

et des modules interactifs d'ECD	quotidienne de l'acquisition d'une IN.	d'occurrence d'IN.
<b>Evaluation des modèles</b>	Les modèles d'analyse et de conception de cette tâche nécessitent une révision.	Les modèles d'analyse et de conception des deux cas d'utilisation analysés ont été validés dans le cadre de cette activité d'évaluation.
Elaboration des maquettes avancées d'IHM	Une maquette électronique de l'IHM d'affichage de la connaissance extraite (pourcentage + texte + image d'alerte) est préparée dans cette itération.	<i>Sans objet dans cette itération</i>
Spécifier l'architecture	L'architecture initiale de ce module est constituée du cas d'utilisation d'intégration de la connaissance.	L'architecture du module est enrichie par l'ajout du cas d'utilisation de génération des solutions possibles.
<b>Evaluation de l'architecture</b>	L'architecture du module sera acceptée si on prend en compte les révisions demandées.	Les utilisateurs ont accepté l'architecture courante du module.
<b>Implémentation</b>		
Codage des fonctionnalités	<i>Sans objet dans cette itération</i>	La fonctionnalité d'affichage de la connaissance de prédiction d'occurrence d'IN est codée en tenant compte des modèles d'analyse et de conception.
Codage des Interfaces Homme-Machine	<i>Sans objet dans cette itération</i>	Préparation de l'IHM d'affichage de la connaissance de prédiction d'occurrence d'IN en tenant compte de la maquette électronique d'IHM.
Construction du prototype	<i>Sans objet dans cette itération</i>	Assembler le fichier cs et celui resx d'affichage de la la prédiction quotidienne d'occurrence d'IN. Le premier prototype de ce module permet d'intégrer la connaissance de prédiction quotidienne d'occurrence d'IN et de l'afficher sous forme de pourcentage, texte et image d'alerte si le risque est grand.
<b>Test</b>		
Test des fonctionnalités	<i>Sans objet dans cette itération</i>	L'exécution de ce premier prototype de gestion des connaissances satisfait les futurs utilisateurs.
Test des IHM	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
Correction des erreurs de fonctionnalités	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
Correction des problèmes d'interaction HM	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
Validation du prototype	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>

Durant la phase d'élaboration, nous avons pu identifier, analyser et concevoir la majorité des cas d'utilisation du module de gestion de connaissances. Nous avons aussi commencé le développement d'un premier prototype du module en question.

#### 4.4.2.3 Phase de construction

C'est le moment où l'on construit le produit. L'architecture du module de se métamorphose en produit complet ou le produit contient tous les cas d'utilisation, en accord avec les utilisateurs.

**Tableau 4.8** : déroulement de la phase de construction pour le module de gestion des connaissances

Activités	Itération 1	Itération 2
<b>Capture des besoins</b>		
<i>Sans objet dans cette phase</i>		
<b>Analyse et conception</b>		
Spécification des tâches automatiques	<i>Sans objet dans cette itération</i>	<i>Sans objet dans cette itération</i>
Spécification des tâches interactives	Analyse des tâches de validation et d'annulation d'un choix effectué par un décideur parmi les solutions proposées.	Révision de l'analyse du cas d'utilisation de prise la décision pour donner la main à l'utilisateur pour ajouter des commentaires relatifs à la prise de la décision.
Spécification des interfaces HM	Analyse de l'IHM du cas d'utilisation de prise de décision	Révision de l'analyse de l'IHM du cas d'utilisation de prise de décision pour l'ajout du champ des commentaires.
Conception des modules automatiques d'ECD	Conception des tâches de validation et d'annulation d'un choix effectué par un décideur.	Révision de la conception de l'IHM du cas d'utilisation de prise de décision pour l'ajout du champ des commentaires.
Conception des interfaces HM et des modules interactifs d'ECD	Conception de l'IHM du cas d'utilisation de prise de décision pour l'ajout du champ des commentaires.	Révision de la conception de l'IHM du cas d'utilisation de prise de décision pour l'ajout du champ des commentaires.
<b>Evaluation des modèles</b>	Les utilisateurs veulent ajouter des commentaires lors de la prise d'une décision.	La révision des modèles d'analyse et de conception du cas d'utilisation de prise de décision a été acceptée.
Elaboration des maquettes avancées d'IHM	Réalisation d'une maquette de l'IHM du cas d'utilisation de prise de décision pour l'ajout d'un champ des commentaires.	<i>Sans objet dans cette itération</i>

Spécifier l'architecture	Il s'agit de spécifier l'architecture finale du module de gestion des connaissances.	<i>Sans objet dans cette itération</i>
<b>Evaluation de l'architecture</b>	L'architecture finale est acceptée par les futurs utilisateurs.	<i>Sans objet dans cette itération</i>
<b>Implémentation</b>		
Codage des fonctionnalités	Coder la fonctionnalité de génération des solutions possibles en fonction de la connaissance découverte par le module de fouille de données.	Coder les fonctionnalités de validation et d'annulation du cas d'utilisation de prise de décision.
Codage des Interfaces Homme-Machine	L'IHM de génération des solutions possibles a été implémentée.	Ajout du champ de commentaires et des boutons de validation et d'annulation.
Construction du prototype	Le code de la fonctionnalité de génération des solutions et celui de son IHM ont été assemblés pour la tester. Le nouveau prototype est maintenant incrémenté par la fonctionnalité de génération des solutions possibles au problème de décision.	Le prototype construit dans le cadre de cette itération est le prototype du module de gestion des connaissances entier.
<b>Test</b>		
Test des fonctionnalités	L'exécution de la fonctionnalité implémentée de la génération des solutions possibles pour l'aide à la décision s'est bien déroulée.	Lors de l'exécution de ce prototype, les développeurs ont trouvé quelques erreurs dans les transactions avec la base de données.
Test des IHM	Les zones de saisie semblent inefficaces pour les utilisateurs.	<i>Sans objet dans cette itération</i>
Correction des erreurs de fonctionnalités	<i>Sans objet dans cette itération</i>	Correction du problème de transaction avec la base de données.
Correction des problèmes d'interaction HM	Là où c'est possible, les zones de saisies ont été remplacées par des boutons radio ou des cases à cocher.	<i>Sans objet dans cette itération</i>
Validation du prototype	<i>Sans objet dans cette itération</i>	Le prototype de gestion des connaissances est validé par le concepteur et le développeur.

Durant la phase de construction, nous avons terminé l'implémentation des cas d'utilisation du module de gestion de connaissances.

#### 4.4.2.4 Phase de transition

Le module est développé, il est en version bêta. L'utilisateur représentatif teste les interactions Homme-Machine et les fonctionnalités et détecte les anomalies et défauts. La phase de transition, suppose des activités comme la formation des utilisateurs médecins, la notion d'un service d'assistance et la correction des anomalies constaté.

**Tableau 4.9** : déroulement de la phase de transition pour le module de gestion des connaissances.

Activités	Itération 1
<b>Capture des besoins</b>	
<i>Sans objet dans cette itération</i>	
<b>Analyse et conception</b>	
<i>Sans objet dans cette itération</i>	
<b>Implémentation</b>	
<i>Sans objet dans cette itération</i>	
<b>Test</b>	
Test des fonctionnalités	Il s'agit, à ce niveau, de vérifier la fiabilité du prototype, de tester sa tendance à traiter les exceptions en vue de surmonter les anomalies qui peuvent être rencontrées. Les utilisateurs ont voulu aussi exprimer leur volonté de tester les résultats de la prévention des IN dans le service de réanimation. Un test de plusieurs cas de patients à la file a généré des erreurs d'implémentation.
Test des IHM	Aucun problème d'IHM n'est détecté dans cette itération.
Correction des erreurs de fonctionnalités	Correction des erreurs d'actualisation lors de passage d'un cas de prédiction à un autre.
Correction des problèmes d'interaction HM	<i>Sans objet dans cette itération</i>
Validation du prototype	Validation des résultats de prévention des IN dans le service de réanimation et donc du module de gestion des connaissances d'où l'acceptation finale du module de gestion des connaissances.

## 4.5. Conclusion

Ce chapitre a proposé l'application de notre approche centrée utilisateur de conception (ExUP) à un cas concret dans le domaine médical. Ce système a été développé en collaboration avec l'équipe médicale du Centre Hospitalier Universitaire Habib Bourguiba à Sfax visant la lutte contre les infections nosocomiales. L'étude de ses caractéristiques a montré son aspect dynamique. Pour cette raison, il est considéré comme un système interactif d'aide à la décision dynamique médicale basé sur l'ECD (SIADDM/ECD).

Le système est décomposé en quatre modules (saisie et stockage des données, fouille des données, évaluation des modèles et gestion des connaissances). Les deux premiers modules et une partie du troisième ont été développés lors des travaux antérieurs (auxquels nous avons participé). Nous avons utilisé l'ExUP pour développer en premier lieu une nouvelle application du module de fouille de données avec une nouvelle technique dynamique qu'est la technique des réseaux bayésiens dynamiques, et en deuxième lieu le module de gestion des connaissances.

Nous avons vu que les modules ont été développés et acceptés par les utilisateurs finaux ; il est possible d'en déduire une première validation de la démarche proposée. Une évaluation globale de cette démarche et un ensemble de perspectives font l'objet du chapitre suivant.



## Chapitre 5 : Evaluation de l'ExUP et perspectives de recherche

### 5.1. Introduction

Ce chapitre propose une évaluation globale de la démarche de développement ExUP (**Extended Unified Proces**) proposée et appliquée dans un contexte médical. Ainsi, à partir du développement d'un SIADDM/ECD pour la lutte contre les IN montré dans le chapitre précédent, un bilan d'évaluation sera présenté dans la première section du chapitre. Cette évaluation concerne d'une part les aspects méthodologiques de l'ExUP sous l'angle des interactions homme-machine et d'autre part l'acceptation des utilisateurs du SIADDM/ECD mis en place au sein du service de réanimation de l'hôpital Habib Bourguiba de Sfax, Tunisie.

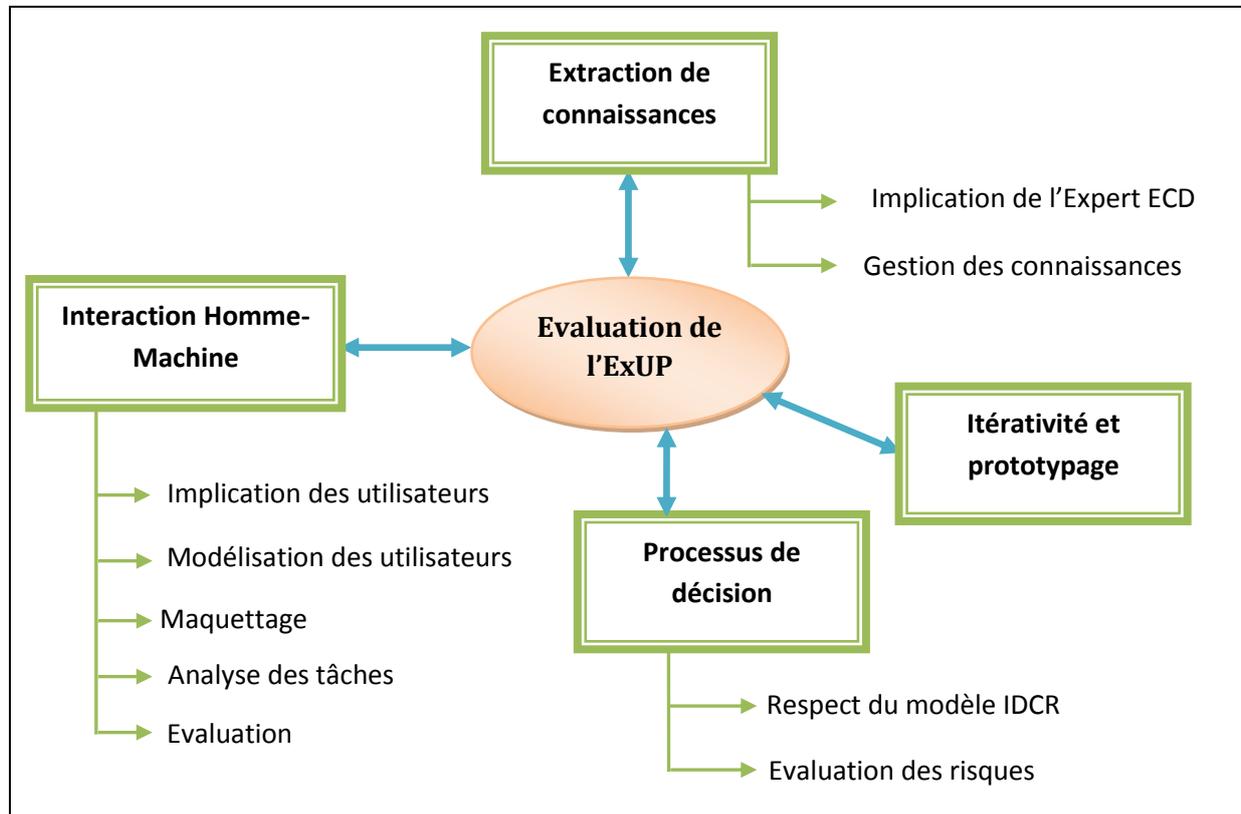
La seconde section sera axée sur nos perspectives de recherche concernant l'extension de la démarche proposée, l'intégration des techniques de visualisation dans les différents modules du SIADDM/ECD, ainsi que l'évaluation de la qualité des données. Pour l'intérêt potentiel que présentent les prédictions instantanées d'occurrences des IN aux médecins, une autre perspective sur le plan expérimental sera également envisagée en vue du développement d'un Système Interactif d'Aide à la Décision Instantanée Médicale basé sur l'ECD.

### 5.2. Bilan d'évaluation globale de la démarche

L'objectif de cette section est de mettre en avant quelques résultats provenant de l'évaluation globale de notre contribution par rapport aux objectifs fixés *a priori*. En fait, l'intitulé de ce mémoire résume de lui-même les objectifs que nous nous étions fixés : définir une démarche centrée utilisateur pour la conception de SIAD basés sur un processus d'Extraction de Connaissances à partir de Données. Pour répondre à ces objectifs nous avons proposé l'ExUP (cf. chapitre 3) qui a été utilisée pour le développement d'un SIADDM/ECD dans le domaine médical (cf. chapitre 4).

Le principal objectif de l'ExUP est de concevoir et développer des SIAD/ECD centrés utilisateur. Pour cette raison, nous avons enrichi les activités du processus unifié sous l'angle des interactions homme-machine tout en tenant compte des spécificités des tâches d'ECD. L'évaluation de la démarche de l'ExUP va se baser principalement sur un ensemble de

critères visibles en figure 5.1. La sélection de ces critères a été faite selon quatre dimensions : *l'interaction homme-machine, le processus de décision, l'aspect itératif de développement et l'extraction de connaissances*. Ce sont les dimensions sur lesquelles a été basée la proposition de la démarche ExUP. Les conclusions ainsi que les résultats de cette évaluation seront présentés dans le tableau 5.1.



**Figure 5.1** : critères d'évaluation de la démarche

L'évaluation de la démarche ExUP a commencé dès les premières itérations de développement des modules du SIADDM/ECD développés dans le cadre de cette thèse. En fait, des évaluations centrales et terminales dans chaque activité nous ont permis de contacter les médecins du service de réanimation (utilisateurs) et l'expert ECD, de prendre leur avis et de tirer des conclusions sur la démarche proposée.

Pour chaque critère, nous donnerons, dans le tableau 5.1, nos conclusions suite à l'évaluation établie durant l'utilisation de l'ExUP pour le développement des modules de fouille de données et de gestion des connaissances du SIADDM/ECD.

**Tableau 5.1** : bilan d'évaluation de la démarche

Critère	Conclusions	Résultat <sup>16</sup>
<b>Interaction Homme-Machine</b>		
Implication des utilisateurs	Les utilisateurs (médecins) ont participé d'une manière continue à la majorité des activités de l'ExUP. Lors du développement du SIADDM/ECD, ils ont donné leurs remarques et souhaits quel que soit le stade de développement.	2
Modélisation des utilisateurs	Les utilisateurs du SIADDM/ECD sont des médecins. Leur comportement en tant que décideurs a été précisément décrit. Leurs stratégies et/ou tactiques de décision les mieux appropriées ont été bien définies dès le début du développement du module « Gestion des connaissances ».	2
Maquettage	<p>Les premières maquettes réalisées pour la conception des deux derniers modules du SIADDM/ECD, ont énormément simplifié la compréhension des besoins des utilisateurs (par exemple la maquette présentant les solutions possibles et qui est visible dans la figure 4.5, cf. chapitre précédent).</p> <p>De leur part, les maquettes avancées nous ont permis de commencer à entrer dans les détails des IHM. Ces maquettes ont été présentées aux médecins pour une évaluation précoce des IHM. Elles ont aussi aidé le concepteur à élaborer certains diagrammes de séquence.</p>	2
Analyse des tâches	<p>Les tâches automatiques et interactives de modules de fouille de données et de gestion des connaissances du SIADDM/ECD ont été définies et réparties et l'ordre dans lequel elles s'enchaînent a été établi. Les évaluations des tâches et des scénarios d'usages fournies ont montré que l'ExUP permet d'inciter à décrire et analyser les tâches.</p> <p>Nous avons utilisé le diagramme de communication UML pour l'analyse des tâches. Nous avons constaté que ce type de diagramme est adapté au PU mais pas suffisamment à l'ExUP qui est orienté utilisateur. Pour cette raison, nous proposons dans des travaux futurs d'intégrer un formalisme issu de domaine de l'IHM pour décrire les arbres de tâches. Un exemple est le formalisme MAD (Méthode Analytique de Description) [Scapin 89].</p>	1
Evaluation	<p>Les évaluations sont préconisées dans toutes les activités de l'ExUP. Elles avaient pour objectif de diminuer les risques de mauvais choix de conception impliquant les utilisateurs et les experts ECD. Elles ont permis également de prévoir et d'effectuer des feedbacks si nécessaire. A titre d'exemple, nous mentionnons le cas d'utilisation de prédiction d'occurrence d'IN qui a nécessité de nombreuses réunions d'évaluation avec les médecins et l'expert ECD. En effet, la complexité de l'IHM de ce cas d'utilisation (tâches d'ECD) a été limitée à un sous-ensemble de tâches simples et fréquentes ; l'aide en ligne a été développée et adaptée ; les retours de la fonctionnalité de prédiction (messages confirmation ou erreur) sont devenus fréquents et compréhensibles.</p> <p>Des évaluations sont aussi prévues à la fin des activités (sous forme de tests). Elles ont permis de valider chaque prototype du module SIADDM/ECD concerné en fin d'itération.</p>	2

<sup>16</sup> Le résultat d'évaluation montre le degré global de validation du critère (0= non validé ; 1= partiellement validé ; 2= validé).

<b>Processus de décision</b>		
Respect du modèle IDCR de Simon <sup>17</sup>	Le SIADDM/ECD développé permet actuellement d'assister le médecin/décideur, l'accompagner dans sa compréhension de la situation décisionnelle du patient ; et lui proposer une liste des solutions possibles en fonction de la probabilité d'occurrence d'IN. Ce processus est itératif puisque le médecin peut à tout moment changer les choix qu'il adopte.	2
Evaluation des risques	Les médecins ont constaté, lors de l'utilisation du SIADDM/ECD en service de réanimation, que ce système ne leur permet pas d'évaluer les risques qu'ils prennent lorsqu'ils adoptent telle ou telle solution de décision. Ce critère sera pris en compte dans nos perspectives de recherche.	0
<b>Itérativité et prototypage</b>		
Itérativité et prototypage	<p>À l'issue de chaque itération, nous avons obtenu un prototype testé par les développeurs, les médecins et l'expert ECD. Ce caractère itératif de l'ExUP a permis d'affiner progressivement les spécifications, de simuler les IHM, d'évaluer les solutions retenues, puis d'intégrer les modifications jusqu'à l'obtention du prototype final considéré comme globalement satisfaisant (sauf remarque présentée par la rubrique "d'évaluation des risques") qui est, actuellement, en cours d'utilisation à l'Hôpital Habib Bourguiba de Sfax, Tunisie.</p> <p>Les développements des modules de fouille de données et de gestion de connaissances se sont déroulés en 14 itérations (7 itérations par module). Chaque itération a duré 2 semaines environ. La phase d'élaboration est celle qui nécessite le plus d'itérations. En fait, il a été difficile de définir entièrement les spécifications au début de cette phase, surtout celles concernant la fouille des données et l'intégration de connaissances.</p>	2
<b>Intégration de connaissances</b>		
Implication de l'Expert ECD	Lors du développement des deux derniers modules du SIADDM/ECD, l'expert ECD a bien contribué à l'analyse, la conception et l'évaluation des tâches en relation avec les étapes de processus ECD. En se basant sur son expérience en ECD, il a aidé le développeur dans l'implémentation de l'algorithme basé sur les réseaux bayésiens dynamiques.	2
Gestion des connaissances	<p>Les connaissances extraites par l'outil de fouille de données du SIADDM/ECD sont intégrées pour la génération des solutions possibles au problème de décision. Néanmoins, nous n'avons pas pu traiter tous les aspects relatifs à la gestion des connaissances, tels que les types de connaissances (tacites et explicites), les ontologies, le retour d'expérience, etc.</p> <p>Une autre difficulté rencontrée par les utilisateurs/décideurs est la présentation des résultats de l'application des algorithmes de fouille de données. Ces résultats (des probabilités) sont présentés dans des tableaux. Les utilisateurs proposent d'utiliser des axes de temps, des courbes et des figures représentatives pour faciliter la compréhension de ces résultats et donc l'interprétation des connaissances.</p> <p>Nous visons de prendre en compte ces problèmes dans nos futurs travaux de recherche.</p>	1

<sup>17</sup> Cf. chapitre 1, section 1.1.2

L'acceptation finale du SIADDM/ECD par les médecins du service de réanimation a permis de valider *a priori* l'approche de développement utilisée, ExUP. Pourtant, à partir des évaluations précédemment menées, des limitations ont été constatées. Ces limitations seront considérées comme des objectifs pour des futurs travaux de recherche visant l'amélioration de l'approche ExUP. Nous terminons cette section par l'évaluation de l'ExUP dans son ensemble en fonction des critères d'évaluation présentés ci-dessus (figure 5.2).

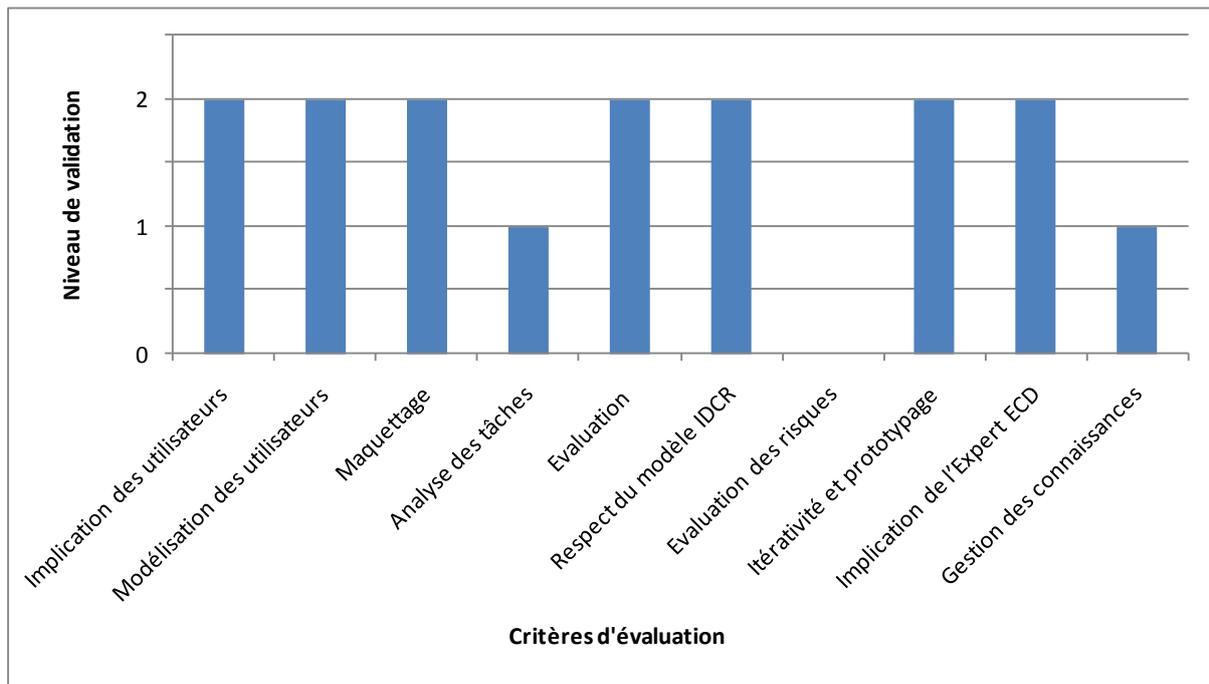


Figure 5.2 : évaluation globale de l'ExUP

L'évaluation globale de la démarche de l'ExUP (cf. figure 5.1) montre que : sept critères sur 10 sont validés (ce qui correspond à 70% de l'ensemble des critères), deux critères sont partiellement validés (20%) et un seul critère qui n'est pas validé (10%). Cette évaluation nous donne des résultats globalement satisfaisants. Nous passons maintenant à l'évaluation empirique du SIADDM/ECD développé.

### 5.3. Bilan d'évaluation empirique du SIADDM/ECD

Dans un but d'amélioration du SIADDM/ECD, qui doit être fiable et répondre aux besoins des médecins, il est important de l'évaluer en termes de qualité de l'interface réalisée et d'utilité des résultats obtenues par le système. On s'intéresse donc ici aux deux dimensions d'évaluation classiques dans le domaine de l'Interaction Homme-Machine que sont « l'utilisabilité » et « l'utilité » [Nielsen 93] et qui feront l'objet des deux sections suivantes.

### 5.3.1 Evaluation de l'utilisabilité du SIADDM/ECD

Cette évaluation consiste à mesurer l'utilisabilité du SIADDM/ECD, avec pour visée de valider la satisfaction des utilisateurs dans la réalisation des tâches. Parmi les différentes méthodes d'évaluation envisageables [Baccino *et al.* 05], nous avons utilisé, en premier lieu, les **tests utilisateurs**. Pour ce faire, nous avons placé l'utilisateur représentatif en situation d'utilisation réelle du SIADDM/ECD en observant les difficultés rencontrées, aussi bien d'ordre fonctionnel que sous l'angle des interactions homme-machine. Ces tests utilisateurs ont été présentés dans le chapitre 4 (cf. tableaux 4.3, 4.4, 4.6, 4.7 et 4.8) ; ils ont permis d'identifier les différents points à perfectionner, ce qui a été réalisé par la suite dans les itérations suivantes.

Dans ce cadre, nous avons utilisé une autre méthode d'évaluation qui est la **grille d'évaluation**. Cette dernière nous a permis d'effectuer une évaluation globale des interfaces homme-machine du SIADDM/ECD ; et ceci selon quatre critères principaux : Facilité d'apprentissage, Cohérence, Efficacité d'utilisation et Prévention et correction d'erreurs dans lesquels nous avons groupé les critères ergonomiques. Cette grille proposée par [Bahloul *et al.* 10] est présentée par le tableau 5.2 : les réponses des utilisateurs représentatifs sont marquées par des croix.

**Tableau 5.2** : grille d'évaluation

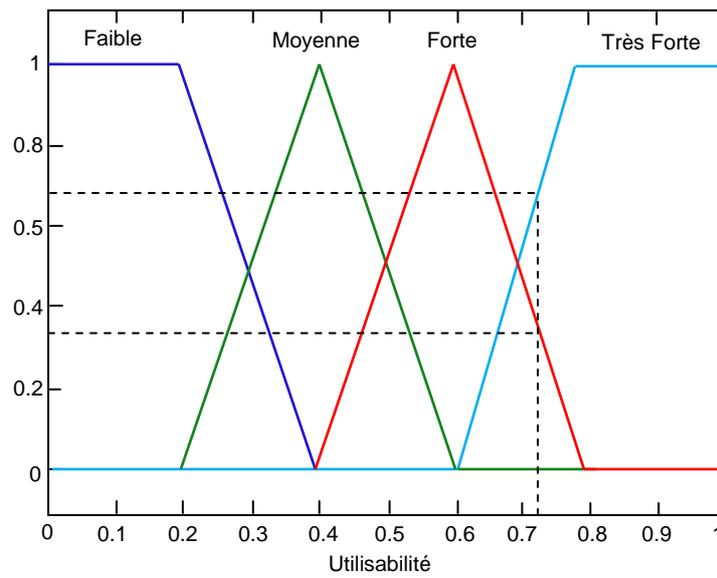
N°	Questions	toujours	souvent	quelquefois	Jamais
<b>A</b>	<b>Facilité d'apprentissage</b>				
1	Est-ce que l'entrée des données est guidée ?	X			
2	Pour chaque champ un label est-il fourni ?	X			
3	Les termes employés sont-ils familiers aux utilisateurs et relatifs à la tâche à réaliser ?		X <sup>18</sup>		
4	Est-ce que chaque libellé est facile à comprendre ?		X		
5	Est-ce que l'utilisation des images rend l'interface plus compréhensible ?	X <sup>19</sup>			
6	L'organisation des items lorsque plusieurs options sont présentées est-elle logique ?	X			
7	Est-il facile de trouver l'information désirée ?	X			
<b>C</b>	<b>Cohérence</b>				
1	Les couleurs sont-elles utilisées de façon cohérente ?	X			
2	Y a-t-il localisation similaire des titres des interfaces ?		X		
3	Le format des champs d'entrée de données est-il toujours le même ?	X			
4	Les libellés, les noms et les abréviations sont-ils cohérents ?	X			
<b>E</b>	<b>Efficacité d'utilisation</b>				
1	Existe-t-il des messages qui informent sur les tâches effectuées ?	X			
2	Est-ce que le curseur est facilement repérable ?		X		
3	Y a-t-il la possibilité de modifier les commandes lors de leurs saisies ?				X
4	Existe-t-il des accès directs vers l'information désirée ?		X		
5	Lors d'une saisie de données chaque action est-elle claire ?	X			
6	Le retour d'information est-il suffisant suite à une action ?	X			
7	Les données qui peuvent être calculées à partir de celles saisies le sont-elles ?	X			
8	Le temps d'exécution est-il court ?		X		
<b>P</b>	<b>Prévention et correction d'erreurs</b>				
1	La saisie des données d'une liste est-elle automatique ?	X			
2	Suite à une action erronée y a-t-il des messages d'erreurs ?	X			
3	Les messages d'erreurs sont-ils explicites ?	X			
4	En cas d'erreur de manipulation, est-il possible de revenir en arrière ?		X		

<sup>18</sup> Ce sont surtout les termes relatifs à la fouille de données.

<sup>19</sup> Les médecins ont exprimé leur volonté d'utiliser d'autres figures pour la présentation des résultats de la fouille des données, ce qui entre dans le cadre de nos perspectives de recherche (cf. plus loin).

Pour évaluer le degré d'utilisabilité, nous avons utilisé la méthode d'évaluation proposée par [Bahloul *et al.* 10] ayant pour entrée les réponses des utilisateurs (grille d'évaluation) et pour sortie quatre variables linguistiques d'utilisabilité : faible, moyenne, forte et très forte.

Cette méthode utilise une base de règles floues. En appliquant ces règles, nous avons trouvé que l'utilisabilité du système est égale à 0.73. En projetant cette valeur sur la courbe, nous nous apercevons que l'utilisabilité est à 34.6% forte et à 64% très forte (cf. figure 5.3).



**Figure 5.3** : évaluation de l'utilisabilité du SIADDM/ECD

L'évaluation de l'utilisabilité (tests d'utilisabilité et grille d'évaluation) a montré que les utilisateurs ont globalement apprécié l'utilisation du SIADDM/ECD. La principale limite réside dans le besoin de visualiser les résultats de la fouille des données autrement, c'est-à-dire avec des figures et des axes de temps.

### 5.3.2 Evaluation de l'utilité du SIADDM/ECD

En parallèle à l'évaluation de l'utilisabilité, il est également nécessaire d'évaluer l'utilité du SIADDM/ECD. Il s'agit de vérifier, en premier lieu, si le SIADDM/ECD peut être d'une grande utilité aux utilisateurs (médecins). En deuxième lieu, des résultats de prédiction seront exposés pour évaluer la performance du système en question.

#### 5.3.2.1 Evaluation des effets du SIADDM/ECD

Cette étude a pour but de déterminer si les utilisateurs (médecins) estiment que le système développé présente de l'intérêt pour la surveillance des infections nosocomiales au sein du

service de réanimation. Un questionnaire a été proposé aux personnels du service de réanimation après un mois d'utilisation du SIADDM/ECD. De ce questionnaire, nous avons tiré les conclusions suivantes.

- *Amélioration de la surveillance et de la prise en charge des IN* : elle s'est faite par les alertes permettant la détection de l'apparition de germes pathogènes dans l'établissement, et par des mécanismes d'identification de patients infectés utilisant des critères de définition pris des infections nosocomiales (Examens Infectieux).
- *Diminution du temps accordé à la surveillance des IN par les hygiénistes* : le temps ainsi gagné peut être consacré aux audits ou à l'enseignement en hygiène hospitalière. De la même façon, le temps consacré par les médecins à la recherche d'informations spécifiques concernant un patient donné est diminué.
- *Amélioration de la qualité des données exploitées* : c'est grâce aux opérations de prétraitement des données, que l'effet d'anomalies est minimisé telles que des erreurs de saisie, des champs nuls, des valeurs aberrantes. Toutefois, il n'y a pas une méthodologie bien déterminée pour l'évaluation de la qualité des résultats de ces opérations de prétraitement effectués dans les SIAD/ECD.
- *Amélioration de la prévention des IN* : les prédictions automatiques quotidiennes de l'occurrence des IN proposent des traitements possibles à appliquer pour lutter contre ces infections. Ce point sera plus détaillé dans la section suivante.

### **5.3.2.2 Expérimentation, résultats et performances**

Dans cette section, l'objectif est d'évaluer le SIADDM/ECD sous l'angle de la prédiction dynamique de l'état du patient. Rappelons que cette prédiction évolue tout au long de l'hospitalisation du patient. Les données du patient sont traitées au fur et à mesure. Elles enrichissent les modèles qui les utilisent pour fournir d'autres prédictions. A chaque série de temps, il est possible de prédire l'état du patient ; cette prédiction sert ensuite d'entrée pour la prédiction de l'état pour le jour suivant. De plus les observations du jour courant sont aussi des entrées pour la prédiction du résultat.

La figure 5.4 montre un exemple de prédiction dynamique effectuée pour un patient qui est resté 20 jours dans le service de réanimation et pour lequel le système a prédit la survenue d'une IN.

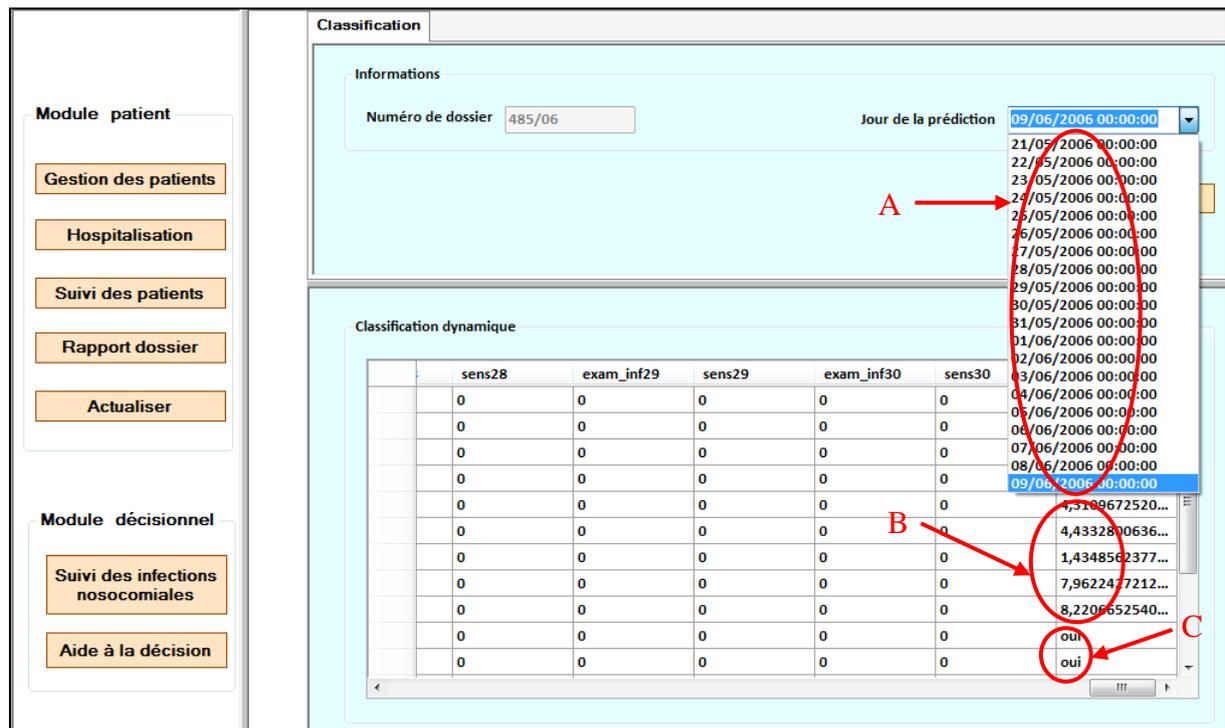


Figure 5.4 : les résultats de la prédiction dynamique pour un patient

Sur la figure 5.4, nous avons présenté trois bulles :

- (1) la première (A) indique les différentes séries de temps relatives à la durée d'hospitalisation du patient ;
- (2) la deuxième (B) montre les différentes valeurs de probabilité d'acquiescer une IN et ceci pour les premiers jours de son hospitalisation. La dernière valeur de cette série de probabilité est de 82% ce qui montre qu'il y a une forte probabilité que le patient attrape une IN.
- (3) après cette durée de temps le patient a effectivement contracté une IN, ce qui est illustré par la bulle C.

Les réseaux bayésiens dynamiques sont appliqués sur des données temporelles et fixes. Nous avons appliqué notre algorithme sur des données réelles provenant du service de réanimation de l'hôpital Habib Bourguiba de Sfax ; nous avons pu extraire des modèles de connaissance et les transformer automatiquement pour obtenir des résultats probabilistes, quantitatifs et qualitatifs de prédiction.

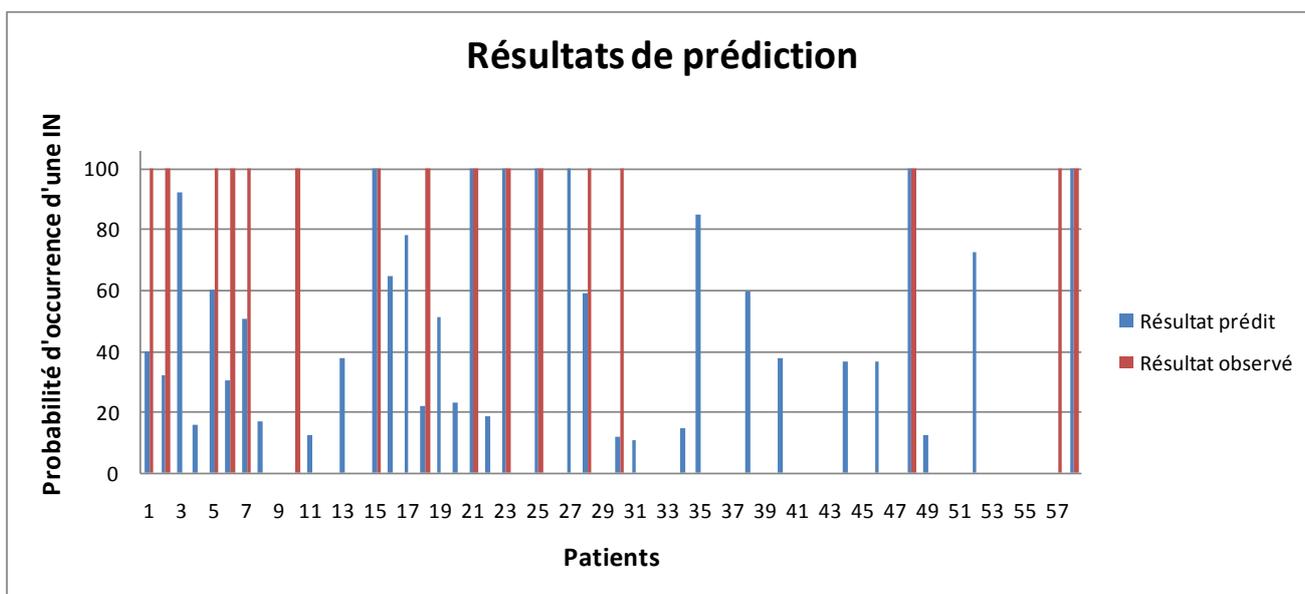
Pour l'évaluation de la performance de notre système, nous avons utilisé une *base de test* qui contient **58** cas (patients). Nous avons obtenu les résultats donnés par la matrice de confusion<sup>20</sup> suivante.

**Tableau 5.3** : la matrice de confusion des résultats fournis par notre réseau bayésien dynamique [Ltifi *et al.* 10b]

Résultats observés \ Résultats prédits	Oui	Non	Total
Oui	9	8	17
Non	7	34	41
Total	16	42	58

Nous avons calculé les taux d'évaluation à partir des résultats de prédiction obtenus par notre structure élaborée par le RBD, nous avons trouvé que le taux de classification était correct à **0.74**, le pouvoir de prédiction positif = **0.56**, le pouvoir de prédiction négatif = **0.81**.

Les résultats donnés par le tableau 5.3 sont représentés par l'histogramme visible en figure 5.5. Ces résultats sont fiables à 74%, ce qui est très encourageant.



**Figure 5.5** : résultats de la prédiction pour une base de 58 patients

Les données issues de l'observation ont confirmé à 74% les prédictions quotidiennes relativement à la population étudiée. Au cours de ces observations, les médecins ont exprimé

<sup>20</sup> Oui : exprime le fait d'avoir une IN ; Non : fait de ne pas avoir une IN ; total : le total des prédictions

le souhait de disposer de prédictions instantanées de l'occurrence d'IN surtout pour les patients présentant une situation dangereuse en service de réanimation.

Dans cette première partie de ce chapitre, nous avons décrit les différentes évaluations réalisées pour évaluer la démarche ExUP ainsi que l'utilisabilité du SIADDM/ECD puis son utilité pour les médecins. Au cours de ces différentes évaluations, nous avons dégagé différentes constatations concernant l'acceptabilité du SIADDM/ECD à partir des analyses des usages par les médecins, d'une grille d'évaluation et des questionnaires remplis par ceux-ci. Nous avons tiré un certain nombre de limitations de la démarche ExUP. Ces limitations concernent principalement l'analyse des tâches, l'évaluation des risques des solutions choisies par le décideur ainsi que la visualisation des informations par les médecins. Une autre limitation constatée lors de l'évaluation du SIADDM/ECD concerne l'évaluation de la qualité des données prétraitées et exploitées pour l'extraction des connaissances.

Après avoir recensé et analysé les différents problèmes rencontrés, il devient possible de faire des propositions pour tenter de résoudre ces problèmes. Celles-ci sont regroupées en un ensemble de perspectives de recherche faisant l'objet de la partie suivante.

## **5.4. Perspectives de recherche**

Après avoir évalué notre approche proposée ExUP et le SIADDM/ECD développé, il reste des points d'évaluation qui font partie de nos perspectives de recherche, aussi bien sur le plan méthodologique que sur le plan expérimental.

### **5.4.1 Perspectives sur le plan méthodologique**

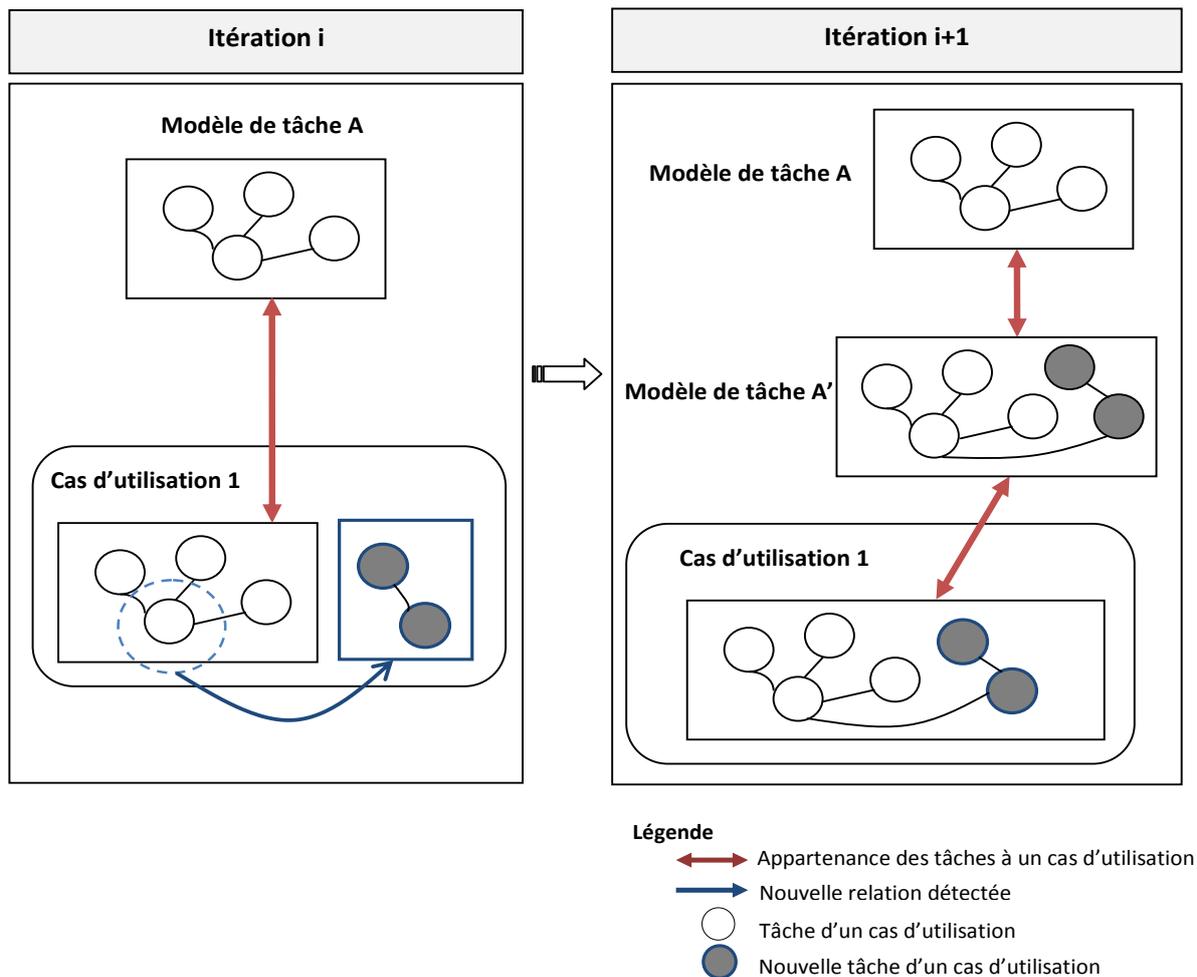
Parmi nos perspectives de recherche, nous visons l'extensibilité et l'enrichissement de la démarche de développement ExUP. Pour ce faire, nous prenons comme point de départ, les problèmes détectés dans la première partie de ce chapitre.

#### **5.4.1.1 Formalisation et modélisation des tâches**

Une amélioration possible de la démarche ExUP est la proposition d'un formalisme d'analyse et de description de tâches. Il peut être intégré au sein de l'activité d'analyse et de conception de l'ExUP. Il doit permettre une description stable et complète des tâches décisionnelles automatiques, manuelles et interactives (objectif de la tâche, buts, conditions d'exécution,

liste de sous-tâches, contraintes fonctionnelles et/ou temporelles, etc.) ; tout en prenant en compte les connaissances et les représentations des utilisateurs. Cette formalisation devra :

1. prendre en compte des recommandations ergonomiques et retenir les aspects intéressants des modèles issus du domaine de l'interaction homme-machine.
2. constituer un bon cadre pour décrire les tâches relatives à l'ECD. Pour ce faire, nous visons la planification d'une étude critique des formalismes existants de modélisation et de description des tâches décisionnelles.
3. Prendre en compte l'aspect itératif de l'ExUP. Rappelons que les fonctionnalités sont exprimées par des cas d'utilisation. Les tâches découlant d'un cas d'utilisation peuvent ne pas être définies dès la première itération de l'ExUP. De même, les relations entre ces tâches peuvent être redéfinies plusieurs fois. C'est ce que nous appelons des raffinements des cas d'utilisation. Le formalisme à proposer doit pouvoir tenir compte de ces raffinements possibles lors de la description des tâches d'un cas d'utilisation ; ce qui peut être illustré par la figure 5.6.



**Figure 5.6** : raffinement possible des modèles des tâches

### 5.4.1.2 Intégration des techniques de visualisation de l'information

Un point important, que nous n'avons pas eu le temps de développer, est la mise en œuvre d'une ou plusieurs techniques de visualisation issues des domaines de l'ECD (visual data mining [Keim 02]) et d'IHM comme proposé par les utilisateurs dans la phase d'évaluation (cf. section 5.1). En fait, l'utilisateur d'un SIAD/ECD est doté d'une capacité à visualiser l'information graphique qui joue un rôle majeur dans ses processus cognitifs (reconnaissance rapide de motifs, couleurs, formes et textures). Il utilise des graphiques afin de mieux appréhender cette information souvent abstraite [Bertin 77] (cf. figure 5.7).

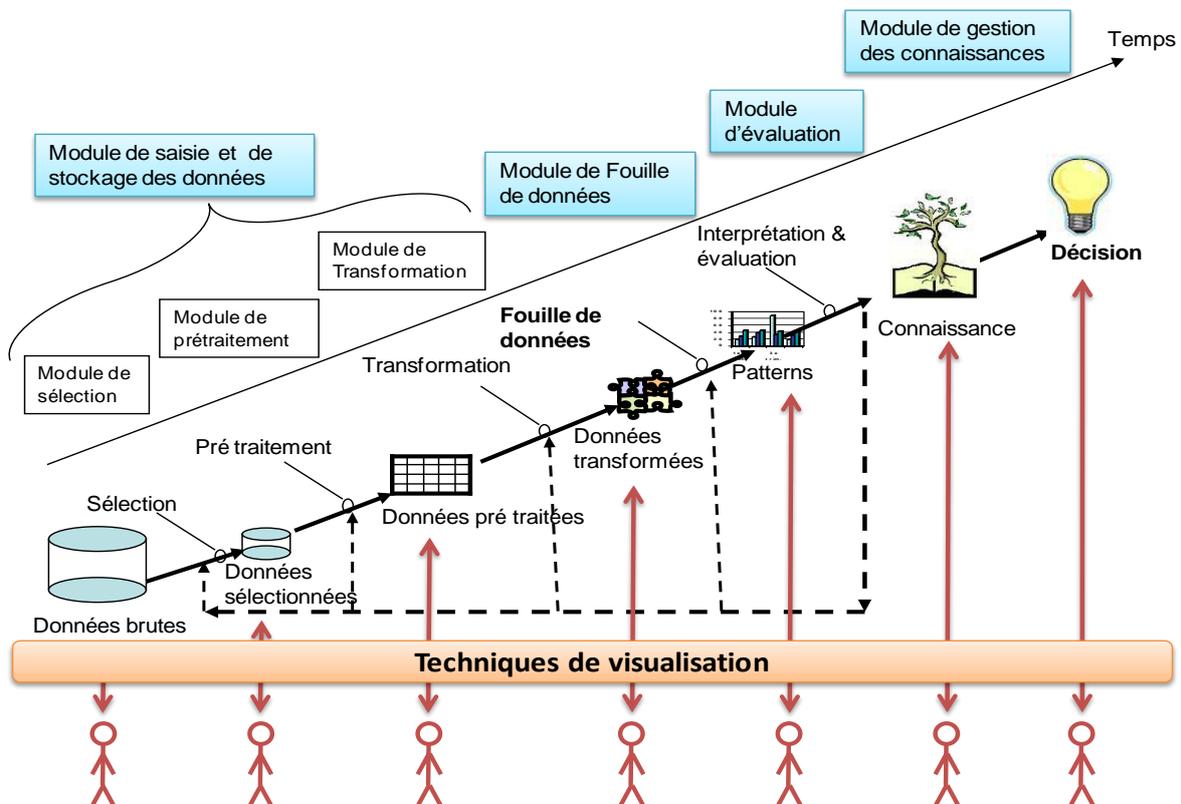


Figure 5.7 : intégration des techniques de visualisation

La visualisation d'information est un domaine de recherche qui se développe et qui a montré son utilité pour faire des découvertes, prendre des décisions, ou trouver des explications [Fekete 05]. L'utilisateur d'un SIAD/ECD a recours à la visualisation pour explorer les tendances qui se dégagent des données ou pour en analyser les détails.

Lorsque les représentations visuelles sont générées, elles ne sont pas forcément statiques ; elles peuvent être dynamiques et l'utilisateur peut les manipuler pour en explorer les données. Un mode de visualisation doit en fait servir autant pour l'exploration que pour la représentation.

Dans ce cadre, nous avons publié un article présentant un état de l'art sur les techniques de visualisation existantes, leurs qualités et leurs faiblesses [Ltifi *et al.* 2009b]. Cette étude a pour but de mieux cerner les difficultés rencontrées par les utilisateurs lorsqu'ils interprètent des informations visualisées. Nous visons, dans nos futurs travaux de recherche, l'intégration des techniques de visualisation pas uniquement pour la visualisation et l'interprétation des résultats de la fouille de données mais dans tous les modules du SIAD/ECD (cf. figure 5.7) :

1. *Pour le module de saisie et de stockage des données* : vu que la manipulation des grands volumes de données est de plus en plus hors de portée des capacités humaines [Hascoët 04], les techniques de visualisation permettent une exploration plus rapide, meilleure et plus adaptée des ensembles de données.
2. *Pour le module de fouille de données* : les techniques de visualisation doivent permettre d'améliorer les techniques d'ECD en augmentant l'implication de l'utilisateur et sa confiance à propos des constatations découvertes. Ce fait devrait contribuer à augmenter la demande dans l'utilisation de techniques de visual data mining et à rendre indispensables celles-ci dans la conjonction avec les techniques automatiques de fouille de données.
3. *Pour le module d'évaluation* : l'idée de base est de présenter les patterns découverts dans une forme visuelle permettant à l'utilisateur (acteur humain) d'interpréter les données, tirer les conclusions ainsi qu'interagir directement avec ces patterns.
4. *Pour le module de gestion de connaissances* : la présentation visuelle des problèmes de décision et des solutions générées par le SIAD/ECD peut orienter le décideur vers la meilleure solution.

Suite à l'état de l'art réalisé sur les techniques de visualisation [Ltifi *et al.* 09b], nous proposons une utilisation possible d'une ou plusieurs techniques<sup>21</sup> pour chaque module du SIAD/ECD (cf. tableau 5.4). Il est important de préciser que ce tableau ne se veut pas du tout exhaustif, mais plutôt représentatif du potentiel du domaine du visual data mining pour les modules classiques en SIAD/ECD.

---

<sup>21</sup> La description des principales techniques a été reportée en annexe A.

**Tableau 5.4 :** application possible des techniques de visualisation pour chaque module du SIADDM/ECD

Module du SIAD/ECD		saisie et stockage de données	Fouille de données	évaluation	gestion de connaissances
Technique de visualisation					
Approche technique : orientée structures de données	Tables de données	X			
	Murs fuyants ou en perspective	X			
	Frises chronologiques	X	X		X
	Nuages de points		X		
	Système de polyligne	X			
	Approche matricielle	X			
	Approche par listes indentées	X			
	Approche par diagrammes	X			
	Cartes d'arbres (treemaps)		X	X	X
	Arbres circulaires (slices)		X	X	X
	Approche hyperbolique		X	X	X
	Approche 3D ou conique		X	X	X
	Approche paysage		X	X	X
	Réseaux de liens hypertextes	X	X		
	Réseaux de co-citation	X	X		
	Réseaux sociaux	X	X		
	Approche de similarités		X	X	X
Approche de pertinence		X	X	X	
Approche interactive : orientée utilisateur	Projection dynamique	X		X	
	Filtrage interactif			X	
	Zoom interactif		X	X	X
	Distorsion interactive		X	X	X
	Liens interactifs et brossage		X	X	X

Ces choix effectués dans le tableau 5.4 vont être appliqués au SIADDM/ECD puis évalués par les médecins du service de réanimation.

### 5.4.1.3 Evaluation de la qualité des données

La découverte de connaissances et la prise de décision à partir de données de qualité médiocre (c'est-à-dire contenant des erreurs, des doublons, des incohérences, des valeurs manquantes, etc.) ont des conséquences directes pour les utilisateurs. Pour cela, le thème de la qualité des données est devenu, depuis ces dernières années, un des sujets d'intérêt émergent dans le domaine de la recherche relativement aux systèmes d'ECD [Cloppet *et al.* 05].

Toutes les applications dédiées à la fouille de données requièrent différentes formes de préparation des données avec de nombreuses techniques de traitement, afin que les données passées en entrée aux algorithmes de fouille ne contiennent pas d'incohérences, de doublons, de valeurs manquantes ou incorrectes.

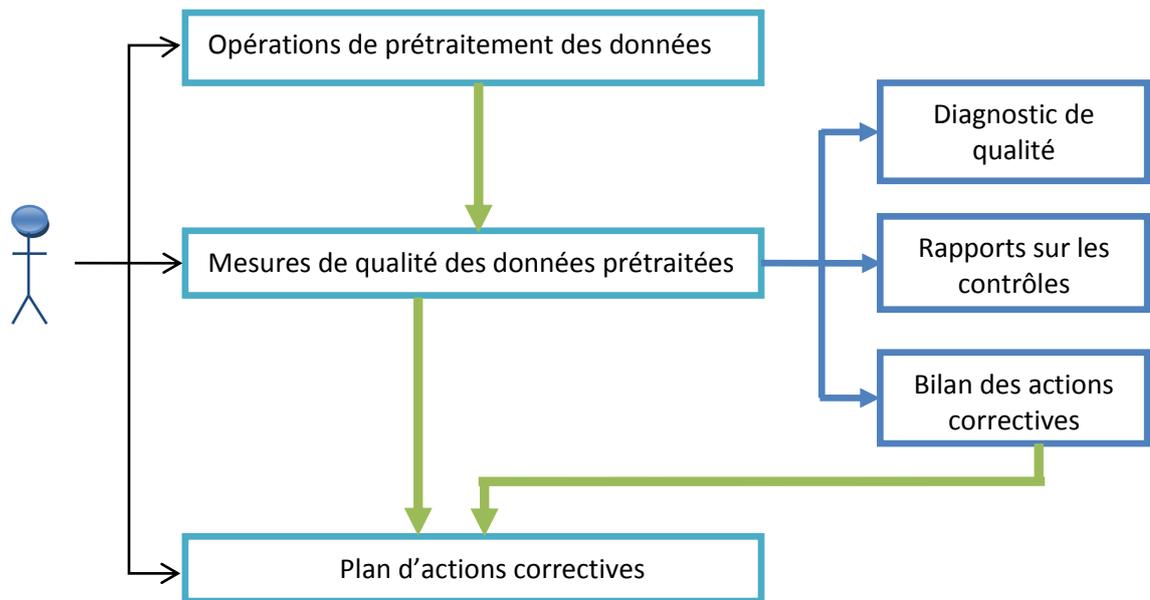
L'évaluation des résultats issus du processus de traitement des données fixes et temporelles, est généralement effectuée par un expert ECD. Cette tâche est souvent très lourde, et un moyen de la faciliter consiste à aider l'expert ECD en lui fournissant des critères d'évaluation sous la forme de mesures de qualité ou d'intérêt des résultats. Ces mesures de qualité doivent être conçues afin de combiner deux dimensions : une dimension objective liées à la qualité des données, et une dimension subjective liées aux intérêts de l'utilisateur.

Nous visons, dans nos futurs travaux de recherche, la proposition d'une méthodologie d'évaluation de la qualité des données fixes et temporelles. Pour ce faire, nous prévoyons de nous consacrer à l'étude des méthodes, des techniques d'analyse et de nettoyage, des approches algorithmiques et des métriques de qualité des données. Une telle étude va nous permettre de comprendre, d'explorer les données, de détecter et corriger les problèmes de qualité des données et donc de mieux préparer la qualité des connaissances extraites à partir de ces données.

Une telle étude doit nous permettre de définir et de décrire les étapes à suivre. Pour chaque étape, il s'agit de répondre aux questions suivantes : *Quelles données évaluer ? Comment les évaluer ? Quand ? Par qui ?*

L'évaluation doit se baser sur une ou plusieurs mesures de qualité. Parmi ces mesures, on peut citer : la complétude (taux de valeurs manquantes), l'exactitude (taux de valeurs correctes), l'actualité (comparaison de la date de saisie à la date courante), la cohérence (taux de données

ne satisfaisant pas à un ensemble de contraintes), la fiabilité et le délai de disponibilité des données [Aze *et al.* 02]. Ce contrôle qualité doit tenir compte également de la technique de fouille de données à utiliser. Le pilotage de la qualité de données fixes et temporelles est une démarche progressive qui va du diagnostic qualité à une certification des données ou un plan d'actions correctives (cf. figure 5.8). Ce projet paraît important pour une bonne découverte des connaissances.



**Figure 5.8** : évaluation de la qualité des données

Nous avons présenté dans cette section nos perspectives de recherche sur le plan méthodologique ; particulièrement celles en relation avec l'analyse des tâches, l'intégration des techniques de visualisation dans les modules du SIAD/ECD ainsi que l'évaluation de la qualité de données. Nous allons maintenant exposer une autre perspective en relation avec les prédictions de l'occurrence des IN dans le service de réanimation.

#### 5.4.2 Perspectives sur le plan expérimental

Une extension de l'algorithme de prédiction des IN pourrait être envisagée. En effet, avec notre système actuel la prédiction est faite offline, c'est-à-dire quotidiennement après la saisie de toutes les données collectées pendant les dernières 24 heures d'hospitalisation pour un patient. Cette prédiction pourrait être améliorée afin d'être effectuée à chaque observation détectée et à chaque instant par notre système qui sera appelé SIADIM/ECD (Système Interactif d'Aide à la Décision Instantanée Médicale basé sur l'ECD).

Pour ce faire, la technique des réseaux bayésiens dynamiques doit être développée pour extraire des modèles instantanés de connaissances en se basant sur les facteurs de risques présentés dans le chapitre 4. Ce modèle doit se baser sur des séries temporelles dont l'unité est l'instant, pendant laquelle une mesure (observation) ou plusieurs sont introduites au SIADIM/ECD.

Une autre amélioration envisagée à ce niveau est l'évaluation des risques de chaque solution générée par le SIADIM/ECD. Des interviews avec les médecins seront planifiées pour définir ces risques. Des courbes (ou autre technique de visualisation d'information (cf. tableau 5.4)) peuvent être aussi tracées pour représenter l'évolution de l'état d'un patient entre les différents instants d'observations pendant une période bien désignée par l'utilisateur (heure, jour, semaine, mois, etc.).

## 5.5. Conclusion

Dans ce chapitre, nous avons présenté une évaluation globale de l'approche ExUP proposée dans le troisième chapitre. En nous basant sur le développement du SIADDM/ECD réalisé dans le chapitre précédent, nous avons effectué plusieurs bilans d'évaluation aussi bien sur les aspects méthodologiques de l'approche que sur les résultats empiriques. Ces bilans ont permis de valider globalement les principes intégrés dans l'ExUP.

Il reste cependant de nombreuses perspectives qui ont été exposées en seconde partie de ce chapitre. Nous visons dans des futurs travaux de recherche d'étendre la démarche d'ExUP par un formalisme d'analyse et description des tâches et par l'intégration des techniques de visualisation dans tous les modules du SIAD/ECD. D'autres perspectives visent à proposer une méthodologie pour l'évaluation de la qualité des données manipulées par les outils de fouille de données. Sur le plan expérimental, nous visons le développement en utilisons l'ExUP, d'un **Système Interactif d'Aide à la Décision Instantanée Médicale** basé sur l'**ECD** (SIADIM/ECD).



## Conclusion générale

Dans ce mémoire, nous nous sommes intéressés à la conception et au développement des Systèmes Interactifs d'Aide à la Décision basés sur un processus d'Extraction de Connaissances à partir de Données (SIAD/ECD). Nous avons constaté qu'il n'existe pas à ce jour une démarche de conception centrée utilisateur précise et appropriée aux SIAD/ECD. Pour cette raison, nous avons proposé une démarche appelée processus unifié étendu (ExUP pour **Extended Unified Process**).

L'originalité de cette démarche proposée repose sur la prise en compte des aspects méthodologiques relatifs à trois domaines de recherche, qui nous semble complémentaires : GL, IHM et ECD. En effet, notre proposition a été basée sur un cadre méthodologique global, issu du GL, qui est le processus unifié. Cette approche a été enrichie sous l'angle des interactions homme-machine d'une part, et des spécificités techniques de l'extraction de connaissances à partir des données de l'autre part.

Dans le premier chapitre de ce mémoire, nous avons présenté le cadre de notre recherche relative au développement des SIAD/ECD. Il est organisé en deux parties. Dans la première, nous avons décrit les caractéristiques des systèmes d'aide à la décision. Nous avons axé nos recherches principalement sur les SIAD que nous considérons avant tout comme des systèmes interactifs. Ces systèmes devraient prendre en compte les attentes des utilisateurs et proposer des interfaces homme-machine correspondant à leurs besoins. Une présentation succincte des systèmes d'aide dans le domaine médical, nous a permis de montrer que notre orientation concerne la mise en œuvre d'un Système Interactif d'Aide à la Décision Médicale (SIADM) permettant de fouiller les données médicales.

Les SIAD qui nous intéressent sont dirigés par les connaissances ; ce qui nous a amené à présenter l'outil décisionnel d'extraction de connaissances à partir des données. C'est dans la deuxième partie de ce chapitre, que nous avons présenté les principaux concepts relatifs à cet outil décisionnel (processus, tâches et algorithmes). Dans ce chapitre, nous avons pu expliquer que les SIAD et l'ECD sont exigeants en matière d'IHM.

Dans le deuxième chapitre, nous avons réalisé un état de l'art sur les approches de conception correspondant aux domaines du GL, l'IHM et des SIAD. Nous avons mis en avant des limites des modèles et méthodes classiques vis-à-vis des IHM ayant conduit aux modèles dits

enrichis. Nous avons remarqué des limites et insuffisances dans ces derniers vis-à-vis du développement des SIAD. Nous avons également présenté les approches de développement spécifiques aux SIAD. L'étude de ces approches, nous a permis de conclure qu'ils ne prennent pas en compte les spécificités relatives à l'ECD.

L'état de l'art sur les méthodes et modèles de développement de systèmes interactifs classiques puis enrichis d'une part, et les développements orientés SIAD d'autre part a permis de mettre en évidence l'absence de démarches précises et appropriées aux SIAD/ECD. Suite à cette étude, nous avons pu déterminer trois types de contraintes à considérer dans notre contribution : (1) l'intégration des aspects de développement issus du domaine de GL, (2) la prise en compte des interactions homme-machine, et (3) la prise en compte des tâches d'extraction de connaissances à partir des données.

C'est en partant de ce constat que nous avons proposé une nouvelle démarche de conception et de développement de SIAD/ECD, intégrant les principes des domaines auxquels ils se rattachent (GL, IHM et ECD).

Dans le troisième chapitre, nous avons décrit notre contribution. C'est une démarche centrée utilisateur pour le développement de SIAD/ECD ; qui prend en considération les aspects du GL et de l'IHM, tout en ayant une vision d'ECD. Nous avons commencé par la décomposition du SIAD/ECD en modules (en se référant aux étapes du processus d'ECD). Dans notre contexte, quatre modules ont été jugés nécessaires et suffisants (saisie et stockage des données, fouille des données, évaluation et gestion des connaissances).

Pour le développement des modules du SIAD/ECD, nous avons proposé un cadre méthodologique global. Il a pris la forme du processus unifié, méthode issue du GL intégrant les aspects itératifs et architecturaux. Ce cadre général de développement se basant sur les trois principes du processus générique unifié (démarche itérative, pilotée par les cas d'utilisation, et centrée sur l'architecture) nous a permis de répondre à la première contrainte de notre contribution. Concernant la deuxième contrainte, nous avons pu enrichir les activités du processus unifié sous l'angle des interactions homme-machine; d'où l'appellation du processus unifié étendu (**Extended Unified Process**). Pour chaque activité de l'ExUP, nous avons présenté des aspects génériques à tous les modules et d'autres spécifiques pour chaque module du SIAD/ECD en se référant aux spécificités techniques des tâches d'ECD, ce qui nous a permis de répondre à la troisième contrainte.

Le quatrième chapitre a traité l'application et la mise en œuvre de l'approche proposée sur une étude de cas relative à la supervision quotidienne des IN acquises dans le service de réanimation du centre hospitalier universitaire Habib Bourguiba à Sfax. Un ensemble de visites et réunions à l'hôpital ont été effectuées afin de définir les besoins des médecins du service. En effet, pour chaque patient hospitalisé, ils ont voulu prendre des décisions dynamiques pour lutter contre l'occurrence des IN, chaque jour de son hospitalisation. De ce constat, nous avons développé un **Système Interactif d'Aide à la Décision Dynamique Médicale** basé sur l'ECD (SIADDM/ECD).

Des modules ont été déjà développés dans le cadre soit de mon travail de mastère soit d'autres travaux de recherche appartenant à l'équipe de travail dont je suis membre. Les modules du SIADDM/ECD que nous avons développés en utilisant l'ExUP sont : le "module de fouille de données" et le "module de gestion des connaissances". Pour le premier module nous avons implémenté la technique des Réseaux Bayésiens Dynamiques (RBD). Cette technique a été utilisée pour modéliser la complexité de notre système, due au fait que les données manipulées sont temporelles. Une validation préliminaire de l'ExUP et du SIADDM/ECD réalisé auprès des utilisateurs a été discutée dans ce chapitre.

Dans le cinquième chapitre, nous avons effectué une évaluation globale de notre approche ExUP et de son application dans le domaine médical (SIADDM/ECD). Cette évaluation a permis de valider globalement les principes de base intégrés dans l'ExUP (l'IHM, le processus de décision, l'aspect itératif de développement et l'ECD) ainsi que l'utilité et l'utilisabilité du SIADDM/ECD par les médecins du service de réanimation. Il reste cependant de nombreuses perspectives qui ont été exposées.

Une de nos premières perspectives serait de procéder à une extensibilité complémentaire de la démarche de l'ExUP en intégrant un formalisme d'analyse de tâches dans les activités d'analyse et de conception de l'ExUP.

Par ailleurs, une autre perspective serait de proposer une méthodologie d'évaluation de la qualité des données fixes et temporelles. Cette évaluation conditionne la qualité des connaissances extraites par les outils de fouille de données. Dans le même état d'esprit, nous viserons l'intégration de techniques de visualisation d'information dans tous les modules du SIADDM/ECD. Cette dernière technologie permet de présenter les données et les connaissances dans une forme visuelle permettant à l'utilisateur d'interpréter les données, tirer les conclusions ainsi qu'interagir directement avec ces données. On considère que les techniques de visualisation peuvent améliorer les techniques d'ECD courantes en augmentant

l'implication de l'utilisateur et sa confiance à propos des constatations découvertes. Une telle méthodologie d'évaluation doit permettre d'étudier l'expérience cognitive et affective des utilisateurs du système d'aide à la décision pour la lutte contre les infections nosocomiales.

Enfin, une dernière perspective sur le plan expérimental serait d'étendre la technique des RBD, utilisé pour la fouille de données, pour effectuer des prédictions à chaque observation détectée et à chaque instant. Le système serait un Système Interactif d'Aide à la Décision Instantanée Médicale basé sur l'ECD (SIADIM/ECD).

## Bibliographie

- [Abed 90] Abed M. *Contribution à la modélisation de la tâche par outils de spécification exploitant les mouvements oculaires : application à la conception et à l'évaluation des interfaces homme-machine*. Thèse de doctorat, Université de Valenciennes et du Hainaut-Cambrésis, France, 1990.
- [Abed 01] Abed M. *Méthodes et modèles formels et semi-formels pour la conception et l'évaluation des systèmes homme-machine*. Habilitation à diriger des Recherches, Université de Valenciennes et du Hainaut-Cambrésis, France, Mai 2001.
- [Abello et Ham 04] Abello J. et Ham F.V. *Matrix Zoom: A Visual Interface to Semi-External Graphs* in Proceedings of the IEEE Symposium on Information Visualization (INFOVIS'04), IEEE Computer Society, pp. 183-190, 2004.
- [Abrial 96] Abrial J.R. *The B BOOK, Assigning programs to meaning*. Cambridge University Press, 779 pages, 1996.
- [Akharraz 04] Akharraz A., *Acceptabilité de la décision et risque décisionnel : Un système explicatif de fusion d'informations par l'intégrale de Choquet*. Thèse de doctorat, Université de Savoie, 2004.
- [Akinde et al. 03] Akinde M.O., Böhlen M.H., Johnson T., et Lakshmanan L.V. Efficient OLAP query processing in distributed data warehouses. *Information Systems*, vol. 28 (1-2), pp. 111-135, 2003.
- [Amalberti et al. 91] Amalberti R., De Montmollin M., Theureau J. (Eds.). *Modèles en analyse du Travail*. Liège : Mardaga, 1991.
- [André 93] André J. *Le cycle de vie en Y*. Arche SQL, 1993.
- [Andrews et Heidegger 98] Andrews K., et Heidegger H. Information Slices: Visualising and Exploring Large Hierarchies using Cascading, Semi-Circular Discs. *InfoVis*, pp. 9-12, 1998.
- [Andrews et al. 04] Andrews K., Kienreich W., Sabol V., Granitzer M. The visualization of large hierarchical documents spaces with InfoSky. *Proceedings of CoData'04 - Workshop on Information Visualization, Presentation, and Design*, Prague, Czeck Republic, 2004.

- [Arnott *et al.* 07] Arnott D., Jirachiefpattana W., et O'Donnell P. Executive information systems development in an emerging economy. *Decision Support Systems*, vol. 42 (4), pp. 2078-2084, 2007.
- [Atkinson *et al.* 09] Atkinson N.L., Saperstein S.L., et Pleis J. Using the Internet for Health-Related Activities: Findings From a National Probability Sample. *Journal of Medical Internet Research*, vol. 11 (1): e4, 2009.
- [Auber *et al.* 03] Auber D., Chiricota Y., Jourdan F. et Melançon G. Multiscale navigation of Small World Networks. *IEEE Symposium on Information Visualisation*, Seattle, GA, USA, IEEE Computer Science Press, pp. 75-81, 2003.
- [Aze *et al.* 02] Aze J., et Kodratoff Y. Evaluation de résistance au bruit de quelques mesures d'extraction de règles d'association. *Extraction de Connaissances et Apprentissage*, vol. 14, pp. 143-154, 2002.
- [Baccino *et al.* 05] Baccino T., Bellino C., et Colombi T. *Mesure de l'utilisabilité des interfaces*, Paris : Hermès, 280 pages, 2005.
- [Bahloul *et al.* 10] Bahloul R., Ben Ayed M., et Alimi A-M. Vers une méthode d'évaluation de Système Interactif d'Aide à la Décision basé sur le processus d'ECD. Atelier *EvalECD'2010 Évaluation des méthodes d'Extraction de Connaissances dans les Données*, pp. 1-12, Hammamet, Tunisie, Janvier 2010.
- [Balbo 94] Balbo S. *Evaluation ergonomique des interfaces utilisateur : un pas vers l'automatisation*. Thèse de doctorat, Université Joseph Fourier, Grenoble 1, 1994.
- [Barthet 88] Barthet M.F. *Logiciels interactifs et ergonomie, Modèle et méthodes de conception*, Dunod informatique, Paris, 1988.
- [Bass *et al.* 98] Bass L., Clements P., et Kazman R., *Software Architecture in Practice*. Addison Wesley, 1998.
- [Ben Ayed *et al.* 10] Ben Ayed M., Ltifi H., Kolski C., Alimi A. A User-centered Approach for the Design and Implementation of KDD-based DSS: A case Study in the Healthcare Domain. *Decision Support Systems*, Vol. 50 (1), pp. 64-78, 2010.
- [Berkhin 02] Berkhin P., *Survey of clustering data mining techniques*. Technical report, Accrue Software, 2002.

[Bertin 77] Bertin J. *La Graphique et le Traitement Graphique de l'information*. Flammarion, 1977.

[Bertin 99] Bertin J. *Sémiologie graphique : les diagrammes, les réseaux, les cartes* (3<sup>ème</sup> Edition). Paris, Editions de l'Ecole des Hautes Etudes en Sciences, 1999.

[Boehm 88] Boehm B. A Spiral Model of Software Development and Enhancement. *IEEE Computer*, vol. 21 (5), pp. 61-72, 1988.

[Boutin 05] Boutin F. *Filtrage, classification et visualisation multi-échelles de graphes d'interactions à partir d'un focus*. Thèse de doctorat, Université Montpellier II, le 29 novembre 2005.

[Boy 91] Boy G. *Intelligent Assistant Systems*. New York: Academic Press, 1991.

[Brachman *et al.* 04] Brachman R. et Levesque H. *Knowledge Representation and Reasoning*. Morgan Kaufmann Publishers Inc, 2004.

[Breedvelt *et al.* 97] Breedvelt I., Paternò F., et Sereriins C. Reusable Structures in Task Models. *Proceedings Design, Specification, Verification of Interactive Systems*, Springer Verlag, pp. 251-265, 1997.

[Brossette *et al.* 98] Brossette S.E, Sprague A.P., Hardin J.M., Waites K.B., Jones W.T., et Moser S.A. Association Rules and Data Mining in Hospital Infection Control and Public Health Surveillance. *Journal of the American Medical Informatics Association*, vol. 5(4), pp. 373-381, 1998.

[Brossette *et al.* 00] Brossette S.E., Sprague A.P., Jones W.T., et Moser S.A. A data mining system for infection control surveillance. *Methods of Information in Medicine*, vol. 39(4-5), pp. 303-310, 2000.

[Bueno-Cavanillas *et al.* 91] Bueno-Cavanillas A., Rodriguez-Contreras R., Lopez-Luque A., Delgado-Rodriguez M., et Galves-Vargas R. Usefulness of severity indices in intensive care medicine as a predictor of nosocomial infection risk. *Intensive Care Medicine*, vol. 17, pp. 336-339, 1991.

[Burnside *et al.* 06] Burnside E.S., Rubin D.L., Fine J.P., Shachter R.D., Sisney G.A., Leung, W.K. Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: initial experience. *Radiology*, vol. 240(3), pp. 666-673, 2006.

- [Clark 06] Clark S.J. A general temporal data model and the structured population event history register, *Demographic research*, vol. 15(7), pp. 181-252, 2006.
- [Checroun 92] Checroun A. *Comprendre, concevoir et utiliser les SIAD*. MASSON, Paris, 168 pages, 1992.
- [Chen *et al.* 05] Chen H., Fuller S.S., Friedman C., Hersh W. *Medical Informatics: Knowledge Management and Data mining in Biomedicine*. Springer, pp. 3-33, 2005.
- [Chen 05] Chen C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific litterature. *Journal of American society for information science and technology*, vol. 57(3), pp. 359-377, 2005.
- [Chevrin *et al.* 07] Chevrin V., Coiturier O., Mephu Nguifo E., et Rouillard J. Recherche anthropocentrée de règles d'association pour l'aide à la décision. *Revue d'Interaction Homme-Machine*, vol. 8 (2), pp. 117-146, 2007.
- [Clark *et al.* 89] Clark P., et Niblett T. The CN2 Induction Algorithm. *Machine Learning*, vol. 3, pp. 261-283, 1989.
- [Cloppet *et al.* 05] Cloppet F., Petit J-M., et Vincent N. *Extraction des connaissances : Etat et perspectives*. Revue des Nouvelles Technologies de l'Information, RNTI-E-5, CÉPADUÈS-ÉDITIONS, 2005.
- [Collins 95] Collins D. *Designing Object-oriented user interfaces*. Readwoods City, CA, Benjamin/Cummings Publishing Inc., 1995.
- [Courbon *et al.* 81] Courbon J.C., Grageof J., et Tomasi J. L'approche évolutive. *Informatique et Gestion*, vol. 21, pp. 29-34, 1981.
- [Couturier *et al.* 06] Couturier O., et Mephu Nguifo E. Une approche anthropocentrée interactive pour l'aide à la décision en marketing bancaire. *Actes de la 18ème conférence Francophone en Interaction Homme-Machine (IHM'06)*, Canada, pp. 253-256, 2006.
- [Curtis *et al.* 94] Curtis B., et Hefley B. A WIMP No More: The Maturing of User Interface Engineering. *ACM Interactions*, vol. 1(1), pp. 23-34, 1994.
- [Dargam *et al.* 06] Dargam F.C.C., et Zaraté P. Decision Support Systems: Methodologies and Applications, *Journal of Decision Systems*, vol. 15 (2-3), pp. 157-158, 2006.

[Darmoni *et al.* 94] Darmoni S.J., Massari P., *et al.* SETH: an expert system for the management on acute drug poisoning in adults. *Computer Methods and Programs in Biomedicine*, vol. 43, pp. 171-176, 1994.

[Degoulet *et al.* 91] Degoulet P., et Fieschi M. Traitement de l'information médicale, *Méthodes et applications hospitalières*. Collection : Manuels Informatiques Masson – Entreprise, 320 pages, 1991.

[DeMarco 79] DeMarco T. *Structured analysis and system specification*. Prentice Hall PTR, Upper Saddle River, NJ, 1979.

[Detyniecki 06] Detyniecki M. *Fusion et Fouille de données Multimédia*. Habilitation à diriger des recherches, Université Pierre et Marie Curie, 2006.

[Diaper 04] Diaper D. Understanding task analysis for Human-Computer Interaction. In Diaper D., N. Stanton (Eds.), *the handbook of task analysis for human-computer interaction*, pp. 5-47, LEA Pub, 2004.

[Dickover *et al.* 77] Dickover M.E., McGowan C.L., et Ross D.T. Software design using: SADT. *Proceedings of the 1977 ACM Annual conference*, ACM NY, USA, pp. 125-133, 1977.

[Doise 92] Doise W. *Dissensions et consensus*. Editions PUF, 1992.

[Dreiseitl et Binder 05] Dreiseitl S., et Binder M. Do physicians value decision support? A look at the effect of decision support systems on physician opinion. *Artificial Intelligence in Medicine*, vol. 33 (1), pp. 25-30, 2005.

[Duribreux-Cocquebert, 1995] M. Duribreux-Cocquebert. *MODESTI : vers une méthodologie interactive de développement de Systèmes à Base de Connaissances*. Thèse de doctorat, Université de Valenciennes et du Hainaut-Cambrésis, 1995.

[Eom 04] Eom S. The Changing Structure of Decision Support Systems Research: An Empirical Investigation through Author Cocitation Mapping. *The 2004 IFIP International Conference on Decision Support Systems*, Prato, Tuscany, pp. 243-251, 2004.

[ESA 91] ESA. *Software Engineering Standards*. European Space Agency, Rapport Issue 2, 1991.

[Fayyad *et al.* 96] Fayyad U.M., Djorgovski S.G., et Weir N. Automating the Analysis and Cataloging of Sky Surveys. *Advances in Knowledge Discovery and Data Mining*, MIT Press, pp. 471-494, 1996.

[Fayyad *et al.* 01] Fayyad U.M, Grinstein G., Wierse A. *Information Visualization in Data Mining and Knowledge Discovery*. Morgan Kaufmann, 2001.

[Fayyad *et al.* 03] Fayyad U.M., Piatetsky-Shapiro G. et Uthurusamy R. Summary from the KDD-03 panel -- Data Mining: The next 10 years. *SIGKDD Explorations Newsletter*, vol. 5 (2), pp. 191-196, ACM Press, 2003.

[Fekete 05] Fekete J.D. *Nouvelle génération d'Interfaces Homme-Machine pour mieux agir et mieux comprendre*. Habilitation à diriger les Recherches, Research Report 1411, Université Paris-Sud, 2005.

[Frawley *et al.* 92] Frawley, W.J., Piatetsky-Shapiro, G., et Matheus C.J. Knowledge discovery in databases: an overview. *Proceedings of AI Magazine*, pp. 57–70, 1992.

[Freeman 97] Freeman E. *The Lifestreams Software Architecture*. Thèse de doctorat, 1997.

[Gabay 01] Gabay J. *Merise et UML pour la modélisation des systèmes d'information*. Dunod, Paris, 2001.

[Gafsi-Moalla 05] Gafsi-Moalla S. *Les infections acquises en réanimation, étude prospective réalisée dans le service de réanimation de Sfax sur une période de 3 mois*. Thèse de doctorat, Faculté de Médecine de Sfax, Tunisia, 2005.

[Garlatti 96] Garlatti S. Tutorial : Multimédia et systèmes d'aide à la décision en situation complexe. In *43<sup>th</sup> meeting of the european working group "Multicriteria Aid for Decisions"*, Brest, 1996.

[Garner *et al.* 88] Garner J.S., Jarvis W.R, Emori T.G., Hogan T.C., et Hugues J.M. CDC definitions for nosocomial infections. *American journal of infection control*, vol. 16 (3), pp. 128-140, 1988.

[Gaynes *et al.* 99] Gaynes RP, Horan TC. Surveillance of nosocomial infections. In: Mayhall CG, editor. *Hospital Epidemiology and Infection Control*. 2<sup>nd</sup> edition. Baltimore, Md: Williams & Wilkins, pp.1285-1317, 1999.

- [Ge *et al.* 07] Ge L., Mourits M.C.M. et Huirne R.B.M. Towards flexible decision support in the control of animal epidemics. *Revue scientifique et technique - Office international des épizooties*, vol. 26 (3), pp. 551-563, 2007.
- [Ghoniem et Fekete 02] Ghoniem M. et Fekete J.-D. Visualisation de graphes de co-activité; par matrices d'adjacence. In *IHM '02: Proceedings of the 14th French-speaking conference on Human-computer interaction*, pp. 279-282, New York, NY, USA, 2002. ACM Press.
- [Ghoniem *et al.* 04] Ghoniem M., Fekete J.-D. et Castagliola P. A Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations. *Proceedings of the 10<sup>th</sup> IEEE Symposium on Information Visualization, Austin, Texas, USA*, pp. 17–24, 2004.
- [Girvan et Newman 02] Girvan M. et Newman M.E.J. Community structure in social and biological networks. *Proceedings of the National Academy Science USA*, vol. 99, pp. 7821-7826, 2002.
- [Gould et Lewis 85] Gould J.D., Lewis C.H. Designing for usability – key principles and what designers think. *Communications of the ACM*, vol. 28(3), pp. 300-311, 1985.
- [Greenberg 04] Greenberg S. Working through task-centered system design. In *the handbook of task analysis for human-computer interaction*, D. Diaper, N. Stanton (Eds.), pp. 49-65, LEA Pub, 2004.
- [Guiard *et al.* 01] Guiard Y., Bourgeois F., Mottet D. et Beaudouin-Lafon M. Beyond the 10-bit barrier: Fitts' law in multiscale electronic worlds. *IHM-HCI*, pp. 573-587, 2001.
- [Hackathorn et Keen 81] Hackathorn R.D. et Keen P.G.W. Organizational Strategies for Personal Computing in Decision Support Systems. *MIS Quarterly*, vol. 5 (3), pp.21-28, 1981.
- [Hackos et Redish 98] Hackos Joann T., Redish Janice C., *User and Task Analysis for Interface Design*, John Wiley & Sons, Inc. 1998.
- [Hascoët et Beaudouin-Lafon 01] Hascoët M. et Beaudouin-Lafon M. Visualisation interactive d'information. *Revue I3*, vol. 1(1), pp. 77-108, 2001.
- [Hascoët 04] Hascoët M. *Visualisation d'information et interaction*. In : Méthodes avancées pour les systèmes de recherche d'informations. M. Ihadjadene. Paris: Hermes, 2004.
- [Hassine *et al.* 02] Hassine I., Rieu D., Bounaas F., et Seghrouchni O. Towards a Reusable Business Components Model. *Workshop on Reuse in Object Oriented Information Systems Design, OOIS 2002*, Montpellier, France, 2002.

[Haley *et al.* 85] Haley RW, Culver DH, White JW, *et al.* The nationwide nosocomial infection rate, A new need for vital statistics. *American Journal of Epidemiology*, vol. 121, pp. 159-167, 1985.

[Hatley *et al.* 87] Hatley D.J., et Pirbai I.A. *Strategies for real-time system specification*. Dorset House Publishing Co., Inc. New York, NY, USA, 1987.

[Hatley *et al.* 91] Hatley D.J., et Pirbai I.A. *Stratégies de spécification des systèmes temps réel SA-RT*. Editions Masson, Paris, 1991.

[Haug *et al.* 03] Haug P.J., Rocha B., et Evans R.S. Decision support in medicine: lessons from the HELP system. *International Journal of Medical Informatics*, vol 69 (2-3), pp. 273-284, 2003.

[Heer et Boyd 05] Heer J. et Boyd D. Vizster: Visualizing Online Social Networks. *IEEE InfoVis 2005*. Minneapolis, Minnesota, pp. 23-25, 2005.

[Hergafi 06] Hergafi L. *Présentation et validation d'un nouveau système pour la surveillance de l'infection acquise en réanimation*. Thèse de doctorat, Faculté de Médecine de Sfax, Tunisia, 2006.

[Herman *et al.* 99] Herman I. M. G., de Ruitter M. M. et Delest M. Latour - A Tree Visualisation System. *Proceedings of the 7<sup>th</sup> International Symposium on Graph Drawing (GD '99)*, Stirin Castle, Czech Republication, pp.392–399, 1999.

[Hix *et al.* 93] Hix D. et Hartson H.R. *Developing user interfaces: Ensuring usability through product & process*. Wiley professional computing, John Wiley & Sons, inc., NY, USA, 1993.

[Holsapple et Whinston 96] Holsapple C.W., et Whinston A.B. *Decision Support Systems: A Knowledge-Based Approach*. St. Paul: West Publishing, 1996.

[Holtzman 89] Holtzman S. *Intelligent Decision Systems*. Addison Wesley, 1989.

[Humbert 06] Humbert P. *Prise en compte des besoins des utilisateurs dans la visualisation des connaissances scientifiques de corpus bibliographiques : Approche dynamique par classification*, Mémoire Présenté et soutenu le 4 juillet 2006 Pour l'obtention du Master Recherche Sciences de l'Homme et de la Société Sciences de l'Information et de la Communication Université Nancy 2.

[IGL 89] I.G.L. Technology. *SADT, un langage pour communiquer*. Eyrolles, Paris, 1989.

- [Inmon 96] Inmon W., *Building the Data Warehouse*. 2<sup>nd</sup> Edition, John Wiley & Sons Inc., 1996.
- [Inselberg et Dimsdale 90] Inselberg A et Dimsdale B. Parallel coordinates: A tool for visualizing multidimensional geometry. In *IEEE Visualization*, pp. 361-378, 1990.
- [ISO13407 99] ISO 13407. *Processus de conception centrée sur l'opérateur humain pour les systèmes interactifs*. 1999.
- [ISO/TR 18529 00] ISO/TR 18529. *Ergonomie - Ergonomie de l'interaction homme/système - - Descriptions des processus cycle de vie centrées sur l'opérateur humain*. 2000.
- [ISO/TR 16982 02] ISO/TR 16982. *Ergonomie de l'interaction homme-système - Méthodes d'utilisabilité pour la conception centrée sur l'opérateur humain*. 2002.
- [Itzhak et al. 89] Itzhak G. et Schmeidler D. Maxmin Expected Utility with a Non-Unique Prior. *Journal of Mathematical Economics*, vol. 18, pp. 141-153, 1989.
- [Jacobson et al. 93] Jacobson I., Christenson M., Jonsson P., et Overgaard G. *Le génie logiciel orienté objet, une approche fondée sur les cas d'utilisation*. Addison-Wesley, USA, ACM Press Books, 1993.
- [Jacobson et al. 99] Jacobson I., Booch G., et Rumbaugh J. *Le processus unifié de développement logiciel*. Eds Addison Wesley New York, 1999.
- [Jensen 96] Jensen, F. *An Introduction to Bayesian Networks*. Springer, Verlag, NY, 1996.
- [Jinyan et al. 06] Jinyan L., Qiang Y. et Ah Hwee T. (Editors). Data Mining for Biomedical Applications. *Proceedings of PAKDD 2006 Workshop, BioDM 2006*, Singapore. Lecture Notes in Bioinformatics, vol. 3916 (LNBI 3916), Springer-Verlag, 2006.
- [Johansson et al. 93] Johansson H., McHugh P., Pendlebury A., et Wheller W. *Business Process Reengineering: Breakpoint Strategies for Market Dominance*, Wiley, Chichester, Angleterre, 1993.
- [Jones 90] Jones C.B. *Systematic software development using VDM*. (2nd ed.), Prentice-Hall, Inc., Upper Saddle River, NJ, 1990.
- [Kallel et al. 05] Kallel H., Bouaziz M., Ksibi H., Chelly H., Hmida C.B., Chaari A., Rekik N., et Bouaziz M. Prevalence of hospital-acquired infection in a Tunisian Hospital. *Journal of Hospital Infection*, vol. 59, pp. 343-347, 2005.

- [Kantardzic 03] Kantardzic M. *Data Mining: Concepts, Models, Methods, and Algorithms*. John Wiley & Sons, 2003.
- [Kapusova 04] Kapusova D. *Visualisation de l'information dans le portail STAF18*, 2004.
- [Keim 02] Keim D. Information Visualization and Visual Data Mining. *IEEE Transactions on Visualization and Computer Graphics*, vol. 8(1), pp. 1-8, 2002.
- [Kleinberg et al. 01] Kleinberg E., van de Wetering H., et al. Botanical Visualization of Huge Hierarchies, *IEEE InfoVis'01*, pp. 87-94, 2001.
- [Knaus et al. 81] Knaus WA, Zimmerman JE, Wagner DP, Draper EA et Lawrence DE. APACHE: acute physiology and chronic health evaluation: a physiologically based classification system. *Critical Care Medicine*, vol. 9, pp. 591-597, 1981.
- [Kodratoff et al. 01] Kodratoff Y., Napoli A., Zighed D. *L'extraction de connaissances, fouille de données et intelligence artificielle*. Bulletin AFIA ECD, numéro 46/47, 2001.
- [Kolski 97] Kolski C. *Interfaces Homme-Machine, application aux systèmes industriels complexes*. (2<sup>ème</sup> édition revue et étendue), Editions Hermès, Paris, 1997.
- [Kolski 98] Kolski C. A call for answers around the proposition of an HCI-enriched model. *ACM SIGSOFT Software Engineering Notes*, vol. 23 (3), pp. 93-96, 1998.
- [Kolski et al. 01] Kolski C., Ezzedine H., et Abed M. Développement du logiciel : des cycles classiques aux cycles enrichis sous l'angle des IHM. In Kolski C. (Ed) Analyse et conception de l'IHM, *Interaction Homme-Machine pour les SI*, vol. 1, pp. 23-49, Hermès, Paris, 2001.
- [Krzysztof et Moore 02] Krzysztof J.C. et Moore G.W. Uniqueness of medical data mining. *Artificial Intelligence in Medicine*, Vol. 26 (1-2), pp. 1-24, 2002.
- [Kulikowski 82] Kulikowski C.A., et Weiss S.M. Representation of Expert Knowledge for Consultation: The CASNET and EXPERT Projects. Chapter 2 in Szolovits, P. (Ed.) *Artificial Intelligence in Medicine*. Westview Press, Boulder, Colorado. 1982.
- [Kuo 98] Kuo F.Y. Managerial intuition and the development of executive support systems. *Decision Support Systems*, vol. 24 (2), pp. 89-103, 1998.
- [Lajnef et al. 05] Lajnef, M., Ben Ayed, M. et Kolski, C. Convergence possible des processus du data mining et de conception-évaluation d'IHM : adaptation du modèle en U. *Proceedings of IHM'05, International Conference Proceedings Series*, ACM Press, Toulouse, pp. 243-246, 2005.

- [Lamping et al. 95] Lamping, J., Rao, R., Pirolli, P. A focus + context technique based on hyperbolic geometry for visualizing large hierarchies. *Proceedings of ACM Human Factors in Computing Systems (CHI'95)*, pp. 401–408, 1995.
- [Laprie et al. 95] Laprie J.C., Arlat J., Blanquart J.P., et al. *Guide de la sûreté de fonctionnement*. Cépaduès Editions, Rapport No 94090, 1995.
- [Larman 05] Larman C. *UML 2 et les Design Patterns*. Editions Pearson Education, 2005.
- [Larvet 94] Larvet P. *Analyse des systèmes : de l'approche fonctionnelle à l'approche objet*. InterEditions, Paris, 1994.
- [Lauritzen et al. 99] Lauritzen, L., Cowell G., Dawid, P.A., Spiegelhalter, D.J. *Probabilistic Networks and Expert Systems*. Springer-Verlag New York, Inc., 1999.
- [Lavrac 99] Lavrac N. Machine learning for data mining in medicine. *Proceedings of the Seventh Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making (AIMDM)*, Aalborg, Denmark, pp. 57-60, 1999.
- [Lavoie 07] Lavoie B. *Raisonnement non monotone*. Synthèse de lectures, Programme de Doctorat en Informatique Cognitive, Université du Québec à Montréal. 11 pages, 2007.
- [Lebraty 98] Lebraty J.F. Technologies de l'information et Décision : L'essor des Datawarehouses. *Systèmes d'Information et Management*, vol. 3 (1), pp. 95-118, 1998.
- [Lefébure et al. 01] Lefébure R., et Venturi G. *Data Mining*. Editions Eyrolles, 2001.
- [Lemieux et al. 06] Lemieux F., et Desmarais M.C. RUP et conception centrée sur l'utilisateur : une étude de cas. *Actes de la conférence ERGO-IA '06*, Biarritz, pp. 9-37, 2006.
- [Lepreux et al. 03] Lepreux S., Abed M., et Kolski C. A human-centred methodology applied to decision support system design and evaluation in a railway network context. *Cognition, Technology and Work*, vol. 5, pp. 248-271, 2003.
- [Lepreux 05] Lepreux S. *Approche de Développement centré décideur et à l'aide de patrons de Systèmes Interactifs d'Aide à la Décision*. Thèse de doctorat, Université de Valenciennes et du Hainaut-Cambrésis, France, 2005.
- [Leroy et al. 07] Leroy G., et Chen H. Introduction to the special issue on decision support in medicine. *Decision Support Systems*, vol. 43 (4), pp. 1203-1206, 2007.
- [Lévine et al. 89] Lévine P., et Pomerol J.C. *Systèmes interactifs d'aide à la décision et systèmes experts*, Hermès, 1989.

[Long *et al.* 90] Long J.B., Denley I. *Evaluation for practice*, tutorial. Ergonomics society annual conference, 1990.

[Lortholary *et al.* 95] Lortholary O., Fagon J. Y., Buu A., Hoi, Slama M. A., Pierre J., Giral P., Rosenzweig R., Gutmann L., Safar M., et Acar J. Nosocomial acquisition of multiresistant *Acinetobacter baumannii*: risk factors and prognosis. *Clinical Infectious Diseases*, vol. 20, pp. 790-796, 1995.

[Ltifi 07] Ltifi H. *Vers une approche de développement de Système Interactif d'Aide à la Décision basé sur le Data Mining*. Mastère de recherche, Ecole Nationale d'Ingénieurs de Sfax, Tunisie, 2007.

[Ltifi *et al.* 08] Ltifi H., Ben Ayed M., Kolski C., Alimi A-M. Prise en compte de l'utilisateur pour la conception d'un SIAD basé sur un processus d'ECD. D. Galaretta, P. Girard, J.C. Tucoulou, M. Wolff (Ed.), *Actes de la Conférence ERGO'IA 2008, L'humain au cœur des systèmes et de leur développement (Bidart/Biarritz, 15-17 octobre)*, ESTIA, pp. 85-92, octobre, 2008.

[Ltifi *et al.* 09] Ltifi H., Ben Ayed M., Lepreux S., Alimi A.M. Survey of Information Visualization Techniques for Exploitation in KDD. *The 7<sup>th</sup> ACS/IEEE International Conference on Computer Systems and Applications*, Rabat, Morocco, pp. 218-225, 2009.

[Ltifi *et al.* 10] Ltifi H., Ben Ayed M., Kolski C., Alimi A. Démarche centrée utilisateur pour la conception de SIAD basés sur un processus d'ECD, application dans le domaine de la santé. *Journal d'Interaction Personne-Système (JIPS)*, vol. 1(1), pp. 1-25, 2010.

[Ltifi *et al.* 10b] Ltifi H., Kolski C., Ben Ayed M., Alimi A. Human-centred design approach applied to Medical Dynamic DSS. *The 11<sup>th</sup> IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems* (August 31 - September 3), Valenciennes, France, septembre.

[Lucas *et al.* 00] Lucas P.J.F., de Bruijn N.C., Schurink K., Hoepelman I.M. A Probabilistic and decision-theoretic approach to the management of infectious disease at the ICU. *Artificial Intelligence in Medicine*, vol. 19(3), pp. 251-279, 2000.

[Lucas 01] Lucas, P. Bayesian networks in medicine: A model-based approach to medical decision making. *Proceedings of the EUNITE Workshop on Intelligent Systems in Patient Care*, Vienna, pp. 73-97, 2001.

- [Lussier *et al.* 07] Lussier Y.A., Williams R., Li J., Jalan S., Borlawsky T., Stern E. et Kohli I. Partitioning knowledge bases between advanced notification and clinical decision support systems. *Decision Support Systems*, vol. 43 (4), pp. 1274-1286, 2007.
- [Marckmann 01] Marckmann G. Recommendations for the Ethical Development and Use of Medical Decision-Support Systems. *Medscape General Medicine*, 2001. Available at: <http://www.medscape.com/viewarticle/408143>
- [Marca 88] Marca D.A., et McGowan C.L. *SADT -- Structured Analysis and Design Technique*. New York, McGraw-Hill, 1988.
- [Martin 91] Martin J. *Rapid Application Development*. MacMillan, New York, 1991.
- [McDermid *et al.* 84] McDermid J., et Ripkin K. *Life cycle support in the ADA, environment*. Cambridge University Press, Cambridge, UK, 1984.
- [Médina 00] Médina R. *L'Extreme Programming*. [www.design-up.com](http://www.design-up.com), 2000.
- [Meinadier 02] Meinadier, J.-P. *Le métier d'intégration de systèmes*. Hermes Science Publications, Paris, France, 2002.
- [Mellor et Ward 85] Mellor S., et Ward P. *Structured Development for Real-Time Systems*. EnglewoodCliff, NJ : Prentice-Hall, 1985.
- [Meredith 03] Meredith R. Using the Internet to support the patient treatment decision process. *Journal of Decision Systems*, vol. 12 (2), pp. 209-224, 2003.
- [Merrer *et al.* 01] Merrer J., De Jonghe B., Golliot F., *et al.* Complications of femoral and subclavian venous catheterization in critically ill patients: a randomized controlled trial. *The Journal of the American Medical Association*, vol. 286 (6), pp.700-707, 2001.
- [Millet 90] Millet P. Coopération homme-machine : exemple de la téléopération. *Actes des Journées du GR Automatique*, Strasbourg, 1990.
- [Montmain *et al.* 03] Montmain J., et Penalva J.M. *Etat de l'art sur les théories de la décision et méthodologies de l'approche système*. Unité de recherche sur la complexité, Centre de recherche LGI2P, 2003.
- [Morley 06] Morlay C. *Management d'un projet système d'information : Principes, techniques, mise en œuvre et outils*. 5<sup>ème</sup> Edition, Édition Dunod, Paris 2006.

[Morton 71] Morton M.S. *Management Decision Systems; Computer-based support for decision making*. Boston, Division of Research, Graduate School of Business Administration, Harvard University, 1971.

[Mouloudi 07] Mouloudi A. *Intégration des besoins des utilisateurs pour la conception de systèmes d'information interactifs, application à la conception d'un système d'information voyageurs multimodal (SIVM)*. Thèse de doctorat, Université de Technologie de Compiègne, France, 2007.

[Mouloudi *et al.* 11] Mouloudi A., Morizet-Mahoudeaux P., Valentin A. RAMSES: A Method for the Design Process of Interactive Information Systems. *International Journal of Humam-Computer Interaction*, vol. 27(2), pp. 107-130, 2011.

[Moussa *et al.* 06] Moussa F., Kolski C., et Riahi M. Analyse des dysfonctionnements des systèmes complexes en amont de la conception des IHM : apports, difficultés, et étude de cas. *Revue d'Interaction Homme Machine (RIHM)*, vol. 7 (2), pp. 79-111, 2006.

[Myers 95] Myers B. A., User Interface Software Tools. *ACM Transactions on Computer-Human Interaction*, vol. 2 (1), pp. 64-103, 1995.

[Nanci *et al.* 01] Nanci D. et Espinasse B. *Ingénierie des systèmes d'information : MERISE, 2ème génération*, Vuibert, Paris, 2001.

[Nendaz 05] Nendaz M., Charlin, B., Leblanc, V., Bordage, G., *et al.* Le raisonnement clinique : données issues de la recherche et implications pour l'enseignement. *Pédagogie Médicale*, vol. 6 (4), pp. 235-254, 2005.

[Nielsen 93] Nielsen J. *Usability engineering*. Academic Press, Boston, 1993.

[Norman 86] Norman D.A. Cognitive engineering. In D.A. Norman & S.W. Draper (Eds.), *User centred system design: new perspectives on human computer interaction*, Hillsdale N.J., Elbraum, 1986.

[Nussbaumer 05] Nussbaumer A. *Hierarchy Browsers: Integrating Four Graph-Based Hierarchy Browsers into the Hierarchical Visualization System (HVS)*. Master's Thesis, Institute for Information Systems and Computer Media (IICM), Graz University of Technology, Austria, 2005.

[Nykanen 00] Nykanen P. *Decision Support System from a Health Informatics Perspective*. Thèse de doctorat, University of tampere, 2000.

- [Octo 02] Octo Technology. *Architecture de systèmes d'information*. Livre Blanc, 2002.
- [Ordonez *et al.* 01] Ordonez C., Omiecinski E., de Braal L., Santana C.A., Ezquerra N., *et al.* Mining constrained association rules to predict heart disease. *Proceedings of the First International Conference on Data Mining (ICDM)*, IEEE Computer Society Press, San Jose, USA, pp. 433-440, 2001.
- [Paetz *et al.* 04] Paetz J., Arlt B., Erz K., *et al.* Data quality aspects of a database for abdominal septic shock patients. *Computer Methods and Programs in Biomedicine*, vol. 75, pp. 23-30, 2004.
- [Paolaggi *et al.* 01] Paolaggi, J. B., Coste, J. *Le raisonnement médical : de la science à la pratique clinique*. Paris : Éditions ESTEM, 267 pages, 2001.
- [Paternò 99] Paternò, F. *Model Based Design and Evaluation of Interactive Applications*. Springer Verlag, Berlin, 1999.
- [Pearl 88] Pearl, J. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufman, San Mateo, California, 1988.
- [Pelayo *et al.* 09] Pelayo, S., Bernonville, S., Kolski, C., et Beuscart-Zéphir, M. Applying a human factors engineering approach to healthcare IT applications: example of a medication CPOE project. In J.G. Mcdaniel (Eds.), *Advances in Information Technology and Communication in Health*, IOS Press, pp. 334-339, 2009.
- [Pelleaumail 86] Pelleaumail P. *La méthode Axial*. Editions d'organisation, Paris, 1986.
- [Pena-Reyes et Sipper 00] Pena-Reyes C.A. et Sipper M. Evolutionary computation in medicine: an overview. *Artificial Intelligence in Medicine*, vol. 19, pp. 1–23, 2000.
- [Pirollo et Rao 96] Pirollo P. et Rao R. Table Lens as a Tool for Making Sense of Data. *Proceedings of the AVI '96 Workshop*, Gubbio, Italy, June 1996.
- [Pirolli *et al.* 03] Pirolli P., Card S. K. et Van Der Wege M. M. The effects of information scent on visual search in the hyperbolic tree browser, *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 10(1), pp.20-53, 2003.
- [Pittet *et al.* 93] Pittet D, Li N, et Wenzel RP., Association of secondary and polymicrobial nosocomial bloodstream infections with higher mortality. *European journal of clinical microbiology & infectious diseases*, vol. 12, pp. 813-819, 1993.

[Power 02] Power D.J. *Decision support systems: concepts and resources for managers*. Westport, Conn., Quorum Books, 2002.

[Quinlan 86] Quinlan J. R. Induction of Decision Trees. *Machine Learning*, vol. 1 (1), pp. 81-106, 1986.

[Quinlan 93] Quinlan J.R. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.

[Reiter 80] Reiter R. A Logic for Default Reasoning. *Artificial Intelligence*, vol. 13 (1-2), pp. 81-132, 1980.

[Rekimoto et Green 93] Rekimoto J. et Green, M. The Information Cube: Using Transparency in 3D Information Visualization. *In Proceedings of the Third Annual Workshop on Information Technologies & Systems*, pp. 125-132, 1993.

[Rettig 94] Rettig M. Prototyping for tiny fingers. *Communications of the ACM*, vol. 37 (4), pp. 21-27, 1994.

[Rialle 95] Rialle V. Cognition and decision in biomedical artificial intelligence: From symbolic representation to emergence. *Artificial Intelligence and Society*, vol. 9 (2), pp. 138-160, 1995.

[Riesbeck 89] Riesbeck C., Shank R. *Inside Case-Based Reasoning*. Lawrence Erlbaum, 1989.

[Ritchie 90] Ritchie S.G. A knowledge-based decision support architecture for advanced traffic management. *Transportation research, Part A*, vol. 24 (1), pp. 27-37, 1990.

[Robert 03] Robert J.M. Que faut-il savoir sur les utilisateurs pour réaliser des interfaces de qualité ? In G. Boy (Ed.), *Ingénierie cognitive, IHM et cognition*, Hermès Lavoisier, Paris, pp. 249-283, 2003.

[Robertson et al. 91] Robertson G.G., Mackinlay J.D., et Card, S.K. Cone Trees: Animated 3D Visualizations of Hierarchical Information. *Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology*, New Orleans, Louisiana, United States, ACM Press, pp. 189-194, 1991.

[Robertson et Mackinlay 93] Robertson G. G. et Mackinlay J. D. The Document Lens. *Proceedings of ACM Symposium on User Interface Software and Technology (UIST'93)*, pp. 189-194, 1993.

- [Roques 04] Roques P. *UML 2 par la pratique*. Editions Eyrolles, 311 pages, 2004.
- [Roques 05] Roques P. *UML 2 par la pratique : Etudes de cas et exercices corrigés*. 4<sup>ème</sup> Edition, Editions Eyrolles, Paris 2005.
- [Ross 77] Ross D.T. Structured Analysis: A language for communicating ideas. *IEEE Transactions on Software Engineering*, vol. 3 (1), pp. 16-34, 1977.
- [Royce 70] Royce W. *Managing the development of large software systems: Concepts and techniques*. WESCON technical papers, 1970.
- [Ruault 02] Ruault JR. UML and interactive systems, another step forward. In Kolski C., Vanderdonckt J. (Eds.), *Computer-aided Design of User Interface III*, pp. 243-256, 2002.
- [Rumbaugh et al. 99] Rumbaugh J., Jacobson I., et Booch G. *The Unified Modeling Language Reference Manual*. Addison-Wesley Professional, Har/Cdr edition, 1999.
- [Sabah 90] Sabah G. *L'intelligence Artificielle et le langage, représentation des connaissances*. Hermès, Paris, 358 pages, 1990.
- [Sage 91] Sage A.P. *Decision Support System Engineering*. Wiley, New York, 1991.
- [Scapin 86] Scapin D. L. *Guide Ergonomique de Conception des Interfaces Homme-Machine*, Rapport Technique INRIA n°77, 1986.
- [Scapin 89] Scapin D., et Pierret-Golbreich C. Towards a method for task description. *Actes de la conférence Working With Display Units (WWDU'89)*, pp. 371-380, 1989.
- [Schneider-Hufschmidt 93] *Adaptive User Interfaces*. Ed. by Schneider-Hufschmidt M., Kühme T., Malinkowski U., Netherlands: Elsevier Science Publishers, 1993.
- [Shneiderman 92] Shneiderman B. Tree Visualization with Tree-Maps: 2-D space-Filling Approach. *ACM Transactions on Graphics*, vol. 11(1), pp. 92-99, 1992.
- [Shneiderman 09] Shneiderman B. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Eds Addison Wesley; 5<sup>ème</sup> edition, 624 pages, 2009.
- [Silvestre et al. 96] Silvestre P., et Verlhac D. Stratégie de conception des systèmes d'information. *Techniques de l'ingénieur*, Paris, France, pp. 1-24, 1996.
- [Simon 77] Simon H.A. *The new science of management decision*. Prentice Hall, New Jersey, 1977.

- [Spivey 89] Spivey J.M. *The Z Notation: A Reference Manual*. Prentice-Hall, Englewood Cliffs, N. J., 1989.
- [Soula *et al.* 87] Soula G, Thirion X, Gullana JM, Navez I, Vialettes B, San Marco JL, Ledoray V. Une validation multicentrique du système expert PROTIS. In: *Comptes rendus des journées d'Avignon, Les systèmes experts et leurs applications*, pp. 239-255, 1987.
- [Stacey *et al.* 07] Stacey M., et McGregor C. Temporal abstraction in intelligent clinical data analysis: A survey. *Artificial Intelligence in Medicine*, vol. 39, pp. 1-24, 2007.
- [Stapleton 97] Stapleton J. *DSDM, Dynamic System Development Method*. Addison Wesley, Harlow, 1997.
- [Swartout 83] Swartout WR. XPLAIN: A System for Creating and Explaining Expert Consulting Programs. *Artificial Intelligence*, vol. 21, pp. 285-325, 1983.
- [Swayne *et al.* 98] Swayne DF, Cook D, et Buja A. XGobi: interactive dynamic data visualization in the X Windows System. *Journal of Computational and Graphical Statistics*, vol. 7(1), pp. 113-130, 1998.
- [Szolovits 85] Szolovits P. Types of Knowledge as Bases for Reasoning in Medical AI Programs. *Proceedings of the International Conference on Artificial Intelligence in Medicine*, Pavia, Italy, 1985.
- [Tardieu *et al.* 83] Tardieu H., Rochfeld S., et Coletti R. *La méthode MERISE*. Organisations, 1983.
- [Tardieu *et al.* 91] Tardieu H., Rochfeld S., et Colletti R. *La méthode Merise, principes et outils*. 2<sup>ème</sup> édition, Editions d'Organisation (tome 1), Paris, 1991.
- [Thuraisingham 00] Thuraisingham B. A Primer for Understanding and Applying Data Mining. *IEEE IT Professional*, vol. 2 (1), pp. 28-31, 2000.
- [Turban 93] Turban E. *Decision Support and Expert Systems*. Macmillan, New York, 1993.
- [Turban *et al.* 95] Turban E., et Walls J.G. Executive information systems - a special issue. *Decision Support Systems*, vol. 14 (2), pp. 85-88, 1995.
- [Turban *et al.* 98] Turban E., et Aronson J.E. *Decision Support Systems and Intelligent Systems*. Upper Saddle River, Prentice-Hall, New Jersey, 1998.
- [Turban *et al.* 08] Turban E., Aronson J. E , Liang T., Sharda R. *Decision Support and Business Intelligence Systems*. Pearson Education, 2008.

- [van Wijk et van de Wetering 99] van Wijk J. J. et van de Wetering H. Cushion Treemaps: Visualization of Hierarchical Information, *Information Visualization*, Washington, USA, pp. 73-78, 1999.
- [van Ham et van Wijk 03] van Ham F. et van Wijk J. J. Beamtrees: compact visualization of large hierarchies, *Information Visualization*, vol. 2(1), pp. 31-39, 2003.
- [Veenstra *et al.* 99] Veenstra D.L., Saint S., Sullivan S.D. Cost-effectiveness of antiseptic-impregnated central venous catheters for the prevention of catheter-related bloodstream infection. *The Journal of the American Medical Association*, vol. 282, pp. 554-560, 1999.
- [Verduijn *et al.* 07] Verduijn M., Sacchi L., Peek N., Bellazzi R., de Jonge E., et de Mol B.A.J.M. Temporal abstraction for feature extraction: A comparative case study in prediction from intensive care monitoring data. *Artificial Intelligence in Medicine*, vol. 41, pp. 1-12, 2007.
- [Vidal 06] Vidal S. *Visualisation de l'information : un panorama d'outils et de méthodes*. Dossier de synthèse, 2006.
- [Vickoff 09] Vickoff J-P. *Méthode AGILE, les meilleures pratiques Compréhension et mise en œuvre*. Editions QI, 2009.
- [Victor 97] Victor S. *L'Informatique décisionnelle*. *Hermes Sciences Publicat.* 126 pages, 1997.
- [Villmann *et al.* 00] Villmann T., Wieland H., et Geyer M. Data mining and knowledge discovery in medical applications using self-organizing maps. *Proceedings of the International Symposium on Medical Data Analysis (ISMDA)*, LNCS vol. 1933, Frankfurt am Main, Germany, Springer, Berlin, pp. 138-151, 2000.
- [Watson *et al.* 93] Watson H.J., et Walls J. Executive information systems. *Proceeding of the Twenty-Sixth Hawaii International Conference on System Sciences*, vol. 3, pp. 204-205, 1993.
- [Watson *et al.* 98] Gray P., et Watson H.J. *Decision Support in the datawarehouse*. Prentice Hall PTR, 1998.
- [Wesphal *et al.* 98] Wesphal C., et Blaxton T. *Data Mining Solutions*. John Wiley, New York, 1998.
- [Williges 87] Williges R.C., Williges B.H., Elkerton J., Software Interface Design. In Salvendy G. (Ed.), *Handbook of Human factors*, pp. 1416-1449, New York, Wiley, 1987.

[Zadeh 65] Zadeh L.A. Fuzzy sets. *Information and Control*, vol. 8 (3), pp. 338–353, 1965.

[Zhou 03] Zhou Z. Three perspectives of data mining. *Artificial Intelligence*, vol. 143 (1), pp. 139-146, 2003.

[Zhu *et al.* 07] Zhu X., Davidson I. *Knowledge Discovery and Data Mining: Challenges and Realities*. Hershey, New York, pp 163-189, 2007.

[Zighed *et al.* 02] Zighed D., Rakotomalala R. Data Mining. *Techniques de l'ingénieur*, Editions Techniques de l'Ingénieur, vol. H3 744, pp. 1-26, 2002.

# Annexe



## Annexe A. Techniques de visualisation de l'information<sup>22</sup>

### A.1. Introduction

La visualisation d'information s'inspire de divers thématiques de recherche, la visualisation de l'information, selon [Kapusova 04], combine des aspects de la visualisation scientifique, des interfaces homme-machine (*human-computer interfaces*), de l'exploitation de données (*data mining*), de l'imagerie et des graphiques. Pour [Fekete 05], la visualisation de l'information s'est détachée de trois domaines connexes : l'interaction Home-machine, l'analyse statistique et la cartographie, et la visualisation scientifique.

### A.2. Classification des techniques de visualisation

Il existe actuellement de nombreuses techniques de visualisation et d'interaction qui ont été définies. Le but essentiel de cette section consiste à dresser un panorama de ces techniques. Pour présenter la majorité de techniques de visualisation décrites dans la littérature, nous adoptons la classification présentée par le tableau A.1. Ces techniques sont résumées ci-après.

**Tableau A.1** : classification des techniques de visualisation

<b>Techniques de visualisation</b>		
<b>Approche technique : orientée structures de données</b>		
Données linéaires	Tables de données	
	Murs fuyants ou en perspective	
	Frises chronologiques	
Données multidimensionnelles	Nuages de points	
	Système de polyligne	
Hiérarchies (structures arborescentes)	Approche matricielle	
	Approche par listes indentées	
	Approche par diagrammes	
	Approche surfacique	Cartes d'arbres (treemaps)
		Arbres circulaires (slices)
	Approche hyperbolique	
	Approche 3D ou conique	

<sup>22</sup> Cette synthèse a fait l'objet d'un article publié dans une conférence internationale [Ltifi *et al.* 09].

	Approche paysage
Réseaux	Réseaux de liens hypertextes
	Réseaux de co-citation
	Réseaux sociaux
Modèle vectoriel	Approche de similarités
	Approche de pertinence
<b>Approche interactive : orientée utilisateur</b>	
Projection dynamique	
Filtrage interactif	
Zoom interactif	
Distorsion interactive	
Liens interactifs et brossage	

### A.3. Description des techniques

#### A.3.1 Approche technique

##### A.3.1.1 Données linéaires

1. **Les tables de données** : les techniques concernées permettent une visualisation de larges ensembles de données sous forme tabulaire [Pirollo et Rao 96].
  - Avantages : on peut avoir une vue d'ensemble, filtrer, détailler, réarranger en sous-groupes, trier les données et voir les relations entre les objets [Vidal 06].
  - Inconvénients : on ne peut extraire ou exporter les données d'intérêt une fois identifiées [Vidal 06].

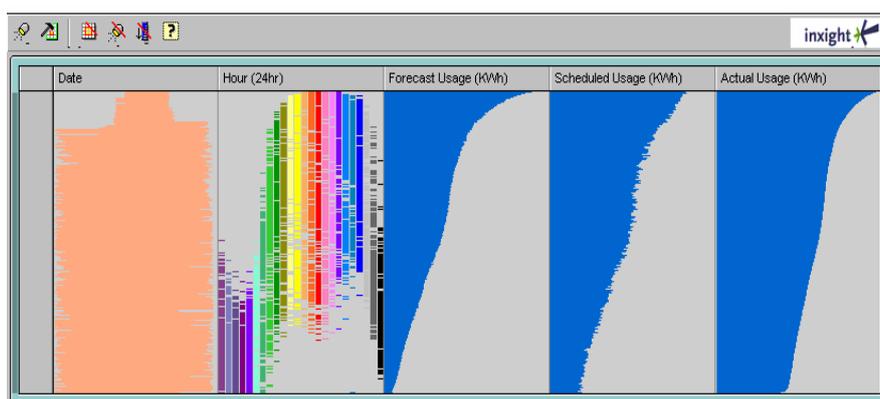


Figure A.1 : table de données « Table lens » [Pirollo et Rao 96]

2. **Les murs fuyants ou en perspective** : les données sont présentées chronologiquement sur plusieurs panneaux (trois en général).

Le panneau central se trouve en premier plan, les autres panneaux de droite et de gauche sont en arrière plan, ce qui donne un effet de perspective correspondant à la notion de

"focus+context". Ce qui est en avant correspond à une période donnée, la partie gauche correspond à ce qui est antérieur à cette période et la partie droite à ce qui est postérieur [Robertson et al. 91].

- Avantages : il y a un suivi temporel des événements. Ainsi on peut avoir une meilleure résolution sur les périodes et les événements d'intérêt tout en gardant une vue d'ensemble [Vidal 06].
- Inconvénients : il faut savoir où chercher dans la zone temporelle [Vidal 06].

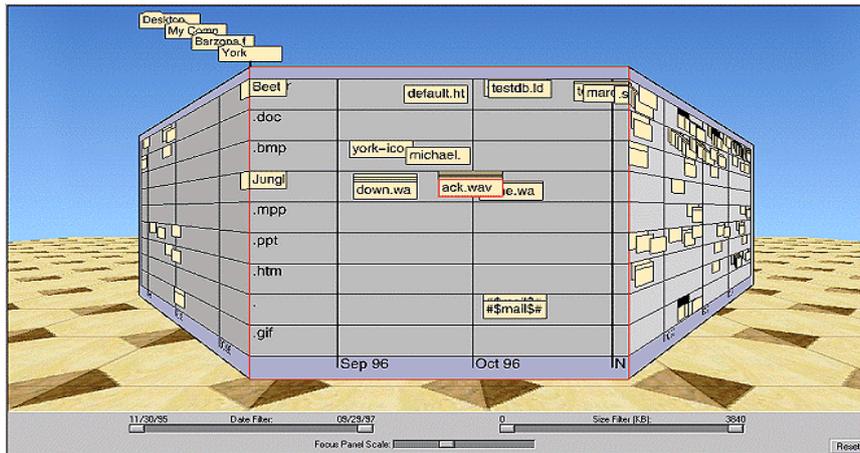


Figure A.2 : mur fuyant [Robertson et al. 91]

3. **Frises chronologiques** ("Times lines") : Ce type de technique permet d'avoir un suivi temporel selon plusieurs points de vue. Ce sont des visualisations d'enregistrements ou documents ou événements selon un ordre chronologique [Freeman 97].

- Avantages : il est possible d'effectuer un suivi temporel des événements.

### A.3.1.2 Données multidimensionnelles

Ce sont des ensembles de données qui se composent de plus de trois attributs. Ils ne permettent pas une visualisation simple à deux ou trois dimensions [Keim 02]. L'objectif des techniques de visualisation de ce type de données est de trouver des transformations géométriques permettant de réduire le nombre de dimensions à 2 ou 3 afin de faciliter l'interprétation et l'analyse par l'utilisateur final.

Parmi ces techniques, on parle de :

1. **Nuages de points** : il s'agit de représenter des données avec des nuages de points [Swayne et al. 98].

- Avantages : possibilité de filtrage des informations.
- Inconvénients : la multiplicité des points entraîne une difficulté de lecture.

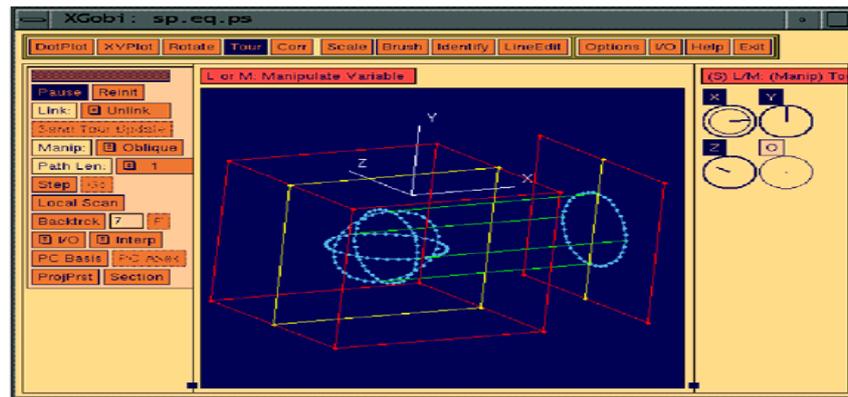


Figure A.3 : nuages de points [Swayne et al. 98]

2. **Système de polyligne** : il s'agit de représenter chaque élément de la base par une polyligne dans un système à N axes, N étant le nombre de dimensions de la donnée. Chacun des axes représente l'une des dimensions des données. Parmi les outils des plus connus, on peut citer le parallel coordinates [Inselberg et Dimsdale 90].

- Avantages : ce diagramme permet de visualiser simultanément un nombre de dimensions important. Il représente des modèles des tendances et des corrélations. Le but de ce diagramme est de découvrir un phénomène qui est difficilement détectable [Vidal 06].
- Inconvénients : cette visualisation nécessite une certaine familiarisation. Il faut jouer sur tous les paramètres pour faire apparaître des vues pertinentes [Vidal 06].

Parallel Coordinates Visualization Applet

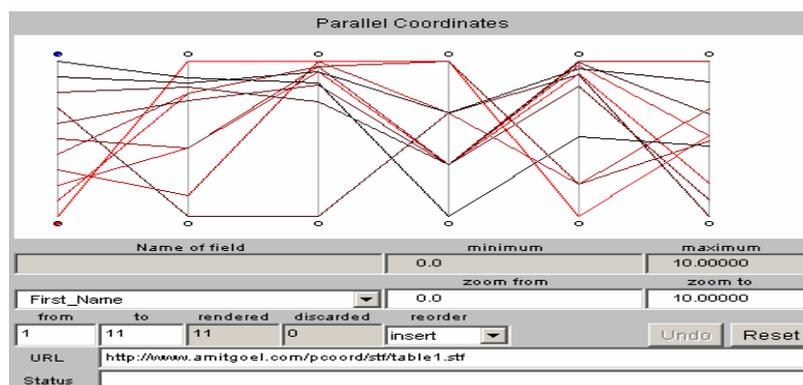


Figure A.4 : système de polyligne, coordonnées parallèles [Inselberg et Dimsdale 90]

### A.3.1.3 Hiérarchies

Les hiérarchies sont des informations organisées hiérarchiquement via des structures arborescentes avec des liens ou connexions entre nœuds parents et nœuds enfants (arbres, graphes, etc.). Ce type de visualisation est le plus représenté du fait de la simplicité de

compréhension par l'utilisateur, d'une part, et surtout du fait qu'il s'agisse du système d'organisation des informations le plus éprouvé dans le monde, d'autre part [Nussbaumer 05].

La visualisation hiérarchique présente une plus grande diversité des approches que dans les autres types de visualisations :

**1. Approche matricielle :** Il s'agit de matrices de booléens (initialisés à faux) où les sommets de l'arborescence sont disposés en lignes et en colonnes. Pour chaque tuple du graphe, l'intersection de la ligne et de la colonne correspondante contient la valeur « vrai ». Pour augmenter la richesse de la représentation, il est possible de remplacer les booléens par des valeurs numériques et ainsi, il devient possible de représenter des relations binaires valuées [Ghoniem et al. 04].

- Avantages :
  - Cette approche permet d'éviter les problèmes d'occlusion [Ghoniem *et al.* 04]
  - Ghoniem et Fekete [Ghoniem et Fekete 02] montrent que la représentation par matrices d'adjacence permet de déceler clairement la structure sous-jacente d'un graphe.
  - Les matrices permettent d'identifier des modèles (structures sous-jacentes) dans l'espace sous réserve d'appliquer un certain nombre de permutations entre les lignes et les colonnes [Bertin 99]. Cette propriété est particulièrement intéressante pour le datamining.
  - Les matrices permettent notamment de passer à l'échelle et de traiter de très grands graphes [Abello et Ham 04].



**Figure A.5 :** représentation matricielle [Abello et Ham 04]

- Inconvénient : les représentations matricielles sont peu utilisées en raison de [Ghoniem *et al.* 04] :

- La difficulté d'appréhender chaque entité indépendamment des autres, et
- La difficulté de représenter de nombreuses relations non binaires avec une seule matrice.

**2. Approche par listes indentées :** Les informations sont disposées dans une liste. La profondeur d'indentation est proportionnelle à la profondeur du document dans la hiérarchie [Robertson et Mackinlay 93].

- Avantage : simplicité de compréhension et facilité d'utilisation [Vidal 06].
- Inconvénient : elles ne parviennent pas à donner une vue d'ensemble de l'espace pour des arborescences de grande taille : l'arbre n'est jamais entièrement visible car le nombre de nœuds affichés simultanément est forcément limité par l'espace disponible [Hascoet et Beaudouin-Lafon 01] [Vidal 06].

### **3. Approche par diagrammes**

L'arbre est représenté par un diagramme<sup>23</sup> dans un espace en 2D constitué de nœuds et de liens. La principale difficulté de cette approche est de déterminer le meilleur placement des entités de l'arborescence. Pour cela, de nombreux algorithmes existent [Herman et al. 99]. La plupart de ces algorithmes ont pour objectif de mettre en valeur la structuration hiérarchique de l'espace et/ou de l'optimiser. Parmi les outils de cette approche, on peut citer le *Space Tree* [Hascoet et Beaudouin-Lafon 01].

- Avantages : l'utilisation de l'espace 2D pour le placement des nœuds est optimisée par rapport aux listes indentées. Pour cette raison, cette approche est utile pour la visualisation d'arbres de petite et moyenne taille puisqu'on voit toute la hiérarchie [Vidal 06].
- Inconvénients : la difficulté de définir un algorithme idéal de placement d'arbre [Hascoet et Beaudouin-Lafon 01].

---

<sup>23</sup> Les diagrammes sont aussi nommés représentations « nœud-lien »

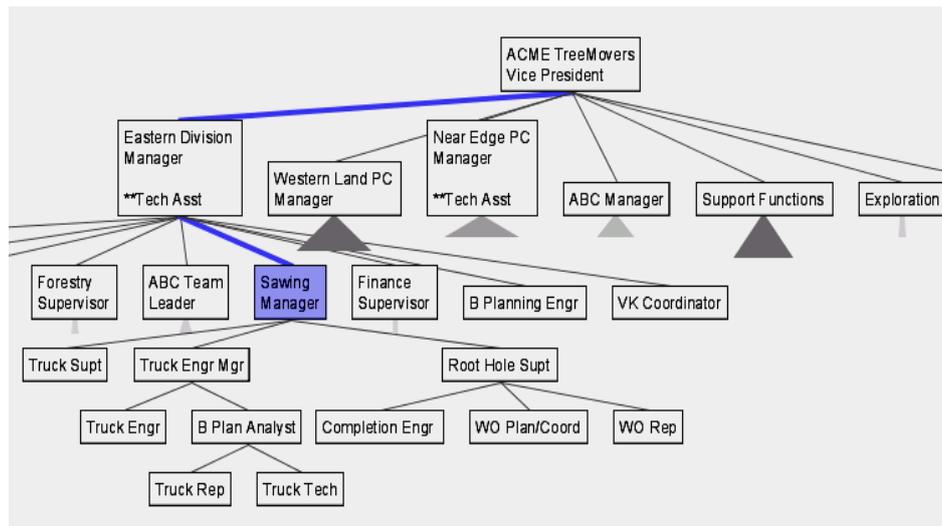


Figure A.6 : exemple d'outils d'approche par diagrammes : Space Tree [Hascoet et Beaudouin-Lafon 01]

#### 4. Approche surfacique (représentation par pavage)

a. **Cartes d'arbres (treemaps) et ses variantes** [Shneiderman 92] [Hascoet et Beaudouin-Lafon 01] : il s'agit de visualiser un arbre en découpant une surface donnée (en général rectangulaire) proportionnellement à chaque sous arbre. Chaque rectangle représente un nœud ou feuille de l'arbre et sa surface est proportionnelle à son importance dans l'arbre.

Parmi les évolutions des treemaps, ayant pour objectif de faciliter la perception de la structure, on peut citer :

- « Cushion Treemaps » : une évolution des treemaps qui rajoute une impression de relief sur les rectangles [van Wijk et van de Wetering 99]
- « Beamtrees » : une autre évolution des treemaps en 3D qui représente les nœuds par des cylindres et place chaque sous-nœud perpendiculairement par rapport à son père [van Ham et van Wijk 03].

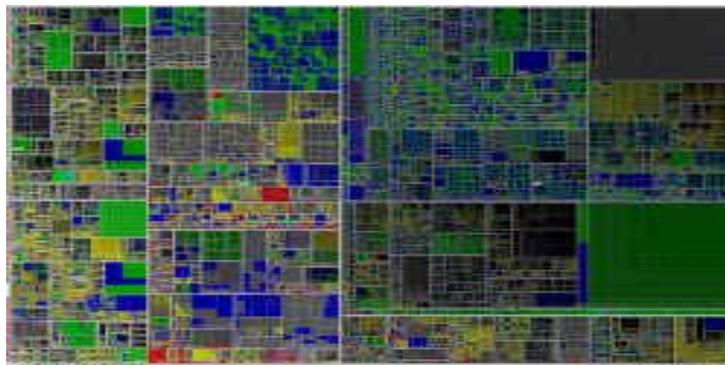
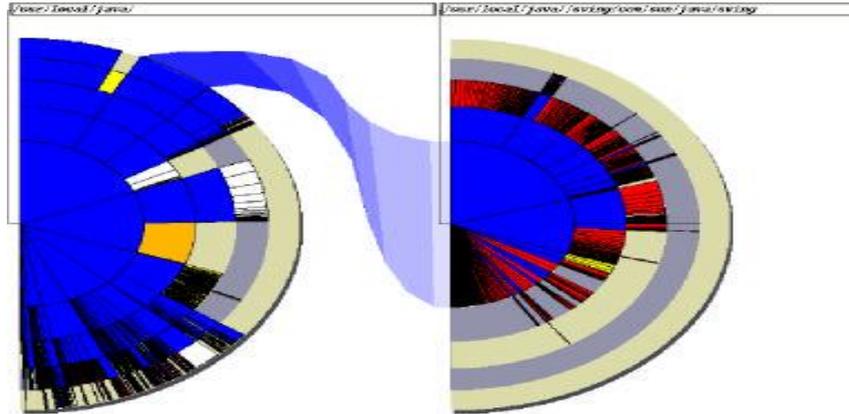


Figure A.7 : treemaps [Hascoet et Beaudouin-Lafon 01]

- b. Arbres circulaires (slices) :** on les appelle encore les disques d'informations [Andrews et Heidegger 98], la hiérarchie est dépliée radialement (chaque niveau de l'arbre est représenté par un anneau que les nœuds se partagent et chaque nœud correspond à une portion du disque dont l'angle est proportionnel à son importance dans l'arbre), le sommet de la hiérarchie étant au centre. Plus on s'éloigne du centre, plus on descend dans la hiérarchie. Les angles et couleurs correspondent à des attributs des données.



**Figure A.8 :** arbres circulaires [Andrews et Heidegger 98]

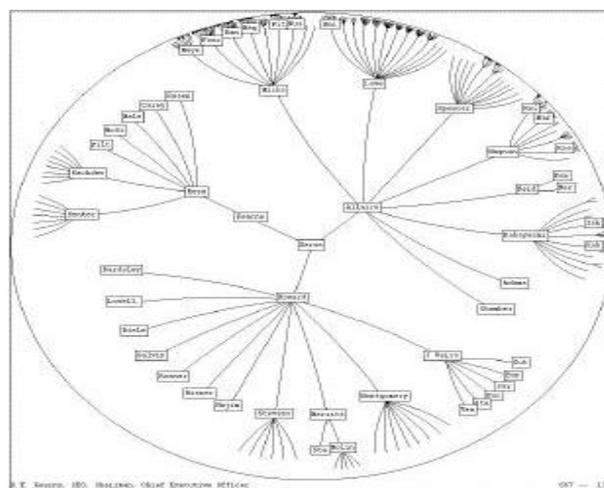
Cette approche présente les avantages et les inconvénients suivants [Hascoet et Beaudouin-Lafon 01] :

- Avantages :
  - Permet de visualiser de grands arbres tout en évitant les problèmes d'occlusion
  - Visualisation rapide de l'ensemble de l'arbre pour déceler certains éléments particuliers.
  - Utilisation des couleurs et des motifs de fonds en vue de représenter certaines caractéristiques des nœuds.
  - Les arbres circulaires permettent de mieux percevoir la structure arborescente des éléments que les Treemaps et de faciliter l'interprétation ensembliste de la structure.
- Inconvénients :
  - plus un élément est situé profondément dans la hiérarchie, plus la surface du rectangle correspondante est petite donc très peu visible dans la vue d'ensemble.

- Pour les arbres circulaires, l'affichage de texte à l'intérieur d'un nœud est quasiment impossible car les nœuds ont la forme de secteurs d'anneaux.
- Pour les treemaps, le texte est invisible lorsque les rectangles sont très étirés<sup>24</sup>.

**5. Approche hyperbolique** : c'est une approche radiale qui se base sur une géométrie non euclidienne [Lamping et al. 95]. Les nœuds de l'arbre sont placés de la même manière sur des cercles concentriques dont le rayon est calculé proportionnellement à la profondeur du nœud dans l'arbre (les plus centrés étant les plus proches de la racine). Elle se caractérise par le fait que le périmètre des cercles hyperboliques croît exponentiellement avec leur rayon ce qui permet de visualiser des arbres contenant un grand nombre de feuilles.

- Avantages :
  - les techniques de cette approche permettent de mieux visualiser les nœuds en périphérie par rapport à un diagramme radial [Vidal 06].
  - Les arbres hyperboliques ont la capacité d'aider les utilisateurs à « deviner et anticiper » les sous-éléments de chaque élément [Pirolli *et al.* 03].
- Inconvénients : l'utilisateur non entraîné doit faire des efforts cognitifs importants pour reconstituer le contexte suite à une application d'une technique de visualisation interactive ("fisheye") [Hascoet 04].



**Figure A.9** : arbre hyperbolique [Lamping et al. 95]

---

<sup>24</sup> Il existe des variantes des treemaps permettant d'optimiser les proportions des rectangles représentant les nœuds de façon à pouvoir afficher des informations textuelles dans ces rectangles.

**6. Approche 3D ou conique :** l'arbre 3D permet d'avoir une vue d'ensemble plus réaliste avec un effet de perspective [Robertson et al. 91]. Il s'agit d'associer à chaque nœud le sommet d'un cône, et on arrange ses fils autour de la base circulaire du cône qui est généralement transparent. L'arbre conique peut être horizontal ou vertical.

- Avantages :

- Les informations possédant trois dimensions sont facilement visualisables.
- Cette approche permet d'afficher des arbres de grande taille de façon intuitive [Hascoet et Beaudouin-Lafon 01].
- "context + focus"<sup>25</sup> [Vidal 06] : cette approche permet à l'utilisateur de garder une vue d'ensemble de l'arbre ainsi que pouvoir visualiser les cônes de même niveau situés en arrière plan. Chaque cône d'intérêt est représenté avec les détails d'un nœud pour garder la vision de l'ensemble.

- Inconvénients :

- Un premier inconvénient des techniques de cette approche est lié à l'usage de la 3D et notamment aux problèmes d'occlusion (dû à l'enchevêtrement des liens, dès que la taille du graphe ou la densité des liens augmente un peu), de manipulation et d'orientation qui lui sont inhérents [Hascoet et Beaudouin-Lafon 01].
- Il reste difficile de bien distinguer les détails. En effet, les cônes étant les uns devant les autres, même avec la transparence [Vidal 06], ce qui rend difficile pour l'utilisateur d'explorer l'arborescence visuellement et d'interagir avec ses éléments constitutifs.

---

<sup>25</sup> Contexte : une vue globale / Focus : une vue plus locale / Contexte+focus : combinaison des vues globales et des vues locales

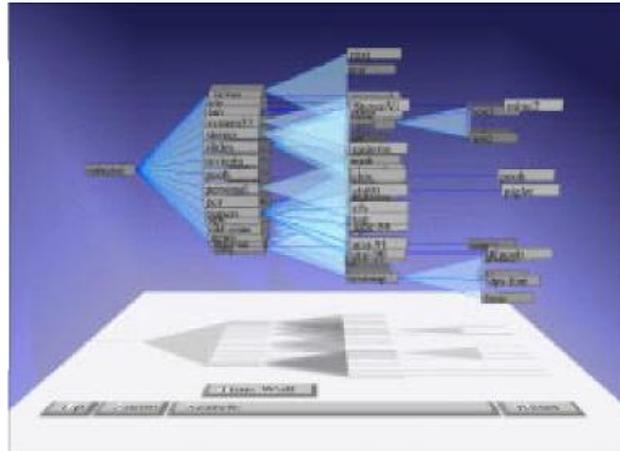


Figure A.10 : arbre conique horizontal [Robertson et al. 91]

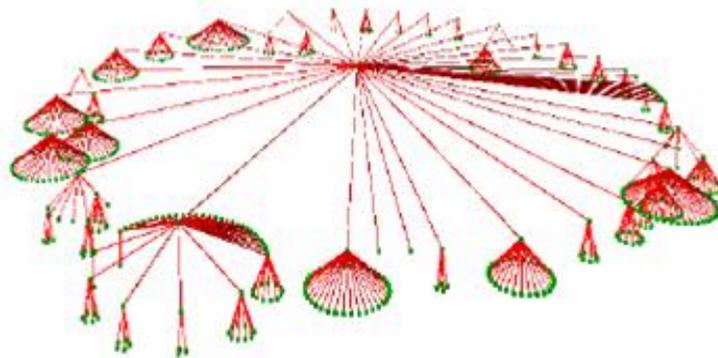


Figure A.11 : arbre conique vertical [Robertson et al. 91]

## 7. Approche paysage (Landscape)

Les nœuds de la hiérarchie sont placés dans un paysage virtuel en 3D où les éléments similaires sont placés proches physiquement l'un de l'autre. Les nœuds sont représentés par des attributs physiques tels que la hauteur du bloc ou la couleur.

Parmi les exemples de cette approche :

- *Les cubes d'informations* : ils proposent une représentation de hiérarchie en 3D par imbrication de cubes [Rekimoto et Green 93].
- *Visualisation botanique* : représentation de hiérarchies avec une métaphore virtuelle en 3D dont les feuilles de la hiérarchie sont représentées par des fruits ce qui permet d'enrichir la représentation avec l'exploitation de nombreuses variables visuelles [Kleinberg *et al.* 01].
- Avantages :
  - Représentation qui met en valeur la structure hiérarchique

- vision de l'arbre dans son entier et survol dynamique permettant divers points de vue ainsi qu'une prise de connaissance des caractéristiques des nœuds (attributs, couleurs, etc.) [Vidal 06].
- Inconvénients : le survol d'un paysage en 3D peut nécessiter des ressources importantes en mémoire [Vidal 06]. Et comme toutes les représentations en 3D, l'utilisateur est confronté aux problèmes d'occlusion.

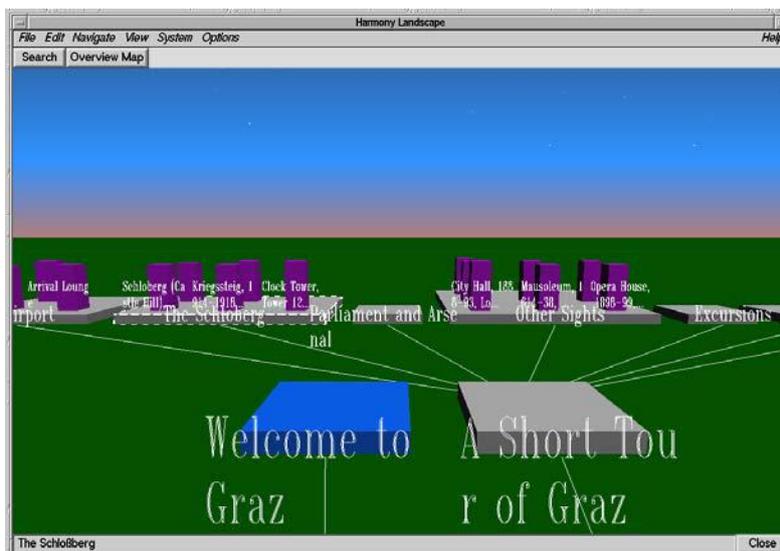


Figure A.12 : cubes d'informations [Rekimoto et Green 93]

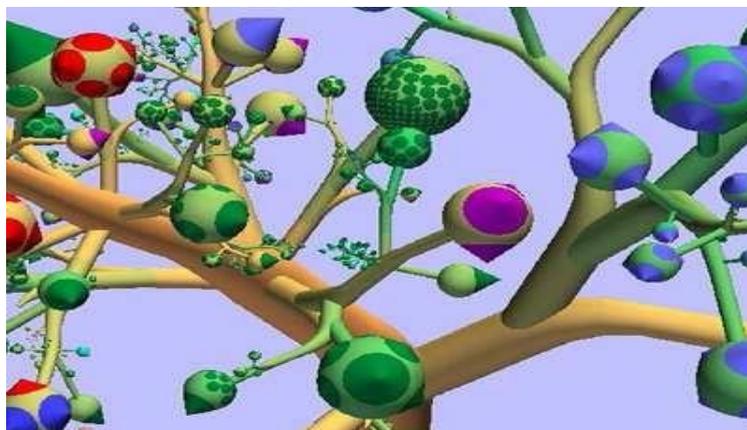


Figure A.13 : visualisation botanique [Kleinberg *et al.* 01]

#### A.3.1.4 Réseaux

Il s'agit de visualiser des ensembles de données sous forme de réseaux. Ces réseaux d'objets ne sont pas obligatoirement d'ordre hiérarchique (par exemple les réseaux de co-citations). Ce mode de visualisation a pour objectif de faire gagner en perspicacité à la lecture d'une structure qui peut se composer de nombreuses données élémentaires [Humbert 06].

Les travaux d'Auber *et al.* [Auber et al. 03], ceux de Newman et Girvan [Girvan et Newman 02] ou de Boutin [Boutin 05] proposent des méthodes permettant d'identifier la structure de communautés d'un réseau.

On distingue trois types de réseaux :

- a) **Réseaux de liens hypertextes** : réseaux de pages web [Chen 05]
- b) **Réseaux de co-citation** : réseaux de documents [Chen 05]
- c) **Réseaux sociaux** : réseaux de personnes et organisateurs d'idées ou de connaissances en fonction de concepts [Heer et Boyd 05].

- Avantages :

- souple au niveau de l'interactivité.
- Représentation qui permet de révéler des structures sous-jacentes dans les données et plus particulièrement les composantes connexes et fortement connexes [Heer et Boyd 05].
- La visualisation des réseaux de co-citation permet d'appréhender l'importance d'un domaine scientifique, détecter les principaux acteurs de ce domaine ainsi que les tendances et les émergences grâce à la prise en compte de la dimension temporelle [Vidal 06].

- Inconvénients :

- Pour certains outils, on ne peut pas aller au-delà de deux niveaux de hiérarchie.
- Les réseaux de co-citations volumineux sont vite illisibles [Vidal 06].

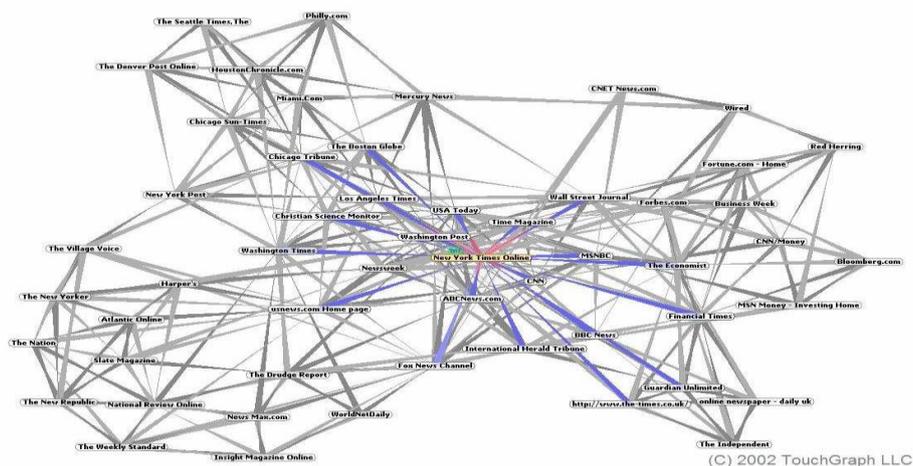


Figure A.14 : exemple de réseau de liens hypertextes [Chen 05]

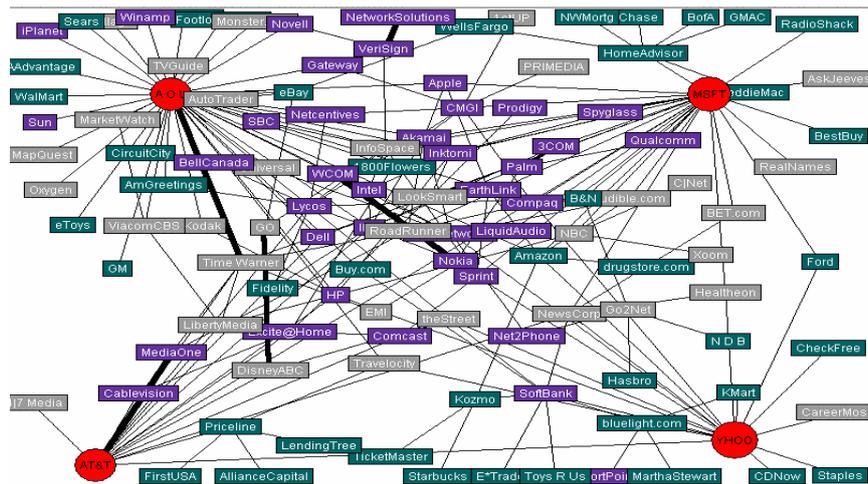


Figure A.15 : exemple de réseau social [Heer et Boyd 05]

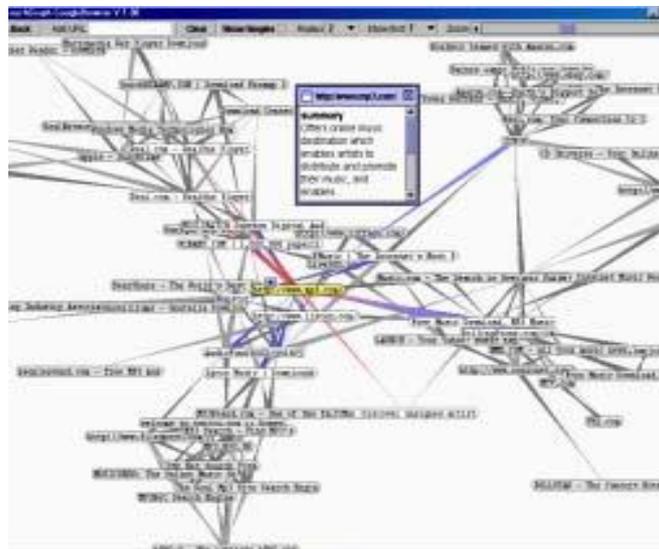


Figure A.16 : exemple de réseau de co-citation [Chen 05]

### A.3.1.5 Modèle vectoriel

La représentation des informations se fait par des vecteurs (sortes d'index). Les coordonnées de ces vecteurs dépendent du modèle utilisé. Des procédés de calcul des coordonnées issus de ce modèle permettent de définir des mesures de similarités ou de pertinence entre différents « morceaux » d'information [Hascoet et Beaudouin-Lafon 01].

**1. Approche de similarités :** Il s'agit de placer les données similaires – au sens de la mesure de similarité – de telle sorte qu'ils soient voisins. Le placement des nœuds fait appel à des algorithmes comme l'algorithme de Voronoï [Andrews et al. 04]

- Avantages [Vidal 06] :
  - Bonne vue d'ensemble de la représentativité des informations.
  - Classification automatique et non supervisée des informations.

- Inconvénients [Vidal 06] :
  - Les graphiques générés peuvent être déroutants.
  - Navigation aveugle avant de trouver la zone de l'information d'intérêt.

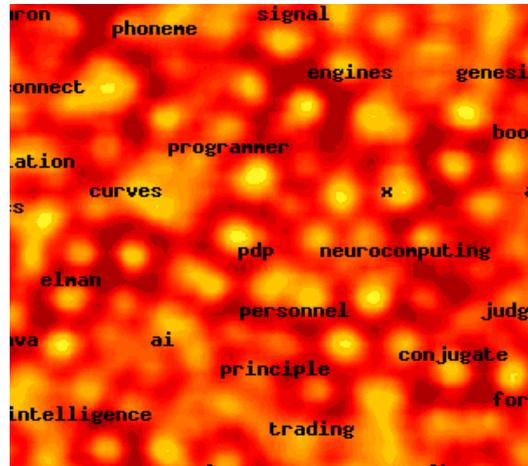


Figure A.17 : approche de similarité [Andrews et al. 04]

## 2. Approche de pertinence

Son principe consiste à utiliser les points d'un espace 2D ou 3D pour représenter les informations ; ces points sont placés dans l'espace en fonction de points de référence. Le placement de ces points est calculé de telle sorte que la distance euclidienne des points-informations aux points-références illustre la pertinence des informations par rapport aux références [Hascoet et Beaudouin-Lafon 01]

- Avantages : navigation facile puisque on est directement sur ce qui nous intéresse [Vital 06].
- Inconvénients : certaines techniques utilisant cette approche sont difficiles à appréhender [Vital 06].

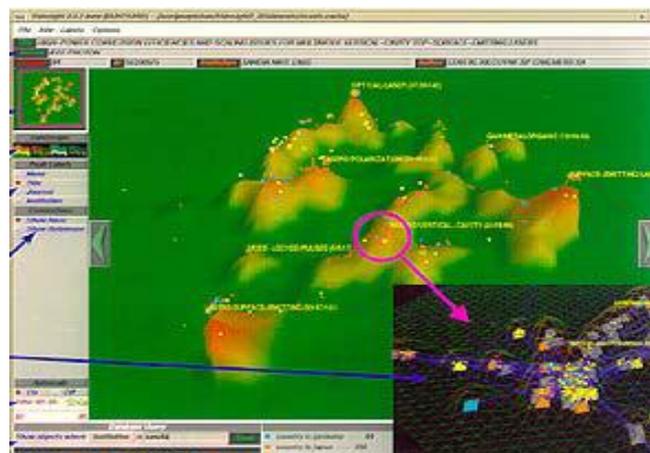


Figure A.18 : approche de pertinence [Hascoet et Beaudouin-Lafon 01]

### **A.3.2 Approche interactive**

Le concept de visualisation de l'information n'a de sens que dans la simplification qu'elle apporte à l'utilisateur dans sa démarche d'exploration et de recherche d'information dans de grands ensembles de données [Humbert 06]. De ce fait, pour une exploration effective de données, il semble nécessaire d'utiliser certaines techniques d'interaction et de distorsion. Ces dernières sont développées par Daniel Keim [Keim 02] qui distingue d'une part entre :

- Les techniques d'interaction permettent à l'analyste de données d'interagir directement avec les visualisations et de changer dynamiquement les visualisations selon les objectifs de l'exploration.
- Les techniques de distorsion aident dans le processus d'exploration de données par la concentration en détails tout en préservant une vue d'ensemble de données.

Et d'autre part, entre les qualificatifs "dynamique" et "interactif" selon que les modifications apportées à la visualisation des données soient effectuées automatiquement ou manuellement (l'utilisateur final pouvant agir directement) :

#### **1) Projection dynamique**

Il s'agit de changer dynamiquement les projections afin d'explorer un ensemble de données multidimensionnelles [Keim 02].

- Avantages : le traitement dynamique des données à visualiser.
- Inconvénients : cette technique présente une difficulté dans le traitement des données de haute dimensionnalité [Keim 02].

#### **2) Filtrage interactif**

Il s'agit de diviser interactivement l'ensemble de données dans des segments et de se concentrer sur les sous-ensembles intéressants. Ceci peut être fait en choisissant directement le sous-ensemble désiré (browsing) ou en spécifiant des propriétés du sous-ensemble désiré (querying) [Keim 02].

- Avantages : approche de visualisation graduée, permettant d'amplifier dynamiquement les détails autour d'un centre d'intérêt tout en gardant le contexte global de la vue [Hascoet et Beaudouin-Lafon 01].
- Inconvénients : exploration difficile des données [Hascoet et Beaudouin-Lafon 01].

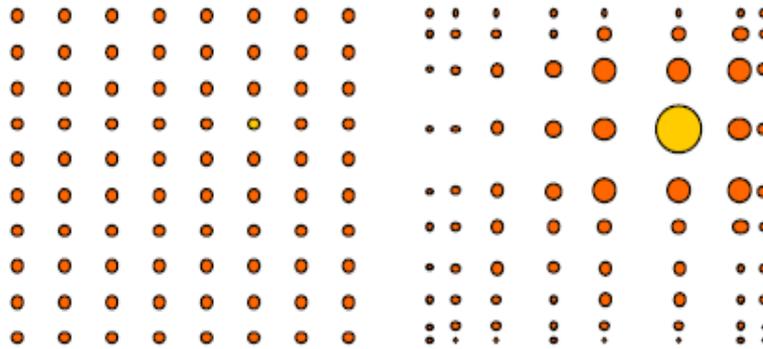


Figure A.19 : filtrage dynamique : application de fisheye [Keim 02]

### 3) Zoom interactif

L'idée de cette technique est de partir d'une vue globale des données et de permettre l'affichage des détails selon différentes résolutions [Keim 02].

- Avantages : le temps de pointage dans une interface zoomable suit exactement la même loi de Fitts que pour le pointage normal sur écran [Guiard et al. 01]. La navigation dans un espace zoomable est donc très efficace [Hascoet et Beaudouin-Lafon 01].
- Inconvénients : risque de se perdre dans une interface zoomable contenant une grande quantité d'information [Hascoet et Beaudouin-Lafon 01].

### 4) Distorsion interactive

L'idée de base est de montrer des parties de données avec un niveau élevé de détail tandis que d'autres sont montrées avec un niveau moindre de détail. Les techniques courantes de distorsion sont des déformations hyperboliques et sphériques qui sont souvent employées sur des hiérarchies ou des graphiques [Keim 02]. Ces techniques utilisent une sorte de loupe (fisheye) déformante que l'on promène à son gré sur l'ensemble des données.

- Avantages : les distorsions permettent de favoriser un focus tout en conservant le contexte [Boutin 05].
- Inconvénients : les distorsions interactives qui s'appliquent aux arborescences radiales et/ou hyperboliques sont relativement perturbatrices pour un œil non averti ; ce qui fait perdre à l'utilisateur le contexte de l'espace d'information [Vidal 06].

### 5) Liens interactifs et brossage

L'idée est de combiner des méthodes différentes de visualisation pour surmonter les imperfections des techniques simples [Keim 02].

- Avantages : cette approche permet de surmonter les défauts des techniques utilisées seules [Keim 02].

#### **A.4. Conclusion**

Un ensemble représentatif de techniques de visualisation a été présenté. Néanmoins l'efficacité d'une technique de visualisation est difficile voire impossible à évaluer de façon absolue. Tout au plus peut on espérer l'évaluer par rapport à un contexte particulier, en prenant en compte la tâche de l'utilisateur et les caractéristiques des données. Elles offrent un ensemble de perspectives relativement aux SIAD/ECD, mais de nombreuses études sont encore à mener.

## **Résumé**

Ce mémoire contribue au domaine de l'interaction homme-machine (IHM), en proposant un cadre méthodologique global en vue de la conception et la réalisation de systèmes interactifs d'aide à la décision basés sur l'extraction de connaissances à partir de données (SIAD/ECD). Dans ce travail, nous mettons l'accent sur les aspects humain et itératif du développement des SIAD, et en tenant compte des spécificités des systèmes d'ECD.

Le cadre méthodologique proposé consiste à étendre les activités du processus unifié sous l'angle de l'IHM pour le développement des différents modules d'un SIAD/ECD. Cette méthodologie, nommée processus unifié étendu, a été appliquée en vue de démontrer ou valider ses principes fondateurs sur une étude de cas dans le domaine médical, relative à la prise de décision dynamique pour la lutte contre les infections nosocomiales (SIADDM/ECD) dans un service de réanimation d'un hôpital situé à Sfax, Tunisie.

Une évaluation globale de la démarche suivie d'un ensemble de perspectives de recherche termine le mémoire.

## **Mots clés**

Système Interactif d'Aide à la Décision, Extraction de Connaissances à partir de Données, Interaction Homme-Machine, Processus Unifié, Infection Nosocomiale.

## **Abstract**

This thesis presents a contribution to the human-computer interaction (HCI) field, which consists of a proposal of a global framework towards the design and the realization of interactive Decision Support Systems based on Knowledge Discovery in Databases technology (KDD-based DSS). In this framework, we highlight the human and iterative aspects of the DSS development, and taking into account the specificities of the KDD systems.

The proposed methodological framework consists in extending the activities of the Unified Process under the angle of the HCI for the development of the various modules of a KDD-based DSS. This methodology, named Extended Unified Process (ExUP), has been applied in order to demonstrate or validate its founder principles on a medical case study relative to the dynamic decision making for the fight against nosocomial infections (KDD-based MDDSS) in the Intensive care unit of a hospital situated in Sfax, Tunisia.

A global evaluation of the approach, as well as research perspectives, ends the thesis.

## **Key-words**

Decision Support System, Knowledge Discovery in Databases, Human-Computer Interaction, Unified Process, Nosocomial Infection.