

Le projet GDFR

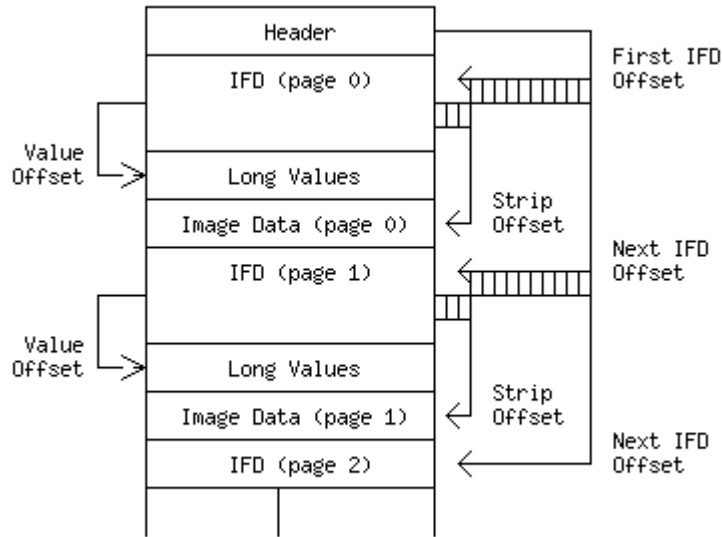
Description et implications de la BnF

Laurent Duploux
Département des systèmes d'information

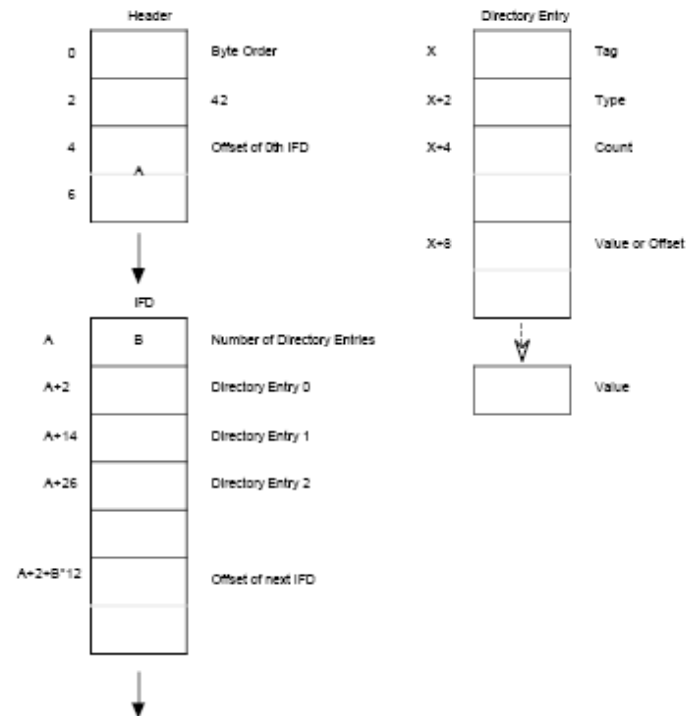
- Description
 - Une définition des formats
 - Un modèle de données
 - Le modèle collaboratif proposé
 - La gestion des identifiants
- Perceptives d'implication de la BnF
- Répertoires de format et SPAR

- *AIM* [Abstract Information Model]
 - L'unité d'échange d'information. Ex : Fichier TIFF
- *FCS* [Format Coded Set]
 - Décomposition d'un AIM en élément d'information structurée sémantiquement consistant. Ex : cf. page suivante
- *FEF* [Format Encoding Form]
 - Décomposition d'un *FCS* en séquence d'unités de données. Ex : cf. page suivante
- *FES* [Format Encoding Scheme]
 - Décomposition d'un *FEF* en une série de bits. Ex : un entier non signé codé sur 32 bits (LONG).

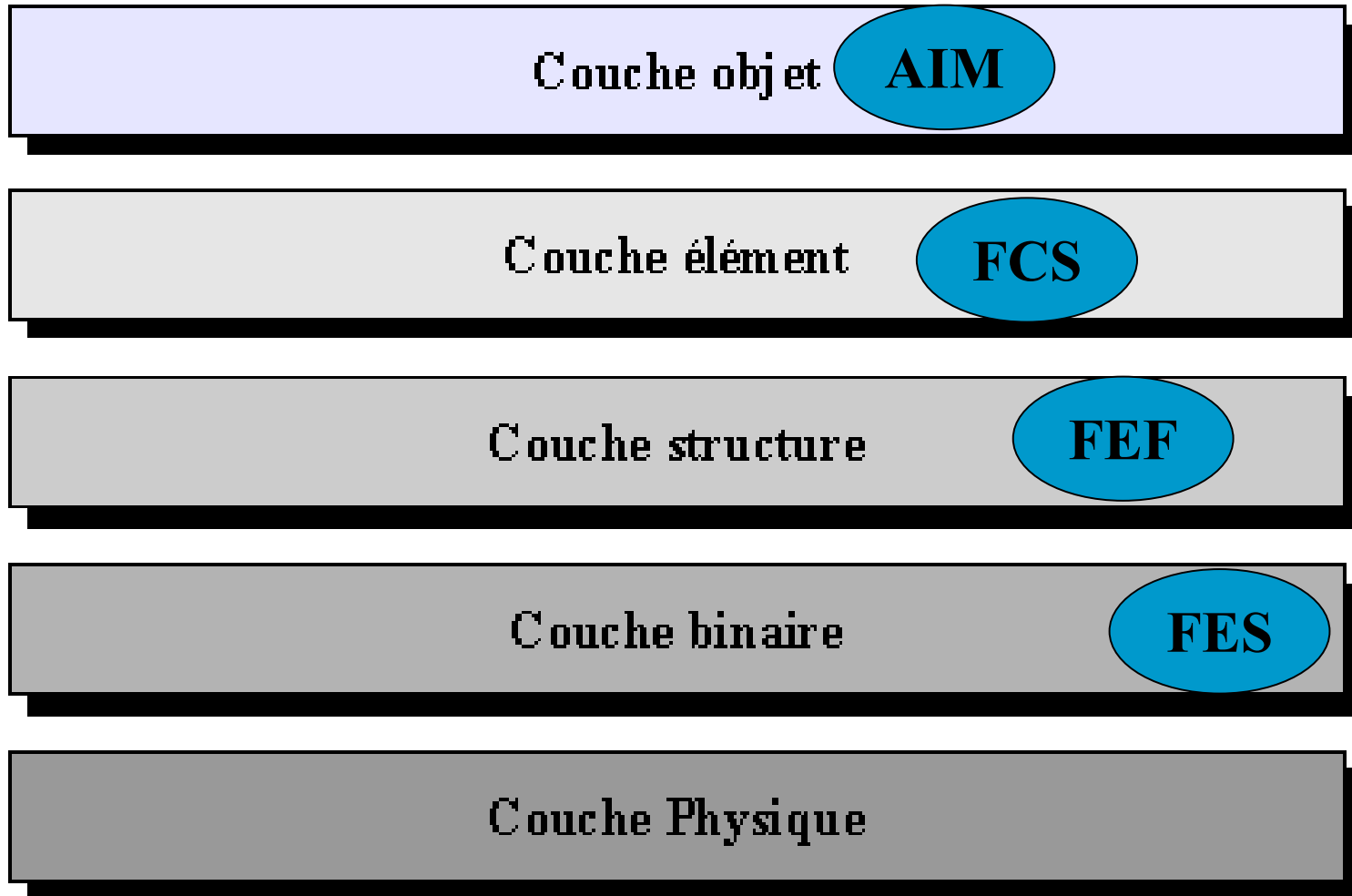
Structure d'un fichier TIFF



Offset	Description	Type	Value
0	Byte Order	Short	0x4949 (II)
2	Version	Short	42
4	Offset of 0th IFD	Long	0x 0000 0008



Structure de l'information numérique (PIN)



■ Relations

→ Subtype

❖ Example ASCII \subset UTF-8 (“ASCII is a *subtype* of UTF-8”)

→ Extension

❖ Example DNG \Leftarrow TIFF 6.0 (“DNG is an *extension* of TIFF 6.0)

→ Containment

❖ Example WAVE $<$ μ -law audio (“WAVE can *contain* μ -law
audio”)

→ Equivalence

❖ Example DXF (ASCII) \equiv DXF (binary) (“DXF (ASCII) is *equivalent*
to DXF (binary)”)

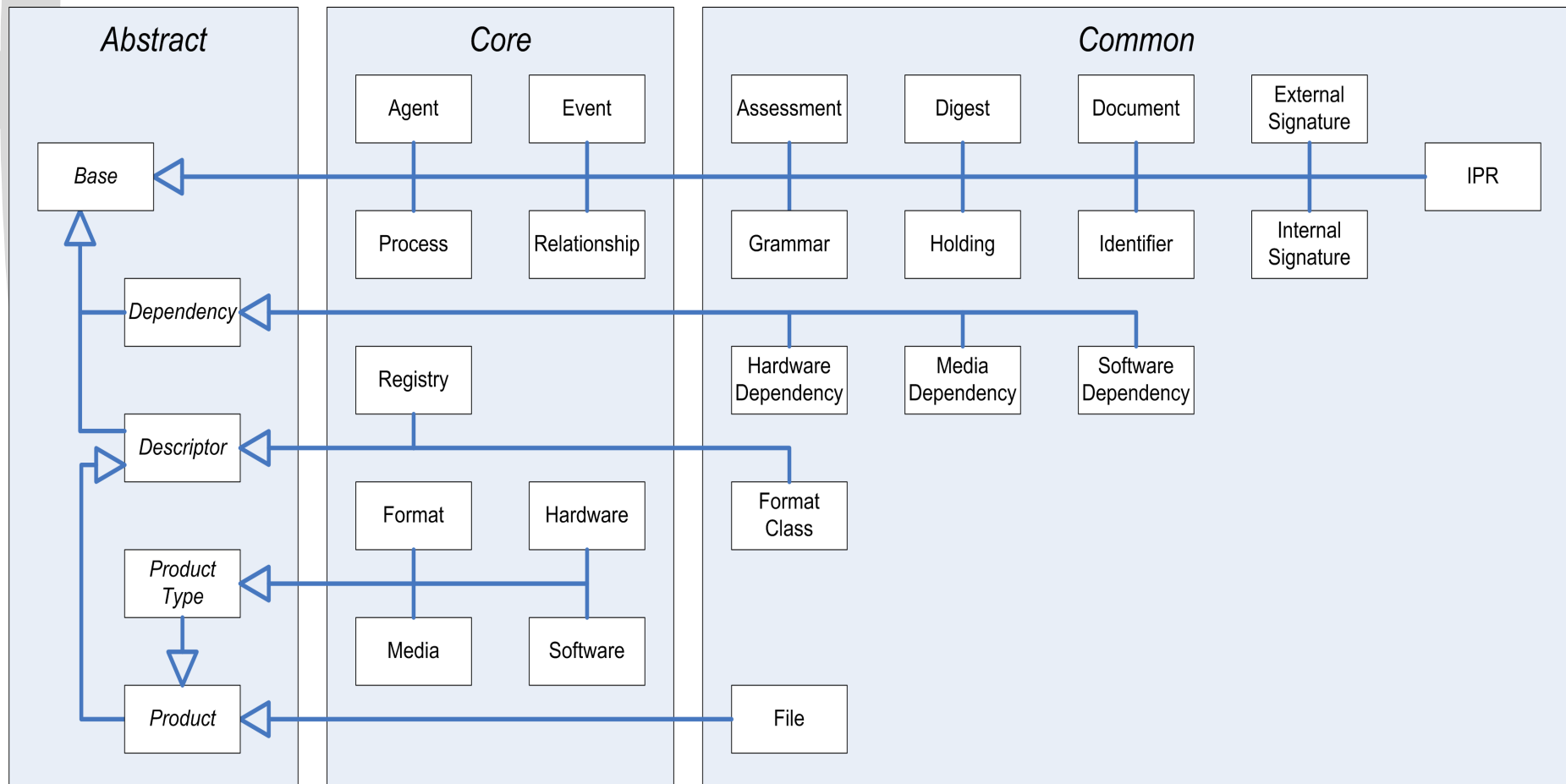
→ Version

❖ Example Word 6.0 \leftarrow Word 97 (“Word 6.0 is a *new*
version of Word 97”)

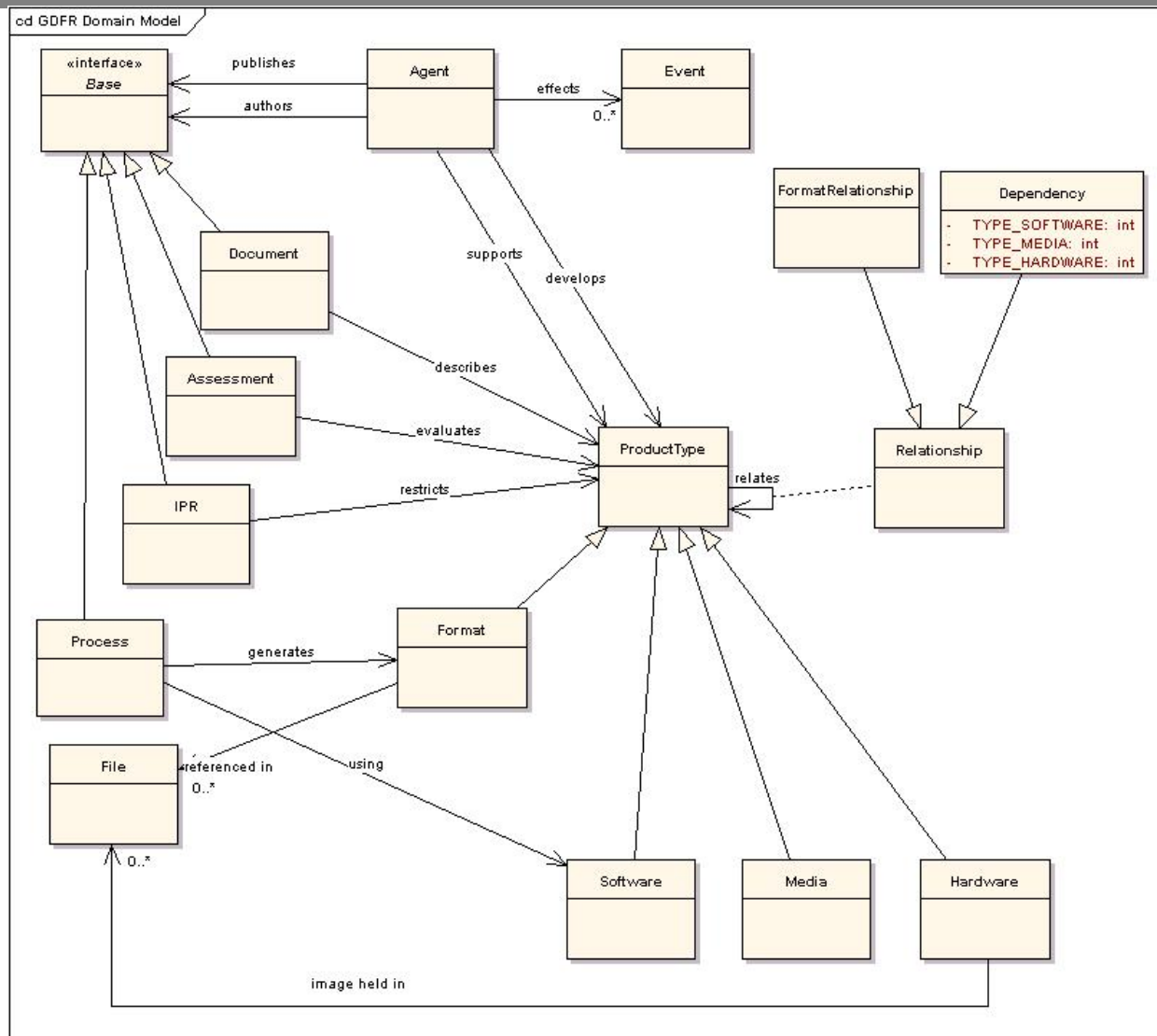
→ Affinity

❖ Example SPIFF \approx JPEG (“SPIFF has an *affinity* to
JPEG”)

Un modèle de données



Un modèle de données



Un modèle de données : l'entité format

Property	Type	Cardinality	Description	Example/note	
GDFR identifier	Identifier	1	Globally unique persistent identifier assigned by the GDFR	“info:gdfr/format/123”	
Class	FormatClass	1..*	Format classification		
Relationship	Relationship	0..*	Relationship to other format		
Disclosure	Enumeration	0..1	Level of public disclosure		
			Full	Full disclosure	
			Partial	Partial disclosure	
			None	No disclosure	
			Other	Other or unknown level	Note required
Orientation	Enumeration	1	Format orientation:		
			Binary	Binary only	
			Mixed	Binary and text	
			Text	Text only	
			Not known	The format orientation is not known	
Byte order	Enumeration	1	Byte order prescribed by the format:		
			Big-endian	Big-endian only	
			Either	An instantiated byte stream can use either big- or little-endian byte ordering	“TIFF”
			Little-endian	Little-endian only	
			Mixed	Both big and little-endian byte ordering can be mixed in the same instantiated byte stream	“WAVE”
			Not applicable	The format is byte-oriented rather than word-oriented	
			Not known	The byte order is not known	
Internal signature	InternalSignature	0..*	Internal signature, e.g. magic number	“MMx00\x2a”	
External signature	ExternalSignature	0..*	External signature, e.g. file extension	“tif”	
Reference file	File	0..*	Reference file		
Grammar	Grammar	0..*	Formal grammar describing the format syntax		
Assessment	Assessment	0..*	Assessment of the format		
Properties inherited from the ProductType entity					
Type	Enumeration	1	Product type:		
			Abstract	Abstract product family that can be instantiated by a “concrete” product	
			Concrete	Concrete product that belongs to an “abstract” product family	

Un modèle de données : l'entité format

Properties inherited from the Product entity				
Alias	String	0..*	Alternative product name	
Version	String	0..1	Product version	
Documentation	Document	0..*	Product documentation	
Identifier	Identifier	0..*	External product identifier	
Release date	DateTime	0..1	Product release date	
Withdrawal date	DateTime	0..1	Product withdrawal date	
Software dependency	SoftwareDependency	0..*	Product software dependency	
Hardware dependency	HardwareDependency	0..*	Product hardware dependency	
Media dependency	MediaDependency	0..*	Product storage media dependency	
Developer	Agent	0..*	Agent that developed the product	
Support	Agent	0..*	Support or maintenance agent for the product	
Rights	IPR	0..*	Intellectual property rights associated with the product	Products in the public domain shall not define a Rights property
Status	Enumeration	1..*	Product status:	

Active
Discontinued
Proposed
Recommendation
Standard
Supported
Other
Not known

Other product status
Product status not known

Note required

Properties inherited from the Descriptor entity					
Name	String	1	Name		
Description	String	0..1	Description		
Properties inherited from the Base entity					
System identifier	<i>implementation dependent</i>	1	Unique persistent identifier		
Administrative status	Enumeration	1	Administrative status of the instantiated entity:		
			Active		
			Deleted		Provenance note required
			Deprecated	Submitted for review for deletion	Provenance note required
			Provisional	Submitted for review for addition	
Provenance note	String	0..1	Provenance note	Required if Administrative status = Deleted or Deprecated	

Un modèle de données : l'entité grammar

Property	Type	Cardinality	Description	Example/note	
Type	Enumeration	1	Grammar notation type:		
			ABNF	Augmented Backus-Naur Form, RFC 4234	
			BNF	Algol-60 Backus-Naur Form	
			BSDL	Bitstream Syntax Description Language	
			DFDL	Data Format Description Language	
			DTD (SGML)	SGML Document Type Definition (ISO 8979)	
			DTD (XML)		
			EAST	CCSDS EAST	
			RELAX	ISO 19757-2 RELAX NG schema	
			Schema	W3C XML Schema	
			Schematron	ISO 19757-3 Schematron schema	
	Other		Note required		
Value	String	1	Expression of the grammar		
Identifier	Identifier	0..*	Identifier associated with the grammar		

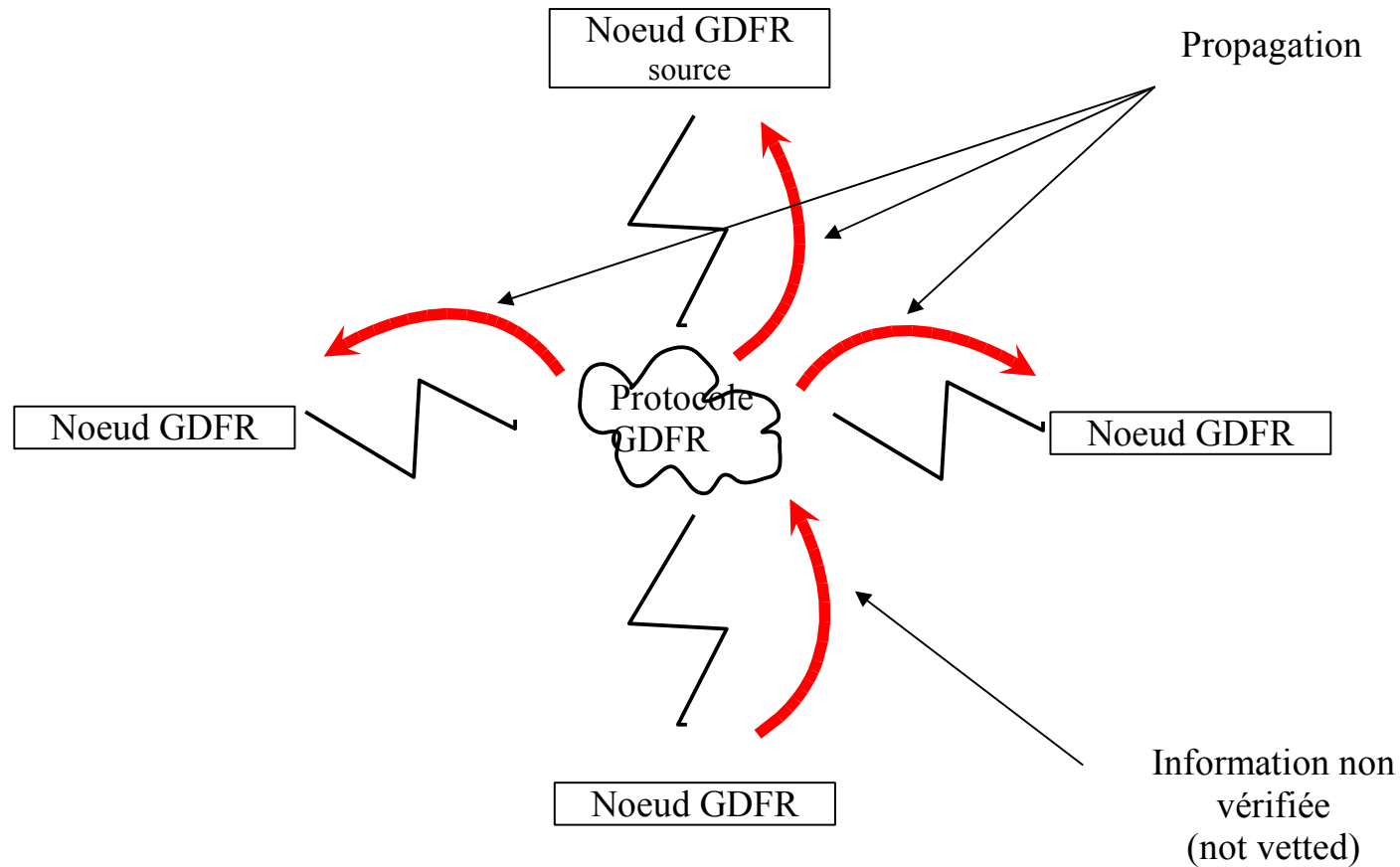
Un modèle de données : l'entité grammar

Properties inherited from the Base entity				
System identifier	<i>implementation dependent</i>	1	Unique persistent identifier	
Administrative status	Enumeration	1	Administrative status of the instantiated entity:	
			Active	
			Deleted	Provenance note required
			Deprecated	Submitted for review for deletion Provenance note required
			Provisional	Submitted for review for addition
Provenance note	String	0..1	Provenance note	Required if Administrative status = Deleted or Deprecated
Source	Agent	1	Agent responsible for the most recent version of the entity	Duplicative of data also captured in the Event entity
Creation date	DateTime	1	Date and time at which the entity was created	
Modification date	DateTime	0..1	Date and time at which the entity was most recently modified	
Note	String	0..1	Optional informative note	Required if Type = Other

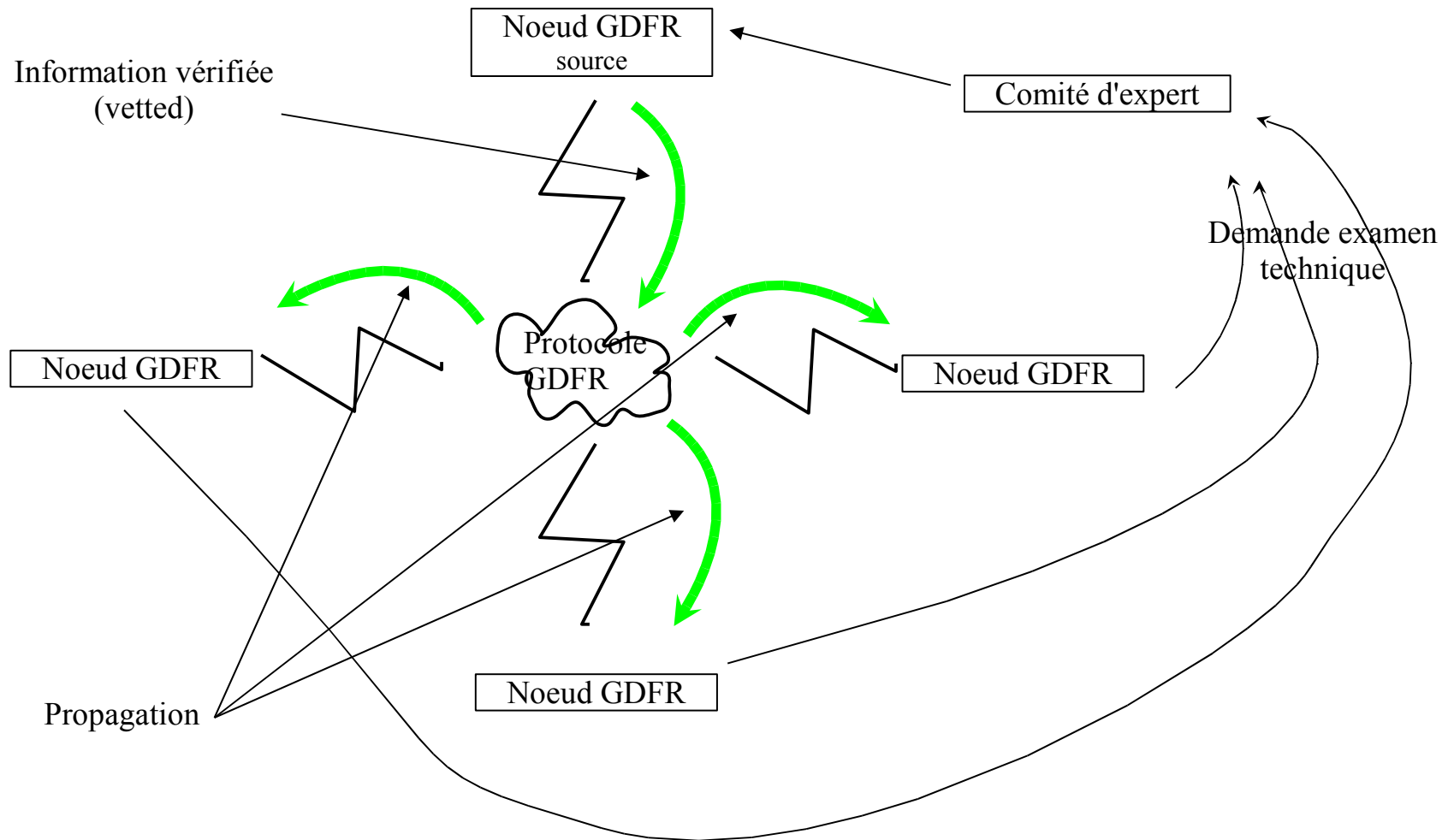
Le modèle collaboratif proposé : principe général

- Global Digital Format Registry (GDFR) fournit des services distribués pour enregistrer, découvrir, et mettre à disposition des informations de représentation sur les formats numériques.
 - une collection organisée de manière centrale de l'information.
 - les services (stockage, découverte, mise à dispo.) sont distribués.
- Architecture distribuée :
 - nouveau format de représentation peuvent être introduit de 2 façons : vetted and non-vetted.
- The GDFR comprend 4 catégories de services:
 - Maintenance: ajouter, reviser, mise à jour, enregistrer des informations de représentation
 - Administratif: administration local d'un nœud
 - Synchronisation : gestion des communication inter nœud
 - Utilisateur: découvrir, et mettre à disposition pour des humains ou des agents

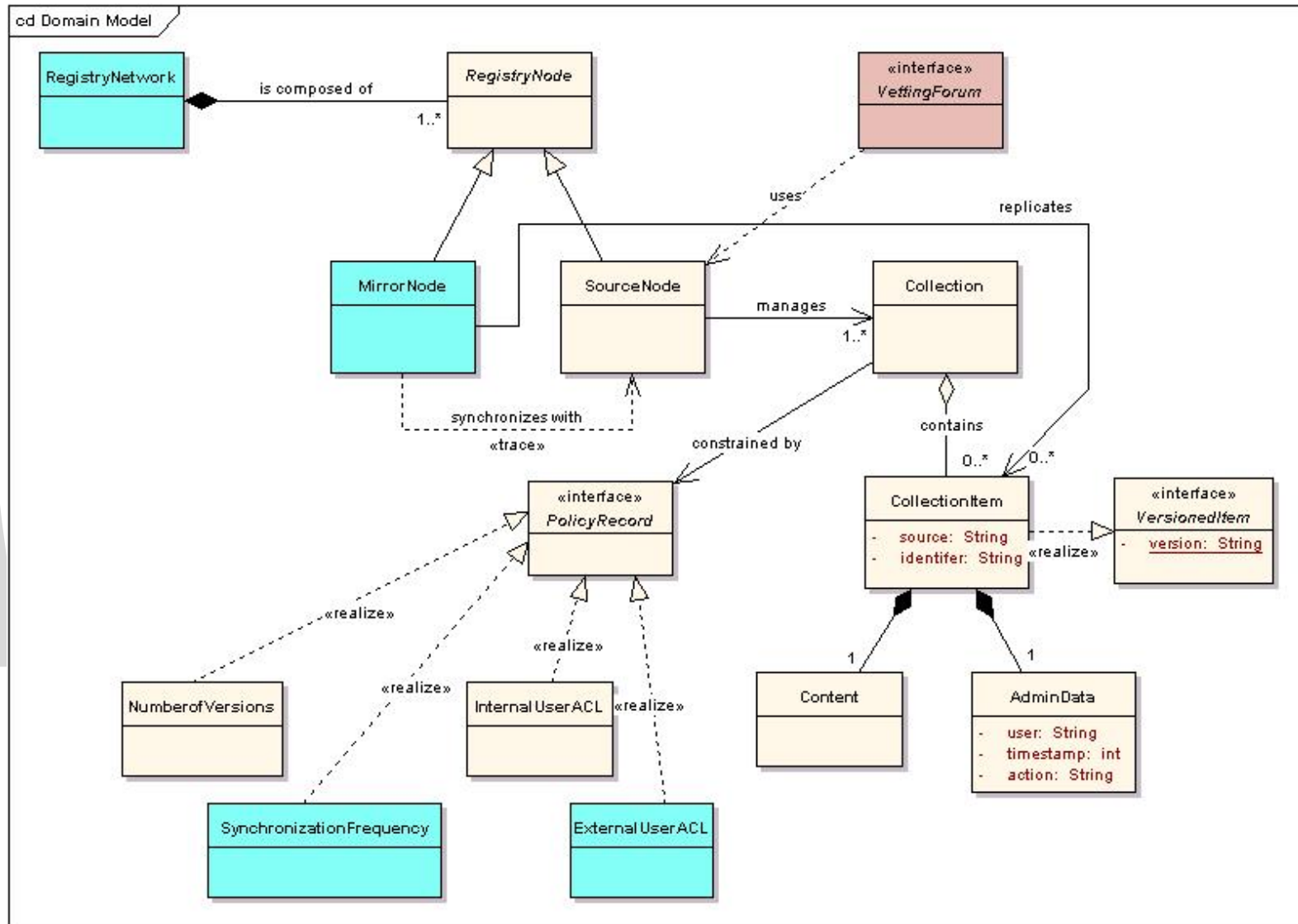
Le modèle collaboratif proposé : propagation



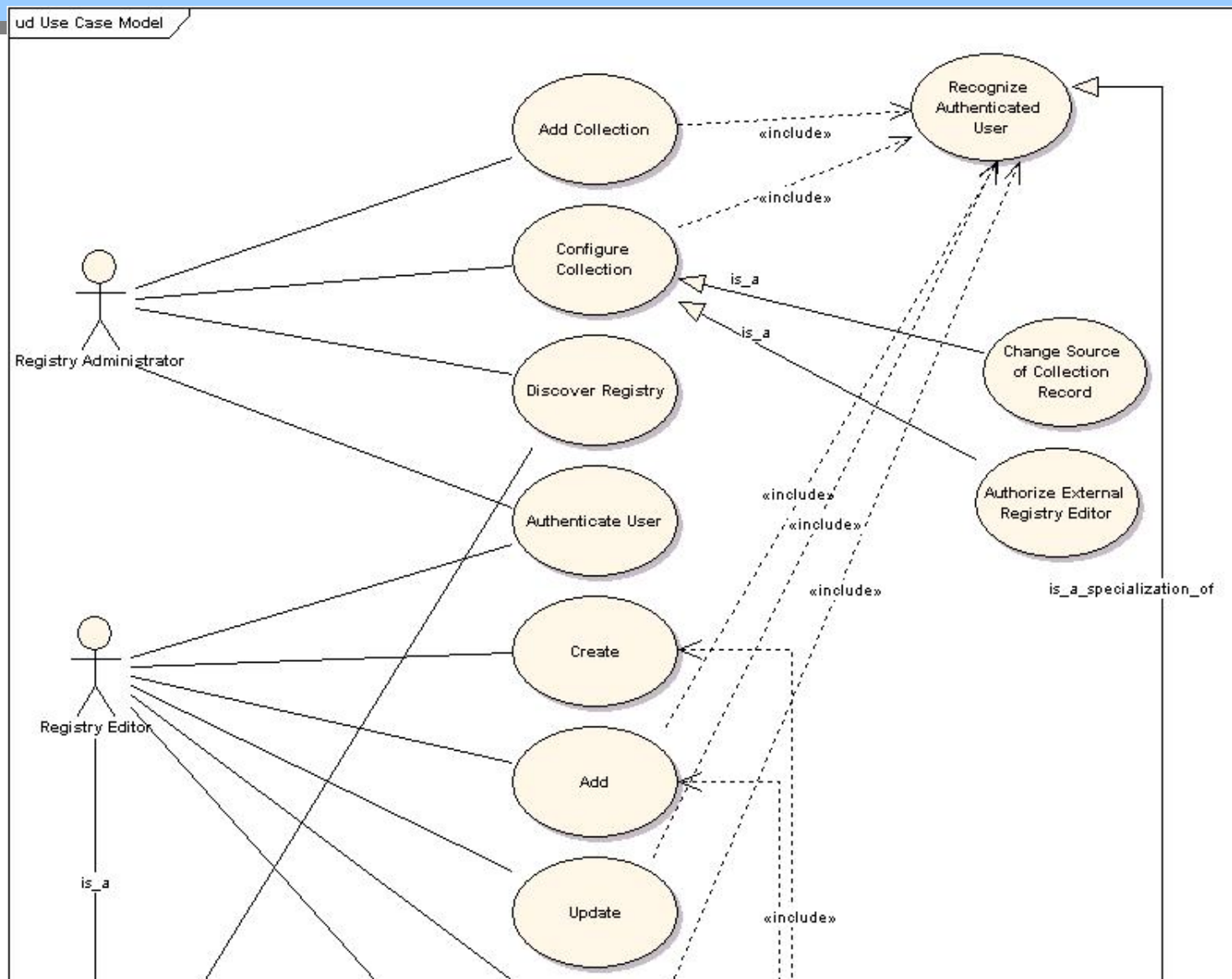
Le modèle collaboratif proposé : propagation



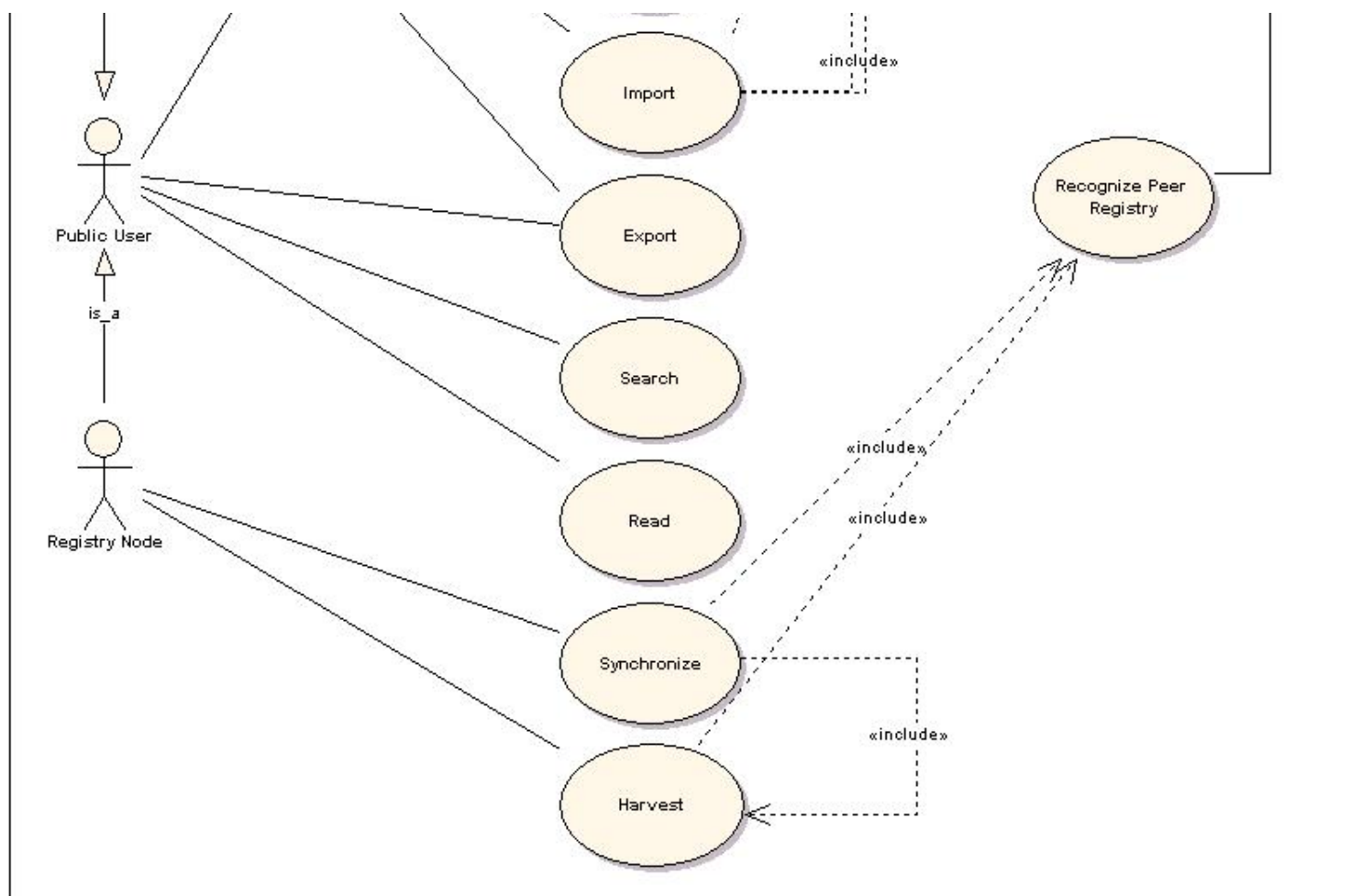
Le modèle collaboratif proposé



Le modèle collaboratif proposé : cas d'utilisation



Le modèle collaboratif proposé : cas d'utilisation



Le modèle collaboratif proposé : interfaces

Business function	Description	Candidate Access Method
Discover	Find relevant registry and access instructions.	UI; human readable url; OpenURL
Search	Returns registry entries based on query	UI; SRU/SRW; Opensearch
Lookup (resolve)	Returns registry entries based on one high order identifier	UI; human readable url; OpenURL
Export	Transfer all or part of registry entries to specified destination	FTP, OAI or browser function, encoded in XML
Import	Transfer registry entries or data elements into registry	FTP, OAI or browser function, encoded in XML
Update	Create , delete or change registry entities	UI; SRU Update; batch script
Subscribe/notify	Request notification, updates or data from registry on a periodic basis	UI; RSS, Atom

■ Basé sur les info- URI :

- ❖ info-URI = info-scheme ":" info-identifiant ["#" fragment]
- ❖ info-identifiant = namespace "/" identifiant
- ❖ namespace = scheme
- ❖ identifiant = *(pchar / "/")
- ❖ Ex : info:sici/0363-0277(19950315)120:5%3C%3E1.0.TX;2-V ou info:ddc/22/eng//004.678

■ Avec le scheme : gdfr

- ❖ Ex : **info:gdfr/format/123**

- Implication le projet GDFR :
 - Etre nœud source
 - Participation à l'élaboration du modèle (G3, SPAR-Réa, etc.)

- Autres solutions
 - WIKI structuré et sécurisé ?
 - Constitution d'un réseau européen ?
 - ❖ Pertinence ?
 - ❖ Quels partenaires ?
 - ❖ ...

Aller plus loin : SPAR et GDFR

