RSACONFERENCE CHINA 2011 2011 信息安全国际论坛

the adventures of alice book

基于OWL的电子记录存储与取证 技术研究

**Electronic Record Storage and Forensic Technology** 

Research based on OWL

演讲人: 龙毅宏

职务:首席技术顾问

北京天威诚信电子商务服务有限公司

#### 大 纲

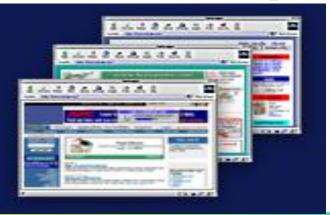
- 二 背景
- **OWL** (Web Ontology Language)
- □ 基于OWL的记录定义与描述
- □ 基于OWL的记录存储与取证
- □ 基于OWL的记录存储与取证系统

# 电子记录(Electronic Record)

- 记录的电子形式
- · 多为数字记录(digital record)
- 电子记录的应用与种类

- 电子订单(purchasing order)、电子合同(contract)、电子病例(medical record)、监控录像(monitor video)、

操作日志(operation log)





章。分報		施爾爾公河爾以 也打定改逐名產單。原先由施爾爾蓋單,其次开始於頓齊蓋 中进行產章,比如維爾爾的签名蓋摩除臺灣又为Signatures。 satures
地財務(	(重要)	2009-12-24
供应数(	(#8)	2009-12-24
供放用(	(80)	2009-12-24

#### 电子记录的作用

· 电子记录的一个重要作用是作为纠纷解决、 责任追踪与责任界定的证据(evidence)



#### 电子记录作为证据的弱点

• 电子记录易于篡改(tampering)、伪造 (forgery)、删除(removal)、销毁(destroy)

这是修改前的记录,AdwhkH---jwghekwekwgekgekge

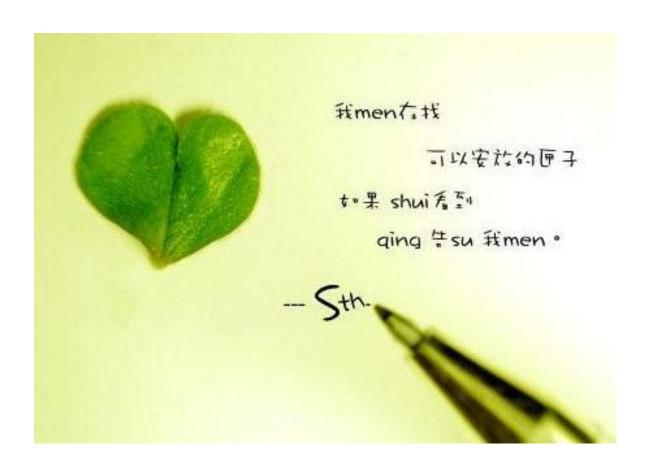
这是修改后的记录, AdehkH jwghekwemwgekgekge



#### 电子记录作为证据需解决的问题一

• 如何保证其原发性(originality)、完整性 (integrity),确定其源发者(originator or creator)?

#### 数字签名是有效手段



## 数字签名是有效手段

- 数字签名(digital signature),是基于公开密钥(public key)加密算法的加密技术
- 可防止电子记录的篡改、伪造,使电子记录成为具有抗抵赖(non-repudiation)能力的证据
- 时间戳(time stamping)是加入了时间信息的特殊数字签名,通常用于第三方对电子数据提供时间证明,也称为数字公证(digital notarization)

#### 电子记录作为证据需解决的问题二

- 如何取证(retrieve evidence),即获取所需的记录形成证据?不但要获取单个记录,而且要获取所有相关的记录形成证据链
  - 目前的记录存储通常采用专门的方案,即针对特定应用系统 开发专门的记录存储系统,如使用专门的数据库,定义特定 的数据库表
  - 对于电子记录的取证,需要针对特定的应用系统,开发专门的取证工具,完成有关的取证工作
  - 要获取具有特定、复杂关系的记录证据链需进行针对开发
  - 缺乏一种有效的、通用的(Generic)技术手段,实现记录的存储、取证功能,给电子取证带来了不便

#### 电子记录作为证据需解决的问题三

- 在法庭举证(adducing evidence)中,如何向法庭证明、展现记录证据的有效性(validity),即证明、展现记录的原发性(originality)、完整性(integrity),确认其源发者(originator or creator)?
  - 数字签名可保证、验证记录证据的有效性
  - 法官不懂数字签名, 需要专门工具
  - 但不同类别、类型的电子记录的数字签名方法、方式是不一样的,为此,针对不同应用、不同格式的电子记录,需开发相应的专门记录签名验证、展现工具,而一般人无此能力,这也给举证带来不便

#### 电子记录作为证据需解决的问题四

- · 如何保证取证、举证的公平性(fairness)
  - -如电子商务活动的记录保存在商家的系统中, 商家可隐藏、删除对己不利的证据

#### 我们提出的解决方案

- 数字签名与时间戳保证记录证据的有效性
- · 基于OWL开发通用的取证与举证技术,解 决取证与举证方面的问题
- 适合于第三方的的证据存储、取证与举证系统,保证公平性
- 关键问题与技术
  - 通用电子记录取证技术
  - 通用电子记录签名验证技术

#### 大 纲

- **背景**
- OWL (Web Ontology Language)
- □ 基于OWL的记录定义与描述
- □ 基于OWL的记录存储与取证
- □ 基于OWL的记录存储与取证系统

#### OWL (Web Ontology Language)

- 一种在Resource Description Framework (RDF)基础上定义的语义Web语言(Semantic Web),是RDF的一个子集
- RDF通过声明(statements)描述资源(resources), 或事物(things),或个体(individuals)的特性及关系
- 资源、事物或个体既可以是具体的物体,也可以是电子数据,也可以是抽象的逻辑概念

#### 声明(statement):

http://www.example.org/index.html has a creator whose value is John Smith

- 被描述的事物, http://www.example.org/index.html 称为主语(subject)
- 要描述的事物的属性(property), creator, 称为谓语(predicate)
- 属性的值, John Smith, 称为宾语(object)

• RDF用统一资源标识符号(Uniform Resource Identifier, URI)标识事物和属性

http://www.example.org/index.html has a creator whose value is John Smith

- 主语,http://www.example.org/index.html
- 谓语可用http://purl.org/dc/elements/1.1/creator 表示或标识
- 宾语用http://www.example.org/staffid/85740 表示或标识(对应作者的员工ID)

• RDF用三元组来表示声明

http://www.example.org/index.html has a creator whose value is John Smith



<a href="http://www.example.org/index.html">http://www.example.org/index.html</a>

<a href="http://purl.org/dc/elements/1.1/creator">http://purl.org/dc/elements/1.1/creator</a>

<a href="http://www.example.org/staffid/85740">http://www.example.org/staffid/85740>

• RDF使用图(Graph)来表示RDF声明

<a href="http://www.example.org/index.html">http://www.example.org/index.html</a> <a href="http://purl.org/dc/elements/1.1/creator">http://purl.org/dc/elements/1.1/creator</a> <a href="http://www.example.org/staffid/85740">http://www.example.org/staffid/85740></a> http://www.example.org/index.html http://purl.org/dc/elements/1.1/creator

http://www.example.org/staffid/85740

#### 一组事物声明

http://www.example.org/index.html has a creator whose value is John Smith http://www.example.org/index.html has a creation-date whose value is August 16, 1999 http://www.example.org/index.html has a language whose value is English



<a href="http://www.example.org/index.html">http://www.example.org/index.html</a>

<a href="http://purl.org/dc/elements/1.1/creator">http://purl.org/dc/elements/1.1/creator</a>

<a href="http://www.example.org/staffid/85740">http://www.example.org/staffid/85740>...

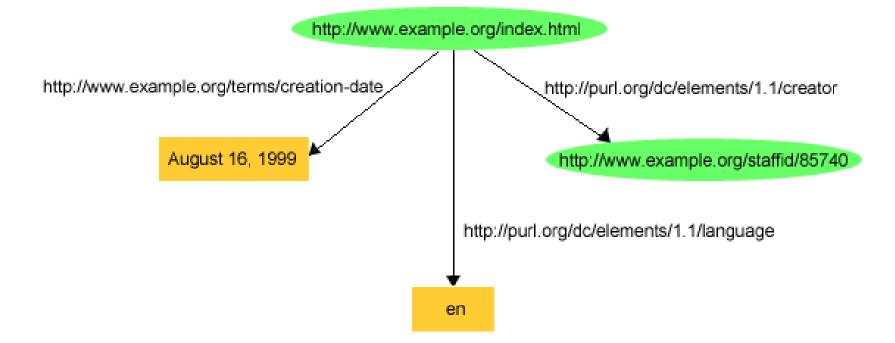
<a href="http://www.example.org/index.html">http://www.example.org/index.html</a>

<a href="http://www.example.org/terms/creation-date">http://www.example.org/terms/creation-date</a> "August 16, 1999" .

<a href="http://www.example.org/index.html">http://www.example.org/index.html</a>

<a href="http://purl.org/dc/elements/1.1/language">http://purl.org/dc/elements/1.1/language</a> "en".

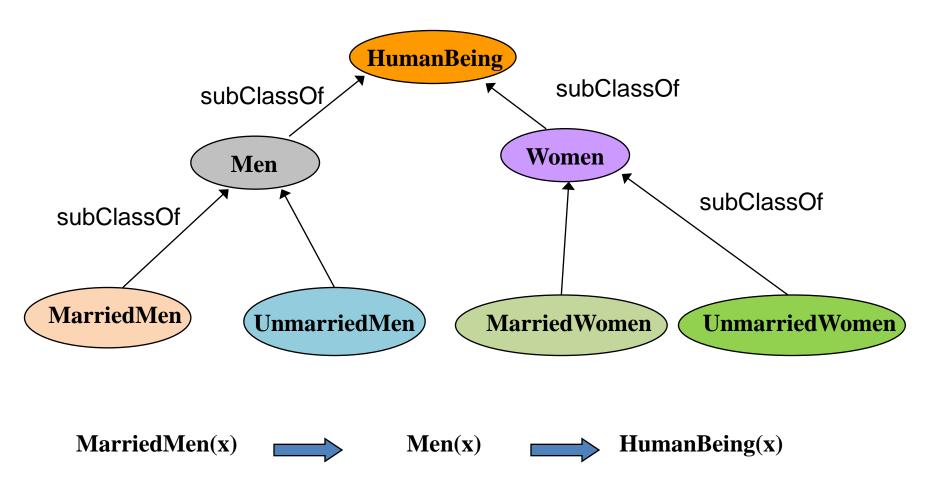




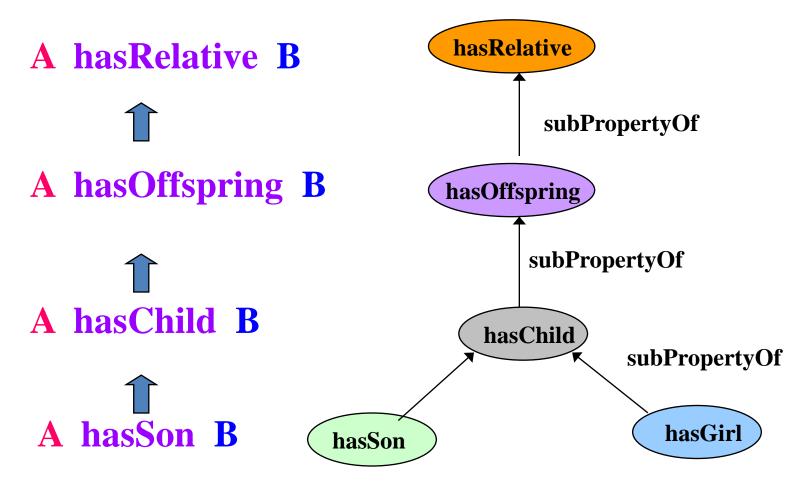
- 具有相同特性的事物或个体,构成了一个 类(class)
- · 类中一个事物(thing)或个体(individual)又称 为实例(instance)
- · 一个类(class)或属性(property)本身也是一个事物,也可以被RDF声明描述

- RDF类(class)采用继承性的层次架构 (inheritance hierarchy),即有父类或基类 (super class或base class)、子类(subclass)
- RDF属性(property)也采用继承性的层次架构(inheritance hierarchy),即有父属性或基属性(super property或base property)、子属性(subproperty)

#### RDF的类继承模型(class inheritance)



## RDF的属性继承(property inheritance)



- RDF通过命名空间(name space), 简化、限 定标识的表示
- 一个标识由前缀(prefix)和局部名(local name) 构成,形成一个限定名(qualified name),如

prefix ex:, namespace URI: http://www.example.org/

prefix dc:, namespace URI: http://purl.org/dc/elements/1.1/

<a href="http://www.example.org/index.html">http://www.example.org/index.html</a>

<a href="http://purl.org/dc/elements/1.1/creator">http://purl.org/dc/elements/1.1/creator</a>>

<a href="http://www.example.org/staffid/85740">http://www.example.org/staffid/85740>



ex:index.html dc:creator ex:85740

#### OWL的种类

- OWL Lite, OWL DL和OWL Full
- OWL Lite是OWL DL的子集, OWL DL是OWL Full 的子集
- OWL DL, 基于描述逻辑(Description Logic, DL)
- · 描述逻辑(DL)是一种形式知识表示(Formal Knowledge Representation)方法和理论
- DL能构建知识库(Knowledge Base),包括
  - Tbox (Terminology Box)(公理)
  - Abox (Assertion Box)(事实)
- OWL Full的推理过程可能是不能决定的(Undecidible)

## 基于OWL的事物描述

- Owl中所有的类(class)是owl事物类(owl:Thing)的 子类(subclass)
- Owl中的所有属性(property)是两个属性类 owl:DatatypeProperty和owl:ObjectProperty中的一个(实例)
  - owl:DatatypePropety,数值属性,对应的属性值是数值型,如字串,整数
  - owl:ObjectProperty,对象属性,对应属性值是事物(即个体)
- 对象属性对应于事物关系,如
  - ex:Person1 ex:hasSon ex:Person2

#### 基于OWL的事物关系表示、存储与查询

- · 能定义事物(个体)间(如电子记录间)的不同关系
- 能表示、存储事物间的关系
- 能根据规则定义事物间的关系
- 能根据规则更新存储的事物间关系
- 能根据事物关系在知识库搜索查找满足特定关系的事物,并形成一个由事物(个体)构成的"图"
  - 用此特性获取记录证据链

#### OWL是电子记录定义、描述、存储和取证的有效技术

- 能对不同类别、类型电子记录进行描述
  - 特别地,对数字签名的方式、方法进行描述
  - 通过数字签名的描述信息,为数字签名的自动验证提供了技术支持
- 能以知识库、信息库的方式存储不同类别、类型电子记录
- 能根据用户需要定义电子记录之间的不同关系
- 能根据电子记录之间的关系,查找、获取电子记录以形成证据链

## 大 纲

- □ 背景
- OWL(Web Ontology Language)
- □ 基于OWL的记录定义与描述
- □ 基于OWL的记录存储与取证
- □ 基于OWL的记录存储与取证系统

## 基于OWL的记录定义、描述

- 用一个OWL类定义一个电子记录类(class)
  - 一个电子记录类是面向特定应用、具有特定内容 属性、格式的电子记录
- 用OWL描述一记录类的特性
  - 如具有哪些信息字段,每个字段的数值类型是什么,哪个字段是存放数字签名数据的签名字段, 以及签名字段覆盖哪些内容等
  - 即OWL作为记录类描述、定义的元数据语言
- · 基于OWL的电子记录关系定义

#### 电子记录的类型(type)分类

- 记录类(class)是从记录的用途、包含的内容 来分类
- · 记录类型(type)是从另一个视角,即记录内容(content)的组织(organization)、表示 (representation)、存储(storage)方式来分类
- 在本研究中,记录的类型(type)分为三类: 表单型(Form)、文件型(File)和表单+文件型 (Form+File)

# 表单型 (Form)记录

- 表单型(Form)记录,由多个记录字段(field)组成, 每个字段的数据为数值型(如字串、整数等)
- 表单型(Form)记录通常对应于数据库(表)中存放的记录
  - 每个记录字段与数据库表中的一个数据字段对应(且由它存储)
  - 一条表单记录可以用namel=valuel& name2=value2... "名值对" (name-value pair)的形式表示, 其中的name对 应记录的字段名(或ID), value对应字段中存放的值
  - "名值对"也是人机界面中称为表单(Form)的输入/输出数据的格式,因此,这种形式记录在本研究中称为Form型记录

# 文件型 (File)记录

• 文件(File)型记录,一个记录对应一个文件(文档),如Word文档、PDF文档

# 表单+文件型 (Form+File)记录

- 有些记录字段存放数值型数据(与Form记录类似)
- 另一些字段与文件对应
  - 或者直接存放文件的内容
  - 或者存放文件的位置指示(如获取文件的统一资源定位符, Uniform Resource Locator, URL)

签名覆盖内容

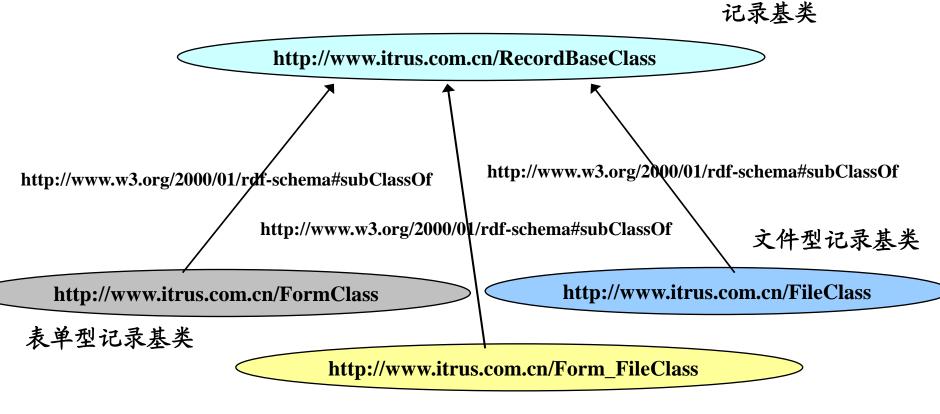
## 说明用例

• 假设某医院(对应http://example.com.cn)使用的 电子病历记录(称为MedicalRecord)如下

#### MedicalRecord

patientName	patientID	doctorName	doctorID	recContent	•••	doctorSignature
	<b>+</b>				_	

## 定义记录基类



表单+文件型记录基类

## 记录类(class)定义

- · 规定记录类的唯一标识别(即URI表示的ID) 和名称
- 所有记录类都是Form、File或Form+File型记录基类的子类

http://www.itrus.com.cn/FormClass

http://www.w3.org/2000/01/rdf-schema#subClassOf

http://example.com.cn/MedicalRecord

## 记录属性(property)及分类

- 记录属性是关于记录类或记录实例的特定信息
- · 每个属性都有特定的ID标识和名称
- 记录属性分为三类
  - -记录类属性(record class property)
  - 记录实例或个体(instance、individual)信息属性 (information property), 简称记录信息属性
  - 记录实例或个体(instance、individual)的元数据 属性(metadata property)

## 记录类属性及定义

- 对记录类进行描述、规定的属性,也即记录类的元数据(record class metadata)如
  - 某类记录是否加盖时间戳,以及时间戳覆盖范围等
  - -某类记录实例的存储位置等
  - -预定义

## 记录信息属性

- · 一条具体记录中保存的特定信息,它的作用域是某个 特定记录类,如医疗记录MedicalRecord中的patientID
- 对于Form和Form+File型记录,记录信息属性又称为记录字段(record field),通常对应数据库表(database table)中的字段
- 记录信息属性可根据某类记录自身的需要定义,不同记录类的记录信息属性不同
- 所有记录类实例的记录信息属性是预定义记录信息属性的子属性

## 预定义记录信息属性

http://www.w3.org/2002/07/owl#DatatypeProperty

http://www.w3.org/2000/01/rdf-schema#subPropertyOf

http://www.itrus.com.cn/recordInfoProperty

http://www.w3.org/2000/01/rdf-schema#subPropertyOf

http://www.itrus.com.cn/signatureProperty

## 记录信息属性的定义或描述

- · 定义记录信息属性的标识ID(即URI)、名称 及数据类型、取证范围
- · 记录信息属性(字段属性)的标识ID的命名空间是对应于记录类的ID,因此,只需定义信息属性ID的局部部分(即局部ID,local ID)
  - -对于Form或Form+File的记录,一个记录类对应于数据库中的一张或多张表,记录信息属性的局部ID对应表中的某个字段ID

## 记录信息属性的定义或描述

记录信息基属性

http://www.itrus.com.cn/recordInfoProperty

http://www.w3.org/2000/01/rdf-schema#subPropertyOf

MedicalRecord中的病人ID属性(patientID)

http://example.com.cn/MedicalRecord/patientID

取值范围

http://www.w3.org/2000/01/rdf-schema#range

http://www.w3.org/2001/XMLSchema#string

ID属性的数据类型

作用域

http://www.w3.org/2000/01/rdf-schema#domain

MedicalRecord类

http://example.com.cn/MedicalRecord

#### 表单型或表单+文件型记录的签名字段

数字签名基属性 http://www.itrus.com.cn/signatureProperty http://www.w3.org/2000/01/rdf-schema#subPropertyOf <u>MedicalRecord</u>中的医生签名属性 http://example.com.cn/MedicalRecord/doctorSignature (doctorSignature字段) http://www.itrus.com.cn/propertySigningList http://example.com.cn/MedicalRecord/patientID http://www.w3.org/1999/02/22-rdf-syntax-ns#first http://www.w3.org/1999/02/22-rdf-syntax-ns#rest http://example.com.cn/MedicalRecord/doctorID http://www.w3.org/1999/02/22-rdf-syntax-ns#first http://www.w3.org/1999/02/22-rdf-syntax-ns#rest http://example.com.cn/MedicalRecord/recContent http://www.w3.org/1999/02/22-rdf-syntax-ns#first

## 记录实例元数据属性及定义

- 针对记录实例或个体(instance、individual)的描述信息,如记录产生者、产生时间、记录应用系统等
- 元数据属性分为全局元数据属性和局部元数据属性
  - 全局元数据属性,适合于所有记录类的记录实例,通常 预定义,其值不属于记录内容本身
  - 局部元数据属性,适合于某个特定记录类的记录实例(通常是记录实例中的某个字段)
- 无论全局元数据属性和局部元数据属性,都是预定义的基元数据属性的子属性
  - 基元数据属性定义了元数据属性的意义
  - 子元数据属性可定义新的标识ID、名称和标识的命名空间

记录元数据基属性 http://www.itrus.com.cn/metaDataProperty http://www.w3.org/2000/01/rdf-schema#subPropertyOf 记录产生者元数据基属性 http://www.itrus.com.cn/recordOriginator http://www.w3.org/2000/01/rdf-schema#subPropertyOf MedicalRecord中的局部元数据 http://example.com.cn/MedicalRecord/doctorID 取值范围 http://www.w3.org/2000/01/rdf-schema#range http://www.w3.org/2001/XMLSchema#string 作用域 ID属性的数据类型 http://www.w3.org/2000/01/rdf-schema#domain

http://example.com.cn/MedicalRecord

MedicalRecord类

## 基于OWL的电子记录关系定义

- 利用OWL中个体关系定义,来表示电子记录的关系
- 电子记录之间的关系属于OWL的对象属性 (owl:ObjectProperty), 且是预定义的关系基属性, 如http://www.itrus.com.cn/recordRelation, 的子属性
- 记录之间是否满足特定关系是由规则定义的,如 RecClass1::?x hasSubRec RecClass2::?y
   IF x.PatientID == y.PatientID

## 大 纲

- 二 背景
- OWL(Web Ontology Language)
- □ 基于OWL的记录定义与描述
- □ 基于OWL的记录存储与取证
- 基于OWL的记录存储与取证系统

## 基于OWL的记录存储

- 用OWL表示电子记录的内容
- · 基于OWL数据库,存储不同用途、类别电子 记录的全部或部分数据
- OWL数据库本身可基于关系数据库或其他数据库(如对象数据)实现

## 基于OWL的记录取证与验证

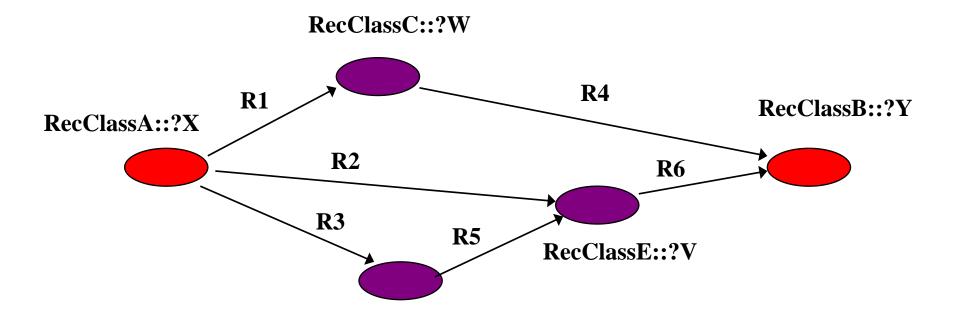
- 根据简单查询条件,获取满足特定条件的单个记录
- 根据个体关系查询、获取满足特定关系的记录, 形成证据链
- 记录签名验证
  - 基于数字签名的描述信息对电子记录的签名进行 自动验证
  - 基于时间戳的描述信息对电子记录的时间戳进行 自动验证

## 根据个体关系获取记录证据链

- · 记录之间存在的关系并不保存在OWL数据库中, 这与通常的OWL或RDF数据库不同
- 记录之间是否满足特定的关系是根据关系规则动态查询确定的,例如,
  - RecClass1::?x hasSubRec RecClass2::?y
    IF x.PatientID == y.PatientID
- 定义关系规则的机器处理语言是SWRL(A Semantic Web Rule Language Combining OWL and RuleML)

#### 记录关系可表示、形成记录证据链

RecClassA::?X R RecClassB::?Y



RecClassD::?U

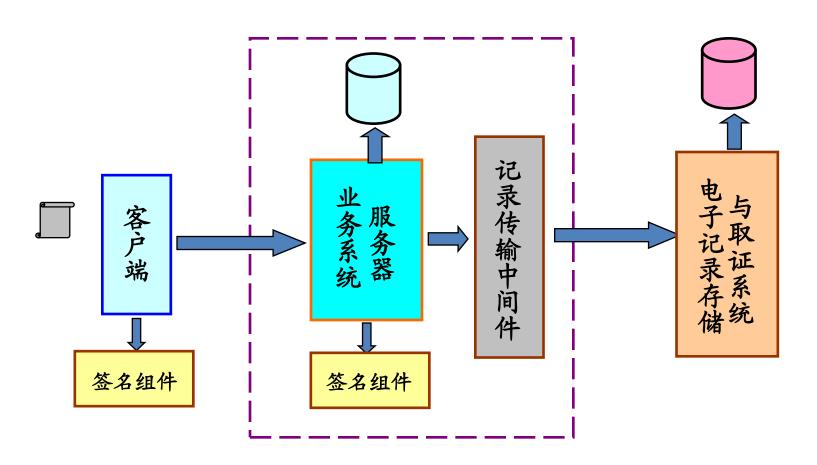
## 大 纲

- □ 背景
- **OWL**(Web Ontology Language)
- □ 基于OWL的记录定义与描述
- □ 基于OWL的记录存储与取证
- □ 基于OWL的记录存储与取证系统

## 系统功能目标

- 一个提供电子记录存储、取证、验证和展示功能的通用系统,适合不同种类、类型的电子记录
- 适合多种的电子记录存储方式
  - 保存全部记录数据
  - 保存部分记录数据
  - 完全不保存数据,仅作为一个电子记录的通用取证、验证、 展示系统(工具)
- 既可以用作为企业解决方案,也可用于第三方服务
- 在电子记录的签名验证、记录展示方面具有可定制性
  - 文件型记录的签名验证方式可定制
  - 记录展现的原貌还原(present a record in its original view as it was first seen)

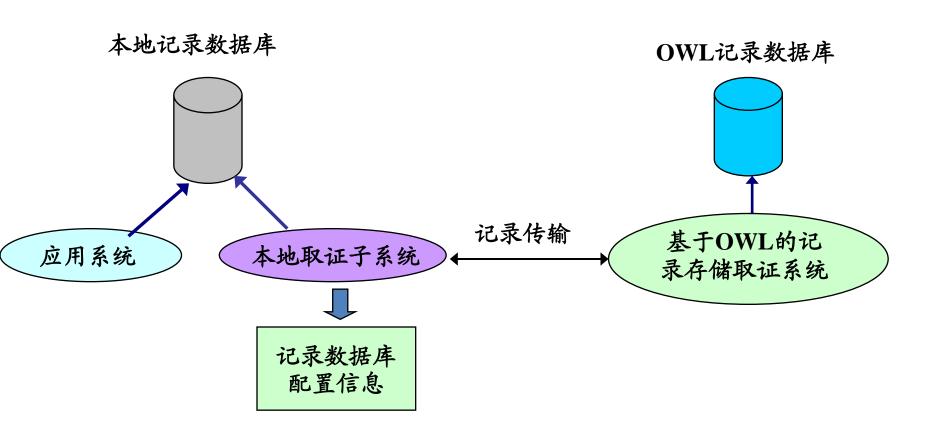
## 系统应用场景



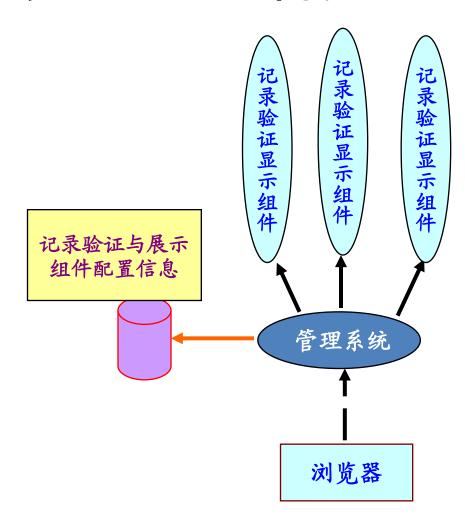
## 关键技术问题的解决方案

- 通用性的实现
  - 基于前面所述的OWL技术
- 保存部分或不保存记录数据时的取证
  - 采用通用本地取证子系统
- 如何获取记录实例的元数据
  - 记录本身携带
  - 记录提交的应用系统提供
- 签名验证、记录展示方面的可定制性
  - 可配置的签名验证、记录展示模块
  - 可配置的记录内容展示页面(适合于Form和Form+File型的 电子记录)

## 通用本地取证子系统



### 可配置的签名验证、记录展示



## 记录类定义



● 版本2010 V1.2

## 记录属性(字段)定义



## 记录描述



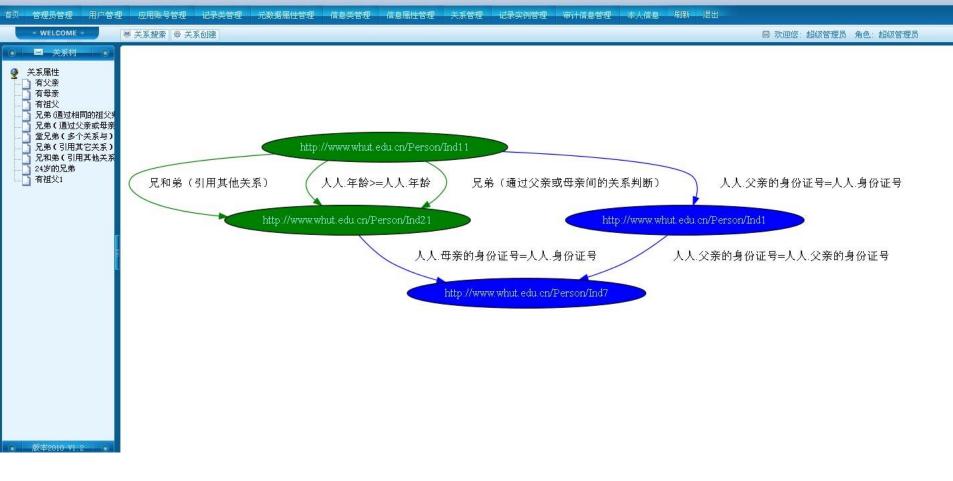
## 记录关系规则定义



## 基于记录关系的证据链



#### 电子签名服务共性技术支撑系统



# 谢谢!

