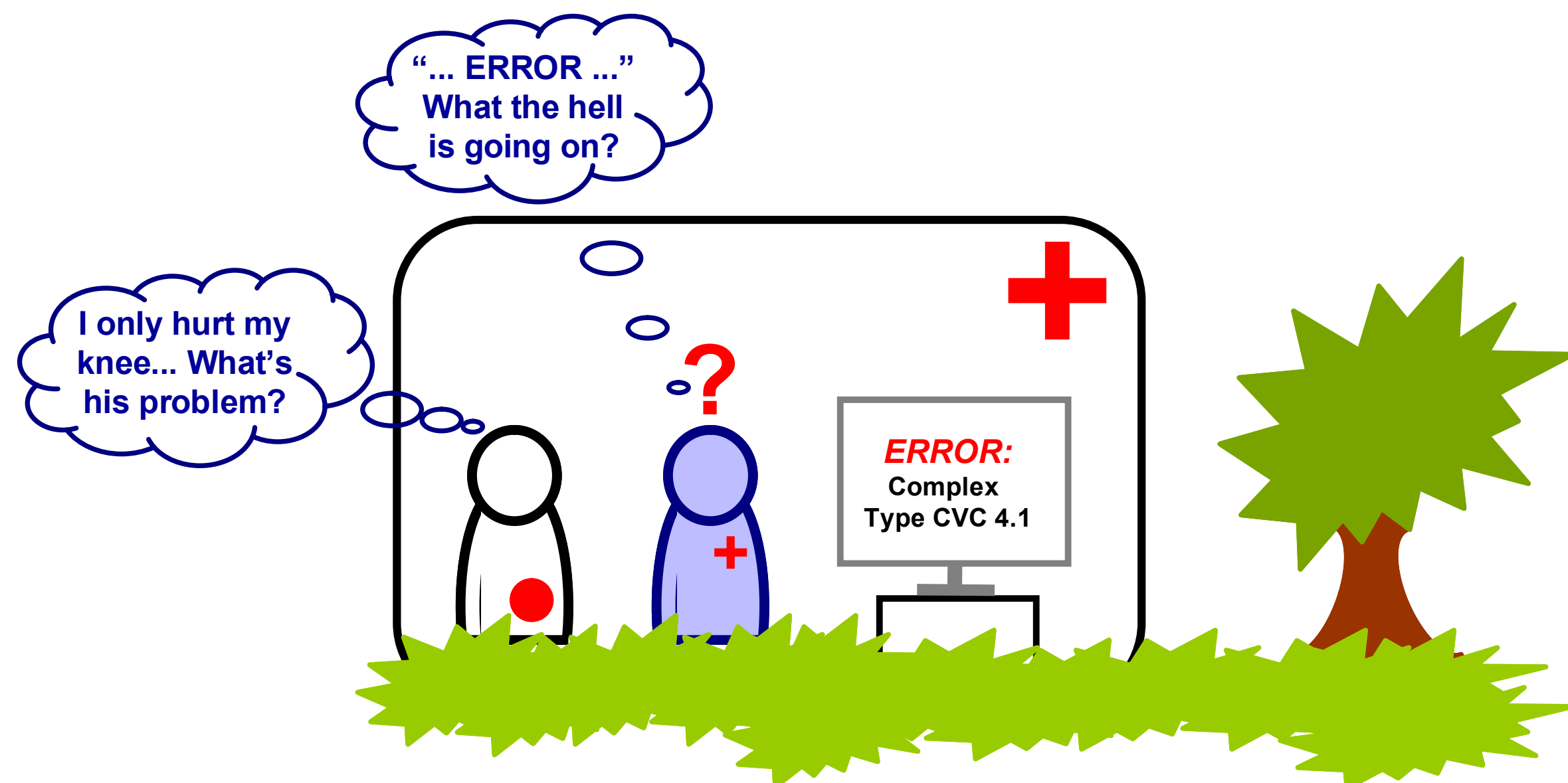


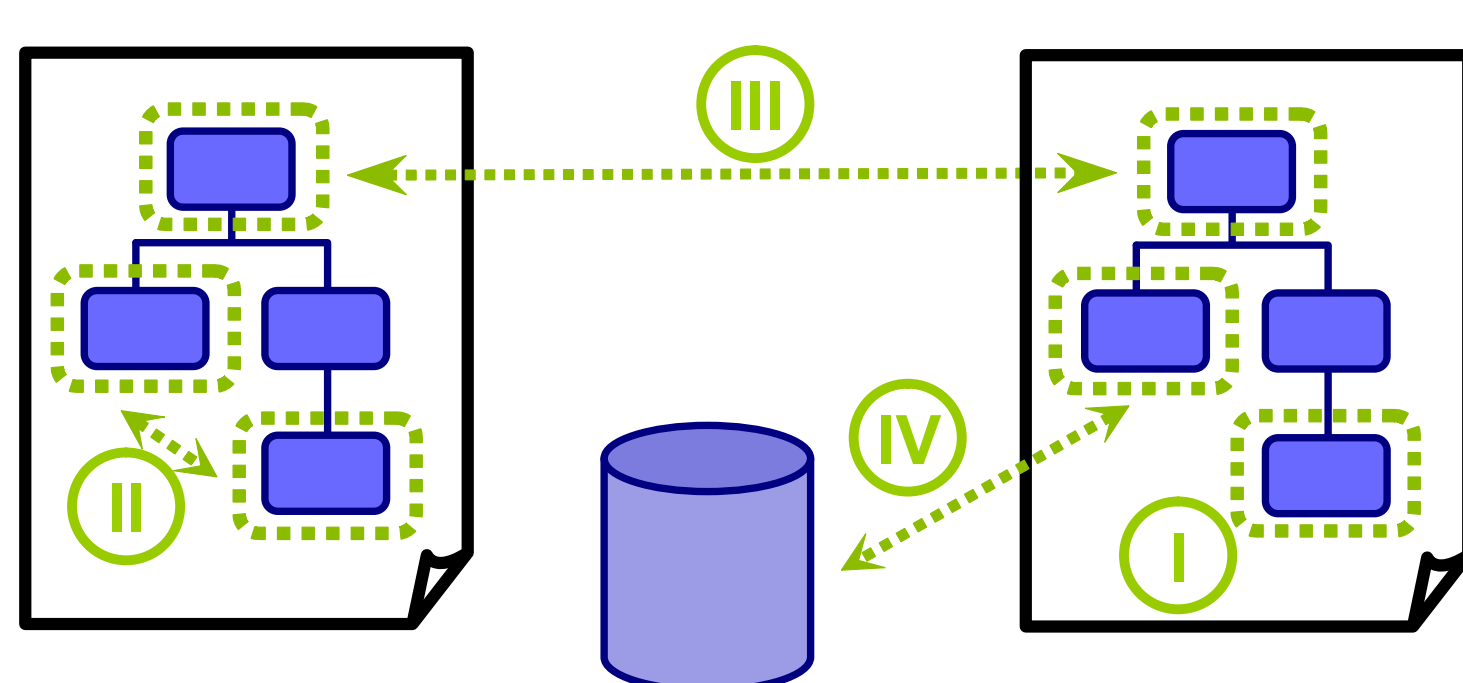
Motivation



Problems, Goals and Requirements

- **Problems:**
 - Healthcare information represented electronically may contain erroneous information
 - Humans can't evaluate masses of healthcare information
- **Goal:**
 - Identify sources where erroneous information may occur
 - Identify technology, including hardware and software, that can be used to validate healthcare information to avoid erroneous information
- **Requirements:**
 - Constraint definitions should be separated from application and data
 - Implementation should be in an SOA Environment
 - Quality of Error messages that result from validation failures should be improved

Constraint Classification



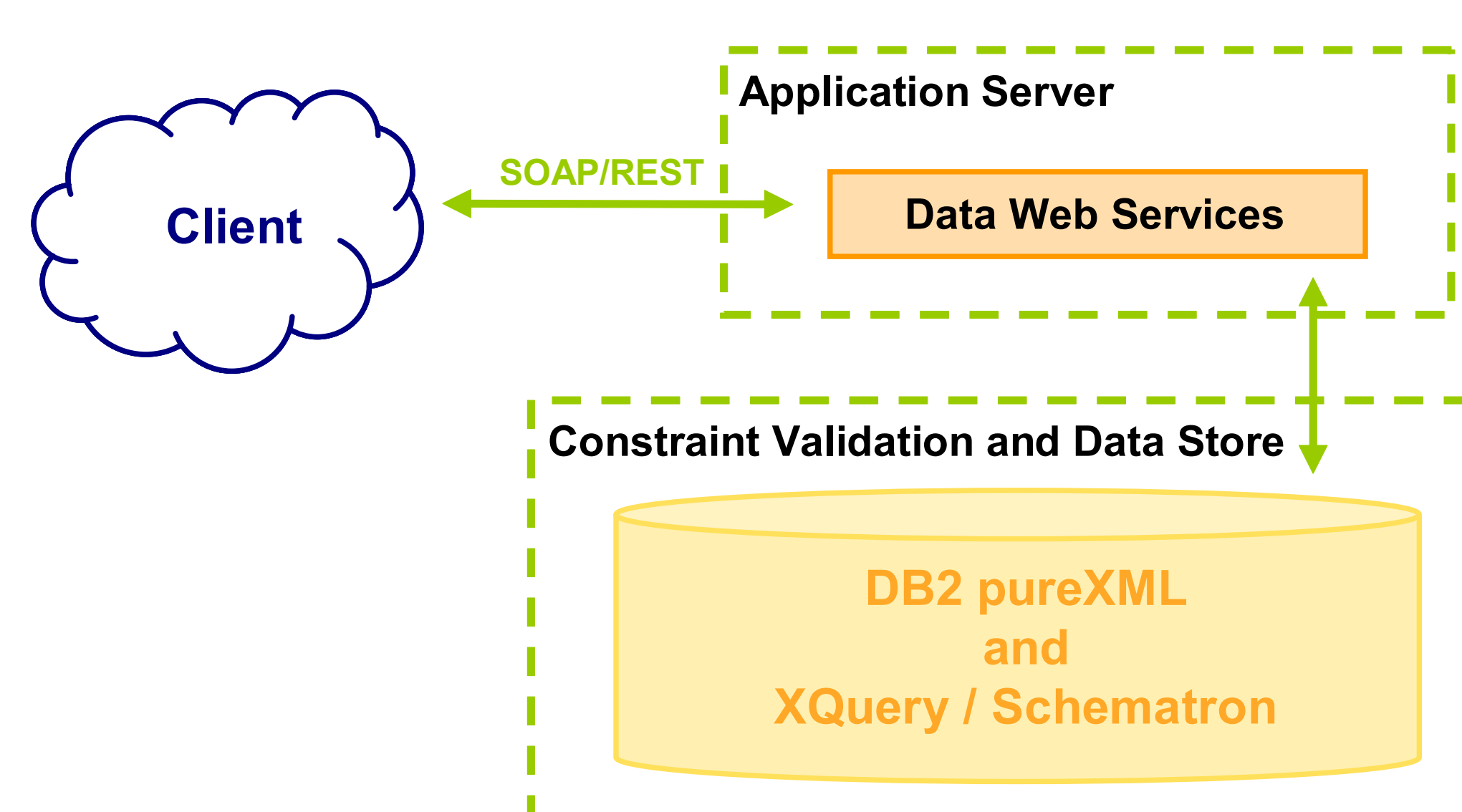
- **Type I:** Structural Requirements
- **Type II:** Reference data within XML document instances
- **Type III:** Reference data across XML document instances
- **Type IV:** Reference data between XML document instance and non-XML data source

Constraint Notations and Technology

- **Notations**
 - **XML Schema:** defines the structure of XML document instances
 - **XQuery:** queries the content of XML document instances
 - **Schematron:** defines assertions on the content of XML document instances
- **Technology**
 - **DB2 pureXML:** database system managing relational and XML data natively
 - **Data Web Services:** allow to expose database operations as Web Services
 - **WebSphere DataPower SOA Appliance:** hardware to process XML

Evaluation

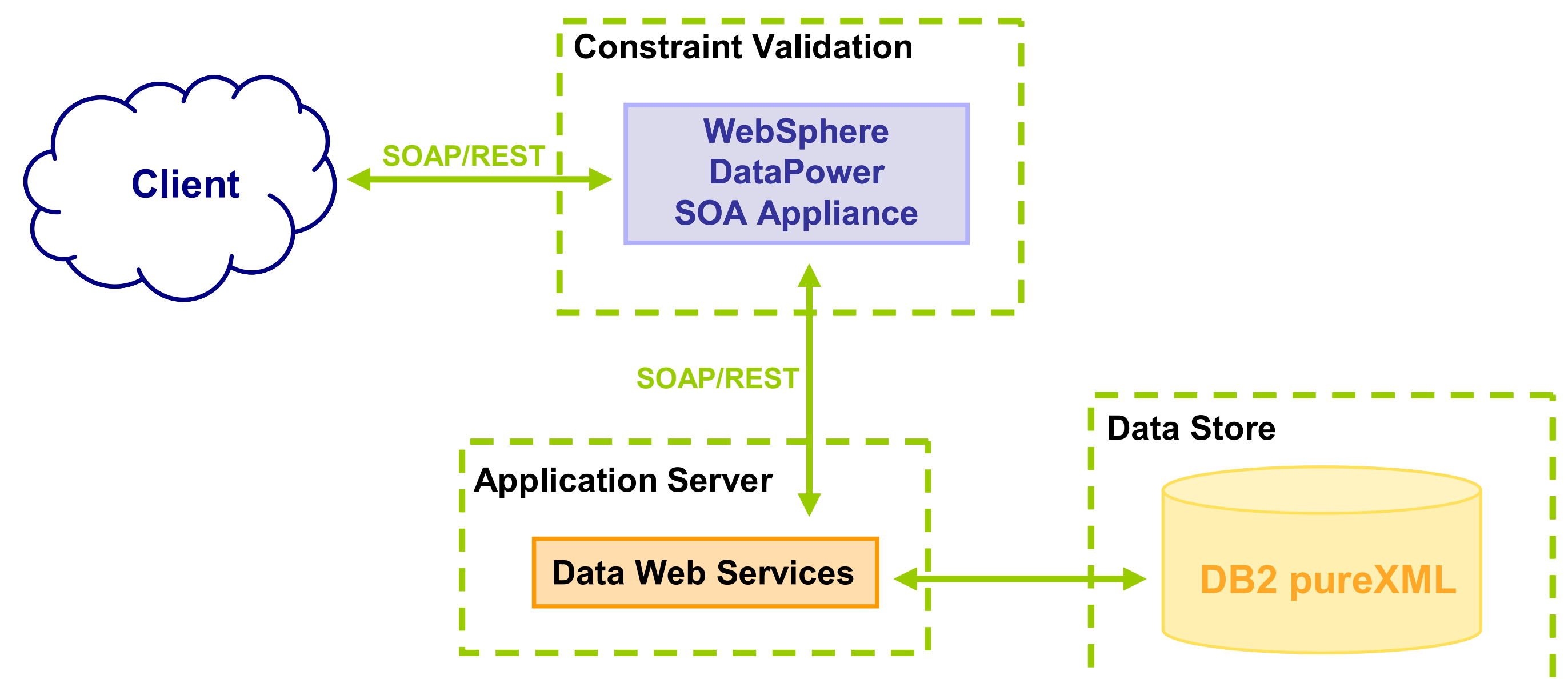
Architecture of Approach 1 and Approach 2 utilizing DB2 pureXML



Approach 1: XML Schema and XQuery [1] [2]

Approach 2: XML Schema and Schematron [1] [3]

Architecture of Approach 3 utilizing WebSphere DataPower SOA Appliance



Approach 3: XML Schema and Schematron [1] [4]

Comparison

	DB2 pureXML (Approach 1)	DB2 pureXML (Approach 2)	WebSphere DataPower SOA Appliance (Approach 3)
Constraint Definition	XML Schema and XQuery	XML Schema and Schematron	XML Schema and Schematron
Constraint Types Covered	Type I	✓	✓
	Type II	✓	✓
	Type III	✓	X
	Type IV	✓	X
Constraint Mechanism	implementation of diverse stored procedures	implementation of diverse stored procedures	configuration only
Constraint Validation	parallel	sequential	sequential
Constraint Violations	store valid as well as invalid information	store valid as well as invalid information	store valid as well as invalid information
Error Messages	customizable	customizable	customizable

Outlook

- **Other industries** next to the healthcare environment also define specific XML-based formats for information exchange (e.g. finance, government, business)
- **Trend** to store information in the same format as it is exchanged ("what you exchange is what you store")
- **Recommendation** to pursue the development of **constraint notations** that may be **applied** to various **industry formats**
- **Development of Schematron extension** that allows to define **Type III** and **Type IV** constraints

Literature

- [1] Susan Malaika, Christian Pichler. **Universal Services for pureXML using Data Web Services.** In: IBM developerWorks, May 2008
- [2] Keith Wells, Susan Malaika, Christian Pichler. **XForms and DB2 pureXML.** In: IBM developerWorks, May 2008
- [3] Susan Malaika, Christian Pichler. **WebSphere DataPower and DB2 pureXML, Part 1: XML schema and content validation using WebSphere DataPower and DB2 pureXML.** In: IBM developerWorks, June 2008
- [4] Susan Malaika, Christian Pichler. **WebSphere DataPower and DB2 pureXML, Part 2: DB2 pureXML as an audit log for WebSphere DataPower.** In: IBM developerWorks, June 2008