Information Flow in the Federal Enterprise Redux:

Governing Federations, Sharing Information and Ensuring Privacy

richard.murphy@gsa.gov

Abstract: Today, government agencies face three fundamental challenges: governing federations, sharing information and ensuring privacy. To be effective, governance teams must better understand the structural and behavioral characteristics of the systems they govern. In this paper, I describe these characteristics as three qualitative governance assessment criteria called the Natural Laws of Federation. Given these laws, I describe an approach to sharing information among members of a federation based on Barwise and Seligman's Information Flow and Goguen's Theory of Institutions. Finally, I propose a policy interaction model as a next generation social contract and briefly describe the Citizen Privacy Service which ensures privacy, a timely public policy issue, in a federation based on the U.S. Privacy Act of 1974.

Each State, in ratifying the Constitution, is considered as a sovereign body, independent of all others, and only to be bound by its own voluntary act. In this relation, then, the new Constitution will, if established, will be a federal, and not a national constitution.

James Madison, The Federalist Papers, #39

Governing Federations: To deliver the best value to citizens for their tax dollars, cooperating programs across government agencies are engaged in cross-organizational initiatives. Through these initiatives, agencies hope to achieve efficiency in shared understanding and shared funding. But, along with the benefits of shared understanding and shared funding, come shared risk and limits on control, especially where agreements to cooperate post-date appropriations. Today, cross-organizational initiatives are wide spread and include organizations typically associated with hierarchical command and control such as defense, intelligence and law enforcement agencies. Because of limits on shared understanding and span of control, governance teams require knowledge of the structural and behavioral characteristics of complex adaptive systems which provide insights into a more effective governance model, often called a federation in U.S. civilian agencies. A wide range of literature is available on complex adaptive systems ranging from scientific research in the Emergence, Organization and Dynamics of Living Systems[1] to more popularized writings such as "Out of Control[2]."

1. SantaFe Institute, http://www.santafe.edu/, 2007

2. Out of Control, Kevin Kelly, 1994

Information sharing, a cross-organizational initiative, is limited by its association with heirarchical governance models. The Natural Laws of Federation are three qualitative assessment criteria, derived from complex systems theory, that provide insights into the structural and behavioral characteristics of effectively governing federations. The Natural Laws of Federation are:

1. Law of Approximation: Model theoretic and axiomatic semantics of governance structures must closely approximate the characteristics of the systems we govern. And federations are graphs and scale-free networks, not hierarchies.

2. Law of Emergence: Discovery precedes determinism. We don't all need to agree first, agree on everything, or agree on the same thing. And we need to share information beyond our span of control.

3. Law of Generativity: We'll know more tomorrow than we do today. Advancement, growth and sustainability require tolerance, independent invention, free extension, language mixing and partial understanding.

With these natural laws as background, I now explain how to share information among a federation of three agencies: the Defense Intelligence Agency (DIA), the Department of Homeland Security (DHS) and the Department of Justice (DOJ).

There is the Intentional Interpretant, which is a determination of the mind of the utterer; the Effectional Interpretant, which is a determination of the mind of the interpreter, and the Communicational Interpretant, or say the Cominterpretant, which is a determination of that mind into which the minds of the utterer and the interpreter have to be fused in order that any communication should take place. This mind may be called the commens. It consists of all that is and must be, well understood between utterer and interpreter, at the outset, in order that the sign in question should fulfill its function. This I proceed to explain.

Charles Sanders Peirce, Spring 1906

Sharing Information Among Members of a Federation: Today, sharing information among government agencies has elevated priority in cases of national and homeland security. Despite the rich history of information theory, government agencies lack sufficient theoretical background to effectively share information. In this section, I develop an information sharing scenario and show how members of a federation can better understand information sharing. The scenario begins with the restricted structural and behavioral characteristics of Information Flow[3], then uses Institutions[4] to better represent information sharing in a federation.

3. Information Flow: The Logic of Distributed Systems, Jon Barwise and Jerry Seligman, 1996

4. Information Integration in Institutions, Joseph Goguen, 2005

Information Flow as Metaphor: Information Flow provides the key metaphor for shared understanding in an information sharing environment. The connotation that information is mobile suggests that information stored in a remote system can be understood locally. Cell phone ring tones provide a timely example of information flow. Consider the statement: "The espionage ring tone carries the information that Ginny was the person calling Rick." Here, the cell phone is the remote system in which Rick assigned the espionage ring tone to calls from his girlfriend, Ginny. Locally, when Rick hears the espionage ring tone, it carries the information that his girlfriend is calling because of his prior knowledge that he assigned that ring tone to Ginny. More generally, we claim that a's being of type α carries the information that b is of type β.

Information Flow Principles: The cell phone ring tone scenario works because of the restrictions on the distributed system, composed of remote and local components. The principles of information flow (*IF*) describe these restrictions.

First Principle (P1): *IF* results from regularities in a distributed system

Second Principle (P2): *IF* crucially involves both types and their particulars (tokens)

Third Principle (P3): It is by virtue of regularities among connections that information about some components of a distributed system carries information about other components

Fourth Principle (P4): The regularities of a given distributed system are relative to its analysis in terms of information channels

Developing the Scenario: Consider information sharing among DIA, DHS and DOJ. If the distributed systems used by these agencies were to adhere to the principles of information flow, information sharing would not be difficult. Unfortunately that is not the case because the remote and local systems typically differ in their use of languages, logics, models and theories. And on cross-organizational initiatives, shared risk as well as limits on shared understanding and span of control cause these agencies to function as a federation. Before extending the system to support what one might expect from the Natural Laws of Federation, I develop a detailed analysis of the structural and behavioral characteristics of the system. The elements which define these structural and behavioral characteristics are: classifications, constraints, local logics, infomorphisms, institutions and channels.

Classifications formalize what we typically think of as components or modules in distributed systems. More formally, classifications are defined as follows:

Definition 1. *A classification* $A = \langle A, \sum_A, \vDash_A \rangle$ *consists of a set* A *of objects to be classified called tokens of* A*, a set* $\sum_A$ *of objects used to classify the tokens, called the types of* A*, and a binary relation* $\vDash_A$ *between* A *and* $\sum_A$ *that tells one which tokens are classified as being of which types.*

In the scenario, three classifications representing local and remote components in the distributed system are: 1) a set of Unified Modeling Language (UML) artifacts held in a Meta Object Facility (MOF) at DIA, 2) a set of Web Ontology Language (OWL) artifacts held in a triple store at DHS, and 3) a set of XML Topic Maps (XTM) held in a meta-data repository at DOJ.

Figure 1 - Three Classifications Representing Local and Remote Components

MOF	Triple Store	Repository
$\sum_M$	$\sum_T$	$\sum_R$
$\vert\vDash_M$	$\vert\vDash_T$	$\vert\vDash_R$
M	T	R

We can also formalize the constraints, or regularities, in the distributed system. It is these constraints in the ring tone example that allow information to flow.

Definition 2. *Let A be a classification and let $\langle \Gamma, \Delta \rangle$ be a sequent of A. A token a of A satisfies $\langle \Gamma, \Delta \rangle$ provided that if a is of type α for every $\alpha \in \Gamma$ then α is of type α for some $\alpha \in \Delta$. We say that Γ entails Δ in A, written $\Gamma \vdash_A \Delta$, if every token a of A satisfies $\langle \Gamma, \Delta \rangle$, If $\Gamma \vdash_A \Delta$ then the pair $\langle \Gamma, \Delta \rangle$ is called a constraint supported by the classification A.*

Implicit in the constraints are assumptions concerning knowledge, perception and belief in the distributed system as embodied in a local logic. In the scenario, the constraint that both the model theoretic and axiomatic semantics of the remote and local systems are identical and that local logics are sound and complete does not hold. For the federation to share information, I first define local logics and infomorphisms, then extend the scenario with institution morphisms.

Definition 3. *A local logic $\mathfrak{L} = \langle A, \vdash_{\mathfrak{L}}, N_{\mathfrak{L}} \rangle$ consists of a classification A, a set $\vdash_{\mathfrak{L}}$ of sequents (satisfying certain structural rules) involving the types of A, called the constraints of $\mathfrak{L}$, and a subset $N_{\mathfrak{L}} \sqsubseteq A$, called the normal tokens of $\mathfrak{L}$, which satisfy all the constraints of $\vdash_{\mathfrak{L}}$. A local logic is sound if every token is normal; it is complete if every sequent that holds of all normal tokens is in the consequence relation $\vdash_{\mathfrak{L}}$.*

An infomorphism allows information to flow across classifications that satisfy the constraints, or regularities of the system.

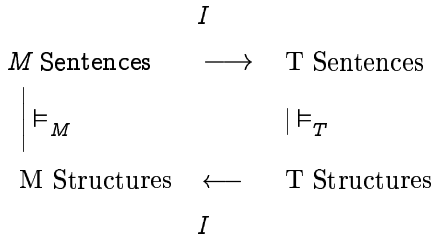
Definition 4. *If $A = \langle A, \sum_A , \vDash_A \rangle$ and $C = \langle C, \sum_C , \vDash_C \rangle$ are classifications then an infomorphism is a pair $f = (f^{\wedge}, f^{\vee})$ of functions satisfying the analogous biconditional: $f^{\vee}(c) \vDash_A \alpha$ iff $c \vDash_C f^{\wedge}(\alpha)$ for all tokens c of C and all types α of A. An infomorphism is represented concisely as $f: A \rightleftarrows C$*

Because the classifications and local logics differ in their model theoretic and axiomatic semantics, we require Institutions for information to flow.

Definition 5. *An Institution consists of an abstract category* Sign*, the objects of which are signatures, a functor* Sen: Sign $\rightarrow$ Set*, and a functor* Mod: Sign^{op} $\rightarrow$ Set*. Satisfaction is then a parameterized relation* $\models_\Sigma$ *between Mod(*Σ*) and Sen (*Σ*), such that the following condition holds for any signature morphism* ψ: $\Sigma \rightarrow \Sigma'$*, any* Σ'*-model* M'*, and any sentence e,* $M' \Vdash_{\Sigma'} \psi(e)$ iff $\psi(M') \Vdash_\Sigma e$ *where* $\psi(e)$*abbreviates* Sen$(\psi)(e)$ *and* $\psi(M')$ *abbreviates* Mod$(\psi)(e)$.

Where infomorphisms allow for structure preserving transformation, institution morphisms preserve satisfiability through semantic preserving transformation. Structure preserving transformation can result in a loss of decideability. Consider the transformation from DIA's UML MOF to DHS's OWL triple store. A UML (Class, Association, Class) structure classifies as ALHOIN(D) in description logic[5]. This structure transforms to the OWL structure (Class,Property,Class), classified in description logic as OWL Full.

Figure 2. Semantic preserving transformation between DIA and DHS.



Now that two members of our federation can share information, we need to define an information channel to generalize information sharing.

Definition 6. *An information channel consists of an indexed family* $C = \{ f_i : A_i \rightleftarrows C \}_{i \in I}$ *of institution morphisms with a common codomain C, called the core of the channel.*

In addition to allowing DOJ, or any member of the federation, to participate in one channel, the Natural Laws of Federation, or complex systems theory, suggest that we need may need more than one channel, or that we allow members to change channels, to more effectively share information.

Institutions, algebraic specifications, semiotics, and theorem provers using techniques such as composition by colimit[6], unskolemization and connections[7] represent tomorrow's approach to information sharing. Currently, knowledge transfer from the academic research community to the public sector is slow and without performance-based incentives, the private sector currently limits investment in these essential developments to sharing information.

---

5. A Description Logic for Use as the ODM Core, Lewis Hart, 2004

6. Composition by Colimit and Formal Software Development, Douglas R. Smith, 2006

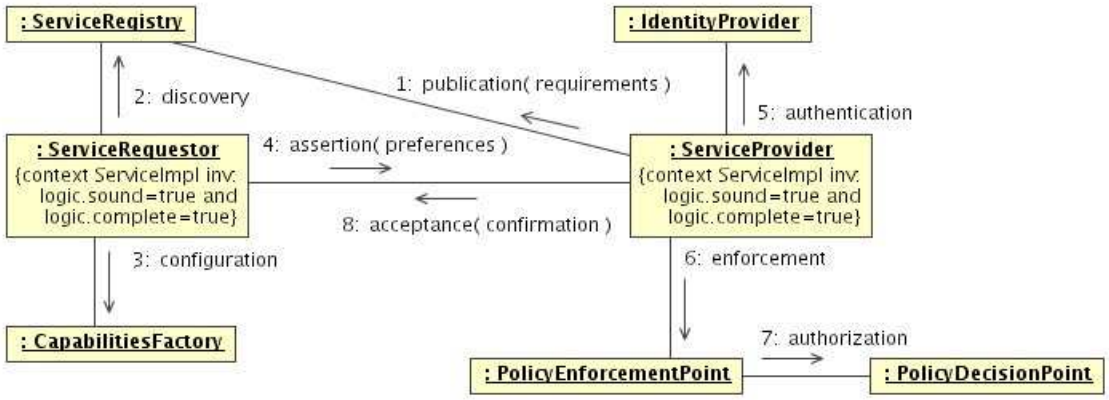
7. Constructing Specification Morphisms, Douglas R. Smith, Kestrel Institute, 1992

The reason why men enter into society, is the preservation of their property; and the end why they chuse and authorize a legislative, is, that there may be laws made, and rules set, as guards and fences to the properties of all the members of the society, to limit the power, and moderate the dominion, of every part and member of the society.

John Locke, Second Treatise on Government

Ensuring Privacy in Federations: Today, the unintended disclosure of personally identifiable information represents a risk to citizens and to government agencies. The ability to ensure privacy among members of a federation is essential to preserve public trust when sharing information. Figure three illustrates a next generation social contract in the collection, use, maintenance and disposition of personally identifiable information in government agencies.

Figure 3 - Policy Interaction Model - A Next Generation Social Contract



The Citizen Privacy Service (CPS) (see http://us-privacy.sourceforge.net/), an OSERA (see http://www.osera.gov/) reference implementation, is an asynchronous component that plugs into an Enterprise Service Bus to provide highly scaleable, policy decision and policy enforcement points based on the U.S. Privacy Act of 1974. In response to a request to disclose a citizen's personally identifiable information, CPS checks the satisfiability of the request and returns to the service requestor the first order logic reasoning steps which support its conclusion. These reasoning steps, or information provenance, provide a high assurance, policy enforcement capability for agencies of the U.S. Federal government.