

Digital Identities Can Tame the Wild, Wild Web

Adarbad Master

Controlled data exchange is the next important growth phase of the Web and related technologies. Digital identity technology provides a foundation to realistically model real-world entities, their attributes, and the exchange of data between them as individual data authorities in a trusted data exchange environment.

What if you were asked to share your financials while visiting your doctor for a bad case of the flu? Or conversely, asked to hand over your medical history before the bank agrees to grant you a home loan? In the real world, an individual plays many different roles in the course of his daily activities. People purchase homes where they are first the “buyer” and then the “owner,” become an “employee” to pay back the money they borrowed for the house, and so on. These roles define the relationships people have with society.

Fortunately, society has done a reasonable job in imposing sensible information exchange limits on such interactions. But with the Web, the story is completely different.

The phenomenal growth of the Web has resulted mainly from the medium’s potential and capacity for information access. Social and technology critics, however,

now point to the lack of information protection and inadequate access restrictions as the Internet’s weakest links. The problem of confidently and securely moving data across network trust boundaries remains mostly unsolved. Today, most information attained via the Web is both publicly and anonymously accessible. While this may prove to be a non-issue in existing use cases, this uncontrolled dissemination of information stands in the way of harnessing Internet technologies for the exchange of sensitive data.

There are few recognized standards to govern data sharing on the Internet, which is one of the main reasons why identity management solutions have topped the list of what IT managers are concerned about today. According to the April 2004 Heat Index, identity management, user provisioning, and single sign-on have emerged as the top-three information security spending priorities for Fortune 1000 companies.¹

This shift in spending can most likely be attributed to the fact that IT managers are aware that on the Web, almost anyone can be anonymous. Those who dare to share private information often fall victim to malicious intruders, or to ambiguous and unenforceable data exchange policies.

ADARBAD MASTER is chief technology officer at Epok, the leading provider of trusted data and identity exchange solutions. Master’s areas of expertise lie in Internet services, open systems deployment, scalable distributed computing and large database system design and management. He can be reached via e-mail at amaster@epok.net.

The benefits of cross-domain data exchange can tame the wild Web.

Businesses and individuals complain that information shared on today's Web is essentially information given away. The benefits of cross-domain data exchange can tame the wild Web.

BENEFITS OF TRUSTED DATA EXCHANGE

What if the exchange of data could be strictly controlled? This article describes technology standards and solutions that significantly lessen the risks associated with inter-organizational data sharing. Furthermore, expanding the network trust boundaries between organizations will bring the following benefits:

- The business value of sensitive "back-office" data would increase
- Data once trapped in "silos" can be made available to trusted partners
- Orchestrating data from disparate systems becomes less complicated and less expensive
- New communications channels with supply chains will emerge
- Enabling technology standards will evolve toward ubiquity

PROBLEM ANALYSIS

Many facets of today's Web do not lend themselves to the private and secure exchange of sensitive or otherwise non-public information. The exchange of business, financial, political, and personal data thus requires either the employment of private networks or new technologies that enable the controlled exchange of data across network trust boundaries. The limitations of the Web are further amplified by the emergence of mobile data networks and by the obvious shortcomings of current extranet environments, in which data exchange is unduly burdensome and extremely expensive to sustain. Compounding the complexities of the extranet is the tremendous diversity of data types that exist in a typical multi-organizational data-sharing environment. While the advent of XML and Services Oriented Architecture (SOA) does show promise for data comprehension, a

complete solution to the controlled exchange of sensitive data will require:

- Global standards to identify all data exchange participants (receivers and senders)
- Mechanisms for explicitly addressing data for nonambiguous exchange
- The ability for organizations to share data in a universally comprehensible format
- Standards for referencing data owned by trusted partners
- A methodology for governing data exchange and a set of rules defining related obligations
- Mechanisms to ensure that the data shared is the most up-to-date data available

Much time has been spent developing file formats and protocols for data exchange, but not much has been spent considering the ownership of exchanged information and foundations of interorganizational trust. Using standards like Secure Socket Layer (SSL), one can virtually guarantee that data cannot be intercepted, but one cannot prove that the receiving party is entitled to the data. Nor can one restrict how data is used once it has left one's domain. Greater control of data exchange can be achieved through new standards and identity management solutions.

What is needed today is an abstraction layer above the data that provides a flexible platform to negotiate and establish trust between communication endpoints. This data framework would bring identity control, coordination, and orchestration to Web services and allow location independence, loose coupling, and dynamic interoperability of SOA components.

DIGITAL IDENTITY AS THE FOUNDATION OF DATA AUTHORITY

Organizations also have information they protect, particularly if they intend to share it. An enterprise may share sales pipeline status with its supply-chain partners for just-in-time inventory, but it certainly would not want this information to be exposed to competitors. Or an organization

might share summaries of the quality of its partners' supplies, but not disclose more specific quality information to all partners. Such limits are typically based on corporate policy, not personal relationships. Such information belongs to the organization, not to any single individual employee or partner, and organizations must take steps to prevent unwanted disclosure.

Both people and organizations alike take steps to establish a foundation of trust between the parties involved in a relationship. In low-risk situations, this is accomplished by social norms; but for important matters, the parties use explicit agreements: contracts. Contracts specify the nature and details of the relationship and record the parties' agreement with its terms. For example, when two organizations formalize a partnership to supply parts, both parties agree to a price per part, supply or delivery timeframes, length of relationship, payment terms, required quality levels, etc. If the contract is broken, the civil court system can be used to recover damages from the party in breach, giving both parties an incentive to keep the agreement.

Establishing this level of trust in the digital world is challenging. As the Web has grown, it has focused largely on enabling rather than protecting or restricting communication. Relationships defined in real-world contracts become unclear and difficult to manage in cyberspace. The Web did not evolve to reflect this aspect of the real world and, as a consequence, information is far less controlled on the Web than in the physical realm. This is the problem that digital identity is intended to solve. Digital identity is a way of defining the roles and interactions between individuals and organizations to help developers build systems that represent these interactions in a fashion consistent with the real world.

Digital identity is a method to computationally represent real-world people, organizations, or objects as authorities of their attributes for the purpose of secure information access and exchange. The following standards and technologies have been

developed to expand the use of digital identity — in concert with Web Services — to increase the level of control in cross-domain data exchange.

OPEN SPECIFICATIONS

New emerging technology standards out of the OASIS (Organization for the Advancement of Structured Information Standards) help expand the scope of using digital identities to secure communications between Web Services endpoints.

Extensible Resource Identifier (XRI) is a specification that allows unambiguous identification of data objects within an enterprise, a federation, or a global community. XRI Data Interchange (XDI) is a specification that allows data authorities (i.e., identities) to negotiate the secure exchange of information. More established identity protocols and standards, such as Liberty Alliance, focus on the unification of access and exchange of attributes of a single principal whose data is scattered in multiple federated identities dispersed among service providers. Liberty is more problem-focused than some of the other approaches and, while it solves those well, may not be as applicable to problems in network identity that do not fit its model.

Security Assertion Markup Language (SAML), from OASIS, is another well-established protocol to securely exchange authentication and authorization information in a federated network and is generally well adopted, even among opposing camps.

Microsoft and IBM are attempting to provide low-level, modular building blocks that provide the basis for interoperable, identity-based Web Services, but they are not particularly open and face strong market resistance.

Harmonization of these protocols is inevitable, but the end result is entirely unpredictable at this point. Consequently, interoperability outside a fairly closed community is unachievable in the short term. However, XRI/XDI are standards that may well answer key issues in the Liberty framework, and inasmuch as the latter is willing

As the Web has grown, it has focused largely on enabling rather than protecting or restricting communication.

to incorporate foreign work, may well have a path to be merged.

Extensible Resource Identifier (XRI)

In January 2003, the Extensible Resource Identifier (XRI) Technical Committee was formed at OASIS to create a new Uniform Resource Identifier (URI)-compatible identifier suited to the requirements for digital identity. The XRI specification was completed in January 2004, and the Technical Committee unanimously approved it in advance of its submission for consideration as an OASIS standard.

XRIs lay the foundation for digital identity by providing a way to identify resources. XRIs are an extended version of URIs that include several features unavailable in URIs or other traditional identifiers. XRIs are “location independent.” The content of an XRI is decoupled from the network location of any data or services associated with the XRI. This means, among other things, that access to a resource identified by an XRI is not necessarily limited to a particular network location or protocol.

XRI syntax provides the ability to indicate that some segments of the identifier are long-lived “primary keys” (a concept borrowed from database technology) while others are human-meaningful identifiers that may be subject to change.

XRIs also incorporate the concept of cross-reference, which provides the ability to “contain” other URIs or XRIs in an XRI identifier. Conceptually, this is similar to quoting someone else’s text to refer to it. Cross-references allow a well-formed URI or XRI to identify not only a concept or resource, but also to identify that concept or resource relative to another concept or resource.

The XRI namespace features unlimited delegation. While URI schemes rely on DNS delegation, XRIs have the ability to use abstract (non-DNS) names or identifiers that can contain a wider set of characters and strings. The XRI specification defines several global namespaces for abstract non-

DNS identifiers (e.g., “@,” “=,” and “+,” which are for organizational, personal, and general names, respectively). XRIs are built from the ground up to use Unicode for internationalization, allowing data exchange across language boundaries.

The XRI specification defines a resolution protocol that converts an abstract XRI into the set of concrete locations and interaction protocols supported by the identified resource. That is, an XRI identifies a resource abstractly, while the result of resolution identifies a resource concretely. XRI resolution can be viewed as the discovery of network endpoints associated with an XRI, together with the services and data offered by those endpoints. XRI resolution is iterative and delegated, somewhat like DNS, but occurs at a higher level using XML documents retrieved via HTTP.

XRI Data Interchange (XDI)

XRI Data Interchange (XDI) is a continuation of the work on XRI. A new OASIS technical committee was formed in February 2004 to build on the features of XRI addressing to establish a standard schema and XML-based protocol for cross-context data sharing, linking, and synchronization. XRI will feature a simple, flexible, and extensible model for security, privacy, and permission management. Work on the XDI standard is currently underway, and the technical committee expects to publish draft specifications in late 2004.

Extensible Name Service (XNS)

XNS is a comprehensive open specification that provides a flexible, interoperable method for establishing and maintaining persistent digital identities and the relationships between identities. The protocol provides services for registering and resolving identity addresses, defining and managing digital identities as XML documents, conducting and protecting identity transactions, and linking and synchronizing identity attributes. Distinguishing characteristics of XNS are its peer-to-peer nature and its focus

on data sharing and relationship management.

XNS was originally managed by the XNS Public Trust Organization, also known as XNSORG. In 2003, the specification was contributed to OASIS (<http://www.oasis-open.org/>) and became the basis for the above two related standards efforts: XRI and XDI.

Liberty Alliance Project (LAP)

Liberty, much like the protocols above, deals with issues fundamental to identity management: federation and data exchange. In this context, federation typically describes the process of linking various user identities that already exist in disparate systems. Once identities are linked, a user can move seamlessly between systems (single sign-on) or transfer data from one system to another. The LAP effort has been divided into several phases:

Phase 1 is essentially an extension of SAML that defines a federation protocol, pseudonymous SSO, and authentication discovery. Federation protocol describes the way existing accounts at multiple locations can be linked, managed, and potentially broken.

Phase 2 defines an infrastructure that allows an attribute consumer to discover a user's attribute provider for a class of attributes and retrieve those attributes.

Phase 3 will focus on defining the data and messages required to solve more targeted business problems, expressing them in terms of the framework defined in phase 2.

The Liberty Alliance and its members are contributing to its evolution with industry leading technologists in the Identity Federation field. Incorporation of concepts from the LAP and other protocols into vendors' application infrastructures is a commitment that leads to continuous roadmap revisions. Harmonization of the various protocols in this space is inevitable, but the end result is entirely unpredictable at this point. Consequently, interoperability outside a fairly

closed community is unachievable in the short run. However, the ideal strategy is to provide a platform that will allow carriers and enterprises to create data exchange applications within their community and yet remain nimble enough to incorporate the appropriate standards and best practices as they become available

TRUSTED DATA EXCHANGE

To achieve the level of control required to exchange data on the Internet, any proposed solution must identify, label, present, link, govern, and synchronize data.

Identify

A fundamental component of controlled data exchange is the ability to identify all the participants involved in the exchange. Knowing with whom the data will be shared is the first step. The open standards identified above can be leveraged to extend the addressing capabilities of the Internet, allowing any entity to be represented on the network. XRIs can serve as unique, persistent identifiers that ensure an exclusive network location for each participant in a data exchange. The use of the term "participants" means any entity, including data.

XRI's roots illuminate how this new standard helps define data authorities. XRI evolved from XNS addressing, which was originally envisioned as the next step beyond the Domain Name Service (DNS). DNS resolves hostnames (e.g., www.example.com) into the numeric addresses (e.g., 192.1.2.3) computers need to establish connections to the named hosts. DNS also maintains and distributes other information about Internet domains and hosts, such as which mail servers are used to deliver mail to particular domains. Whereas DNS lets one name only a limited number of things (e.g., host names and mail servers), XNS and thus XRI allows one to name anything that will resolve to a globally unique digital identity. For example, the XRI address @example would represent the data authority (or digital identity) of Example Incorporated. XRI provides the means of addressing

Liberty deals with issues fundamental to identity management: federation and data exchange.

XRI provides non-reassignable identifiers that uniquely and permanently identify individuals and organizations as the data authority.

an identity by name or by some other identifier.

Identify: XRI provides non-reassignable identifiers that uniquely and permanently identify individuals and organizations as the data authority.

Label

When sharing sensitive data, there is simply no room for error or miscommunication. A way to specify exactly what data one intends to share, with no room for misinterpretation, is imperative for trusted data exchange. Labels assist in the location of data and eliminate ambiguity in both data sharing and data governance as agreements between participants are established (see “Govern” below).

Data labeling extends the identifier to address all data that will be shared, allowing data authorities to distinguish each piece of data they control and to label it unambiguously. The label is used to indicate what data will be exchanged and where it is located. Labels bring a new level of control and management to all data under the data authority’s realm.

Data labeling provides additional capabilities. Persistent labels can be used to ensure that changes to the label’s name do not hinder access to the information. Data versioning provides a way of maintaining a history of a data element’s value and accessing any of its previous states.

Label: Extends the XRI identifier to unambiguously address data elements (i.e., single values, collections, files, etc.).

Present

Exchanging data among various participants calls for a widely recognized standard data format. This is the promise of XML, and a fundamental requirement for controlled data exchange.

Enterprise information systems maintain data in a wide variety of formats across systems and domains. Leveraging XML to present data from various data sources in a

single view results in significantly simplified data integration and improved control of data sharing.

In addition, the use of XML as the common data format lowers the barriers for many organizations that would benefit from exchanging data. It simplifies the integration process and dramatically reduces related IT expenses, thereby enhancing the value of both internal and external data sharing.

Present: The use of XML to define data in a globally comprehensible format.

Link

Data sharing begins with establishing a relationship, or “link,” between data authorities. The XNS and XDI standards provide this linking capability.

The link is the conduit between digital identities that are sharing data. The link, and the data shared across it, is exclusive to the relationship. The link provides assurance that only the intended data is provided, and only to the correct party.

Link: The foundation for data exchange that establishes the relationship between two data authorities for sharing data.

Govern

The exchange of information between identities is a core concept of both real-world relationships and digital ones. Many kinds of relationships can be established, from a salesperson giving a business card to a customer to an organization providing contractual services to another for a period of time.

Relationships are established by transactions governed by social norms (for low risk) or legal contracts (for high risk). XNS provides services for building contracts between identities, recording the addresses of the data to be shared, along with any privacy constraints.

Govern: Digital identities can establish obligations for data usage within the context of data sharing.

Synchronize

Most federated data-sharing solutions are only federated in the way they send data to external participants. Data recipients are federated but the data sharing environment is centrally controlled and monolithic. In contrast, some may envision an environment in which participants create and sustain their own network identities. This unique concept is the foundation for a truly federated environment in which data can move in both directions between participants — and all participants can access information as it changes.

The value of a highly — federated data exchange environment is quickly becoming more recognized. While data synchronization is becoming a common feature in centralized data stores, the ability to extend synchronization to the federated world enables new and exciting categories of private data usage.

This delivery of these e-offerings will likely drive the creation of federated data stores inside:

- Wireless carriers:* address book and calendar event sharing
- Healthcare provider systems:* patient records
- Criminal justice:* interagency records of offender activity
- Retail:* cross-retailer manufacturer performance data
- Defense:* inter-branch activity monitoring
- Medical research:* multi-source pharmaceutical testing

Synchronize: Governed data actively flows between digital identities to ensure data that is up-to-date.

TRUSTED DATA EXCHANGE IN THE CARRIER

Mobile carriers are struggling with providing an ever-increasing array of data services to a consumer population that has proven to be extremely discriminating in what it consumes. The range of services offered, from static Web-based content to interactive

video services, has resulted in a patchwork of services all performing in a silo fashion on a common infrastructure. The drive to develop integrated *service creation environments* that are more dynamic and responsive to consumer needs has resulted in adoption of platforms such as Web services application framework, identity management systems, and rich-media presentation environments. These, and other components, are being stitched together by carriers into delivery platforms for data services and content.

The solution introduces the concept of control at the data level, creating in essence a carrier data framework (CDF) for the service creation environments (SCE). The CDF provides a carrier with the following benefits:

- Rapid integration of data from any source within the SCE
- Data services platform for access, transfer, and exchange of shared data between SCE components and between SCE and legacy carrier services
- Extension of data about devices, applications, and subscribers from within the carrier through enabler services to third-party content and service providers and affiliate carriers
- Social network applications with high-value reusable subscriber linkage based on controlled data exchange
- Inter-carrier peer-to-peer data services based on subscriber-level control of content distribution and synchronization
- Trusted E-commerce across portals through digital contracts that enable relationship-level controls between subscribers and businesses
- Privacy compliance across trust boundaries is provided by “Terms and Conditions” that contractually obligate privacy controls

CONCLUSION

Controlled data exchange is the next important growth phase of the Web and related technologies. Digital identity technology

provides a foundation to realistically model real-world entities, their attributes, and the exchange of data between them as individual data authorities in a trusted data exchange environment.

Data-sharing applications must be built on servers to enable the controlled exchange of virtually any kind of data. The ideal result is a trusted environment in which participants are identified, where data is explicitly labeled and universally comprehensible,

where obligations for data protection and data handling are clearly defined in the form of digital contracts, and where the data is linked and synchronized so it stays fresh. This is the Data Web — a new way of thinking about data interchange.

Note

1. TheInfoPro (TIP) interviewed 175 Fortune 1000 companies about their purchasing plans for IT security products and services.

© Adarbad Master. Used by permission.