



**Conseil Scientifique**

**Appel d'offres 2009  
Programmes pluridisciplinaires**

**APPEL A PROJETS**

**DATE LIMITE : 18 JUIN 2009**

**DOSSIER DE CANDIDATURE**

**1. RENSEIGNEMENTS ADMINISTRATIFS**

**PROGRAMME PLURIDISCIPLINAIRE CONCERNE**

**Modélisation et simulation du vivant et de l'environnement**

**INTITULE DU PROJET DE RECHERCHE**

**Fouille de données : comment extraire des relations spatio-temporelles de causalité depuis des bases de données environnementales ?**

**MOTS CLES : FOUILLE DE DONNEES, CAUSALITE, ENVIRONNEMENT**

**POLES DE RECHERCHE ET DE FORMATION DE RATTACHEMENT DE L'UM2**

- Modélisation : informatique (TATOO du LIRMM)
- Vivant et Environnement : (OREME, TETIS - Cemagref)

**PORTEUR DE LA DEMANDE**

NOM Prénom

PONCELET Pascal

Corps PR

Adresse

Pascal.poncelet@lirimm.fr

Téléphone 0467418653

UMR 5506 - CC 477, 161 rue Ada, 34392 Montpellier Cedex 5 - France

5. UMR 5506

6. Directeur : NOM Prénom Michel Robert

7. CNRS

### COORDINATEUR DU PROJET (SI DIFFERENT)

NOM Prénom  
Adresse @  
Adresse postale  
Unité de recherche principale concernée (Unité UM2 de rattachement du demandeur)  
8. Sigle - Type et numéro – Intitulé  
9. Directeur : NOM Prénom

Corps  
Téléphone

### AUTRES PARTENAIRES

NOM Prénom  
TEISSEIRE Maguelonne  
Adresse  
Maguelonne.teisseire@teledetection.fr  
Maison de la Télédétection  
500, rue J.F.Breton 34093 MONTPELLIER Cedex 5

Corps DR  
Téléphone 0467548762

### UMR TETIS (Territoires, Environnement, Télédétection et Information Spatiale)

10. Directeur : Pascal Kosuth  
11. Cemagref, CIRAD, AgroParisTech

NOM Prénom  
Bringay Sandra  
Adresse  
bringay@lirmm.fr  
UMR 5506 - CC 477, 161 rue Ada, 34392 Montpellier Cedex 5 - France  
12. UMR 5506  
13. Directeur : Michel Robert  
14. CNRS

MCF UM3

0467418636

NOM Prénom  
ARNAUD Nicolas  
Adresse  
Nicolas.Arnaud@univ-montp2.fr

DR CNRS

### NATURE DE LA DEMANDE exprimée en TTC

Soutien financier :

Demande d'une allocation doctorale :

OUI NON

Demande d'une allocation doctorale environnée:

OUI NON

Décharge d'enseignement (1) : nombre d'heures (2) : montant financier (3) :

Décharge d'enseignement associée au recrutement d'un ½ ATER

OUI NON

(1) La décharge d'enseignement devra être incluse dans le budget, sous réserve de l'accord du directeur de la composante d'enseignement

(2) maximum 1/3 de service, à savoir 64h ETD

(3) 1 heure TD = 40,26 €

## CV DU DEMANDEUR ET DU COORDONATEUR DU PROJET (si différent)

### Identité

Nom : PONCELET

Prénom : PASCAL

Date et lieu de naissance : 11 avril 1966 à Uzès (Gard)

Nationalité : Française

Situation de famille : Marié, trois enfants

Adresse personnelle :

18 rue du Puits, 34830 Jacou

Tél. : 04 67 59 88 94

Adresse professionnelle :

LIRMM

161 Rue Ada

3 4392 Montpellier Cedex 5 , France

Tél. : 04 67 41 85 85

Fax : 04 67 41 85 00

Mél : [poncelet@lirmm.fr](mailto:poncelet@lirmm.fr)

Fonction actuelle :

Professeur à l'Université Montpellier 2 – IUT Béziers  
Responsable du projet de recherche Tadoo au Laboratoire LIRMM

### Diplômes

Habilitation à Diriger des Recherches, « De la Conception à l'Extraction de Connaissances dans les Bases de Données », Université Montpellier II, janvier 2001.

Doctorat en sciences mention Informatique, « Contribution à la conception de bases de données avancées : modélisation, évolution et dérivation », Université de Nice-Sophia Antipolis, Mention Très Honorable, mai 1993

DEA Informatique Fondamentale, Université de Nice-Sophia Antipolis, Mention AB, Septembre 1989

### Postes occupés

Professeur à l'Université Montpellier 2, IUT Béziers depuis novembre 2008.

Professeur à l'Ecole des Mines d'Alès de mai 2001 à octobre 2008.

Maître de Conférences à l'Université Paul Valéry – UFR IV (Sciences Economiques, Mathématiques et Sociales), Montpellier, de septembre 1999 à mai 2001

Maître de Conférences au Département Informatique de l'IUT d'Aix en Provence de septembre 1994 à septembre 1999. Première classe depuis 1998.

ATER (Attaché Temporaire d'Enseignement et de Recherche) au Département Informatique de l'IUT d'Aix en Provence de septembre 1993 à juillet 1994

ATER (Attaché Temporaire d'Enseignement et de Recherche) à l'Ecole Supérieure en Sciences Informatique (ESSI) de l'Université de Nice-Sophia Antipolis

### Activité de recherche

Initiée en 1990, au sein de l'équipe Bases de Données au Laboratoire Informatique, Signaux et Systèmes (I3S) de l'Université de Nice-Sophia Antipolis, mon activité de recherche s'est poursuivie au Laboratoire Informatique de Marseille (LIM) de 1994 à 1996. De 1996 à 2001, elle s'est déroulée au Laboratoire d'Informatique, de Robotique et de Micro-électronique de Montpellier (LIRMM). De 2001 à 2008, elle s'est déroulée au sein du Centre de Recherche LGI2P de l'Ecole des Mines d'Alès et du LIRMM. Depuis 2008 elle se déroule au LIRMM. Dans le cadre de mes recherches, j'ai été amené à travailler sur la conception d'applications avancées et sur l'extraction de connaissances dans de grandes bases de données (fouille de données).

Editions d'ouvrages internationaux : 2

Editions de revues nationales : 2

Publications dans un ouvrage : 6

Publications dans des revues internationales avec comité de lecture : 15

Publications dans des revues nationales avec comité de lecture : 9

Publications dans des conférences internationales avec comité de lecture : 51

Publications dans des conférences nationales avec comité de lecture : 37  
Communications, séminaires et conférencier invité : 13  
Nombre de thèses encadrées : 9  
Nombre de stages de Master Recherche ou DEA encadrés : 13

## **Activités Pédagogiques**

Les enseignements réalisés concernent les domaines suivants : Algorithmique, Langage de Programmation, Bases de Données, Bureautique, Conception Orientée Objet, Data Mining, Interface Homme Machine, Programmation Réseau et développement d'applications Client-Serveur, Systèmes d'exploitation, Systèmes d'Information et Méthodes de Génie Logiciel. Outre mon service statutaire, j'ai été amené à réaliser des vacations dans différents établissements Universitaires (Université Montpellier II, Université Aix en Provence, Ecoles des Mines, ...).

Responsable Cours Bases de Données à l'IUT de Béziers  
Responsable Cours Réseau : couches IP à l'IUT de Béziers  
Responsable Cours Bases de Données à l'EMA, l'Université Montpellier III  
Responsable Cours Fouille de Données à l'EMA, Master Informatique Université Montpellier II, Ecole des Mines de Nantes  
Responsable Cours Informatique pour tous en Deug à l'Université Montpellier III  
Responsable Cours C++ à l'IUT Aix en Provence de 1995 à 1999.  
Co-Responsable de la formation continue C++ au département informatique de l'IUT Aix en Proceve de 1995 à 1999.

## **Activités administratives**

Responsable du projet TaToo (Extraction de Connaissances dans les bases de données : motifs séquentiels et ontologies) (8 permanents, 8 doctorants, 2 associés) au Laboratoire LIRMM depuis 2009.  
Directeur Adjoint du Centre de Recherche LGI2P de l'Ecole des Mines d'Alès de 2001 à 2008  
Responsable Scientifique Projet ANR MIDAS (2008-2011).  
Responsable du projet KDD (Knowledge Discovery for Decision making) (5 permanents, 3 doctorants) au Centre de Recherche LGI2P de l'Ecole des Mines d'Alès de 2001 à 2008.  
Responsable du projet Data Mining (2 permanents, 2 doctorants, 1 étudiant en DEA) dans l'équipe Bases de Données – Systèmes d'Information du LIRMM de 1997 à 2001.  
Co-responsable avec J. M. Petit (LIMOS, Clermont Ferrand) du groupe 3.4 (Fouille de Données) de l'Axe « Modèles de données et conception de systèmes d'Information » du GDR I3 de 2003 à 2007.  
Participation à des Comités de Programmes, Comités de Lectures : > 70  
Participation à des Jurys de thèses : 26  
Participation à des Jurys de HDR : 2  
Présidence de Session lors de Conférences : 10  
Participation à l'organisation de Conférences : 8  
Expertises ANR : > 6 – Expert Projet Européen  
Projets de recherche, de transfert technologique et collaborations industrielles : > 17  
Expert Pôle MIPS  
Membre du Comité Technique Paritaire de l'EMA depuis 2007.  
Membre de la CSE de l'Université Claude Bernard (Lyon) de 1998 à 2001.  
Membre du Conseil Scientifique de l'IUT d'Aix en Provence de 1998 à 1999.

## Liste des publications

### Editions d'ouvrages internationaux

- F. Masegla, P. Poncelet and M. Teisseire (Editors). « Successes and New Directions in Data Mining », IDEA Group Publisher, November 2007, ISBN-13 978-1599046457.
- P. Poncelet, F. Masegla and M. Teisseire (Editors). « Data Mining Patterns: New Methods and Applications », IDEA Group Publisher, August 2007, ISBN-13 978-1599041629.

### Editions de revues nationales

- M. Roche et P. Poncelet (Editeurs). « Fouille de Données d'Opinions », Revue RNTI, Cépadués, à paraître en 2009.
- J. M. Petit et P. Poncelet (Editeurs). « Algorithmes pour la découverte de motifs dans les bases de données ». Revue I3 (<http://www.revue-i3.org/>), Cépadués Edition, 2007.

### Publications dans un ouvrage

- D. Li, A. Laurent and P. Poncelet. « Mining Unexpected Sequential Patterns and Implication Rules ». Chapter in "Rare Association Rule Mining and Knowledge Discovery: Technologies for Infrequent and Critical Event Detection" book in Advances in Data Warehousing and Mining Book Series, (to appear 2009).
- F. Masegla, P. Poncelet and M. Teisseire. « Peer to Peer Usage Analysis ». Chapter in "Encyclopaedia of Multimedia Technology and Networking", Second Edition, M. Pagani (ed.), ISBN 978-1-60566-014-1, August 2008, pp. 1136-1141.
- Mokrane, G. Dray et P. Poncelet. « Fouille de Collections de Documents en vue d'une Caractérisation Thématique de Connaissances Textuelles ». Chapitre dans l'ouvrage « Fouille de Données Complexes », Editions Cépadués, fin 2005 (article sélectionné dans le workshop FDC'05).
- Laurent, P. Poncelet and M. Teisseire. « Fuzzy Data Mining for the Semantic Web: Building XML Mediator Schemas ». Chapter in « Fuzzy Logic and the Semantic Web », Elsevier, 20 pages, ISBN: 0-444-51948-3, December 2005.
- F. Masegla, M. Teisseire and P. Poncelet. « Sequential Pattern Mining: A Survey on Issues and Approaches ». Chapter in « Encyclopaedia of Data Warehousing and Mining », J. Wang (ed.), Information Science Publishing, ISBN: 1-59140-556-2, April 2005, 10 pages.
- M. Teisseire, P. Poncelet and R. Cicchetti. « Events as Behavioural Modeling Drivers ». In « Advances in Object-Oriented Modeling », M.P. Papazoglou, S. Spaccapietra and Z. Tari (Eds.), M.I.T Press, October 2000, ISBN 0-262-16189-3, pp. 41-64.

### Publications dans des revues internationales avec comité de lecture

- D. H. Li, A. Laurent, P. Poncelet and M. Roche. « On Unexpected Phrases in Text Documents ». Intelligent Data Analysis (IDA) Journal, Vol. 14, N. 1, 2010.
- D. H. Li, A. Laurent and P. Poncelet. « Discovering Fuzzy Unexpected Sequences with Concept Hierarchies ». International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems (IJUFKS), accepted 2009.
- F. De Razo Lopez, A. Laurent, P. Poncelet and M. Teisseire. « PIVOT : Equivalence Classes-Based Optimized Generation for Tree Mining ». Intelligent Data Analysis (IDA) Journal, Vol13, N. 4, 2009.
- F. De Razo Lopez, A. Laurent, P. Poncelet and M. Teisseire. «FTMnodes: Fuzzy Tree Mining Based on Partial Inclusion». Fuzzy Sets and Systems Journal, to appear 2009.
- F. Del Razo, S. Sanchez, A. Laurent, P. Poncelet and M. Teisseire « Data structures for efficient tree mining: from crisp to soft embedding constraints » International Journal of Applied Mathematics and Computer Science (AMCS), AMCS Special Issue, Soft computing for information management on the Web. To appear, 2009.
- C. Raïssi, T. Calders and P. Poncelet. "Mining Conjunctive Sequential Patterns". Data Mining and Knowledge Discovery, Springer Verlag (Selected paper from ECML/PKDD 08), Vol. 17, N. 1, August 2008.
- F. Masegla, P. Poncelet and M. Teisseire. "Efficient Mining of Sequential Patterns with Time Constraints: Reducing the Combinations", Expert Systems With Applications, Vol. 40, N. 3, 2008.
- F. Masegla, P. Poncelet, M. Teisseire and A. Marascu. « Web Usage Mining: Extracting Unexpected Periods from Web Logs ». Data Mining and Knowledge Discovery, Springer Verlag, Vol. 16, N. 1, February 2008, pp. 39-65.
- P.A. Laur, J.E. Symphor, R. Nock and P. Poncelet. « Mining Evolving Data Streams for Frequent Patterns ». Patterns Recognition, Elsevier, Vol. 40, N. 2, 2007, pp. 492-503.

- C. Raissi, P. Poncelet and M. Teisseire. « Towards a New Approach for Mining Maximal Frequent Itemsets over Data Stream ». *Journal of Intelligent Information Systems*, Springer, Vol. 28, N. 1, 2007, pp. 23-36.
- P.A. Laur, J.E. Symphor, R. Nock and P. Poncelet. « Statistical Supports for Mining Sequential Patterns and Improving the Incremental Update Process on Data Stream ». *Intelligent Data Analysis (IDA) special issue on Knowledge Discovery from Data Streams*, IOS Press, Vol. 11, N.1, 2007, pp. 28-47.
- F. Masseglia, P. Poncelet and M. Teisseire. « Incremental Mining of Sequential Patterns in Large Database ». *Data and Knowledge Engineering*, Vol. 46, N. 1, 2003, pp.97-121.
- F. Masseglia, M. Teisseire and P. Poncelet. « HDM: A Client/Server/Engine Architecture for Real Time Web Usage ». *Knowledge and Information Systems (KAIS) Journal*, Vol. 5, N. 4, pp. 439-465, October 2003.
- F. Masseglia, P. Poncelet and R. Cicchetti. « An Efficient Algorithm for Web Usage Mining ». *Networking and Information System Journal (NIS)*, Vol. 2, N. 5-6, 1999, pp. 571-603.

### **Publication invitée dans une revue internationale**

- F. Masseglia, P. Poncelet and M. Teisseire. « Using Data Mining Techniques on Web Access Logs to Dynamically Improve Hypertext Structure ». In *ACM SigWeb Letters*, Vol. 8, N. 3, October 1999, pp. 13-19.

### **Publications dans des revues nationales avec comité de lecture**

- D. (Haoyuan) Li, A. Laurent et P. Poncelet. “Extraction de comportements inattendus dans le cadre du Web Usage Mining”. *Revue RNTI – Numéro Spécial « fouille de données complexes »*, Cepadues Editor, to appear 2009.
- J. Rabatel, S. Bringay, P. Poncelet et M. Teisseire. “Aide au diagnostic de pannes guidée par l’extraction de motifs séquentiels”. *Revue RNTI – Numéro Spécial « fouille de données complexes »*, Cepadues Editor, to appear 2009.
- Harb, M. Plantié, M. Roche, G. Dray, F. Troussel. P. Poncelet. “Détection d’opinions : comment déterminer les adjectifs d’opinions d’un domaine donné?”. *Revue Document Numérique*, Hermès Editor, numéro spécial « Les documents et le Web 2.0 », Vol. 11/1-2, pp. 37-61, 2008.
- V. Kapoor, P. Poncelet, F. Troussel et M. Teisseire. « Préservation de la vie privée : recherche de motifs séquentiels dans des bases de données distribuées ». *Revue Ingénierie des Systèmes d’Information (ISI)*, Numéro spécial « Bases de Données Avancées 2006 », Vol.12, pp. 85-107, décembre 2007.
- Raïssi et P. Poncelet. “Random Sampling over Data Streams for Sequential Pattern Mining”, *La Revue Modulad*, Numéro 36, mai 2007 (Special Issue on Data Stream Analysis – revue issue du Workshop de Caserta).
- Mokrane, G. Dray and P. Poncelet. « Catégorisation thématique de documents textuels ». *Information Sciences for Decision Making*, Vol. 22, 2005 (disponible en ligne).
- F. Masseglia, M. Teisseire et P. Poncelet. « Extraction de motifs séquentiels – Problèmes et Méthodes ». *Revue Ingénierie des Systèmes d’Information (ISI)*, Numéro spécial « Extraction et usages multiples de motifs dans de grandes Bases de Données », Vol. 9, N. 3-4, 2004, pp. 183-210.
- P.A. Laur, M. Teisseire et P. Poncelet. « Données Semi Structurées : extraction, maintenance et analyse de tendances ». *Revue Ingénierie des Systèmes d’Information (ISI)*, Numéro spécial « Bases de Données Semi Structurées », Vol. 8, N. 5/6, décembre 2003, pp. 49-78.
- P. Poncelet, M. Teisseire, R. Cicchetti et L. Lakhali. « IFO<sub>2</sub>, une approche pour la conception de bases de données avancées ». *Revue Ingénierie des Systèmes d’Information (ISI)*, Vol. 1, N. 4, décembre 1993, pp. 467-510.

### **Publications dans des conférences internationales avec comité de lecture**

- H. Saneifar, S. Bonniol, A. Laurent, P. Poncelet, M. Roche. « Terminology Extraction from Log Files ». In *Proceedings of the 20th International Conference on Database and Expert Systems Applications (DEXA '09)*, September 2009, Linz, Austria
- J. Rabatel, S. Bringay and P. Poncelet. « SO\_MAD: SensOr Mining for Anomaly Detection in Railway Data ». In *Proceedings of the 9th Industrial Conference on Data Mining (ICDM 2009)*, July 20 - 22, 2009, Leipzig, Germany.
- G. Singh, F. Masseglia, C. Fiot, A. Marascu and P. Poncelet. « Data Mining for Intrusion Detection : from Outliers to True Intrusions ». In *Proceedings of the 13th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2009)*, April 2009, Bangkok, Thailand.
- Y. Pitarch, A. Laurent and P. Poncelet. « Multidimensional Data Streams Summarization Using Extended Tilted-Time Windows ». In *Proceedings of the 5th International Symposium on Frontiers of Information Systems and Network Applications (FINA 2009) (join with AINA 2009)*, May 2009, Bradford, UK

- P. Gupta, C. Raïssi, G. Dray, P. Poncelet and J. Brissaud. « SS-IDS: Statistical Signature based IDS ». In Proceedings of the 4<sup>th</sup> International Conference on Internet and Web Applications and Services (ICIW 2009), May 2009, Venice, Italy.
- D. Li, A. Laurent and P. Poncelet. « Recognizing Unexpected Recurrence Behaviors with Fuzzy Measures in Sequence Databases ». In Proceedings of the Fifth ACM/IEEE International Conference on Soft Computing as Transdisciplinary Science and Technology, October 2008, Cergy Pontoise, France, pp. 37-43.
- A. Harb, M. Plantié, M. Roche, G. Dray and P. Poncelet. « Web Opinion Mining: How to Extract Opinions from Blogs? ». In Proceedings of the Fifth ACM/IEEE International Conference on Soft Computing as Transdisciplinary Science and Technology, October 2008, Cergy Pontoise, France.
- C. Raïssi, T. Calders and P. Poncelet. « Mining Conjunctive Sequential Patterns ». In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 08), September 2008, Antwerp, Belgium.
- M. Plantié, M. Roche, G. Dray and P. Poncelet. « Is a Voting Approach Accurate for Opinion Mining? ». In Proceedings of the 10<sup>th</sup> International Conference on Data Warehousing and Knowledge Discovery (DaWaK 08), Turin, Italy, September 2008.
- D. Li, A. Laurent, M. Roche and P. Poncelet. « Extraction of Opposite Sentiments in Classified Free Format Text Reviews ». In Proceedings of the 19<sup>th</sup> International Conference on Database and Expert Systems Applications (DEXA 08), Turin, Italy, September 2008, pp. 710-717.
- D. Li, A. Laurent and P. Poncelet. « Mining Unexpected Web Usage Behaviors ». In Proceedings of the 8<sup>th</sup> Industrial Conference on Data Mining (ICDM 08), In “Advances in Data Mining, Medical Applications, E-Commerce, Marketing and Theoretical Aspects”, LNCS, Vol. 5507, pp. 283-297, Leipzig, Germany, July 2008, Best Paper Award.
- D. Li, A. Laurent and P. Poncelet. « Discovering Fuzzy Unexpected Sequence with Beliefs ». In Proceedings of the 12<sup>th</sup> International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 08), Malaga, Spain, June 2008, pp. 1709-1716.
- C. Raïssi and P. Poncelet. « Sampling for Sequential Pattern Mining: From Static Databases to Data Streams ». Proceedings of the IEEE International Conference on Data Mining (ICDM 07), Omaha NB, USA, October 2007.
- C. Raïssi, J. Brissaud, G. Dray, P. Poncelet, M. Roche and M. Teisseire. « Web Analysing Traffic Challenge: Description and Results », Proceedings of the Discovery Challenge ECML/PKDD 2007, September 17-21, 2007, Warsaw, Poland.
- C. Raïssi and P. Poncelet. « Random Sampling over Data Streams for Sequential Pattern Mining ». First European Workshop on Data Streams, Caserta, Italy, March 2007, pp. 61-66.
- F. Del Razo Lopez, A. Laurent, P. Poncelet and M. Teisseire. « Fuzzy Tree Mining: Go Soft on your nodes ». Proceedings of the International Fuzzy Systems Association World Congress (IFSA 07), Special Session "The application of fuzzy logic and soft computing in flexible querying", Cancun, Mexico, June 2007, pp. 145-154.
- V. Kapoor, P. Poncelet, F. Trouset and M. Teisseire. « Privacy Preserving Sequential Pattern Mining in Distributed Databases ». Proceedings of the Fifteenth Conference on Information and Knowledge Management (CIKM 2006), Arlington, US, November 2006, pp. 758-767.
- C. Raïssi, P. Poncelet and M. Teisseire. « SPEED: Mining Maximal Sequential Patterns over Data Streams ». Proceedings of the 3rd IEEE International Conference on Intelligent Systems, London, UK, September 2006.
- S. Sanchez, A. Laurent, P. Poncelet and M. Teisseire. « FuzBT: a Binary Approach for Fuzzy Tree Mining ». Proceedings of the 11<sup>th</sup> International Conference of Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 06), Paris, France, July 2006.
- F. Massegli, P. Poncelet and M. Teisseire. « Peer-to-Peer Usage Analysis: a Distributed Mining Approach ». Proceedings of the IEEE 20<sup>th</sup> International Conference on Advanced Information Networking and Applications (AINA 2006), Vienna, Austria, April 2006, pp. 993-998.
- F. Massegli, P. Poncelet, M. Teisseire and A. Marascu. « Web Usage Mining: Extracting Unexpected Periods from Web Logs ». Proceedings of the IEEE 2nd Workshop on Temporal Data Mining (TDM'05). Held in conjunction with ICDM'05, Houston, USA, November 2005.
- P.A. Laur, R. Nock, J.-E. Symphor and P. Poncelet. « On the Estimation of Frequent Itemsets for Data Streams: Theory and Experiments ». Proceedings of the 14th ACM Conference on Information and Knowledge Management (CIKM 2005), Bremen, Germany, October 2005, pp. 327-328.
- P.A. Laur, J.E. Symphor, R. Nock, and P. Poncelet. « Mining Sequential Patterns on Data Streams: A Near-Optimal Statistical Approach ». In Proceedings of the Second International Workshop on Knowledge Discovery in Data Streams joint with the 16th European Conference on Machine Learning (ECML) and the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), Porto, Portugal, October 2005

- F. Del Razo Lopez, A. Laurent, P. Poncelet and M. Teisseire. « RSF – A New Tree Mining Approach with an Efficient Data Structure ». In Proceedings of the joint conference 4<sup>th</sup> Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2005) and 11<sup>th</sup> Eleventh Rencontres Francophones sur la logique floue et ses applications (LFA 2005), Special session on Text Mining and Web Mining, Barcelona, Spain, September 2005.
- P.-A. Laur, J.-E. Symphor, R. Nock and P. Poncelet. « Statistical Supports for Frequent Itemsets on Data Streams ». In Proceedings of the International Conference on Machine Learning and Data Mining (MLDM'05), , Leipzig, Germany, Lecture Notes in Artificial Intelligence (LNAI), Springer Verlag, Vol. 3587, July 2005, 395-404.
- A. Laurent, P. Poncelet and M. Teisseire. « Towards a Fuzzy Approach for Mining XML Mediator Schemas ». In Proceedings of the Workshop on Fuzzy Logic and the Semantic Web, Marseille, France, February 2005, pp. 8-10.
- R. Arezki, P. Poncelet, G. Dray and D.W. Pearson. « PAWebSearch: An Intelligent Agent for Web Information Retrieval ». Proceedings of the International Conference on Advances in Intelligent Systems – Theory and Applications (AISTA'2004), Luxembourg-Kirchberg, Luxembourg, IEEE Conference, November 2004, actes électroniques.
- R. Arezki, P. Poncelet, G. Dray and D.W. Pearson. « Information Retrieval Model based on User Profile ». Proceedings of the 11<sup>th</sup> International Conference on Artificial Intelligence (AIMSA'2004), Varna, Bulgaria, Lecture Notes in Computer Science (LNCS), Springer-Verlag, Vol. 3192, September 2004, pp. 490-499.
- R. Arezki, P. Poncelet, G. Dray and D.W. Pearson. « Web Information Retrieval Based on User Profile ». Proceedings of the 3<sup>rd</sup> International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2004), Eindhoven, the Netherlands, Lecture Notes in Computer Science (LNCS), Springer-Verlag, Vol. 3137, August 2004, pp. 275-278.
- F. Maseglia, M. Teisseire and P. Poncelet. « Pre-Processing Time Constraints for Efficiently Mining Generalized Sequential Patterns ». Proceedings of the 11<sup>th</sup> IEEE International Symposium on Temporal Representation and Reasoning (TIME'04), Tatihou –Normandie, France, July 2004, pp. 87-95.
- R. Arezki, A. Mokrane, G. Dray, P. Poncelet, D. Pearson. « LUCI: A Personalization Documentary System Based on the Analysis of the History of the User's Actions ». Proceedings of the 6<sup>th</sup> International Conference on Flexible Query Answering Systems (FQAS2004), Lyon, France, Lecture Notes in Artificial Intelligence (LNAI), Springer Verlag, Vol. 3055, June 2004, pp. 487-498.
- Mokrane, O. Laouamri, G. Dray et P. Poncelet. « Modélisation Spatio-temporelle des Connaissances d'un Système d'Information Géographique ». Proceedings of the Sciences of Electronic, Technologies of Information and Telecommunications Conference (SETIT'04), Sousse, Tunisie, IEEE Publisher, March 2004, pp. 144-152.
- Mokrane, R. Arezki, G. Dray and P. Poncelet. « Cartographie Automatique du Contenu d'un Corpus de Documents Textuels ». Proceedings of the 7<sup>th</sup> International Conference on the Statistical Analysis of Textual Data (JADT'04), Vol.2, Louvain la neuve, Belgium, March 2004, pp. 816-823.
- P.A. Laur, M. Teisseire, and P. Poncelet. « Web Usage Mining: Extraction, Maintenance and Behaviour Trends ». Proceedings of the 1<sup>st</sup> Indian International Conference on Artificial Intelligence (IICAI'03), Hyderabad, India, IEEE Publisher, December 2003, pp. 14-22.
- P.A. Laur, M. Teisseire and P. Poncelet. « AUSMS: An Environment for Frequent Sub-Substructures Extraction in a Semi-Structured Object Collection ». Proceedings of the 14<sup>th</sup> International Conference on Database and Expert Systems Applications (DEXA'03), Prague, Czech Republic, Lecture Notes in Computer Science (LNCS), Springer-Verlag, Vol. 2736, September 03, pp. 38-45.
- F. Maseglia, M. Teisseire and P. Poncelet. « Real Time Web Usage Mining with a Distributed Navigation Analysis ». Proceedings of the 12<sup>th</sup> International Workshop on Research Issues on Data Engineering (RIDE'02), San Jose, USA, February 2002, pp. 169-174.
- F. Maseglia, M. Teisseire and P. Poncelet. « Real Time Web Usage Mining: a heuristic based distributed miner ». Proceedings of the 2<sup>nd</sup> International Conference on Web Information Systems Engineering (WISE'01), Kyoto, Japan, December 2001, pp. 297-306.
- P.A. Laur, F. Maseglia and P. Poncelet. « Schema mining: Finding Structural Regularity among Semi Structured Data ». Proceedings of the 4<sup>th</sup> European Conference on Principles of Data Mining and Knowledge Discovery (PKDD'00), Lyon, France, Lecture Notes in Artificial Intelligence (LNAI), Springer Verlag, Vol. 1910, September 2000, pp. 498-503.
- P.A. Laur, F. Maseglia, P. Poncelet and M. Teisseire. « A General Architecture for Finding Structural Regularities on the Web ». Proceedings of the 9<sup>th</sup> International Conference on Artificial Intelligence (AIMSA'00), Varna, Bulgaria, Lecture Notes in Artificial Intelligence (LNAI), Springer Verlag, Vol. 2424, September 2000, pp. 179-188.
- F. Maseglia, P. Poncelet and M. Teisseire. « Web Usage Mining: How to Efficiently Manage New Transactions and New Clients ». Proceedings of the 4<sup>th</sup> European Conference on Principles of Data Mining and Knowledge Discovery (PKDD'00), Lyon, France, Lecture Notes in Artificial Intelligence (LNAI), Springer Verlag, Vol. 1910, September 2000, pp. 530-535.

- F. Masseglia, P. Poncelet and R. Cicchetti. « WebTool: An Integrated Framework for Data Mining ». Proceedings of the 10<sup>th</sup> International Conference on Database and Expert Systems Applications (DEXA'99), Florence, Italy, Lecture Notes in Computer Science (LNCS), Springer-Verlag, Vol. 1677, September 1999, pp. 892-901.
- F. Masseglia, F. Cathala and P. Poncelet. « PSP: Prefix Tree for Sequential Patterns ». Proceedings of the 2<sup>nd</sup> European Symposium on Principles of Data Mining and Knowledge Discovery (PKDD'98), Nantes, France, Lecture Notes in Artificial Intelligence (LNAI), Springer Verlag, Vol. 1510, September 1998, pp. 176-184.
- F. Cathala and P. Poncelet. « Preserving Behaviour: Why and How ». Proceedings of the 9<sup>th</sup> International Conference on Advanced Information Systems Engineering (CAiSE'97), Barcelona, Spain, Lecture Notes in Computer Science (LNCS), Springer-Verlag, Vol. 1250, June 1997, pp. 333-346.
- Z. Bellahsene, P. Poncelet and M. Teisseire. « Views for Information System Design without Reorganization ». Proceedings of the 8<sup>th</sup> International Conference on Advanced Information Systems Engineering (CAiSE'96), Crete, Greece, Lecture Notes in Computer Science (LNCS), Springer-Verlag, Vol. 1080, June 1996, pp. 496-513.
- M. Teisseire, P. Poncelet, and R. Cicchetti. « Towards Event-Driven Modelling for Database Design ». Proceedings of the 20<sup>th</sup> International Conference on Very Large Databases (VLDB'94), Santiago, Chile, September 1994, pp. 285-296.
- M. Teisseire, P. Poncelet, and R. Cicchetti. « IFO<sub>2</sub>: A Uniform Approach for Information System Modelling ». Proceedings of the 5<sup>th</sup> International Workshop on the Deductive Approach to Information Systems and Databases (DAISD'94), Lloret - Costa Brava, Spain, September 1994, pp. 33-53.
- M. Teisseire, P. Poncelet, and R. Cicchetti. « Dynamic Modelling with Events ». Proceedings of the 6<sup>th</sup> International Conference on Advanced Information Systems Engineering (CAiSE'94), Utrecht, The Netherlands, Lecture Notes in Computer Science (LNCS), Springer-Verlag, Vol. 811, June 1994, pp. 186-199.
- P. Poncelet, M. Teisseire, R. Cicchetti and L. Lakhal. « Towards a Formal Approach for Object Database Design ». Proceedings of the 19<sup>th</sup> International Conference on Very Large Databases (VLDB'93), Dublin, Ireland, August 1993, pp. 278-289.
- P. Poncelet and M. Teisseire. « Advanced Database Modelling and Design: The IFO<sub>2</sub> Approach ». Proceedings of the 3<sup>rd</sup> International Conference for Young Computer Science (ICYS'93), International Academic Publishers, Beijing, China, July 1993, pp. 79-82.
- M. Teisseire, P. Poncelet and R. Cicchetti. « A Tool Based on a Formal Approach for Object-Oriented Database Modelling and Design ». Proceedings of the 6<sup>th</sup> International Workshop on Computer-Aided Software Engineering (CASE'93), IEEE Publishers, Singapore, July 1993.
- P. Poncelet and R. Lakhal. « Consistent Structural Updates for Object Database Design ». Proceedings of the 5<sup>th</sup> International Conference on Advanced Information Systems Engineering (CAiSE'93), Paris, France, Lecture Notes in Computer Science (LNCS), Springer-Verlag, Vol. 685, June 1993, pp. 1-21.

### **Publications dans des conférences nationales avec comité de lecture**

- L. Vincelas, J.-É. Symphor, A. Mancheron et P. Poncelet. « Spams, une nouvelle approche incrémentale pour l'extraction de motifs séquentiels fréquents dans les data streams ». Actes des 9<sup>èmes</sup> Journées Francophones « Extraction et Gestion des Connaissances » (EGC 2009), Strasbourg, Janvier 2009.
- N. Verma, F. Trouset, P. Poncelet et F. Masseglia « Détection d'intrusions dans un environnement collaboratif sécurisé ». Actes des 9<sup>èmes</sup> Journées Francophones « Extraction et Gestion des Connaissances » (EGC 2009), Strasbourg, Janvier 2009.
- G. Singh, F. Masseglia, C. Fiot, A. Marascu et P. Poncelet. « Collaborative Outlier Mining for Intrusion Detection ». Actes des 9<sup>èmes</sup> Journées Francophones « Extraction et Gestion des Connaissances » (EGC 2009), Strasbourg, Janvier 2009.
- Y. Pitarch, A. Laurent, M. Plantevit et P. Poncelet. « Fenêtres sur cube ». Actes des 22<sup>èmes</sup> Journées Bases de Données Avancées, Octobre 2008.
- D. H. Li, A. Laurent, M. Roche et P. Poncelet. « Recherche de sentiments opposés par une approche floue à partir de textes libres ». Actes de la Conférence Francophone Logique Floue et ses Applications (LFA 08), Lens, France, Octobre 2008, pp. 26-33.
- A. Harb, G. Dray, M. Plantié, P. Poncelet, M. Roche et F. Trouset. « Détection d'Opinions : Apprenons les bons Adjectifs ! ». Actes de l'Atelier Fouille des Données d'Opinions (FODOP 08), conjointement à la Conférence INFORDID 2008, Fontainebleau, France, Mai 2008.
- P. Gupta, C. Raïssi, G. Dray, P. Poncelet et J. Brissaud. « Détection d'intrusions : de l'utilisation de signatures statistiques ». Actes du 5<sup>ème</sup> Atelier Fouille de Données Complexes, Nice, janvier 2008.
- L. Vincelas, J.-É. Symphor, A. Mancheron et P. Poncelet. « Fiasco : un nouvel algorithme d'extraction d'itemsets fréquents » Démonstration Logicielle, Actes des 8<sup>èmes</sup> Journées Francophones « Extraction et Gestion des Connaissances » (EGC 2008), Nice, Janvier 2008.

- J.-E. Symphor, A. Mancheron, L. Vincelas et P. Poncelet. « Le FIA : un nouvel automate permettant l'extraction efficace d'itemsets fréquents dans les flots de données ». Actes des 8<sup>èmes</sup> Journées Francophones « Extraction et Gestion des Connaissances » (EGC 2008), Nice, Janvier 2008.
- D. H. Li, A. Laurent et P. Poncelet. « Découverte de motifs séquentiels et de règles inattendus ». Actes des 8<sup>èmes</sup> Journées Francophones « Extraction et Gestion des Connaissances » (EGC 2008), Nice, Janvier 2008.
- C. Raïssi et P. Poncelet. « Echantillonnage pour l'extraction de motifs séquentiels : des bases de données statiques aux flots de données ». Actes des 8<sup>èmes</sup> Journées Francophones « Extraction et Gestion des Connaissances » (EGC 2008), Nice, Janvier 2008.
- D. H. Li, A. Laurent et P. Poncelet. « Towards Unexpected Sequential Patterns ». Atelier Bases de Données Inductives, Grenoble, Juillet 2007.
- C. Raïssi et P. Poncelet. « Vers une nouvelle approche d'extraction des motifs séquentiels non dérivables ». Actes des 7<sup>èmes</sup> Journées Francophones « Extraction et Gestion des Connaissances » (EGC 2007), Namur, Belgique, Janvier 2007.
- L. Di Jorio, D. Jouve, D. Kraemer, A. Serra, C. Raïssi, A. Laurent, M. Teisseire et P. Poncelet. « Vpsp : extraction de motifs séquentiels dans Weka ». Démonstration dans les 20<sup>èmes</sup> Journées Bases de Données Avancées, Lille, France, Octobre 2006.
- V. Kapoor, P. Poncelet, F. Troussset et M. Teisseire. « Privacy Preserving Sequential Pattern Mining in Distributed Databases ». Actes des 20<sup>èmes</sup> Journées Bases de Données Avancées, Lille, France, Octobre 2006.
- C. Raïssi, P. Poncelet et M. Teisseire. « FIDS : Extraction efficace d'itemsets dans des flots de données ». Actes de l'atelier « Fouille de Données Temporelles » en conjonction avec la conférence EGC 2006, Lille, France, Janvier 2006.
- F. Masegla, P. Poncelet, M. Teisseire et A. Marascu. « Web Usage Mining : extraction de périodes denses à partir de logs ». Actes des 6<sup>èmes</sup> Journées Francophones « Extraction et Gestion des Connaissances » (EGC 2006), Lille, France, Janvier 2006.
- F. Del Razo Lopez, A. Laurent, P. Poncelet et M. Teisseire. « Recherche de sous-structures fréquentes pour l'intégration de schéma XML ». Actes des 6<sup>èmes</sup> Journées Francophones « Extraction et Gestion des Connaissances » (EGC 2006), Lille, France, Janvier 2006.
- F. Masegla, P. Poncelet et M. Teisseire. « Fouille de Données dans les systèmes pair-à-pair pour améliorer la recherche de ressources ». Actes des 6<sup>èmes</sup> Journées Francophones « Extraction et Gestion des Connaissances » (EGC 2006), Lille, France, Janvier 2006.
- C. Raïssi, P. Poncelet et M. Teisseire. « Need for SPEED: Mining Sequential Patterns in Data Streams ». Actes des 21<sup>èmes</sup> Journées Bases de Données Avancées (BDA 2005), Saint Malo, France, Octobre 2005.
- P. Collet, V. Derozier, G. Dray et P. Poncelet. « Analyse du transcriptome de Plasmodium Falciparum par des règles d'associations ». Poster dans les Journées Ouvertes Biologie, Informatique, Mathématiques (JOBIM05), page 60, Lyon, juillet 2005.
- Mokrane, G. Dray et P. Poncelet. « Caractérisation thématique de collections de documents textuels ». Actes des 9<sup>èmes</sup> Journées sur les Systèmes d'Information Elaborée, Ile Rousse, Corse, Juin 2005.
- R. Arezki, P. Poncelet, G. Dray et D.W. Pearson. « Représentation Dynamique de Documents pour une Recherche Documentaire Intelligente ». Actes de la 2<sup>nd</sup> Conférence en Recherche d'Informations et Applications (CORIA'05), Grenoble, Mars 2005.
- P. Collet, V. Derozier, G. Dray, F. Troussset, P. Poncelet et M. Crampes. "Tendances dans les expressions de gènes : Applications à l'analyse du transcriptome de Plasmodium Falciparum". Actes des 5<sup>èmes</sup> Journées Extraction et Gestion des Connaissances (EGC'2005), Paris, France, Janvier 2005.
- A. Mokrane, G. Dray et P. Poncelet. « Fouille de collections de documents en vue d'une cartographie thématique de connaissances textuelles ». Actes du 2<sup>nd</sup> Atelier sur la Fouille de Données Complexes (FDC'05) organisé conjointement avec les 5<sup>èmes</sup> Journées Extraction et Gestion des Connaissances (EGC'05), Paris, France, Janvier 2005
- Mokrane, P. Poncelet and G. Dray. « Visualisation automatique du contenu d'une base de documents textuels via les hyper-cartes d'information ». Actes du 4<sup>ème</sup> Colloque International Veille Stratégique, Scientifique et Technologique (VSST'2004), Toulouse, France, Octobre 2004, pp. 239-250.
- R. Arezki, A. Mokrane, G. Dray, P. Poncelet et D. Pearson. « Modélisation dynamique et temporelle de l'utilisateur pour un filtrage personnalisé de documents textuels ». Actes des 4<sup>èmes</sup> Journées Extraction et Gestion des Connaissances (EGC'04), Clermont Ferrand, France, Revue des Nouvelles Technologies de l'Information (RNTI), Vol. 2, Janvier 2004, pp. 479-484.
- P.A. Laur and P. Poncelet. « AUSMS : un environnement pour l'extraction de sous-structures fréquentes dans une collection d'objets semi-structurées ». Actes des 3<sup>èmes</sup> Journées Extraction et Gestion des Connaissances (EGC'03), Lyon, France, Janvier 2003, pp. 449-459.
- F. Masegla, M. Teisseire et P. Poncelet. « HDM, un module de fouille de données distribué et temps réel ». Actes des Secondes Journées Extraction et Gestion des Connaissances (EGC'02), Vol. 1, N. 4, Hermès, « Extraction des connaissances et apprentissage », Janvier 2002, pp. 393-398.

- F. Masseglia, M. Teisseire et P. Poncelet. « Web Usage Mining : Analyse du comportement des utilisateurs à impact immédiat ». Actes des 17<sup>èmes</sup> Journées Bases de Données Avancées (BDA'01), Agadir, Maroc, Novembre 2001, pp. 325-341.
- F. Masseglia, P. Poncelet et M. Teisseire. « Incremental Mining of Sequential Patterns in Large Databases ». Actes des 16<sup>èmes</sup> Journées Bases de Données Avancées (BDA'00), Blois, France, Octobre 2000.
- F. Masseglia, P. Poncelet et M. Teisseire. « Extraction efficace de motifs séquentiels : le pré-traitement des données ». Actes des 15<sup>èmes</sup> Journées Bases de Données Avancées (BDA'99), Bordeaux, France, Octobre 1999.
- F. Masseglia, P. Poncelet et R. Cicchetti. « Analyse du comportement des utilisateurs sur le Web ». Actes du 17<sup>ème</sup> congrès INFORSID (INFORSID'99), Toulon, France, Juin 1999.
- R. Cicchetti, P. Poncelet et M. Teisseire. « Une aide à la conception et au contrôle de règles actives ». Actes des 12<sup>èmes</sup> Journées Bases de Données Avancées (BDA'96), Cassis, France, Août 1996, pp. 19-34.
- R. Cicchetti, P. Poncelet et M. Teisseire. « Modélisation et validation comportementales ». Actes du 14<sup>ème</sup> congrès INFORSID (INFORSID'96), Toulouse, France, Juin 1996, pp. 407-425.
- M. Teisseire, P. Poncelet, R. Cicchetti, et L. Lakhal. « Conception de bases de données avancées : le projet IFO<sub>2</sub> ». Actes du Congrès AFCET'93, Colloque « Bases de Données », Versailles, France, Juin 1993, pp. 105-114.
- P. Poncelet. « Conception d'Applications Avancées : modèle, mécanisme d'évolution et dérivation ». Actes du 11<sup>ème</sup> congrès INFORSID (INFORSID'93), Lille, France, Mai 1993, pp. 477-496.
- P. Poncelet, M. Teisseire et L. Lakhal. « IFO<sub>2</sub>, modèle et principe pour la conception de Bases de Données Avancées ». Actes des 8<sup>èmes</sup> Journées Bases de Données Avancées (BDA'92), Trégastel, France, Septembre 1992, pp. 320-338.

**VISA DU DEMANDEUR**

PONCELET PASCAL

**VISA DU OU DES DIRECTEUR(S) D'UNITE(S) DE RECHERCHE**

ROBERT Michel

**VISA DU DIRECTEUR DE COMPOSANTE D'ENSEIGNEMENT**  
(Dans le cas d'une demande de décharge d'enseignement)

NOM Prénom

## **2. RENSEIGNEMENTS SCIENTIFIQUES**

**INTITULE DU PROJET DE RECHERCHE**

**Fouille de données : comment extraire des relations spatio-temporelles de causalité depuis des bases de données environnementales ?**

**PRESENTATION DU SUJET**

L'Observatoire des Sciences de l'Univers (OSU) OREME "Observatoire de REcherche Méditerranéen sur l'Environnement" fédère une communauté scientifique forte de 600 chercheurs autour de disciplines aussi diverses et complémentaires que les Géosciences, les Hydrosciences, et l'Ecologie ou l'étude de la Biodiversité terrestre et marine. L'un des principaux challenges de cet OSU réside dans sa capacité à gérer (récolter, intégrer et partager) les gros volumes de données associés à ces disciplines et mettre en évidence des corrélations entre les informations historisées qui ne pouvaient l'être au préalable. A la clé de cette mise en relation des données : la découverte de signaux systématiques permettant de juger de l'effet du changement global et d'en comprendre les mécanismes dans ses effets environnementaux.

Dans ce contexte, la découverte de relations spatio-temporelles est une solution pertinente et adaptée. En effet, parmi des événements (climatiques, hydrologiques, biologiques, écologiques) observés sur différents lieux et avec différentes constantes de temps, ces relations permettront de mieux appréhender la manière dont l'équilibre physique, chimique et biologique des milieux est rompu sous l'effet de forçages naturels (changement global) ou anthropique et donc d'en comprendre les mécanismes. Le système est compliqué par la très grande diversité des observables mesurés (flux de CO2 des garrigues, température des eaux des lagunes, stock en eau souterraine des karsts, fréquences d'épisodes pluvieux catastrophiques...), des méthodes de

mesure et de la fréquence de celles-ci, donnant à des bases de données excessivement hétérogènes. Une approche fréquemment utilisée pour rechercher de tels motifs est d'extraire les corrélations entre séries temporelles de différents lieux géographiques. L'objectif est alors de rechercher les régions qui ont des « comportements semblables » ou en d'autres termes qui possèdent des séries temporelles similaires. Une approche alternative est de convertir ces séries temporelles en séquences d'événements et d'appliquer des techniques de fouilles de données pour extraire des associations intéressantes dans ces séquences d'événements. Parmi les approches de fouilles de données adaptées à un tel contexte, nous pouvons citer la recherche de règles d'association [AgSr94] ou de motifs séquentiels [SrAg96].

En particulier dans [TaSt01], les auteurs proposent de rechercher des corrélations dans des valeurs de mesures de variables écologiques situées à des localisations différentes. Le principe général qu'ils proposent est de transformer les données initiales en transactions adaptées à la problématique des règles d'association ou des motifs séquentiels. Dans ce contexte, ils peuvent alors extraire différents types de motifs. Cependant même si cette approche est intéressante, elle se résume à adapter les données en entrée et à appliquer différents algorithmes existants.

Dans le cadre de cette thèse, il s'agit de poursuivre et de compléter ces travaux prometteurs afin d'offrir aux membres de l'OSU et aux communautés scientifiques intéressées des connaissances utiles pour mieux comprendre comment s'instaure l'équilibre entre milieu et Vivant, et comment il est perturbé par les changements naturels et/ou anthropiques. La découverte de motifs clés, permettant de déterminer les mécanismes de régulation ou de dérégulation des équilibres, est une étape essentielle dans la découverte d'éventuels processus de remédiation. Pour cela, il faudra étudier les méthodes de discrétisation des données les plus en adéquation avec le contexte, éventuellement intégrer une approche basée sur la logique floue. Il sera également nécessaire de prendre en compte l'ensemble des dimensions associées aux données et mesurer leur impact vis-à-vis des algorithmes de fouilles de données étudiés. Une amélioration des méthodes existantes sera alors nécessaire.

Enfin, dans une dernière étape et de manière à offrir des connaissances adaptées en temps réel, la mise à jour de celles-ci doit être proposée afin de tenir compte de nouvelles informations et des missions de l'Observatoire qui doit assurer la pérennité et l'accessibilité complète de ces données. Sur ce dernier point, la proposition devra s'inspirer des approches de fouille de données incrémentales et des travaux sur les flots de données [Muth05], [RaCa08] et [PILa09]. Les différentes propositions seront expérimentées sur des jeux de données réelles et intégrées dans un démonstrateur associé à la plateforme de l'OREME.

[AgSr94] R. Agrawal and R. Srikant. "Fast algorithms for mining association rules". In Proceeding of the 20th VLDB Conference, pages 487--499, Santiago, Chile, 1994.

[Muth05] S. Muthukrishnan Data Streams: Algorithms and Applications (Foundations and Trends in Theoretical Computer Science), Now Publishers Inc, 2005

[SrAg96] R. Srikant and R. Agrawal. „Mining sequential patterns: Generalizations and performance improvements". In Proceedings of the Fifth International Conference on Extending Database Technology, Avignon, France, 1996.

[PILa09] M. Plantevit, A. Laurent and M. Teisseire «OLAP-Sequential Mining: Summarizing Trends from Historical Multidimensional Data using Closed Multidimensional Sequential Patterns» In Annals of Information Systems, special issue in New Trends in Data Warehousing and Data Analysis. To appear, 2009

[RaCa08] C. Raïssi, T. Calders and P. Poncelet. "Mining Conjunctive Sequential Patterns". Data Mining and Knowledge Discovery, Springer Verlag (Selected paper from ECML/PKDD 08), Vol. 17, N. 1, August 2008.