Fondazione RINASCIMENTO digitale

TRUST AND PERSISTENCE OF INTERNET RESOURCES

Maurizio Lunghi
Chiara Cirinnà
Emanuele Bellini

The URL issues

The current use of the simple URL approach brings many and documented risks in a long term vision not only for retrieval and access of resources but also with respect to the loss of reference to the digital documents or the lack of guarantee of authenticity and provenance.

These risks affect:

a) the cultural heritage and research domains, preventing the implementation of reliable citability services, research evaluation, digital preservation, access, etc.,
b) the business domain, preventing the use of purchase services provided on these objects,
c) the public domain (e-gov), slowing down the dematerialization process of Public Administrations.

Cool URI approach

Advantages

-immediate de-reference ability through the protocol HTTP

-Cool URI approach to persistence is based on the URL design (W3C best practices)

- Context information

Dis-advantage

-persistence is not guaranteed in principle by an independent and trustable third party

-the persistence is based uniquely on the commitment of individual institutions

-There is not a strict policy based on Content preservation

The commitment of a single institution is no longer sufficient to ensure neither long term persistence of URIs nor their trustworthiness.

Fondazione RINASCIMENTO digitale

CH and scientific contents need more…

LOD approach enables a wide accessibility of a huge number of data on the Web in a non-proprietary format and it links these data to other datasets (e.g. Genomes or DBpedia) to disambiguate content and to provide a context.

But…

besides retrieving the data or their relations, it is also important to get information about:

authenticity
authority
integrity
provenance
+
persistency over time

A Persistent Identifiers approach

The PI technologies help make stable the reference to digital resources, even if it is well-known that persistency isn't only a technical issue

- No technology can exist indefinitely or guarantee services without a trusted organization behind and a clearly defined policy.

PI systems are meant as:

a) available **technology**

b) trustable **organization**

c) precise **policies** for digital preservation, implemented by the managers of the related user community

A Trusted Persistent Identifier

Concept of contract: Like a **contract** between the final users and the service-providers responsible for the implementation and maintenance of the PI-service and the functionality of the system.

Community commitment: The persistence of a PI depends also on the **commitment of the community** that promotes and uses the identification system for their own resources.

A Trusted NBN-IT

IETF URN based identifiers, NBN namespace IETF RFC 3188 responsibility cleared assigned to National Libraries

NBN:IT:BNCF:12345

NBN:IT is a service of legal deposit and supports 3 types of persistence:

1) *Persistence of the identifier*
2) *Persistence of the association URNs and URLs*
3) *Persistence of the resource referenced by NBN* (backup copy)

The **community** is represented by all institutions that want to deposit contents to the National Library after the signing of an official agreement (**contract**)

Using PI with LOD
- the Den Hague Manifesto -

1) A PI can be an http URI including content negotiation

2) Using LOD vocabularies for schema elements

3) Identifying a minimum set of common elements across space identifiers in scholarly (examples are DOI kernel metadata, DataCite kernel, etc.)

4) To use ‘same as' to help PI interoperability

5) To use PI for subjects and objects in the RDF triples

Recent and Forthcoming initiatives

- **PersID project by the KE**
- **KE meeting on PI for digital objects (14-15 June 2011)**

  Den Hague Manifesto
- **KE meeting on PI for people (13-14 March 2012)**
- **Seminar on global interoperability and linked data in libraries (18-19 June 2012)**
- **Cultural Heritage Online (13 December 2012)**
- **APARSEN - Interoperability Framework …..**

The PersID project

Meta Resolver Service Infrastructure

www.persid.org

Maurizio Lunghi

Fondazione Rinascimento Digitale

The Hague 14-15 June 2011

***Weak points**: a lot of initiatives are on-going but fragmented*

- ✓ A unique PI technology or domain cannot be expected/imposed to all the user communities so we must manage heterogeneity
- ✓ Technology is not the most important challenge – but agreed policies and governance
- ✓ Added value services tailored on different user communities, are still inadequate and there are not cross-domains services

Goals

- ➢ To set up a European infrastructure to resolve the **URN:NBN** namespace (national registers)
- ➢ To establish a **policy for long term sustainability** of an international resolution and discovery service
- ➢ Reference model to describe an **interoperability framework** and provide an integrated entry point
- ➢ To review & update the related URN RFCs

Why NBN?

Trust-based – National Libraries are responsible for the management of the namespaces. The **service** is provided by a **third part** in a neutral way and the **same** for any **user community**.

Digital Preservation – specific for long term preservation applications for digital repositories

Open & Flexible

- suitable for different user communities (cultural, scientific, private ..) with different user requirements (e.g. selection criteria, fragments, rights management, etc ..)
- enables different architectures to share responsibility and sustainability

Trust Architecture

Content – The **National Libraries** and **National Institutions** grant for their contents **certification, authenticity, integrity, updating**, in a very long term vision.

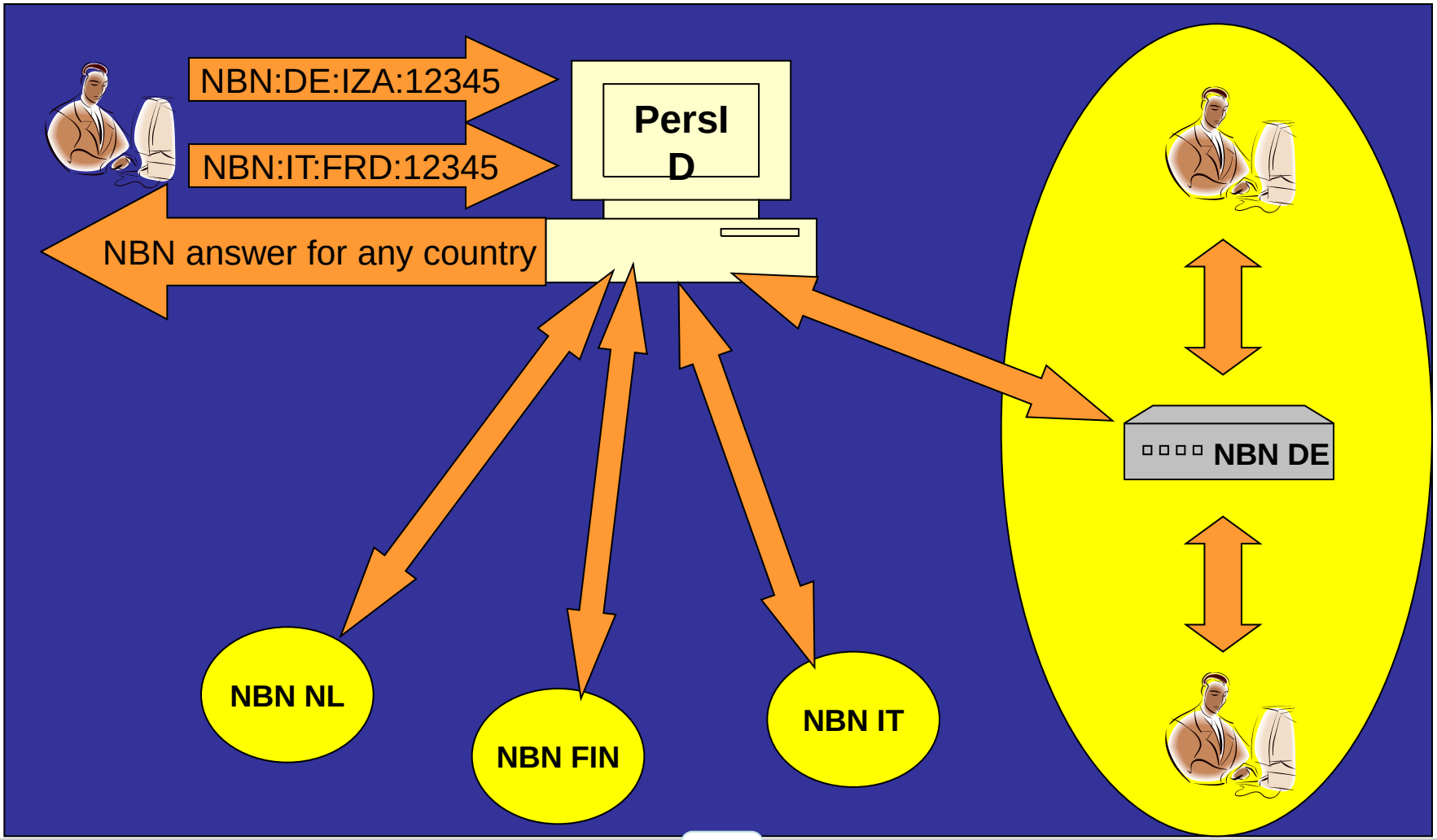
Resolution – The **7 National NBN Registers** in each country grant for the updating and validity of the association between name and info+URL. The Meta Resolver doesn't have a copy of tables but refers always to the National Registers.

Meta Resolver – The central server has a **mirror** and it's in contact with all the National Registers.

Other Resolvers – Some of them are 'trusted' like Handle and DOI or ARK, even if the level of service is not the same.

Outcoming results – II

http://NBN-metaresolver.persid.org



Fondazione RINASCIMENTO digitale

APARSEN
Alliance Permanent Access to the Records of Science in Europe Network

APARSEN

Network of Excellence – 4 years

WP2200
Identifiers and Citability

Maurizio Lunghi
Emanuele Bellini
Paolo Bouquet
Barbara Bazzanella


APA
RSEN
Alliance Permanent Access to the
Records of Science in Europe Network


Science & Technology
Facilities Council


CERN


stm


UNIVERSITÀ DEGLI STUDI
DI TRENTO


UK • DATA
ARCHIVE


In|Con|Tec


Data Archiving and Networked Services
DANS


BRITISH
LIBRARY

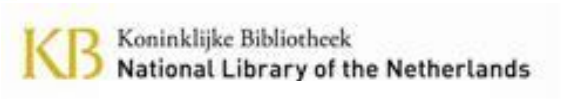

FTK
navigating the networked economy


AIRBUS


INES
Centre Informatique National
de l'Enseignement Supérieur


ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS


inmark
estudiosyestrategias


KB
Koninklijke Bibliotheek
National Library of the Netherlands


esa


HELMHOLTZ
ASSOCIATION


LIBER
LIBER


cini


Fondazione RINASCIMENTO
digitale
nuove
tecnologie
per i beni
culturali


Tessella
Technology & Consulting


LULEÅ
TEKNISKA
UNIVERSITET


dpc


DigitalPreservationCoalition

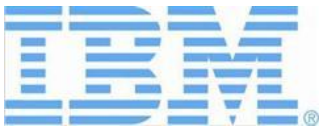

DEUTSCHE
NATIONAL
BIBLIOTHEK


ИКИ


secure
sba-research.org

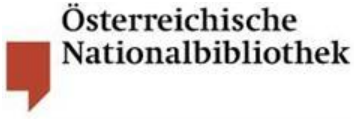

APA
ALLIANCE FOR
PERMANENT ACCESS


GLOBIT

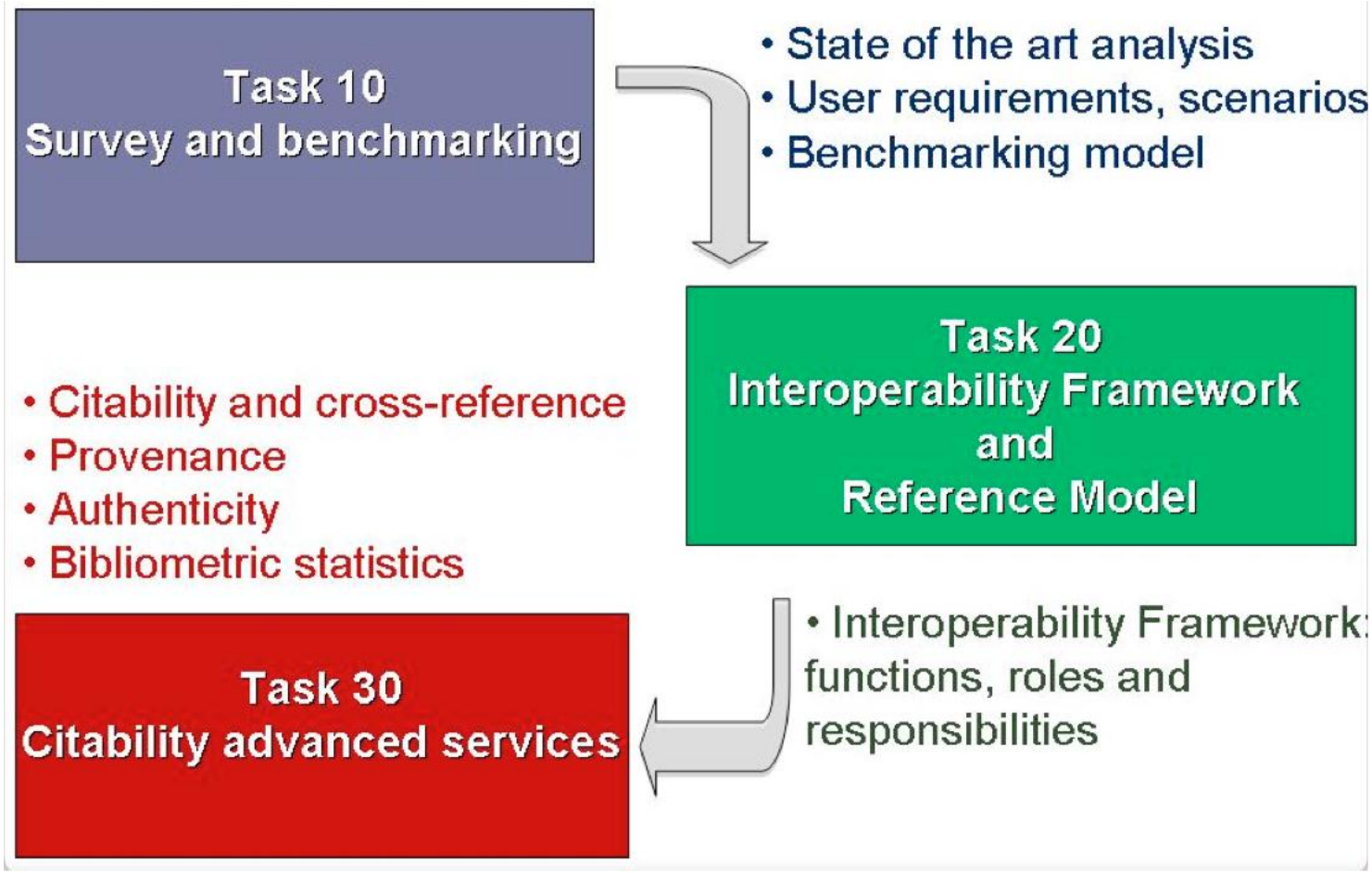

IBM


FORTH
Institute of Computer Science


CSC


Österreichische
Nationalbibliothek

WP 22, IDENTIFIERS AND CITABILITY: OBJECTIVES AND TASK RELATIONS



WP 22: achievements (month 12)

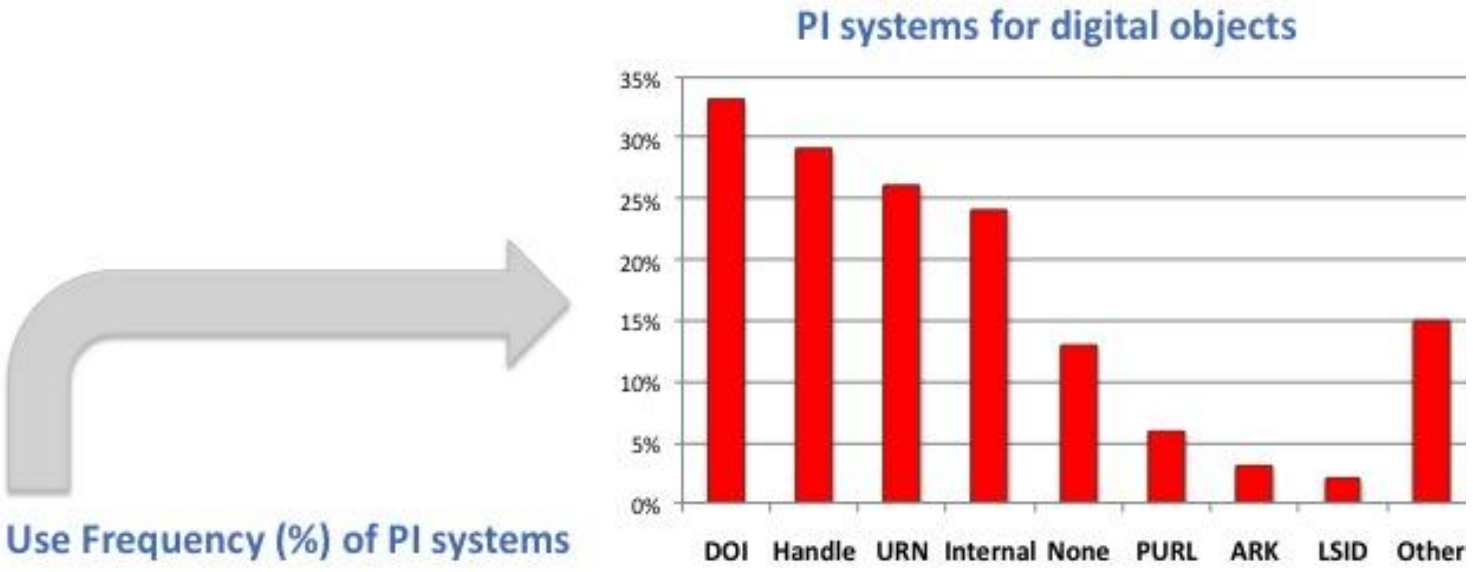
TASK 10: Survey and Benchmarking

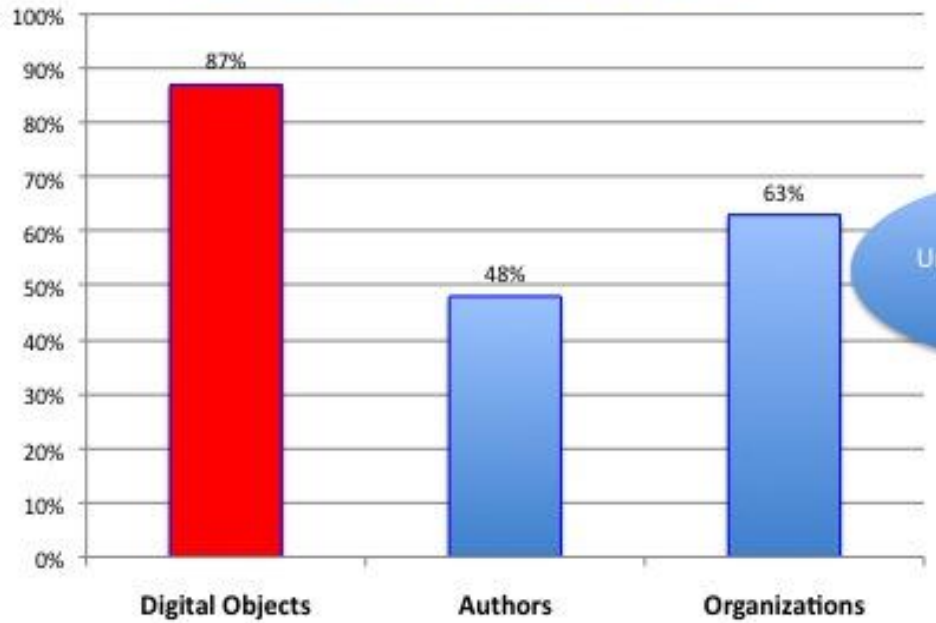
- **State of the art analysis**:
  - PI systems for digital objects and authors: feature analysis
  - PI interoperability: analysis of related projects (e.g. ORCID, PersID, RIDIR, OKKAM…).
- **Survey on PI systems** for digital objects, authors and organizations (103 respondents) ➡ uses and practices, criteria for adoption, limits, user requirements.
- **Benchmark assessment**:
  - Eligibility criteria for the interoperability framework
- Definition of possible **scenarios** about the use of the interoperability framework ➡ **use cases** (who, what, why, where, when, how, issues).

PI SYSTEMS

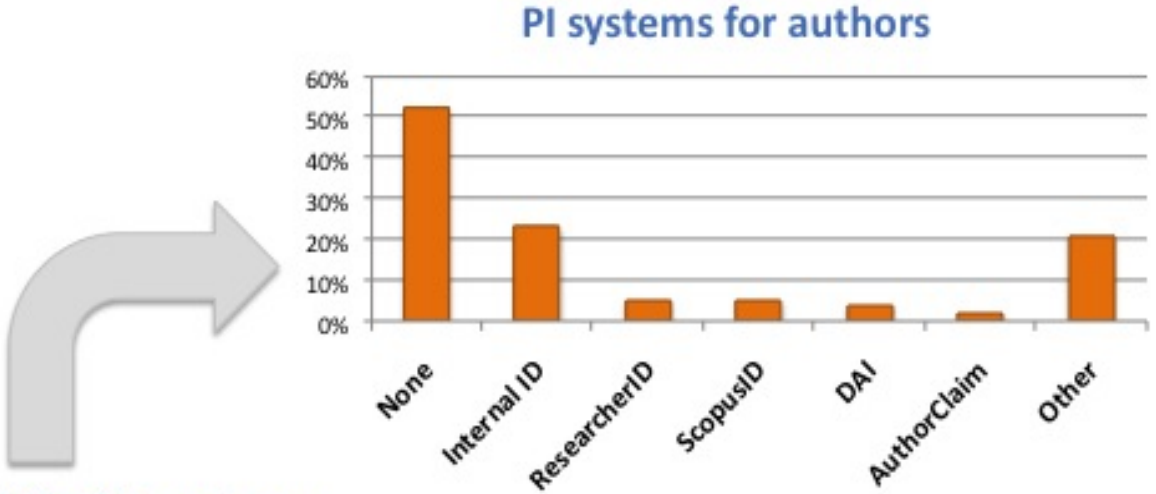
A **COMPLEX** AND **FRAGMENTED** LANDSCAPE:
some evidence from the PI survey

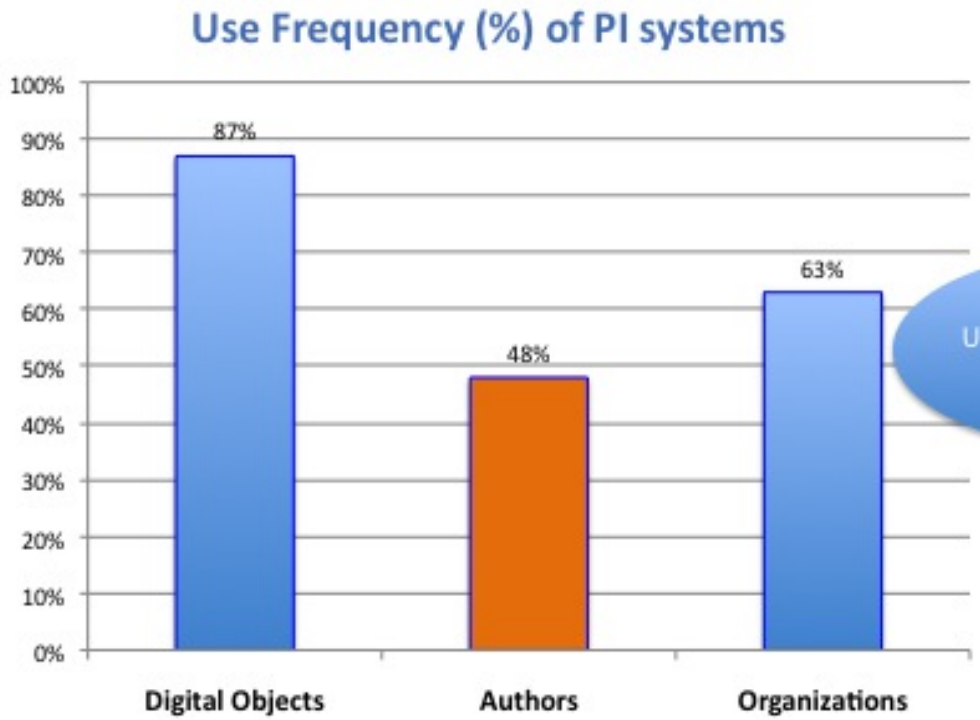
PI systems for **Digital Objects**



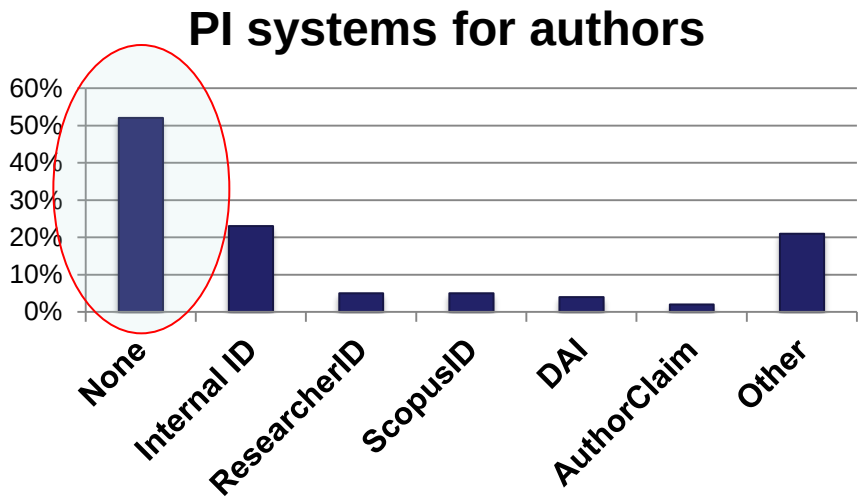


PI systems for **Authors**





PI SYSTEMS FOR AUTHORS

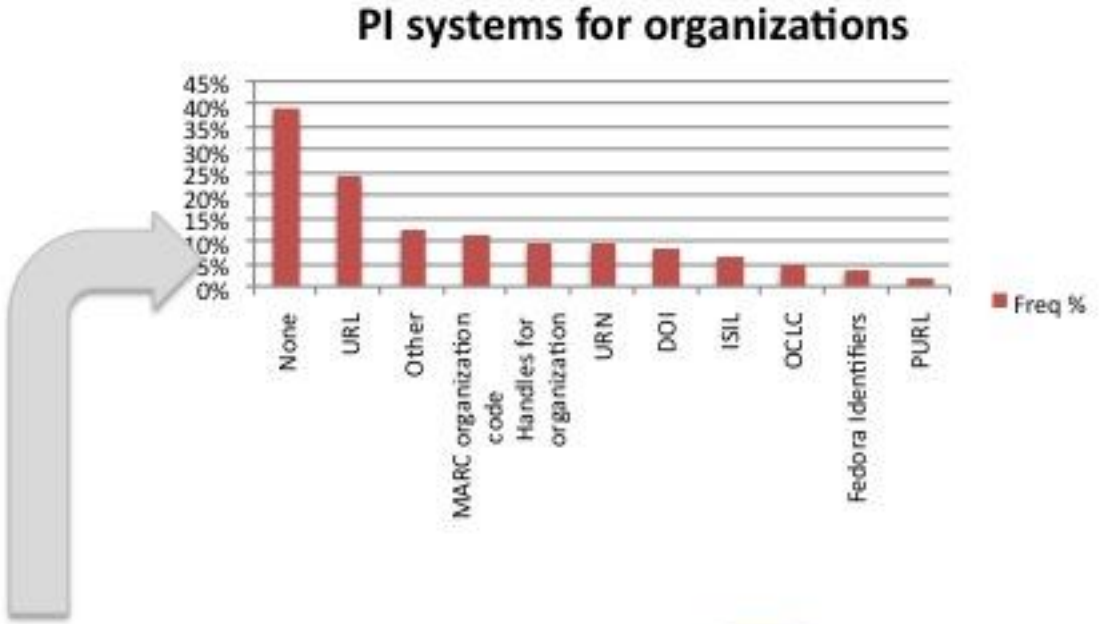


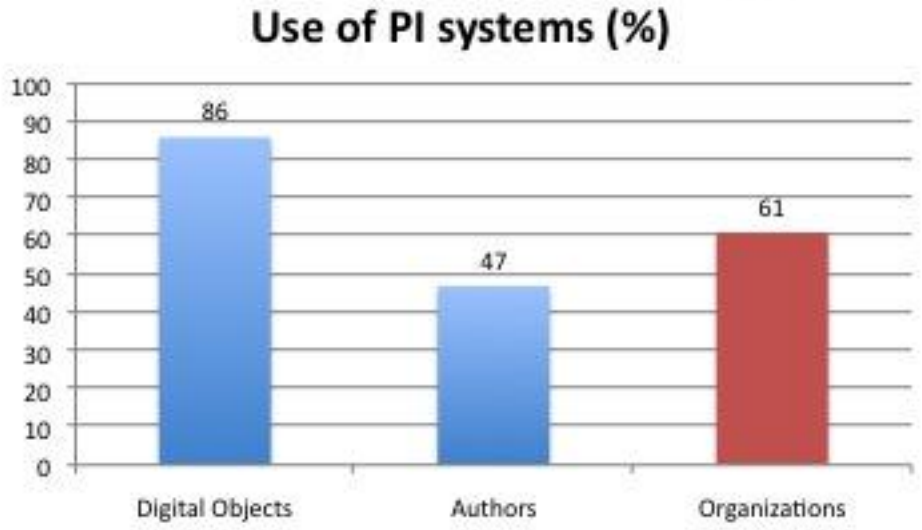
Main Obstacles

Obstacles	Freq	Percent %
It is not a key issue for the organisation	**23**	**22,33**
Authors do not know about (or do not care for)	**19**	**18,45**
Low attractiveness of the service due to low level of adoption	13	12,62
Other (please specify)	11	10,68
National legislation with regard to privacy of personal data	8	7,77
Lack of trust and authority	7	6,8
Total N. of respondents	103	

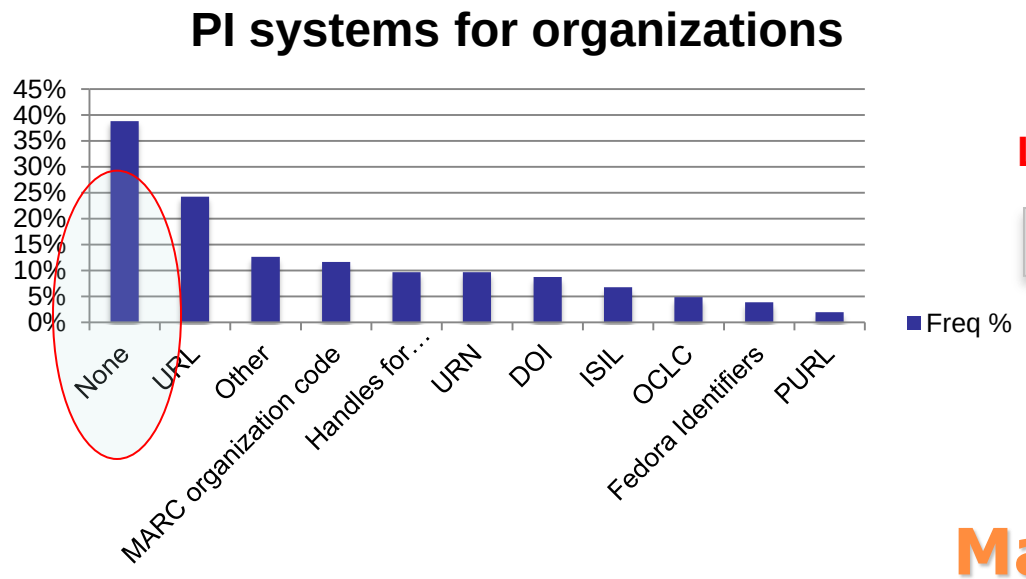
Lack of awareness!

PI systems for **Organizations**





PI SYSTEMS FOR ORGANIZATIONS



Lack of awareness!

Main Obstacles

Value	Freq	Percent %
No enough information about this service	**20**	**19.41**
It is not a key issue for the organization	**19**	**18.45**
Low attractiveness of the service due to low level adoption	6	5.83
Lack of trust and authority	6	5.83
Other	4	3.88
Total N. of respondents	103	

PI SYSTEMS: NEED of TRUST

Requirements	Freq	Percent %
Cross-disciplinary	**83**	**80.5**
Managed by public/government institution	**74**	**71.8**
Nationally not limited	**57**	**55**
Discipline-specific	10	9.7
Other	9	8.7
Nationally limited	5	4.8
Privately managed	7	6.7

Requirements

Cross-boundary systems but…

Factors for Trust

Factors contributing to the trust	Freq	Percent %
Trusted organization running the system	**74**	**71.84**
Methods of verification	**68**	**66.02**
Supported by stable funders	32	31.07
Validation by publishers	31	30.1
Author self-curation	27	26.21
Other	8	7.77
Validation by educators	7	6.8

USER REQUIREMENTS

TECHNOLOGY	%	ORGANIZATION OF THE SERVICE	%	SCOPE	%	NAMING RULE	%
Standard de facto	**53**	**Distributed naming authority**	**49**	**Openness (open to any digital object/actor)**	**82**	**Opaque Identifier**	**55**
Standard de jure	36	Centralized naming authority	38	Closeness (aimed to a particular set of digital objects/actors)	11	Semantic Identifier	35
Open source infrastructure	**88**	**Supported by an institution with a mandate**	**55**	**Cross-community identifier**	**76**	**Deep granularity supported**	**57**
Proprietary infrastructure	4	Supported by a stable funder	34	Community-oriented identifier	16	Low-level granularity	32
Widely adopted	**56**						
Established and mature	36						

From USER REQUIREMENTS to INTEROPERABILITY SCENARIOS

PI services Basic features

PI services advanced features

Value	Freq	Percent %
Citability	**76**	**74**
Global resolution service	**62**	**60**
PI resolution service to the resource	**57**	**55**
Digital Object certification	**55**	**53**
PI resolution service to metadata	**50**	**49**
Association of PI to multiple location (URLs)	41	40
Metrics	31	30
Multiple association name	27	26
Link digital object to dynamic dataset	19	18
Others	3	3

SCENARIONS AND USE CASES

1. Scenarios on **Citability and Metrics services**
2. Scenarios on **Global Resolution Services**
3. Scenarios on **Digital Object Certification**

USE CASES

WHO
WHAT
WHY
WHERE
WHEN
HOW
ISSUES

Scenarios and Use cases: an example

Scenario on unique resolution interface: John wants to find various documents about a theory produced by Mike Mills. The problem encountered by John is that, whenever he tries to use Google to find documents about Mike Mills, he found mostly documents about Mike Mills, the film producer [….]

Who	A cognitive scientist
What	Wishes to discover and explore scientific publications or other information related to a paper, which published the results of a new theory about the role of emotions in decision-making. For instance she want to find: • Other publications on the same theory • Other publications of the same author • Other authors focused on the same research topic The "starting point" paper has its own digital ID.
Why	To explore related content on a topic of interest
Where and When	Online Any time
How	Requires a discovery mechanism to locate resources related to one, which the user had accessed.
Issues	Implies mechanisms for metadata linking though ID.

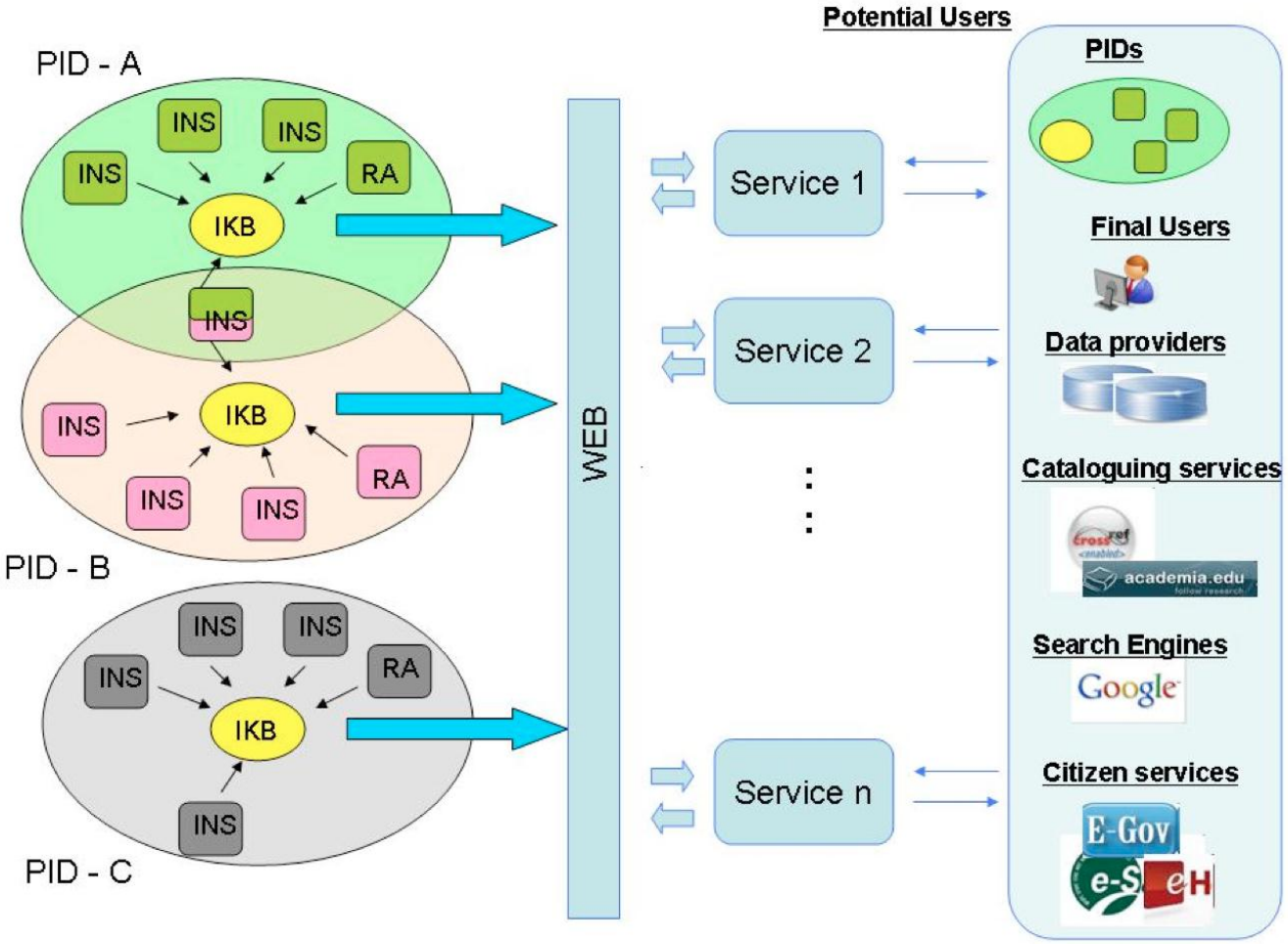
INTEROPERABILITY FRAMEWORK (IF)

GOAL: to set the foundations and identify the basic concepts within the universe of PI systems, for developing appropriate **interoperability solutions and services.**

TRUSTED PI SYSTEM: CRITERIA

1. Having at least one Registration Agency (M).
2. Having one Resolver accessible on the Internet (M).
3. Uniqueness of the assigned PIs within the PI domain (M).
4. Guaranteeing the persistence of the assigned PIs (M).
5. User communities, which implement the PID should implement policies for digital preservation (e.g. trusted digital repositories) (O)
6. Reliable resolution (M).
7. Uncoupling the PIs from the resolver (M).
8. Managing the relations between the PIs within the domain (O).

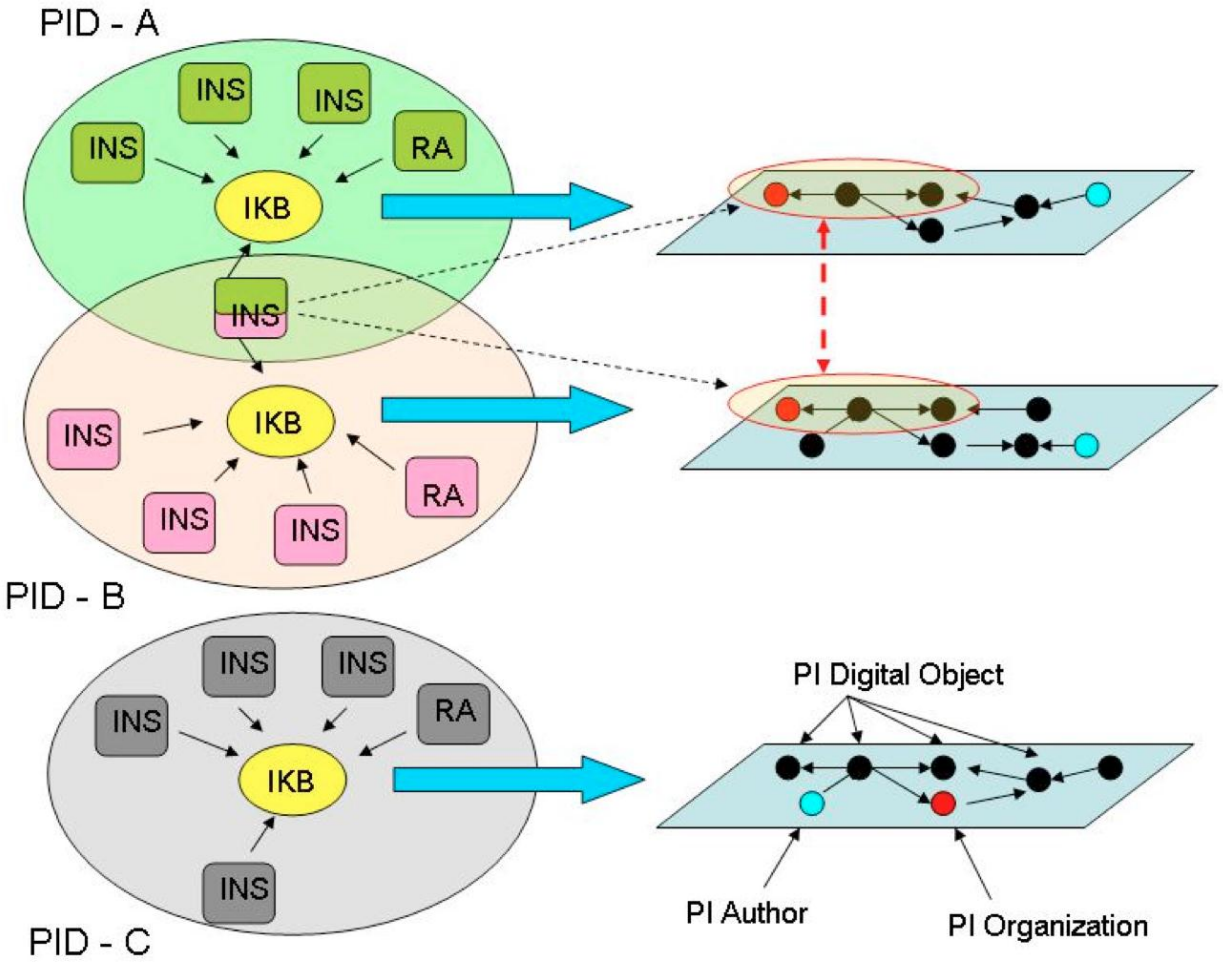
INTEROPERABILITY FRAMEWORK (IF)



Fondazione RINASCIMENTO digitale

APA RSEN
Alliance Permanent Access to the Records of Science in Europe Network

INTEROPERABILITY FRAMEWORK (IF)



INTEROPERABILITY FRAMEWORK (IF): main concepts

- Digital Object
- Author
- Institution
- Persistent Identifier (PI)
- PI Domain (PID)
- Policy
- Resolver
- User

Definition: A Digital Object is any kind of digital resource, which is identified by at least one PI assigned by a trusted PID.

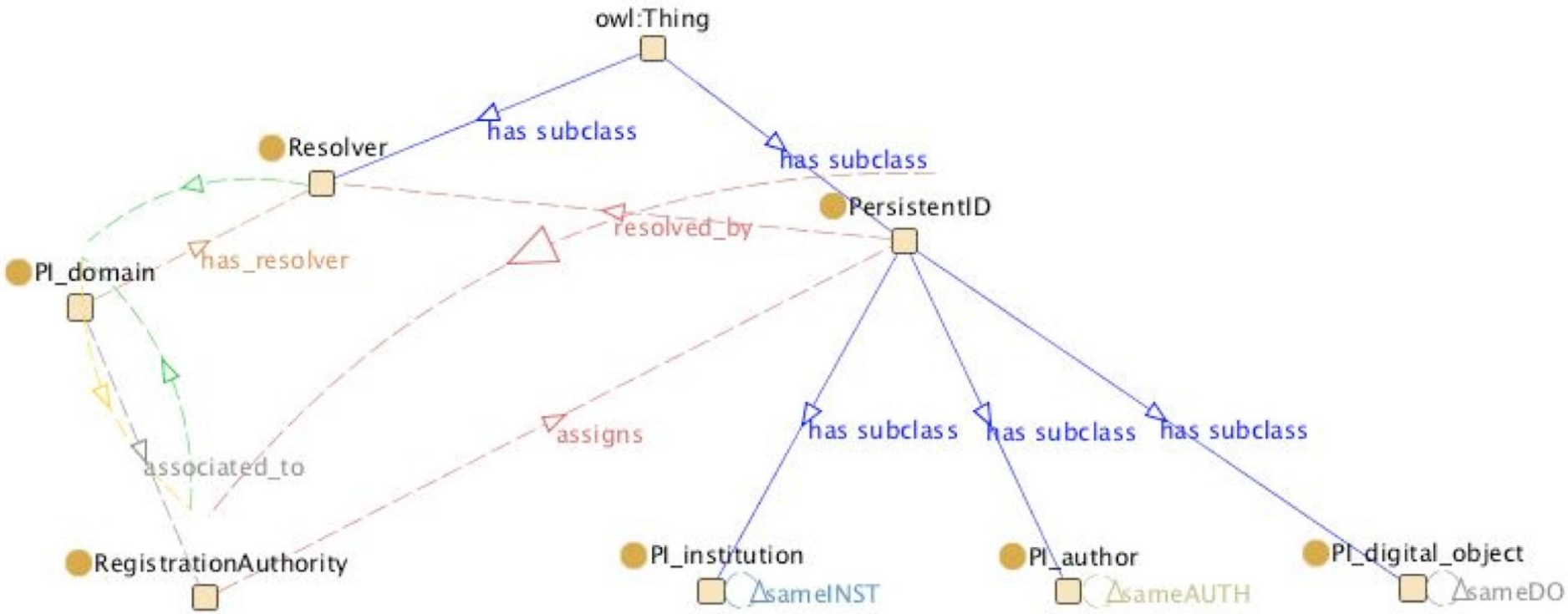
Relationships:

1.Is identified by at least one digital object PI (<hasPI_do>)
2.Is created by an Author (<created_by>);
3.Is related to other digital objects (<related_to>);
4.Is associated to Policies (<associated_to>);
5.Can be described by metadata (<has_metadata>);

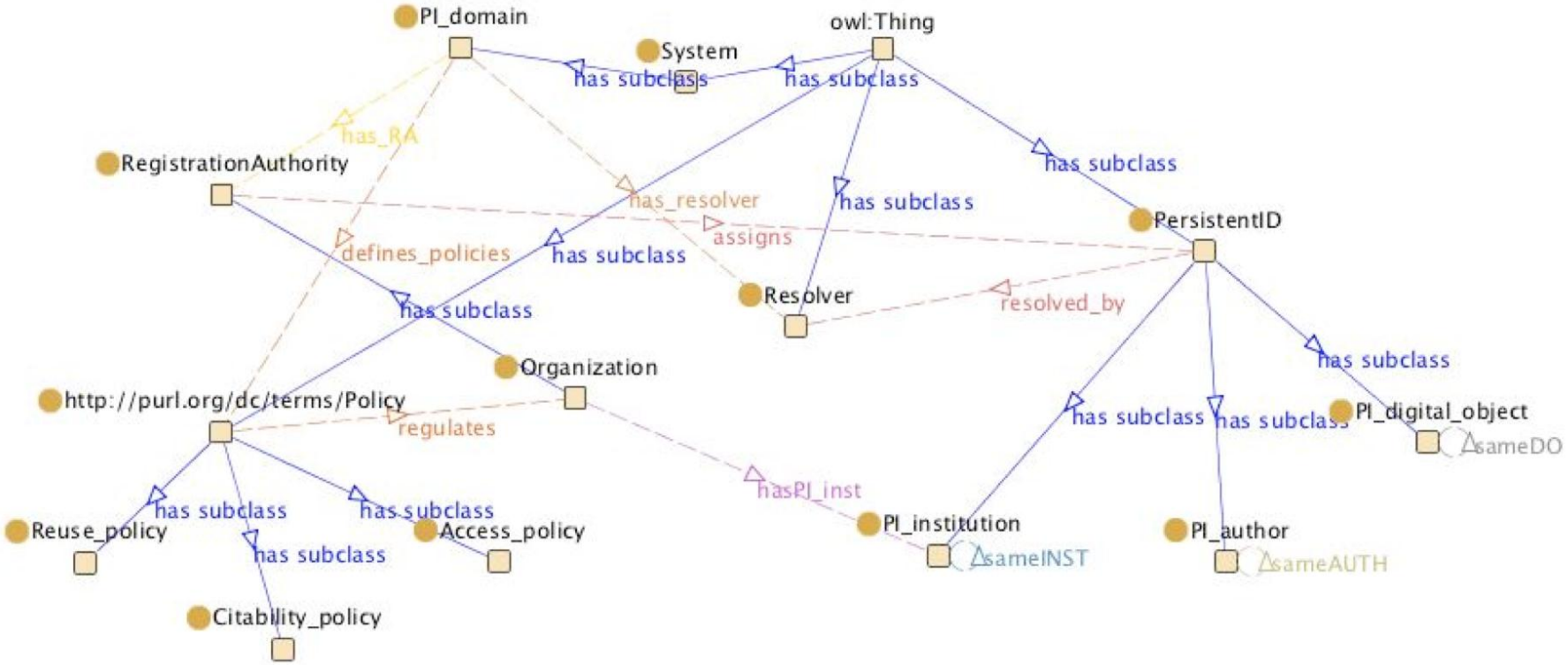
Fondazione RINASCIMENTO digitale

APA RSEN
Alliance Permanent Access to the Records of Science in Europe Network

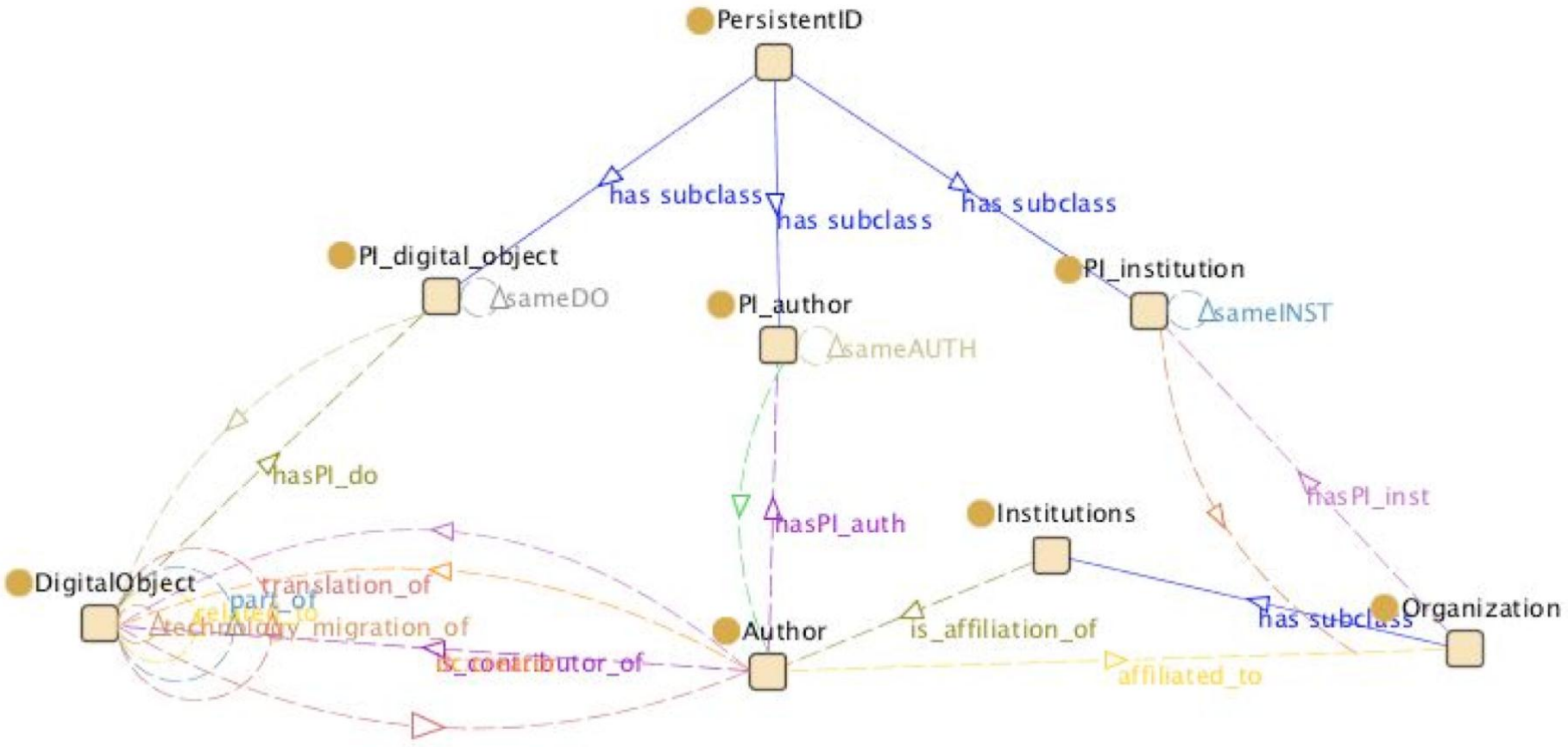
IF: PI and their **RELATIONS**



IF: PI DOMAINS and their RELATIONS



IF: PIs, REFERENTS and their RELATIONS



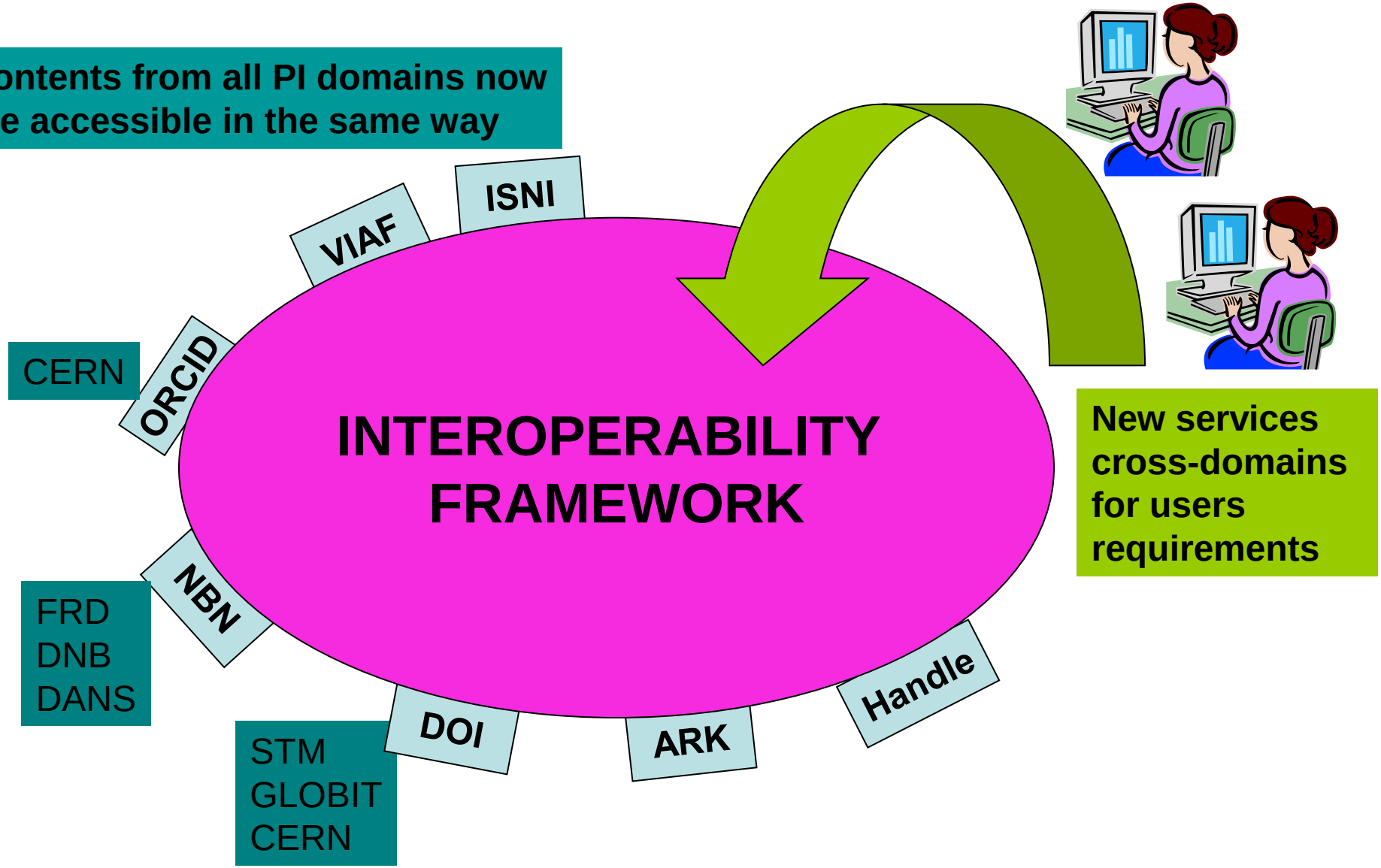
CONCLUSIONS AND NEXT STEPS

Interoperability Framework (IF) for PI systems

1. Validation of the model through a user group with experts on PI
2. Definition and set up of a demonstrator with data from different PI domains and for objects, people and bodies
3. Proposal of few services and development on a cross PI Domains (PIDs) basis

INTEROPERABILITY FRAMEWORK (IF)

Contents from all PI domains now are accessible in the same way



New services cross-domains for users requirements

Fondazione RINASCIMENTO digitale

APA RSEN
Alliance Permanent Access to the Records of Science in Europe Network

Thanks for your attention

Maurizio Lunghi

lunghi@rinascimento-digitale.it