

---

## Curriculum Vitae

**Wang-Chiew TAN**  
Assistant Professor  
Department of Computer Science  
University of California, Santa Cruz

---

### **EMPLOYMENT**

2002 – present      Assistant Professor, UC Santa Cruz

Jan-Mar 2007      Visiting Scientist, IBM Almaden Research Center, CA

July 2004      Visiting Scientist, Institute for Infocomm Research, Singapore

1999 – 2002      Research Assistant, University of Pennsylvania

Summer 1999      Summer Intern, AT&T Laboratories in New Jersey

1998 – 1999      Teaching Assistant, University of Pennsylvania

1997 – 1998      Research Assistant (Dean's Fellow), University of Pennsylvania

1996 – 1997      Software Engineer, Kent Ridge Digital Laboratories and Port of Singapore Authority, Singapore

### **EDUCATION**

2002      Ph.D. in Computer Science, University of Pennsylvania  
Thesis: Data Annotations, Provenance and Archiving  
Advisors: Peter Buneman and Sanjeev Khanna

1999      M.Sc. in Computer Science, University of Pennsylvania

1996      B.Sc. (First Class Honors) in Computer Science, National University of Singapore, Singapore

### **RESEARCH INTERESTS**

Data Provenance, Information Integration

## **AWARDS**

- 2005 2004 IBM Pat Goldberg Memorial Best Paper Award for the paper  
“Composing Schema Mappings: Second-Order Dependencies to the Rescue”  
co-authored with R. Fagin, Ph. G. Kolaitis and L. Popa.
- 2002 Best Paper Award for the paper  
“Keys for XML”  
in Computer Networks vol. 39, no. 5, August 2002  
co-authored with P. Buneman, S. B. Davidson, W. Fan, and C. S. Hara

## **EXTRAMURAL/INTRAMURAL FUNDINGS**

- 2004 National Science Foundation (NSF) Grant. IIS-0430994.  
“Metadata model management: Schema Mappings and Data Exchange”  
PI: Ph. G. Kolaitis, Co-PI: W. Tan
- 2004 National Science Foundation (NSF) CAREER Award. IIS-0347065  
“Tracing the Provenance and Flow of Data”  
PI: W. Tan
- 2003 IBM Shared University Research (SUR) Award.  
(Joint award with Ph. G. Kolaitis.)
- 2002-2003 UCSC General Non-Tenured Faculty Development Award.
- 2002 UCSC Committee on Research: Faculty Research Grant.

## **PATENTS**

- “Quasi Inverses of Schema Mappings”  
R. Fagin, Ph. G. Kolaitis, L. Popa, and W. Tan  
(To be filed), 2007.
- “Sequential Composition of XML Schema Mappings”  
R. Fagin, Ph. G. Kolaitis, L. Popa, W. Tan, and C. Yu  
(Patent Filed No: ARC920050087US1), 2005.
- “Method for Converting Relational Data into XML”  
M. F. Fernández, A. Morishima, D. Suciu, and W. Tan. (US Patent No.: 6,785,673),  
2004.

## INVITED TALKS

2007

- “Debugging Schema Mappings with Routes”, Database Group Seminar, University of Edinburgh.
- “Debugging Schema Mappings with Routes”, K55 Lunch Talk, IBM Almaden Research Center.

2006

- “Debugging Schema Mappings with Routes”, Computer Science Seminar, School of Computing, National University of Singapore.
- “Tools and Techniques for *Understanding* Data Exchange Systems”, Workshop on Information Integration, University of Pennsylvania.
- “A Tool for Debugging Schema Mappings”, Database Group Seminar, Università Roma Tre.
- “A Tool for Debugging Schema Mappings”, Second International Workshop on Exchange and Integration of Data.

2005

- “DBNotes: A Post-It System for Relational Databases based on Provenance”, SIGMOD Program Committee Workshop.

2004

- “An Annotation Management System for Relational Databases: Semantics and Implementation”, IBM Silicon Valley Lab.
- “An Annotation Management System for Relational Databases: Semantics and Implementation”, Computer Science Seminar, School of Computing, National University of Singapore.
- “Tracing the Provenance and Flow of Data”, Stanford Database Seminar, Stanford University.

2003

- “Tracing the Provenance and Flow of Data”, BEA Systems.
- “Tracing the Provenance and Flow of Data”, Computer Science Department, San Jose State University.

2002

- “Data Annotations and Data Archiving”, Database Group Seminar, University of Washington.
- “Archiving Biological Databases” Workshop on Data Derivation and Provenance.
- “Computing Provenance and Annotations for Views”, Workshop on Data Derivation and Provenance.

## **SERVICE TO THE PROFESSION**

- Program Committee Member:
  - 2008
    - International Conference on Very Large Data Bases (VLDB).
  - 2007
    - International Symposium on Database Programming Languages (DBPL).
    - ACM International Workshop on Web Information and Data Management (WIDM).
    - International Conference on Very Large Data Bases (VLDB).
  - 2006
    - ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS).
    - International Conference on Database Systems for Advanced Applications (DASFAA).
    - International Workshop on the Web and Databases (WebDB).
  - 2005
    - International XML Database Symposium (XSym).
    - International Conference on Very Large Data Bases (VLDB).
    - ACM SIGMOD International Conference on Management of Data (SIGMOD).
  - 2004
    - International Workshop on the Web and Databases (WebDB).
    - International Conference on Very Large Data Bases (VLDB).
    - International Conference on Data Engineering (ICDE).
- Reviewer
  - PODS 2007, ICDE 2007, ICDT 2007, CIDR 2007, PODS 2004, VLDB 2003, SIGMOD 2003, SIGMOD 2002, WWW 2002, ICDE 2002, VLDB 2001, DBPL 2001, PODS 2001.
  - Journals: VLDB Journal, ACM TODS, ACM TOIT, WWWJ, JSS, WIDM-DKE.

### **Ph.D. students currently supervised**

Bogdan Alexe

Laura Chiticariu (advanced to candidacy in Spring 2006)

### **Master Projects supervised**

2005	Laura Chiticariu	Computing the Core: Algorithmic Issues.
2005	Gaurav Vijayvargiya	Tracing the Why-Provenance of Data over Schema Mappings.
2003	Xinyu Hua	A Relational Database Annotation Management System.

### **Member of Ph.D. Thesis Committee**

2006	Jonathan Panttaja	Complexity in Databases, Games and Logics
------	-------------------	---

### **Member of M.Sc. Project/Thesis Committee**

2004	Yongqin Xiao	Scale-Invariant Geometric Features and Descriptors for Height Data Registration
2004	Randolph Goldman	A Survey of Logic in Games and Games in Logic
2003	Guozheng Ge	Bamboo: Automatic Generation of Rule-based Software Configuration Management (SCM) Systems
2003	Kai Pan	An Algorithm for Translating DASL Queries into SQL

## PUBLICATIONS

### Scholarly Journals

1. “Peer Data Exchange”  
A. Fuxman, Ph. G. Kolaitis, R. J. Miller, and W. Tan.  
In ACM Transactions on Database Systems (TODS), vol. 31, no. 4, pp. 1454-1498, 2006.  
(A preliminary version of this paper appeared in the 2005 Proceedings of the ACM Symposium on Principles of Database Systems (PODS). See [17].)
  - *Special Issue with selected papers from the 2005 ACM PODS conference.*
2. “An Annotation Management System for Relational Databases”  
D. Bhagwat, L. Chiticariu, W. Tan, and G. Vijayvargiya  
In Very Large Data Bases (VLDB) Journal, vol 14, no. 4, pp. 373-396, 2006.  
(A preliminary version of this paper appeared in 2004 Proceedings of the International Conference of Very Large Data Bases (VLDB). See [20].)
  - *Special Issue with selected papers from the 2004 VLDB conference.*
3. “Composing Schema Mappings: Second-order dependencies to the rescue”  
R. Fagin, Ph. G. Kolaitis, L. Popa, and W. Tan  
In ACM Transactions on Database Systems (TODS), vol 30, no. 4, pp. 994-1055, 2005.  
(A preliminary version of this paper appeared in 2004 Proceedings of ACM Symposium on Principles of Database Systems (PODS). See [19].)
  - *Special Issue with selected papers from the 2004 ACM PODS conference.*
4. “Research Problems in Data Provenance”  
W. Tan.  
In IEEE Data Engineering Bulletin, vol. 27, no. 4, pp. 45-52, 2004.
  - *Invited survey article*
5. “Archiving Scientific Data”  
P. Buneman, S. Khanna, K. Tajima, and W. Tan.  
In ACM Transactions on Database Systems (TODS), vol 29, no. 1, pp. 2-42, 2004.  
(A preliminary version of this paper appeared in the 2002 Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD). See [23].)
  - *Special Issue with selected papers from the 2002 ACM SIGMOD conference.*
6. “Reasoning about Keys for XML”  
P. Buneman, S. B. Davidson, W. Fan, C. Hara, and W. Tan.  
In Information Systems, vol. 28, no.8, pp. 1037-1063, 2003.  
(A preliminary version of this paper appeared in the 2001 Proceedings of the International Symposium on Data Base Programming Languages (DBPL). See [25].)
7. “SilkRoute: A Framework for Publishing Relational Data in XML”  
M. F. Fernández, Y. Kadiyska, A. Morishima, D. Suciú, and W. Tan.  
In ACM Transactions on Database Systems (TODS), vol.27, no.4, pp. 438-493, 2002.

8. “Keys for XML”  
P. Buneman, S. B. Davidson, W. Fan, C. S. Hara, and W. Tan.  
In *Computer Networks*, vol.39, no.5, pp.473-487, 2002.  
(A preliminary version of this paper appeared in the 2001 Proceedings of International Conference on World Wide Web (WWW10). See [26].)  
○ *Best Paper Award*
9. “Publishing Relational Data in XML: The SilkRoute Approach”  
M. F. Fernández, A. Morishima, D. Suciu, and W. Tan.  
In *IEEE Data Engineering Bulletin*, vol.24, no. 2, pp. 12-19, 2001.  
○ *Invited survey article*
10. “SilkRoute: trading between relations and XML”  
M. F. Fernández, D. Suciu and W. Tan.  
In *Computer Networks*, vol. 33, no. 1-6: pp. 723-745, 2000.  
(A preliminary version of this paper appeared in the 2000 Proceedings of the International Conference on World Wide Web (WWW9). See [30].)
11. “A Graphical Interface to Genome Multidatabases”  
W. Tan, K. Wang and L. Wong.  
In *Journal of Database Management*, vol. 9, no. 1, pp. 24-32, 1998.  
(A preliminary version of this paper appeared in 1996 International Workshop on Database and Expert Systems Applications (DEXA Workshop). See [33].)

### **Peer Reviewed Conference and Workshop Articles**

12. “Provenance in Databases”  
P. Buneman and W. Tan.  
In *Proceedings of ACM SIGMOD International Conference on Management of Data (SIGMOD) (Tutorial Track)*, pp. 1171-1173, 2007.
13. “Quasi-inverses of Schema Mappings”  
R. Fagin, Ph. G. Kolaitis, L. Popa, and W. Tan.  
In *Proceedings of ACM Symposium on Principles of Database Systems (PODS)*, pp. 123-132, 2007.  
○ *Full version invited to ACM Transactions on Database Systems (TODS)*.
14. “Debugging Schema Mappings with Routes”  
L. Chiticariu and W. Tan  
In *Proceedings of International Conference on Very Large Data Bases (VLDB)*, pp. 79-90, 2006.
15. “SPIDER: a Schema mapPIng DEbuggeR”  
B. Alexe, L. Chiticariu and W. Tan  
In *Proceedings of International Conference on Very Large Data Bases (VLDB) (Demonstration Track)*, pp. 1179-1182, 2006.

16. “The Complexity of Data Exchange”  
Ph. G. Kolaitis, J. Panttaja, and W. Tan  
In Proceedings of ACM Symposium on Principles of Database Systems (PODS), pp. 30-39, 2006.
17. “Peer Data Exchange”  
A. Fuxman, Ph. G. Kolaitis, R. J. Miller, and W. Tan.  
In Proceedings of ACM Symposium on Principles of Database Systems (PODS), pp. 160-171, 2005.
  - *Full version invited to ACM Transactions on Database Systems (TODS). See [1].*
18. “DBNotes: a Post-It System for Relational Databases based on Provenance”  
L. Chiticariu, G. Vijayvargiya and W. Tan.  
In Proceedings of ACM SIGMOD International Conference on Management of Data (SIGMOD) (Demonstration Track), pp. 942-944, 2005.
19. “Composing Schema Mappings: Second-Order Dependencies to the Rescue”  
R. Fagin, Ph. G. Kolaitis, L. Popa, and W. Tan.  
In Proceedings of ACM Symposium on Principles of Database Systems (PODS), pp. 83-94, 2004.
  - *Full version invited to ACM Transactions on Database Systems (TODS). See [3].*
  - *IBM Pat Goldberg Memorial Best Paper Award.*
20. “An Annotation Management System for Relational Databases”  
D. Bhagwat, L. Chiticariu, G. Vijayvargiya, and W. Tan.  
In Proceedings of International Conference on Very Large Data Bases (VLDB), pp. 900-911, 2004.
  - *Full version invited to Very Large Data Bases Journal (VLDB Journal). See [2].*
21. “Containment of Relational Queries with Annotation Propagation”  
W. Tan.  
In Proceedings of International Symposium on Data Base Programming Languages (DBPL), pp. 37-53, 2003.
22. “On Propagation of Deletions and Annotations through Views”  
P. Buneman, S. Khanna and W. Tan.  
In Proceedings of ACM Symposium on Principles of Database Systems (PODS), pp. 150-158, 2002.
23. “Archiving Scientific Data”  
P. Buneman, S. Khanna, K. Tajima, and W. Tan  
In Proceedings of ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 1-12, 2002.
  - *Full version invited to ACM Transactions on Database Systems (TODS). See [5].*
24. “On Computing Functions with Uncertainty”  
S. Khanna and W. Tan.  
In Proceedings of ACM Symposium on Principles of Database Systems (PODS), pp. 171-182, 2001.

25. “Reasoning about Keys for XML”  
P. Buneman, S. B. Davidson, W. Fan, C. S. Hara, and W. Tan.  
In Proceedings of International Workshop on Data Base Programming Languages (DBPL),  
pp. 133-148, 2001.  
(The full version of this paper appeared in [6].)
26. “Keys for XML”  
P. Buneman, S. B. Davidson, W. Fan, C. S. Hara, and W. Tan.  
In Proceedings of International Conference on World Wide Web (WWW10), pp. 201-210,  
2001.
  - *Full version invited to Computer Networks. See [8].*
27. “Why and Where: A Characterization of Data Provenance”  
P. Buneman, S. Khanna and W. Tan.  
In Proceedings of International Conference on Database Theory (ICDT), pp. 316-330, 2001.
28. “Towards a Query Language for Annotation Graphs”  
S. Bird, P. Buneman and W. Tan.  
In Proceedings of International Conference on Language Resources and Evaluation (LREC),  
pp. 807-814, 2000.
29. “Data Provenance: Some Basic Issues”  
P. Buneman, S. Khanna and W. Tan.  
In Proceedings of Conference on the Foundations of Software Techniques and Theoretical  
Computer Science (FSTTCS), pp. 87-93, 2000.
30. “SilkRoute: Trading between relations and XML”  
M. F. Fernández, D. Suciú and W. Tan.  
In Proceedings of International Conference on World Wide Web (WWW9), pp. 723-745,  
2000. (The full version of this paper appeared in [10].)
31. “Beyond XML Query Languages”  
P. Buneman, A. Deutsch, W. Fan, H. Liefke, A. Sahuguet, and W. Tan.  
In Query Language Workshop (QL'98), 1998.
32. “A Deterministic Model for Semistructured Data”  
P. Buneman, A. Deutsch and W. Tan.  
In Workshop on Query Processing for Semistructured Data and Non-Standard Data Formats,  
1998.
33. “QUICK: Graphical User Interface to Multiple Databases”  
W. Tan, K. Wang and L. Wong.  
In International Workshop on Database and Expert Systems Applications (DEXA  
Workshop), pp. 404-409, 1996.  
(The full version of this paper appeared in [11].)