

Business Artifacts: A Data-centric Approach to Modeling Business Operations and Processes

David Cohn and Richard Hull
IBM T.J. Watson Research Laboratory
Hawthorne, New York, USA
{dcohn,hull}@us.ibm.com

Abstract

Traditional approaches to business process modeling and workflow are based on activity flows (with data often an afterthought) or documents (with processing often an afterthought). In contrast, an emerging approach uses (business) artifacts, that combine data and process in an holistic manner as the basic building block. These correspond to key business entities which evolve as they pass through the business's operation. This short paper motivates the approach, surveys research and its applications, and discusses how principles and techniques from database management research can further develop the artifact-centric paradigm.

1 Introduction

The importance of effective Business Process Management (BPM) increases as the needs for better insight, understanding and efficiency for business operations increases. Classically, most BPM frameworks (e.g., [LRS02, vdAtHKB03]) have used meta-models¹ centered on activity-flows, with the data manipulated by these processes seen as second-class citizens. Another approach [GM05] focuses on the documents that track the business operations, with the process meta-model typically impoverished. For both, associated requirements, business rules, and business intelligence are based on conceptual meta-models only loosely connected to the base model. This disparity adds substantial conceptual complexity to models of business operations and processes, making them hard to understand. This paper focuses on (*business*) *artifacts*, rather than activity-flows or documents. Artifacts combine both data aspects and process aspects into a holistic unit, and serve as the basic building blocks from which models of business operations and processes are constructed. The approach enables a natural modularity and componentization of business operations and varying levels of abstraction. The paper motivates the approach, surveys research and applications, and highlights ways that philosophic underpinnings and selected techniques from database management research can further its development.

Copyright 2009 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, Volume 32, Number 3, September, 2009

¹Following the tradition of UML and related frameworks, we use the terms 'meta-model' and 'model' for concepts that the database and workflow research literature refer to as 'model' and 'schema', respectively.

Artifacts are business-relevant objects that are created, evolved, and (typically) archived as they pass through a business. The artifact type includes both an *information model* for data about the business objects during their lifetime, and a *lifecycle model*, describing the possible ways and timings that tasks can be invoked on these objects. A prototypical artifact is *Air Courier Package*, whose information model would include slots for package ID, sender, recipient, arrival times, delivery time, and billing information. The lifecycle model would include the multiple ways that the package could be delivered and paid for. Artifacts define a useful way to understand and track business operations, such as the locations that the package has passed through and its arrival times, as typically provided to customers.

Since 2003, IBM Research has been developing meta-models, methods, tools, user-centric paradigms, and other technologies in support of the artifact-centric paradigm [NC03, KNI⁺03, SNK⁺08, CDI⁺08, SSRM07, Hul08]. The methods and tools have been successfully applied in various settings [B⁺05, BCK⁺07, C⁺09].

Three key lessons have been learned from the work to date:

1. The artifact-centric approach enables rich, natural communication among diverse stakeholders about the operations and processes of a business, in ways that activity-flow based and document-based approaches have not. This has measurably reduced the time and staff needed to do business transformations, and enabled unexpected new capabilities.
2. The artifact-centric models, even though expressed in a way that business-level people can understand, are *actionable*, i.e., they can be mapped to execution-level models implementable with tools like IBM's WebSphere Process Server [Fer01], and can serve as an organizing foundation for related BPM capabilities, such as business rules, the development of web screens for task performers, and business intelligence.
3. There is a compelling opportunity for research into numerous aspects of the artifact-centric approach. From a core Computer Science perspective, artifacts provide a well-motivated framework that combines data and process in a manageable way; this combination has been largely missing from research on databases and knowledge representation, that has focused largely on data aspects, and also from research on programming languages, software engineering, workflow, and verification, that has focused largely on process aspects. Specific areas for exploration include conceptual modeling (and, in particular, declarative meta-models), design methods, user-centric aspects, systems issues, integrity constraints, views and foundations. Core philosophic perspectives and techniques from database (and management information science) research can make substantial contributions to this field.

The following sections discuss each of these points in more detail.

2 Enabling understanding and communication

This section outlines the artifact-centric approach to modeling business operations, contrasts it with Entity-Relationship modeling in databases, and highlights how it facilitates stakeholder communication.

As detailed in [C⁺09], IBM Research has applied the artifact-centric paradigm to a problem faced by IBM Global Financing (IGF), which operates in more than 50 countries and annually finances over \$40 billion in IT. After 25 years of organic growth, IGF's global operations were essentially in country "silos", each with different procedures. IGF needed operations based on a global standard with disciplined regional variations, that streamlined operations and allowed the business to expand its focus from large-scale loans to include moderately-sized deals. IGF had tried to do so using traditional techniques (e.g., process decomposition, Lean and Six Sigma), but was not succeeding.

IBM Research, working closely with IGF subject matter experts, applied the artifact method to create a high-level model of the IGF operations. This model is focused on three business artifacts:

- **Deal:** The activity around evaluating a client request, negotiating terms and conditions, signing the contract, issuing invoices for the assets to be financed, and tracking payments and completion.

- **Supplier Invoice:** The purchase and shipping of the asset(s) to the client location(s).
- **Asset:** The individual hardware asset(s) when accepted from the supplier, titled to IGF, delivered to the client, used by the client, and finally sold or disposed.

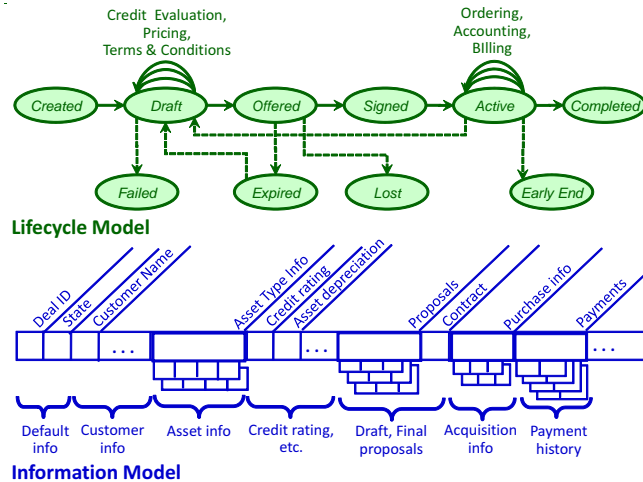


Figure 1: High-level specification of Deal artifact

Figure 1 shows an informal, high-level representation of the information and lifecycle models of the Deal artifact. The information model has slots for information gathered as a Deal artifact instance evolves, including customer details (credit ratings, etc.), types(s) of asset(s), terms and conditions, specific hardware asset(s) acquired, and payment history. The lifecycle model shows the key business-relevant states through which a Deal passes, with transition edges corresponding to tasks performed by specialists. The solid transition edges correspond to the “sunny day” state sequence from Created to multiple Draft versions, through Offered, Signed, multiple loops through Active, and, finally, Completed. The dashed edges show additional potential transitions, some going to additional states. The artifact information model starts out largely

empty, and over the life of the artifact, its attributes are filled in (or overwritten). The first task in a Deal’s Active state creates corresponding Supplier Invoice artifact instance(s); and when each physical asset is accepted by IGF, an Asset artifact instance is created. More generally, in the state-based approach to artifacts, instances interact through message passing as they transition between states. The artifact-based business operations model is being used by IGF to manage operations at both global and local levels. (The full Deal artifact type has about 100 attributes and 70 states.) IGF plans to automate their top-level operations around this model and expects significant efficiency gains.

There are parallels between the artifact approach to business operations modeling and the Entity Relationship (ER) approach [Che76] to modeling the data managed in a business. Both are systematic approaches that use a small set of natural and intuitive constructs. Further (as discussed in Section 3), business artifact specifications are *actionable*, in the same way that ER diagrams are actionable, i.e. the specification can be used to automatically generate an executable system. There is a contrast between how information is typically clustered in artifacts vs. in database schema design and document management systems. With database schemas, there is a tendency to break data into fairly small “chunks”: ER-based techniques use separate entity types and their relationships; normal forms from relational database theory break data apart to avoid update anomalies. This is valuable when data is used by a variety of applications. Similarly, document management systems often focus on the company’s literal document types rather than on the single conceptual entity which multiple document types together represent. In contrast, an artifact information model clusters the various kinds of data which correspond to the stages in the business entity’s lifecycle.

Clustering data based on a dynamic entity that moves through a business’s operations, rather than pieces of its lifecycle, makes a profound difference. As demonstrated in the IGF and other examples, it enables strong communication between a business’s stakeholders in ways that traditional approaches do not. Experience has shown that once the key artifacts are identified, even at a preliminary level, they become the basis of a stakeholder vocabulary. Artifacts enable communication along three dimensions, which we illustrate using the Deal artifact. Along the lifecycle dimension, stakeholders who focus on one part of a lifecycle, say the Draft state, are better equipped to communicate with stakeholders focused on another part, say the Active state. All are talking about the same overall artifact and can confidently discuss attributes that are shared or produced in one part of

the lifecycle and consumed in another. Across the variations dimension, IGF stakeholders from multiple geographies could understand similarities and differences between their respective operations by comparing them to the commonly held artifact model. Communication between stakeholders at different management levels is enhanced because the artifact approach naturally lends itself to a hierarchical perspective. For example, the Deal artifact shown in Figure 1 is easily understood by executives, and a drill down is useful to stakeholders managing the detailed operations.

3 An actionable framework

The artifact-centric framework is actionable along two dimensions. An artifact model expressed in business-level terms can be automatically mapped onto a workflow engine to create a deployed system. Such a model can also be the basis for attaching a variety of traditional BPM capabilities.

There are currently three working implementations of the (state-based) artifact meta-model, each with a different purpose. Two are elements of the tooling associated with the Business Entity Lifecycle Analysis (BELA) *capability pattern* [SNK⁺08], that is part of IBM's Service Oriented Modeling and Architecture (SOMA) method. BELA's FastPath tool lets artifact model designers automatically generate a running model during the design process. It provides a full system shell and preliminary versions of performer web screens. Designers and executives can step through different scenarios, seeing how the artifact model behaves. The second BELA tool can map an artifact model into a workflow that runs on IBM's WebSphere Process Server [Fer01]. This has been used to deploy business processes that operate at a massive scale, with 100s of simultaneous users. The third implementation is the experimental Siena prototype [CDI⁺08, F.F09], that uses a direct architecture. The artifact model is represented as an XML document and execution is performed essentially by a direct interpretation of the XML. This system has been used for rapid prototyping exercises involving small- and medium-size applications, and is available to universities for teaching and research.

As noted, a business artifact is a blend of data and process for a key business-relevant dynamic entity that captures its end-to-end journey. As a result, business artifacts are a natural basis for many BPM suite capabilities. For example, [Lin07] describes a tool for using business rules expressed in OMG's SBVR standard [Obj08] with artifacts. The work shows that the vocabulary provided by artifacts is natural for specifying business rules, and shows how the rules can be mapped into the system to guide task sequencing and prevent rule violations. In the area of web screens for performers, [SMS09] describes how the basic artifact structure is the basis for automated implementation of the screens for carrying out business process tasks. A key enabler here is that the artifact model can include *CRUD* (Create-Read-Update-Delete) permissions in terms of artifact attributes. The rights of performers in a given role can depend on the artifact's state. Business artifacts also provide a natural basis for Key Performance Indicator (KPI) specification, monitoring, and response, because they correspond to the business-relevant entities the KPIs measure. Citation [K⁺07] describes how an artifact-centric model for a supply chain application was used for sense-and-respond monitoring and dashboarding. To summarize, the artifact-centric approach lets many BPM suite capabilities be based on a single model at both conceptual and implementation levels, rather than on several diverse conceptual models.

4 Research challenges

The combination of data and process provided by the business artifact approach raises interesting research issues ranging from conceptual modeling and design, to systems issues, to foundations. The artifact abstraction provides a vehicle for understanding the interplay between data and process in ways not supported by previous Computer Science abstractions. For example, artifacts permit the study of how a broad class of data evolves over time, providing structure and opportunity for application of old techniques and development of new ones. This section highlights challenges that may be of particular interest to the database community. Another survey

of research opportunities is [Hul08].

A central research challenge is to understand the basic building blocks and alternatives for artifact-centric meta-models. In some ways, this is analogous to research into semantic data models in the 70's and 80's [HK87]. Central to this investigation is the diversity of people involved in designing and specifying business operations, ranging from executives to business architects, business analysts, and subject matter experts, and finally to business solution designers. Typically, solution designers are comfortable with detailed artifact models, but the others often prefer high-level requirements, business rules, and scenarios. The relationship between these two levels of specification is analogous to that between semantic data models (including the ER model) and the relational model in database management. Important goals here include formal mechanisms to specify requirements, rules and scenarios, and to map and trace their links to detailed artifact models.

To date, work on the artifact-centric method and artifact meta-models, and also related work [BDW07, RDtHI09], has used a variant of finite state machines to specify lifecycles. Recent theoretical work (e.g., [BGH⁺07, DHPV00, BHS09]), is exploring declarative approaches to specifying the artifact lifecycles following an event-condition-action and/or condition-action style. The Project ArtiFactTM team at IBM Research is developing a first practical artifact-centric meta-model along these lines. The meta-model will incorporate parallelism of human-performed tasks and explicit hierarchy in the lifecycle specification. Declarative approaches promise to enable succinct specification of variations which may arise across differing geographies or customer categories. Also, they may enable the development of multiple perspectives or views on an artifact model or portions of it, which would be useful to executives and subject matter experts. Finally, a declarative approach has already shown itself to be promising as a basis for verification of artifact model properties [BGH⁺07, DHPV00].

Other variations in the meta-model also merit study, including the underlying data meta-model (e.g., XML-based or ontology-based [BDW07]), task models (e.g., CRUD information only, BPEL specifications, or pre- and post-conditions as in semantic web services), and association of tasks to artifacts (e.g., design time as is the tradition, or dynamically at run time). An intriguing direction is to use Active XML [ABM08] as a basis for supporting artifacts, as in [ABGM09].

Similar to database management, the artifact-centric approach enables separation of logical vs. physical concerns. While an artifact's information model may cluster multiple kinds of data and permit users to query and manipulate instances as a unit, they may be physically stored across multiple databases. Further, different parts of an artifact lifecycle might be carried out by different, perhaps legacy, applications or systems. Finally, as discussed in [NC03, ABGM09], it may be beneficial to view artifact instances as traveling between organizations, either conceptually or physically. Against this background of modeling choices, several systems issues need to be addressed. Because of the possibilities of parallel processing and interactions between artifact instances, concurrency control must be provided. The interplay of materialized and virtual data raises traditional problems of fast access, query processing across diverse data sources, and maintaining consistency across redundant copies of data, but in a structured context. If considering large scale deployments, it is useful to study techniques that follow the intended semantics of a declarative artifact model, but enable optimizations according to resource availability. Initial work towards such a framework, reminiscent of the use of the relational algebra as an optimization level under SQL, is reported in [BHS09].

A fundamental and largely unexplored area for artifacts, which received considerable attention in relational databases, is the constellation of design principles and integrity constraints. What is the analog for artifacts of the relational notion of update anomalies and normal forms, and the dependencies used to study them? As noted, normal forms tend to disaggregate data, whereas artifacts encourage clustering of data around an organization's underlying dynamic entities. Citation [LBW07] develops an algorithm that analyzes the input-output properties of different tasks, in order to recommend how data should be clustered to form the key artifacts. It is natural to think in terms of integrity constraints that address the evolution of artifacts; work on dynamic constraints and evolution in the relational model (e.g., [AV89]) can provide a useful starting point. Naturally arising classes of temporal constraints for artifacts may come from part of SBVR [Obj08]

Another unexplored area for artifacts is *views*. This is important, for example, when an artifact-centric model

is used to represent the activity of an *interoperation hub* that facilitates the choreography of multiple business processes. [HNN09]. Conference management sites like EasyChair and ConfTool are such hubs although they are not (currently) artifact-centric. In these applications, stakeholders have access to varying views of the overall system which restrict data and behavioral capabilities. Citation [HNN09] develops a notion of view for state-based artifacts, including projection and selection on the information model, and a form of *condensation* of states for the lifecycle model; an analog for declarative lifecycles remains open. More generally, basic properties such as the interplay of views and integrity constraints, and translating queries and modification requests against views into the base model remain largely unexplored.

Research into foundations underlying the artifact model is at an early stage. Studies of static analysis for state-based artifacts include [GS07, KLW08], and those for declarative lifecycles are in [BGH⁺07, DHPV00]. Citation [FHS09] presents a first study of synthesizing declarative artifact models, and [CGHS09] presents a preliminary investigation into dominance and relative expressive power of such models. Extension of these directions and development of a theory of constraints and views in the context of dynamic behavior, are promising challenges that call for techniques from database theory, finite model theory, and temporal and other logics.

Acknowledgements

This paper is based on work that began at IBM Research in the late 1990's, and picked up momentum in 2003. Many IBMers have made substantial contributions to the effort, and prominent contributors include: Kumar Bhaskaran, Kumar Bhattacharya, Nathan Caswell, Henry Chang, Tian Chao, Fabiana Fournier, Amit Fisher, Fenno (Terry) Heath III, Santosh Kumaran, Mark Linehan, Rong (Emily) Liu, Prabir Nandi, Anil Nigam, Piyawadee Noi Sukaviriya, John Vergo, and Frederick y Wu. Researchers outside of IBM have also contributed directly or indirectly to the larger artifact-centric vision, including Serge Abiteboul, Elio Damaggio, Alin Deutsch, Christian Fritz, Fabio Patrizi, Jianwen Su, and Victor Vianu. Finally, the research by Richard Hull described here is partially supported by NSF grants IIS-0415195, CNS-0613998, and IIS-0812578.

References

- [ABGM09] S. Abiteboul, P. Bourhis, A. Galland, and B. Marinoiu. The AXML Artifact Model. In *Proc. 16th Intl. Symp. on Temporal Representation and Reasoning (TIME)*, 2009.
- [ABM08] S. Abiteboul, O. Benjelloun, and T. Milo. The Active XML project: An overview. *Very Large Databases Journal*, 17(5):1019–1040, 2008.
- [AV89] S. Abiteboul and V. Vianu. A transaction-based approach to relational database specification. *Journal of the ACM*, 36(4):758–789, October 1989.
- [B⁺05] K. Bhattacharya et al. A Model-driven Approach to Industrializing Discovery Processes in Pharmaceutical Research. *IBM Systems Journal*, 44(1), 2005.
- [BCK⁺07] K. Bhattacharya, N. S. Caswell, S. Kumaran, A. Nigam, and F. Y. Wu. Artifact-centered operational modeling: Lessons from customer engagements. *IBM Systems Journal*, 46(4):703–721, 2007.
- [BDW07] M. Born, F. Dörr, and I. Weber. User-friendly semantic annotation in business process modeling. In *Proc. Web Information Systems Engineering - WISE 2007 Workshops*, 2007.
- [BGH⁺07] K. Bhattacharya, C. E. Gerede, R. Hull, R. Liu, and J. Su. Towards formal analysis of artifact-centric business process models. In *Proc. Int. Conf. on Business Process Management (BPM)*, pages 288–304, 2007.
- [BHS09] K. Bhattacharya, R. Hull, and J. Su. A Data-centric Design Methodology for Business Processes. In J. Cardoso and W.M.P. van der Aalst, editors, *Handbook of Research on Business Process Management*. Information Science Publishing, an imprint of IGI Global, Hershey, PA, USA, 2009.
- [C⁺09] T. Chao et al. Artifact-based transformation of IBM Global Financing: A case study, 2009. *Intl. Conf. on Business Process Management (BPM)*, September, 2009.
- [CDI⁺08] D. Cohn, P. Dhoolia, F.F. (Terry) Heath III, F. Pinel, and J. Vergo. Siena: From powerpoint to web app in 5 minutes. In *Intl. Conf. on Services Oriented Computing (ICSOC)*, 2008.

- [CGHS09] D. Calvanese, G. De Giacomo, R. Hull, and J. Su. Artifact-centric workflow dominance. In *Proc. Intl. Conf. on Service Oriented Computing (ICSOC)*, 2009. to appear.
- [Che76] P. P. Chen. The entity-relationship model – toward a unified view of data. *ACM Transactions on Database Systems*, 1:9–36, 1976.
- [DHPV00] A. Deutsch, R. Hull, F. Patrizi, and V. Vianu. Automatic verification of data-centric business processes. In *Proc. Intl. Conf. on Database Theory (ICDT)*, 2009.
- [Fer01] D. Ferguson. IBM Web Services: Technical and Product Architecture Roadmap. IBM Corporation, 2001. <http://www4.ibm.com/software/solutions/webservices/pdf/roadmap.pdf>.
- [F.F09] F.F. (Terry) Heath III and F. Pinel. Siena user’s guide, 2009. In preparation.
- [FHS09] C. Fritz, R. Hull, and J. Su. Automatic construction of simple artifact-based workflows. In *Proc. of Intl. Conf. on Database Theory (ICDT)*, 2009.
- [GM05] R.J. Glushko and T. McGrath. *Document Engineering: Analyzing and Designing Documents for Business Informatics and Web Services*. MIT Press, Cambridge, MA, 2005.
- [GS07] C. E. Gerede and J. Su. Specification and verification of artifact behaviors in business process models. In *Proceedings of 5th International Conference on Service-Oriented Computing (ICSOC)*, Vienna, Austria, September 2007.
- [HK87] R. Hull and R. King. Semantic database modeling: Survey, applications, and research issues. *ACM Computing Surveys*, 19:201–260, 1987.
- [HNN09] R. Hull, N.C. Narendra, and A. Nigam. Facilitating workflow interoperability using artifact-centric hubs. In *Proc. Intl. Conf. on Service Oriented Computing (ICSOC)*, 2009.
- [Hul08] R. Hull. Artifact-centric business process models: Brief survey of research results and challenges. In *On the Move to Meaningful Internet Systems: OTM Confederated Intl. Conf.s*, Monterrey, Mexico, 2008.
- [K⁺07] S. Kapoor et al. Sense-and-respond supply chain using model-driven techniques. *IBM Systems Journal*, 46(4):685–702, 2007.
- [KLW08] S. Kumaran, R. Liu, and F.Y. Wu. On the duality of information-centric and activity-centric models of business processes. In *Proc. Intl. Conf. on Advanced Information Systems Engineering (CAISE)*, 2008.
- [KNI⁺03] S. Kumaran, P. Nandi, F.F. (Terry) Heath III, K. Bhaskaran, and R. Das. ADoc-oriented programming. In *Symp. on Applications and the Internet (SAINT)*, pages 334–343, 2003.
- [LBW07] R. Liu, K. Bhattacharya, and F. Y. Wu. Modeling business contexture and behavior using business artifacts. In *CAiSE*, volume 4495 of *LNCS*, 2007.
- [Lin07] M. Linehan. Ontologies and rules in business models. In *Proc. 3rd Intl. Workshop on Vocabularies, Ontologies and Rules for the Enterprise (VORTE)*, 2007.
- [LRS02] F. Leymann, D. Roller, and M.-T. Schmidt. Web services and business process management. *IBM Systems Journal*, 41(2):198–211, 2002.
- [NC03] A. Nigam and N. S. Caswell. Business artifacts: An approach to operational specification. *IBM Systems Journal*, 42(3):428–445, 2003.
- [Obj08] Object Management Group (OMG). Semantics of Business Vocabulary and Business Rules (SBVR), Version 1.0., January 2008.
- [RDtHI09] G. Redding, M. Dumas, A.H.M. ter Hofstede, and A. Iordachescu. Modelling flexible processes with business objects. In *Proc. 11th IEEE Intl. Conf. on Commerce and Enterprise Computing (CEC)*, 2009.
- [SMS09] N. Sukaviriya, S. Mani, and V. Sinha. Reflection of a year long model-driven business and ui modeling development project. In *Proc. 12th IFIP Conference on Human-Computer Interaction (INTERACT)*, 2009.
- [SNK⁺08] J.K. Strosnider, P. Nandi, S. Kumarn, S. Ghosh, and A. Arsanjani. Model-driven synthesis of SOA solutions. *IBM Systems Journal*, 47(3):415–432, 2008.
- [SSRM07] N. Sukaviriya, V. Sinha, T. Ramachandra, and S. Mani. Model-driven approach for managing human interface design life cycle. In *Proc. Intl. Conf. on Model Driven Engineering Languages and Systems (MoDELS)*, pages 226–240, 2007.
- [vdAtHKB03] W.M.P. van der Aalst, A.H.M. ter Hofstede, B. Kiepuszewski, and A.P. Barros. Workflow patterns. *Distributed and Parallel Databases*, 14(3):5–51, July 2003.