DOI SYSTEM

Persistent Identifier eScience Seminar
Munich, March 27/28

International DOI Foundation

Outline / Key concepts

- "DOI System: what it adds to the Handle System"
- Origin and current position of DOI System activities
- System components
- Data model
- Social infrastructure: "Business model"

The DOI System

- DOI (Digital Object Identifier) System: www.doi.org

- Initially developed from the publishing industry but now wider
  - a non-profit collaboration to develop infrastructure for persistent identification and management of content
  - Approx 2000 user organisations (through agencies)
  - CrossRef (scholarly publishers); EC; science data; major ISBN agencies; etc.

- Currently being standardised in ISO (TC46/SC9)
  - the home of ISBN etc "content identifiers"

- One application of the *Handle System*®
  - adds to it additional features – social and technical infrastructure, policies, metadata management
  - focus on one area of interest (content/intellectual property)
  - offers a specific data model based on indecs (discussed later)
  - DOI System technology equally applicable for **parties** and **licences**

Terminology: the over-used term "identifier"

"Identifier" as numbering schemes

- Registries
- Normally central control, commitment
- Examples: ISBN, EAN bar codes, IANA, ITU phone numbering plans etc
- Normally focus on attributes (metadata)

"Identifier" as syntax specifications

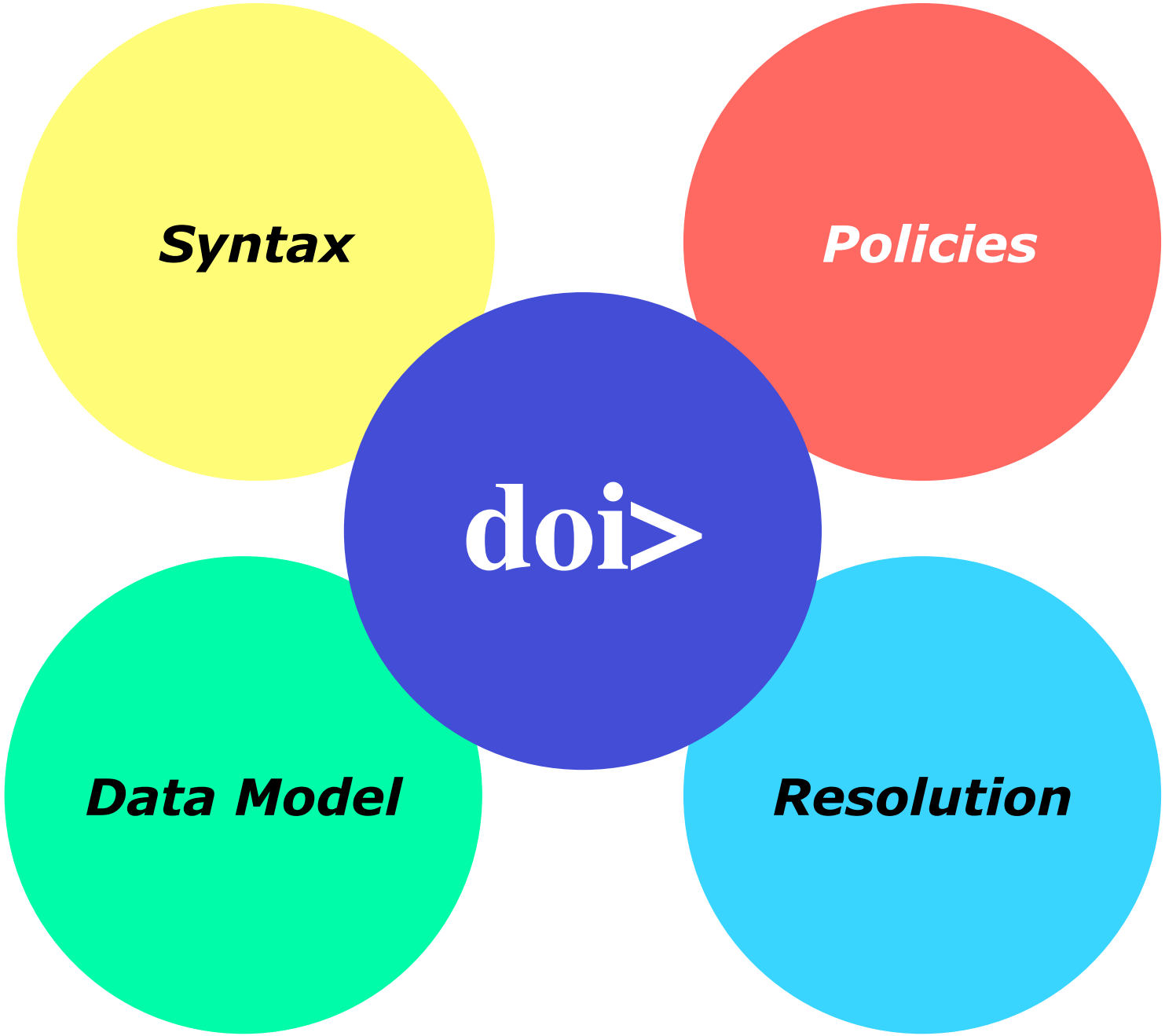
- Normally little central control
- e.g URI (URL); MPEG-21 DII
- Few structured attributes, low barriers to entry
- Some more structured than others: e.g. URN, info URI

Other confusions:

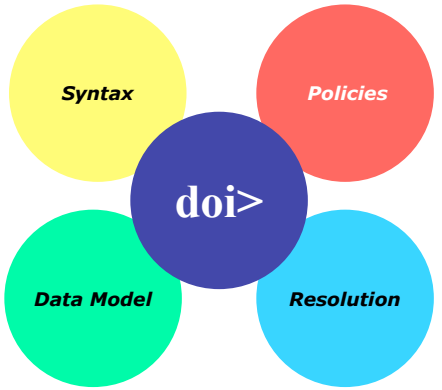
- Some practical systems use both schemes and specifications
- *Representations* and *interactions* between different schemes and specifications:
  - e.g. an ISBN can be expressed as a URL, as an EAN bar code, a DOI name, etc
- Identifier as "system" versus as a "unique label"
- Schemes begin to be used for things outside scope

Interoperability

- We all know our own back yard ("We all know what we mean")
- Q: Why do we want persistent identifiers?
- A: For interoperability
  - Interoperability = the possibility of use in services outside the direct control of the issuing assigner
- "persistence is interoperability with the future"
- We know what we mean, but others may not.
  - Identifiers assigned in one context may be encountered, and may be re-used, in another place (or time) - without consulting the assigner. You can't assume that your assumptions will be known to someone else.
- Interoperability is accelerated through automation:
  - Two key events:
  - 1966: automation of supply chains (ISBN)
  - 1994: automation of sharing resources (WWW)
- Increasing interoperability = increasing chance of breakdown

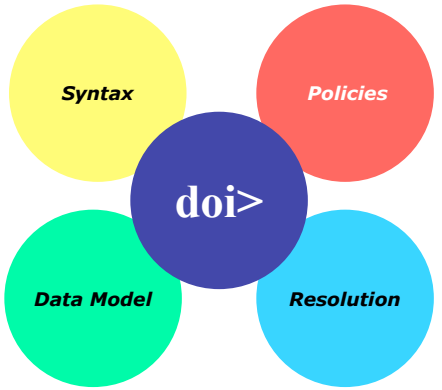

Syntax
Policies
doi>
Data Model
Resolution

DOI name syntax can include **any existing identifier**, formal or informal, of any entity

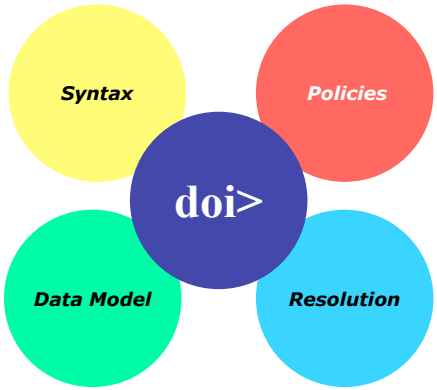


- An identifier "container" e.g.
  - 10.1234/5678
  - 10.5678/978-0-7645-4889-4
  - 10.2224/2007-01-0verview-DOI
- NISO standard Z39.84
- First class object: name
  - Not "intelligent" as a label
  - Cannot tell what it is from looking at the DOI name
- Redirection through *resolution*

- Resolve from DOI name to data
  - initially Location (URL) – *persistence*
- May be to multiple data:
  - Multiple locations
  - Metadata
  - Services
  - Extensible
- Uses the Handle System
  - Implementing URI/URN concept
  - Advantages of granularity, scalability, administrative delegation, security, etc



Resolution allows a DOI name to link to **any & multiple pieces** of current data

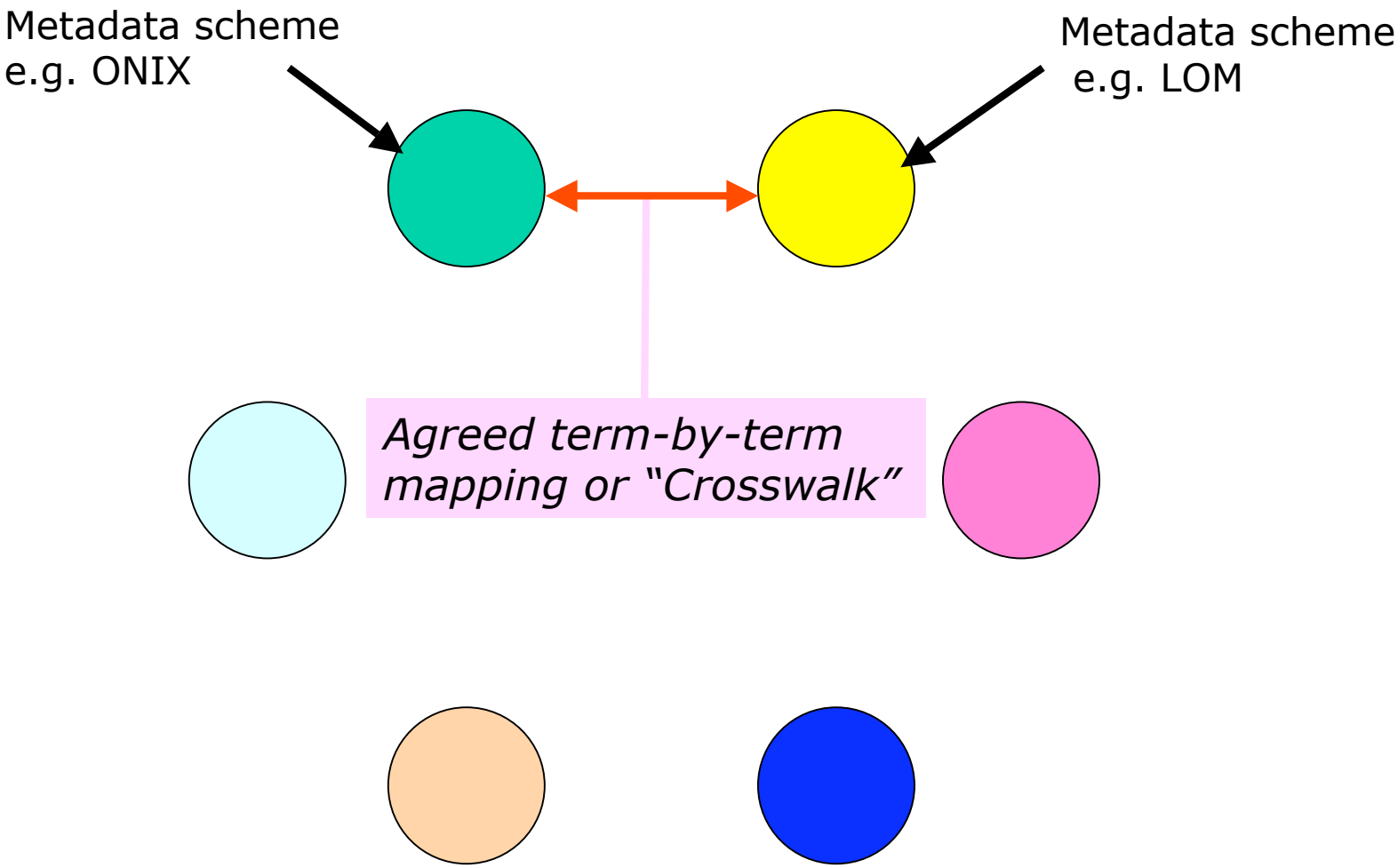


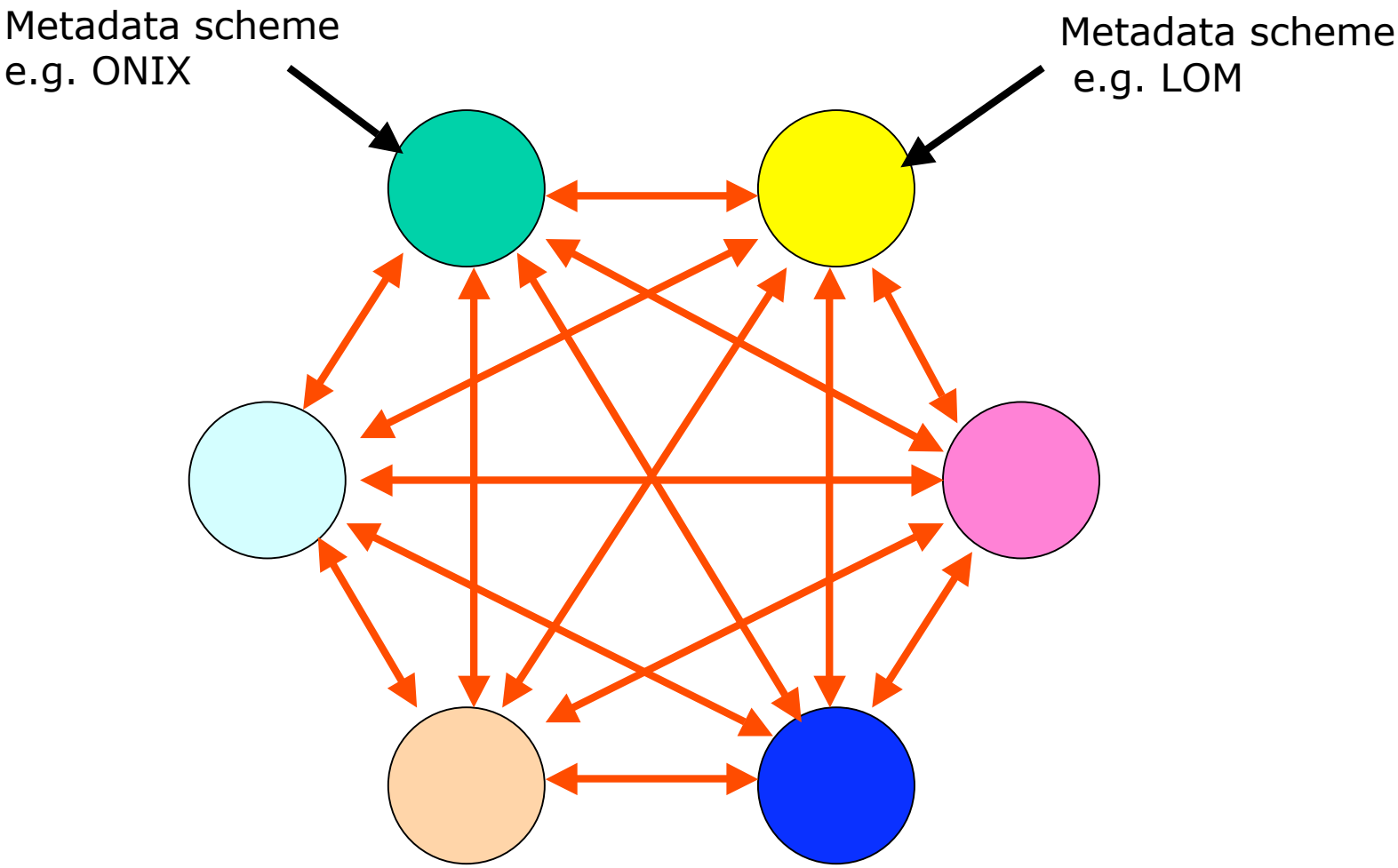
- Metadata tools:
  - a data dictionary to define
  - a grouping mechanism to relate
- Necessary for interoperability
  - "Enabling information that originates in one context to be used in another in ways that are as highly automated as possible".
- Able to use existing metadata
  - Mapped using standard dictionary
  - can describe any entity at any level of granularity

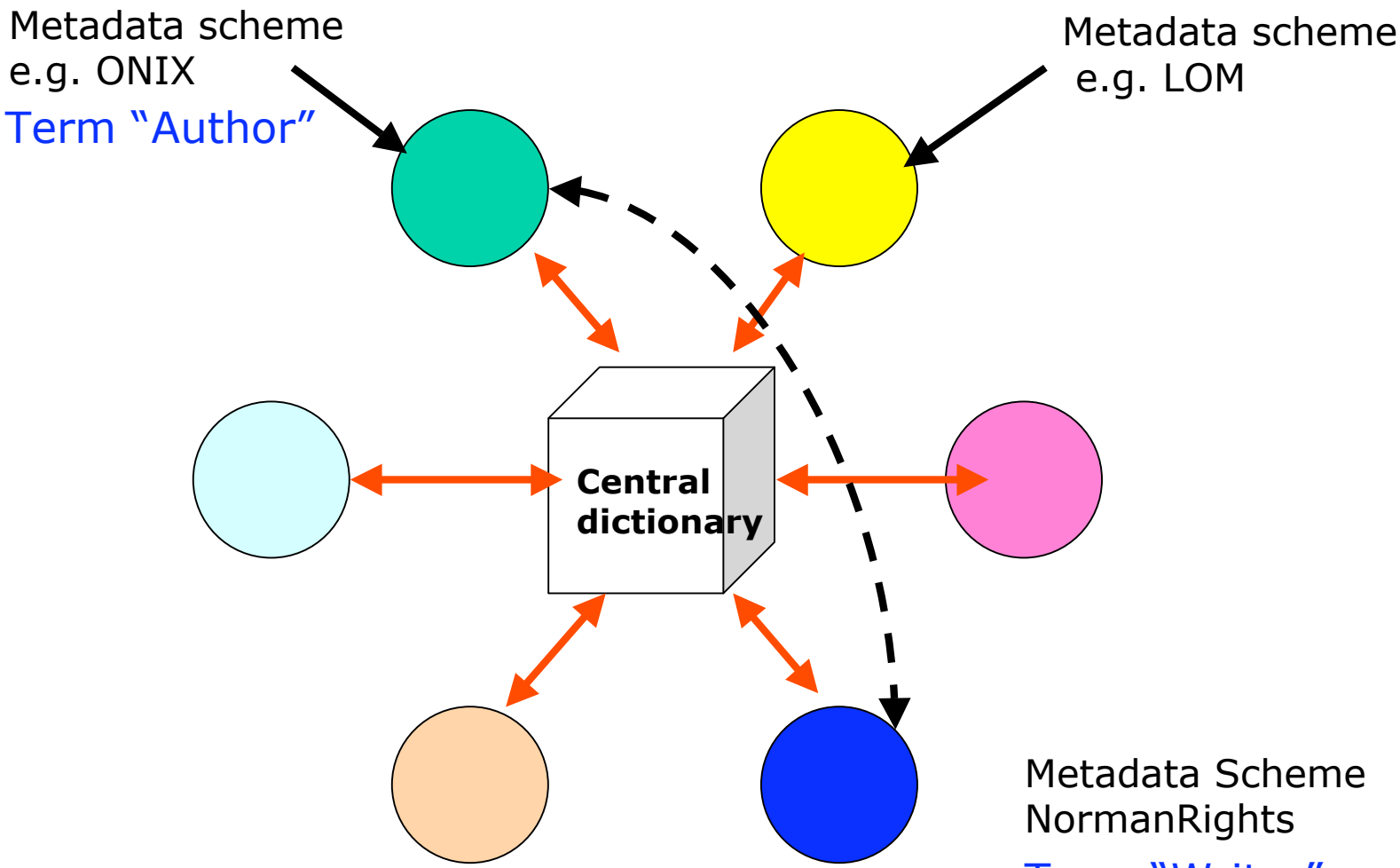
<indecs>
Data Dictionary
+
DOI AP framework

DOI System data model

- The underlying model of how data within the DOI System relates to other data
- Two components
  - Data Dictionary + DOI Application Profile Framework
- Data Dictionary
  - Provides tool for precise *description of entity* through metadata (and mapping to other schemes).
- DOI Application Profile framework.
  - Provides means of *relating entities*: grouping entities and expressing relationships
  - A mechanism for grouping DOI names with similar properties
- DOIs, APs, and DOI System services built using these:
  - have many-to-many relationships: expressed through multiple resolution (handle)
  - may have precise descriptions: expressed through metadata in Data Dictionary


Metadata scheme
e.g. ONIX
Metadata scheme
e.g. LOM
Agreed term-by-term
mapping or "Crosswalk"


Metadata scheme
e.g. ONIX
Metadata scheme
e.g. LOM


Metadata scheme
e.g. ONIX
Term "Author"
Metadata scheme
e.g. LOM
Central
dictionary
Metadata Scheme
NormanRights
Term "Writer"
ONIX:Author = NormanRights:Writer

The <indecs> framework

Interoperability of Data in E-Commerce Systems

- <indecs> project 1998-2000
- <indecs2> 2001-2002 (= MPEG21 Rights Data Dictionary)

- Focus on multimedia rights metadata: recognized that rights and descriptive metadata were inseparable. Produced an event-based reference model/framework (parties, resources, agreements)

- indecs: 50% EC funding + consortium members including:
  - EDItEUR (international book industry standards/ONIX)
  - IFPI (international record industry)
  - MPAA (international film industry)
  - Various copyright societies and associations
  - Various technology providers
  - Library and author representatives
  - International DOI Foundation

- Metadata in networks needs to support interoperability across
  - media (e.g. books, serials, audiovisual, software, abstract works).
  - functions (e.g. cataloguing, discovery, workflow, rights mgmt).
  - levels of metadata (from simple to complex).
  - semantic barriers.
  - linguistic barriers.

The <in*d*ecs> framework

doi>

Principles:

- Unique Identification: every entity should be uniquely identified within an identified namespace.
- Functional Granularity: it should be possible to identify an entity whenever it needs to be distinguished *[1st class].*
- Designated Authority: the author of an item of metadata should be securely identified.
- Appropriate Access: everyone requires access to the metadata on which they depend, and privacy and confidentiality for their own metadata from those who are not dependent on it.
- Definition of metadata: An item of metadata is a relationship that someone claims to exist between two referents *(description).*

Delivered:

- Generic data model of e-commerce
- Applicable to all types of intellectual property
- Specifications for supporting services
- Standardisation proposals
- Documentation at www.indecs.org

Led to:

- Contextual ontology architecture: **contexts, roles, identities**

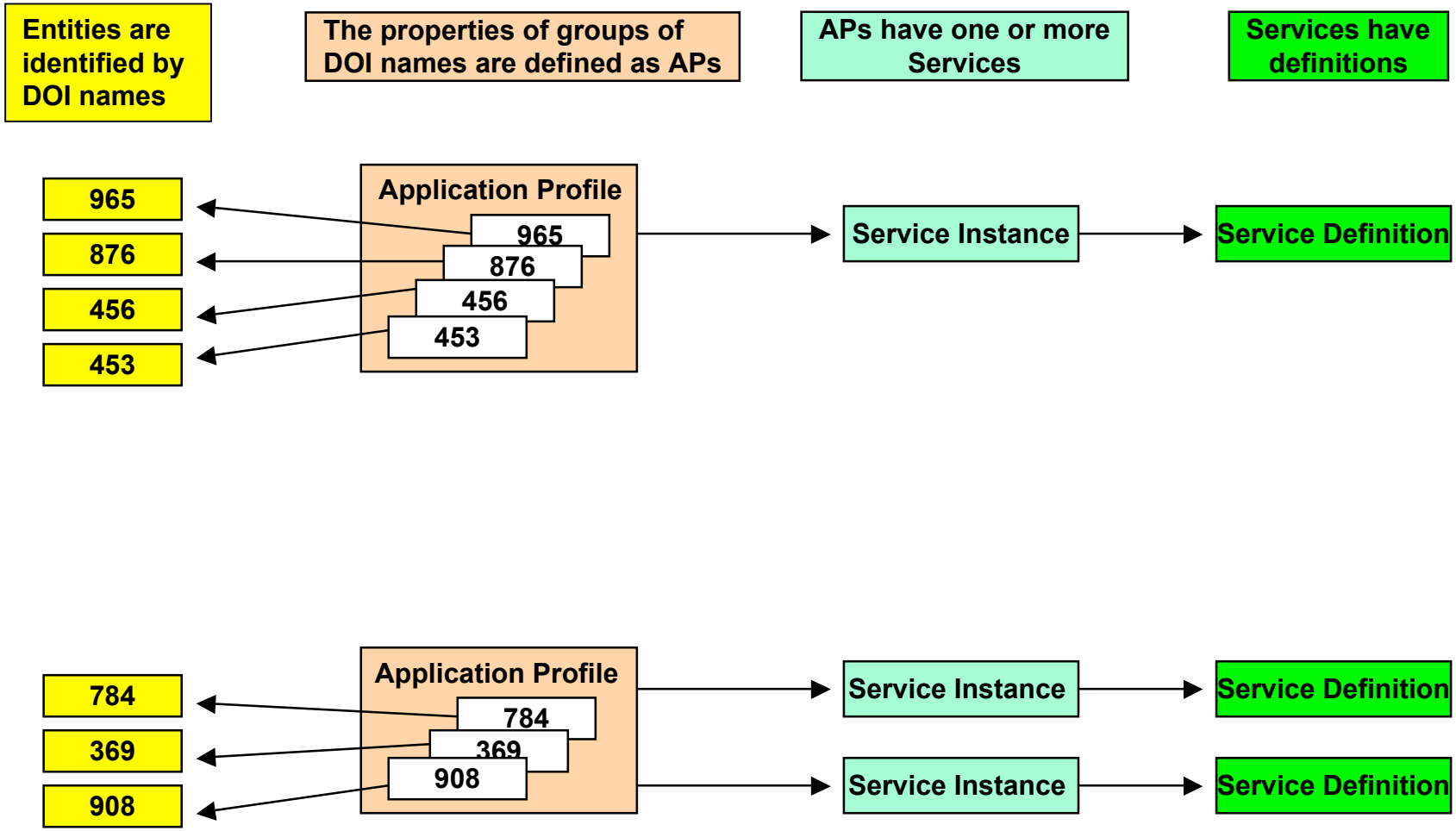
A pointer is not enough

- Precisely what is being named?
  - *Suppose I have here a pdf version of Defoe's "Robinson Crusoe" issued by Norton. I find an identifier – is it of:*
  - All works by Daniel Defoe
  - The *work* "Robinson Crusoe"?
  - The Norton *edition* of "Robinson Crusoe"?
  - The pdf *version* of the Norton edition of…. ?
  - The pdf version of…held on *this server*…?

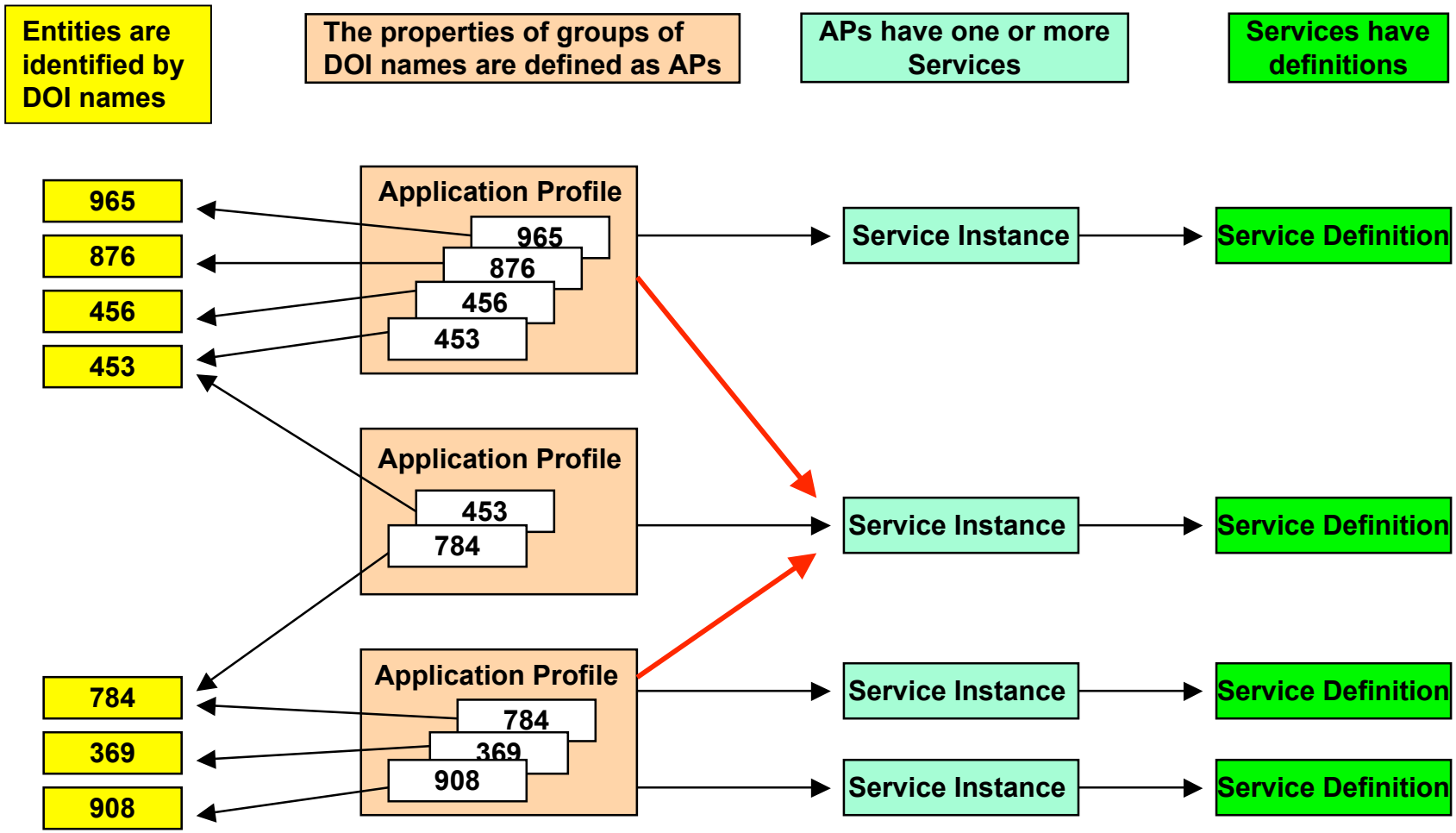
- Most digital objects of interest have compound form, simultaneously embodying several referents.

- Multiple identifiers may be necessary (compare music CDs)

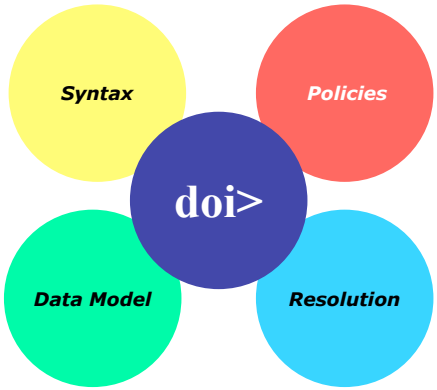
Application Profile (AP) Framework



Application Profile (AP) Framework



- New APs and services may be created or made available
- One change to an AP to affect all DOI names within that AP



DOI System policies allow **any business model** for practical implementations

- Implementation through IDF
  - Governance and agreed scope, policy, "rules of the road" , central tools (dictionary, resolution mechanism)
  - Cost-recovery (self-sustaining)
- Registration agencies ("franchise")
  - Each can develop own applications
  - Use in "own brand" ways appropriate for their community
  - Examples: CrossRef, OPOCE, TIB

DOI System adds value

- Costs: need for human intervention and support of an infrastructure. The DOI System operates on the basis that such costs are borne by the assigner of the DOI name.
  - Number and metadata registration *(maintenance of resolution destinations; declaration of metadata; validation of number syntax and of metadata; liaison with the IDF registry)*
  - Infrastructure *(resolution service maintenance, scaling and further development; customer guidance and outreach; marketing; administration)*
  - Governance *(common "rules of the road"; further development and support of the system)*

- The way in which these costs are recouped depends on the application.

IDF principles

- DOI name free to resolve at point of use
- Costs of assignment paid by assigner
- Allow different business models
  - allow anyone to do what they want
- Distributed system
  - agencies

IDF principles

- IDF to provide governance layer only
  - using a federation of registration agencies
- IDF sets out minimum criteria for registration agencies:
  - rules of the road: not journey maps
  - technical; information management; $
- Does not prescribe details of individual businesses
  - user communities set their own agendas
- Comparable models work well:
  - EAN/UPC; Visa; ISBN etc.

IDF principles

- Variety of potential business models
  - one size does not fit all; need more flexibility
- Variety of naming authority models
  - currently model is one per registrant
  - encourage move to appropriate level
  - allow subdivision (10.1000.2/123)
- RA's will develop and offer applications
  - number registration alone not a "business"

Costs

- For an everyday user:
  - Free: any DOI name may be resolved by anyone
  - No obligations
- For an assigner:
  - Must work through a *Registration Agency*
  - Cost depends on application: DOI registration is bundled in
    - e.g. CrossRef – crosslinking of citations: for a publisher, from $275 per year (2008)
- For a Registration Agency:
  - Must be a full RA member of the *International DOI Foundation*
  - Fees based on volume
- Developing, managing, implementing, standardising, etc:
  - Paid for by International DOI Foundation (open to anyone)

Terms between Registration agencies and IDF

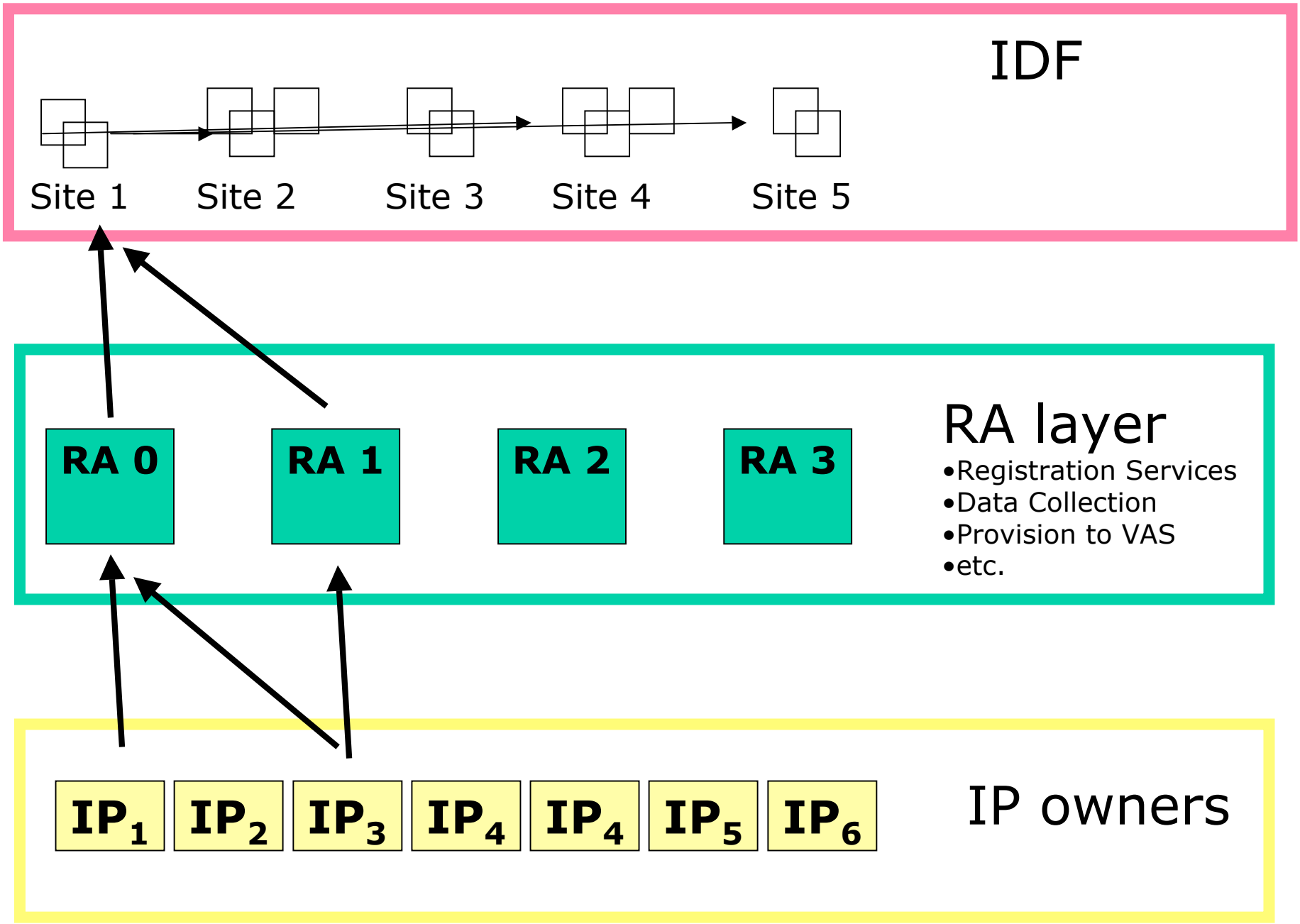
- Common agreement
  - level playing field
  - recognise differences in scale
- "Franchise" model
- Evolving financial model
  - Sliding scale costs per DOI name allocated
  - membership of IDF
  - minimum $20K
- RA working group
  - Letter of Intent
  - Formal agreement

Current RA fee structure

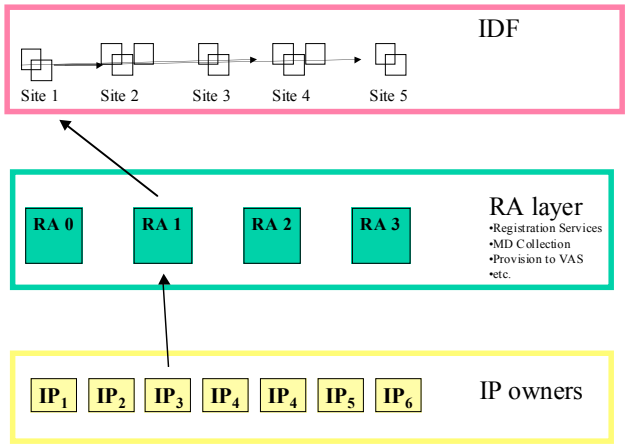
- Fee structure between IDF and RA comprises:
  - Membership in IDF; plus
  - Operating Fees (paid direct to IDF's technical operator), made up of two elements:
    - per DOI name Registration Fee, and
    - Maintenance Fee
- Membership fee: $35,000 per year
- Registration fee:
  - First 250,000 DOI names per year: $20,000 minimum fee
  - Sliding scale up to 5M DOI names/year: from 4 cents to 0.05 cents per DOI name
- Maintenance fee: 1 cent per DOI name up to 5M, 0.5 cents per DOI name thereafter

Under discussion to be changed:

- Cost per DOI name to drop to zero after 5M
- Maintenance fee to be based on a limited time period

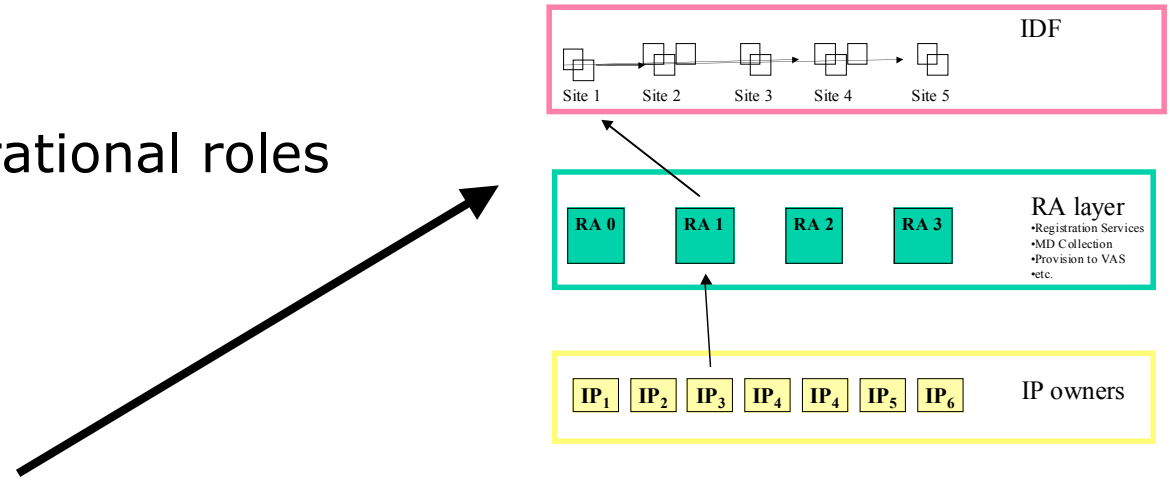

IDF
Site 1
Site 2
Site 3
Site 4
Site 5
RA 0
RA 1
RA 2
RA 3
RA layer
•Registration Services
•Data Collection
•Provision to VAS
•etc.
IP1
IP2
IP3
IP4
IP4
IP5
IP6
IP owners

DOI System operational roles



IP owners: register DOI names with agency

DOI System operational roles

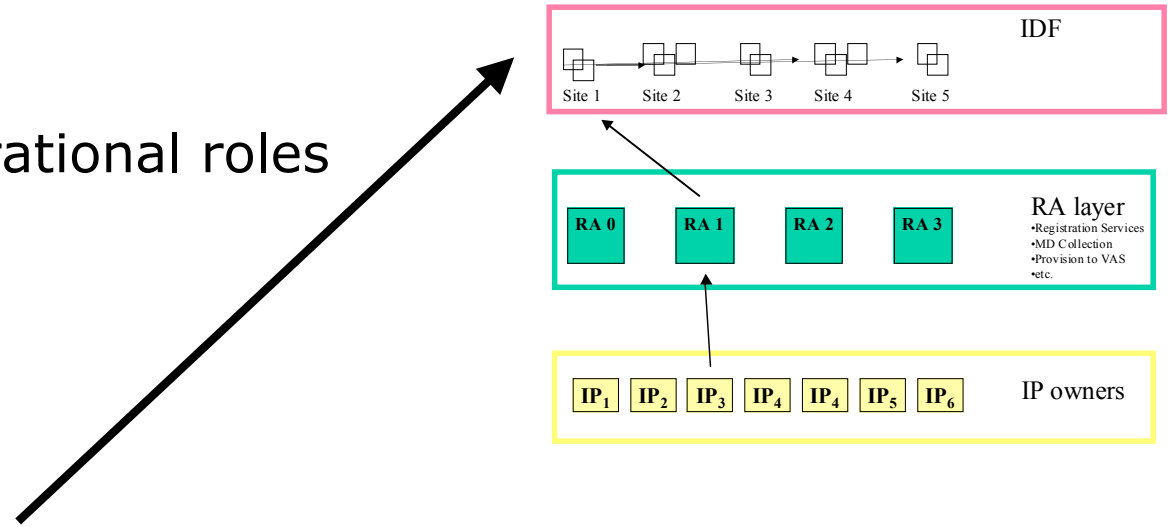


Registration agency:

- agreements with IP owners*
- registration with DOI System (IDF terms)
- metadata collection /added value*
- provision of, or to, Value Added Services by agreement*, etc

* *specific to each RA*

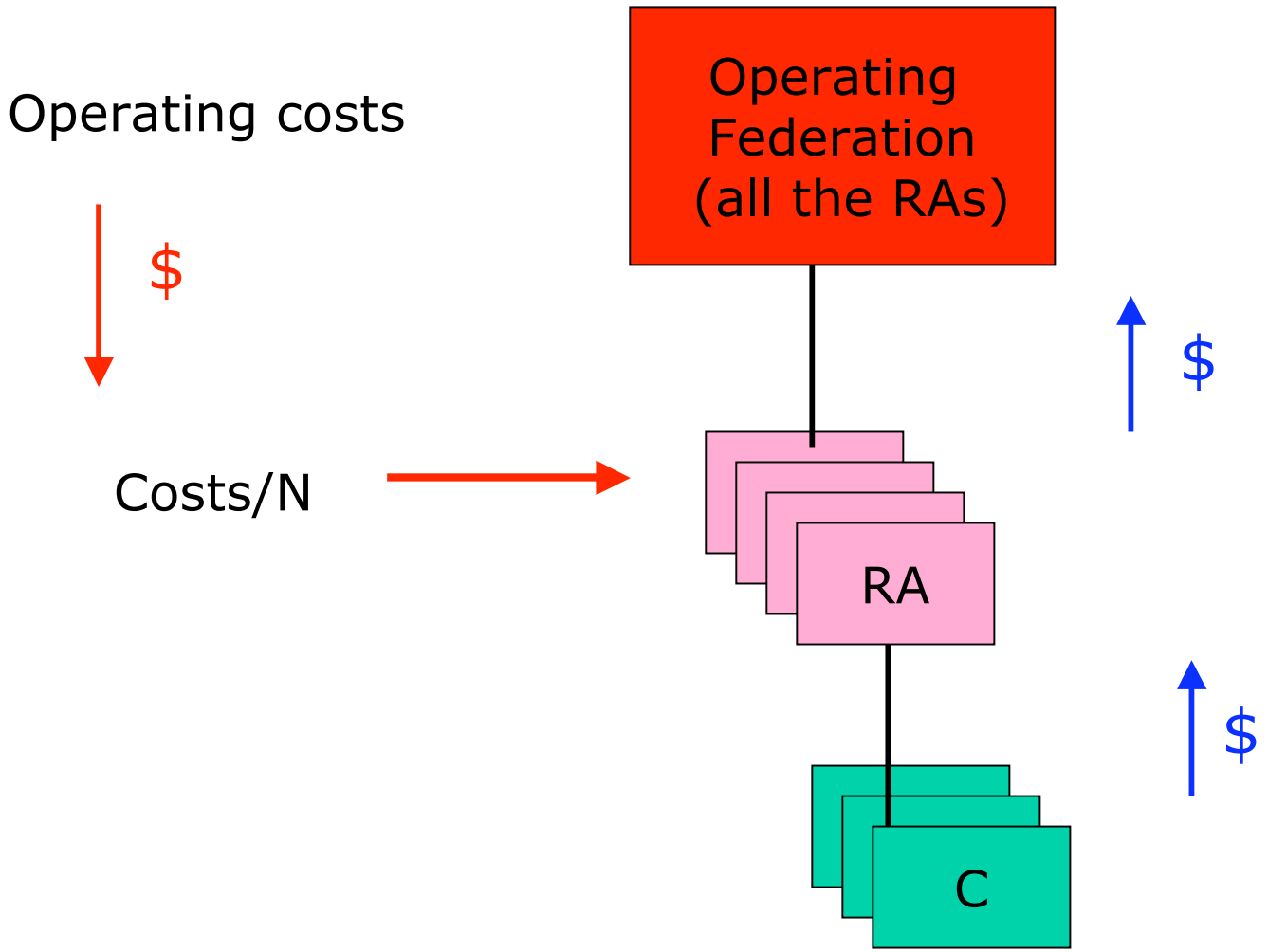
DOI System operational roles



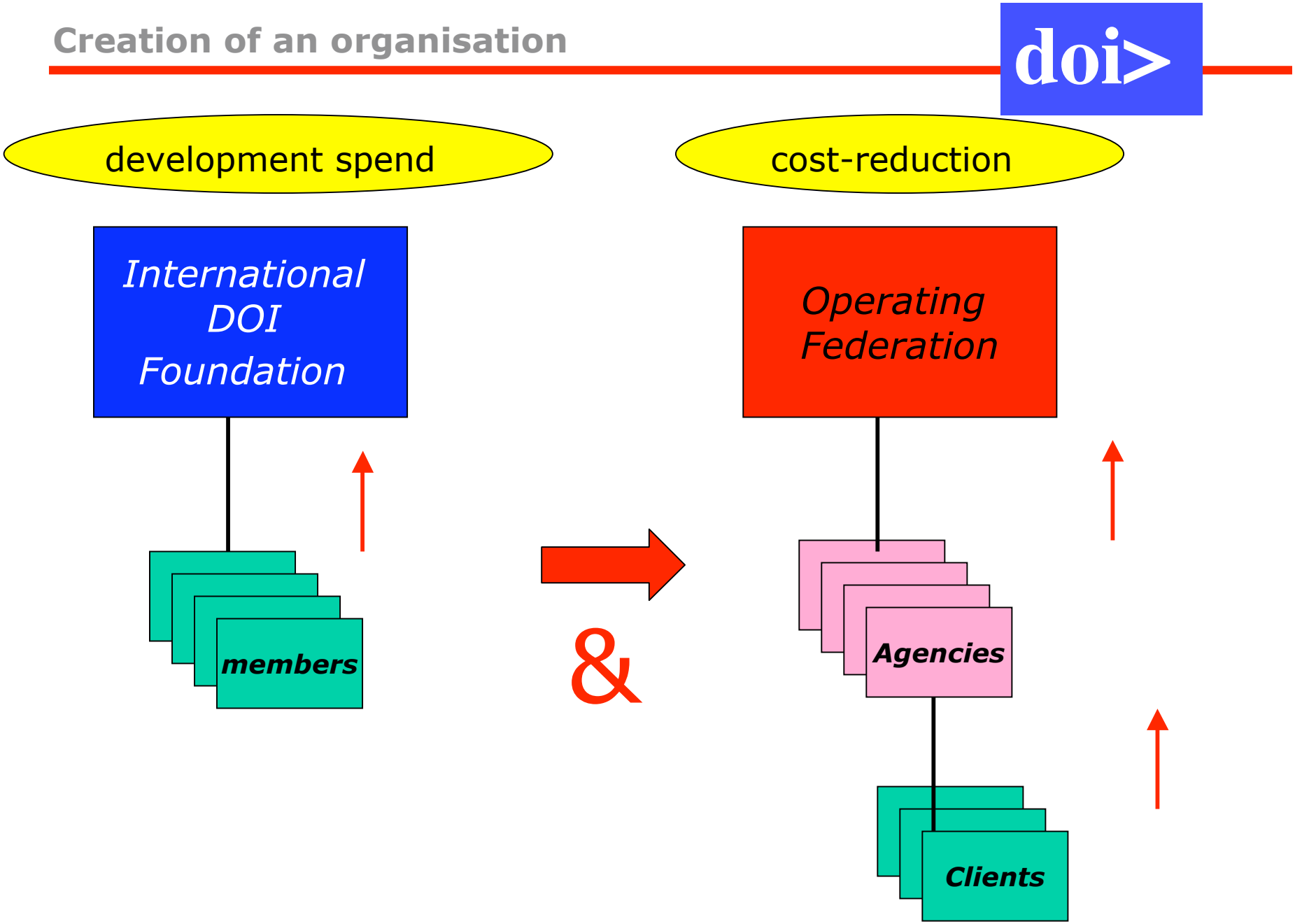
IDF: minimal common agreements

- DOI resolution service
- service integrity
- Data Type Registries
- Policies e.g. archiving, testing, etc

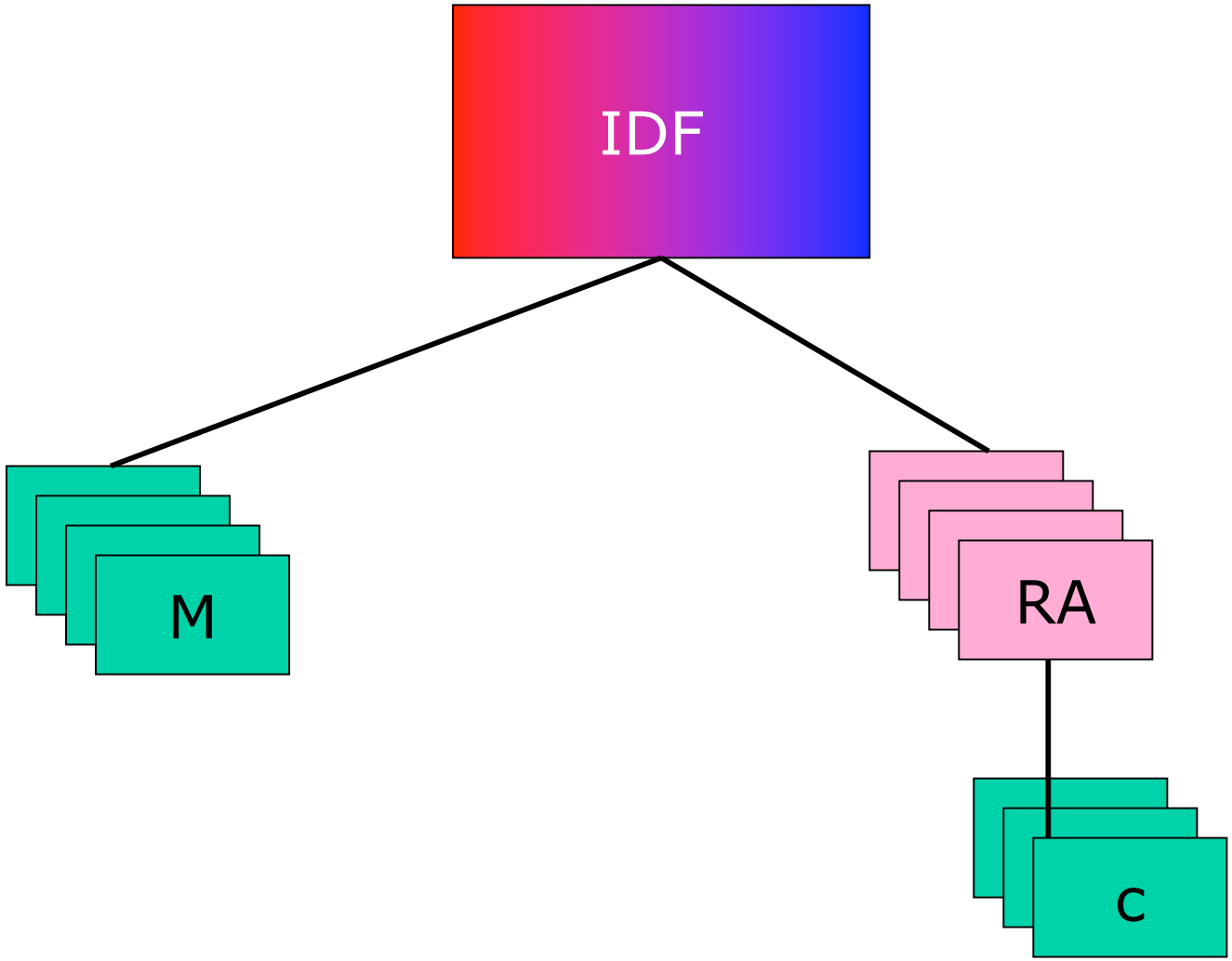
Aim: Registration Agencies run the system



Creation of an organisation



IDF organisation



Issues (1): Functional v national

- Functional (application) agencies
  - applications across borders (e.g. CrossRef)

- National/regional agencies
  - local services
  - language documentation
  - integration with other agencies e.g. ISBN ?
  - e.g. MEDRA (EC)

Issues (2): Value of metadata

- DOI kernel metadata is a small basic set
  - likely not to be of commercial value alone
- Resolution provides “known item”
  - DOI name look-up only
- Metadata is not held in DOI System
  - only a pointer to it
- Metadata promotion maximises value
  - like a catalogue

Issues (3): Make metadata available (syntax)

- On a web page; <meta> tag; etc
  - easy to do, unstructured
- XML (Extensible Markup Language)
  - logical syntax, wide support, needs more to guarantee interoperability
- RDF (Resource Description Framework)
  - Syntax for interoperable semantics; standard still evolving Questions as to acceptability
- Separate database
  - easy but raises issues of access
- Pointer entry (data type) in DOI name record
  - best guarantee of commonality

Issues (4): quality control

- e.g. "Missing" DOI names
  - duplication of prefix; DOI names not entered into directory; citing of early DOI names, etc
- Who will determine rules?
  - May be different guidelines for different areas
  - and who will police them?
- Some checks can be built into system
  - e.g. attempted duplication
- Who defines kernel, and mappings, ensures conformance?
- Rules for user communities?
- Who ensures quality control of content?
- Who is the authority for each metadata element?
- What are the business model implications?

Issues (5): Explaining

- "missionary work"
  - identifiers: precision about what is identified (ISO TC46?)
  - functional granularity; well-formed metadata, etc.
- what can we learn from other efforts?
  - e.g. ISBN
- Best explained by examples: applications

Issues (6): relation to existing services

- Many ISO entities have metadata records
  - ISBN, ISSN, etc - widely used
- May not be consistent with each other
- May not be consistent with indecs mapping
- May not be available for DOI name registration on ideal "do it once" basis
  - commercial considerations of those agencies
- Can metadata be shared?
- Collaboration between RAs

Issues (7): management of standards

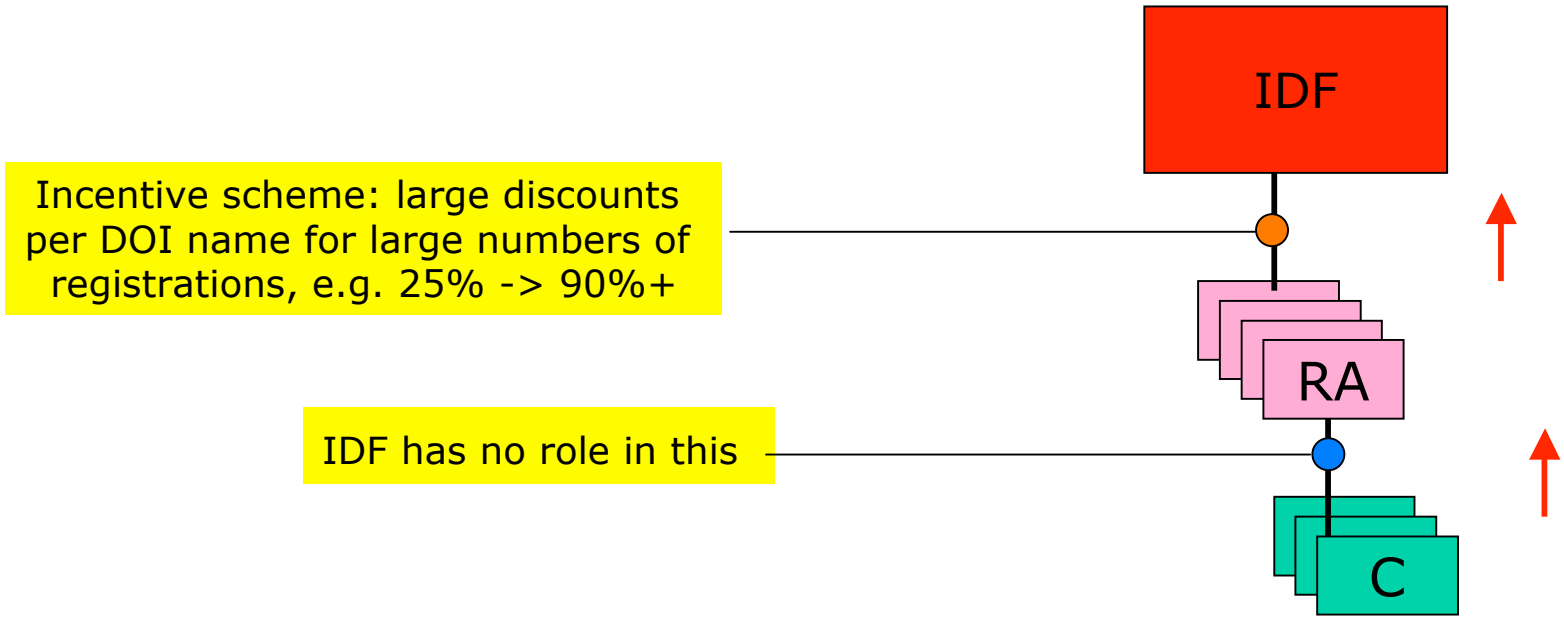
- DOI System and indecs based on open standards
- Who directs evolution?
  - governance structure, maintenance agency (ISO standard)
  - likely not to be of commercial value alone
- Who will invest the resources necessary to make improvements and prevent stagnation?
  - IDF set up as collaborative forum
  - Long term funding and sustainability via funding through use (like bar codes)

Issues (8): governance

- User communities want to control their own activities as far as possible
  - *Principle of Subsidiarity*
  - constitutional rules?

- There must be some common rules
  - *Principle of Interoperability*
  - IDF = operations and standards “council”?

- Focus on enabling current RAs to generate more DOI names
- New RAs in new areas
- Social infrastructure development (RA policies)
- Business model:



Implementation of strategy

- RAs focus on building applications in their existing sectors
  - viability of business models
  - lower costs per DOI name (for volume)

- IDF focus on better tools for RAs:
  - Resolution – e.g. Acrobat plug-in
  - Multiple resolution: DOI-AP framework
  - Semantic interoperability: Data Dictionary
  - and additional tools

DOI System applications

- The main use of the DOI System is not simply to register an identifier
- It is to make use of the identifier in a SERVICE offered to users
- E.g. CrossRef provides bibliographic citation pre-and post-production look-up service across hundreds of publishers
- It uses DOI names as one part of its service
- It has become a de-facto requirement for academic publishing

DOI names work with existing identifier schemes

- *General case*
- ISO standardisation of DOI System
  - "A DOI name is not intended as a replacement for other identifier schemes, but when used with them may enhance the identification functionality provided by those systems with additional functionality…"
- Incorporate the other identifier into the DOI name syntax

  and/or
- Record the other identifier in the DOI name metadata.

- Each scheme retains its autonomy but works together
- ISBN and ISSN have already agreed options

DOI names work with existing identifier systems

- *Representations*
- URL: http://dx.doi.org/10.978.86123/45678
- MPEG-21: DII (Digital Item Identifier)
- URI schemes:
  - Info URI
  - URI
- URN

- *Examples:*

- Scientific data
  - TIB (Registration Agency) is an example
- Biological nomenclature
  - disambiguation and extension of the current taxonomy models: Names-4-Life: (IDF member)
- Clinical Trials
  - identifying specific trials and sub-sets of items
  - UK project currently using DOI names on pilot basis

DOI Registration Agency: TIB

- German National Library of Science and Technology
  - TIB = Technische Informationsbibliothek, Hannover
  - Contact: Dr Jan Brase: Jan.Brase@tib.uni-hannover.de

- *Current:* DOI registration agency for primary data (and other non-commercial scientific information) with other local institutions as its customers.

- *Plan:* In 2008/2009 TIB plans to transit to a new worldwide agency, funded by a federation of local information institutes and libraries.
  - model: in 1999 the publishers funded the independent DOI agency CrossRef
  - Enable easy and cheap access to the DOI system for non-commercial information institutes and libraries worldwide (as Crossef did for STM publishers)

DOI Registration Agency: TIB

- Customers not of TIB but independent partnership in a new organisation. Integration of the DOI registration and DOI system into the *local infrastructure*.

- Current partners:
  - TIB, Germany
  - ETH Zürich Library, Switzerland
  - L'Institut de l'Information Scientifique et Technique (INIST), France
  - Open for other institutions (MPS?)

- ICSTI project: Technical Activities Coordinating Committee (TACC) to start a project on "Integrating numeric data access into textual information and assigning identifiers (e.g. DOIs) to datasets"
  - Partners so far: TIB, CISTI, The British Library, DOE-OSTI, ProQuest, INIST, CODATA

Several DOI Name Tools have been developed, from a variety of sources

http://www.doi.org/tools.html

- Such as plug-ins,
  e.g. Adobe Acrobat plug-in
- At different stages of development or use

http://www.doi.org/tools.html

Google

Tools Available For Use	
Tool	**Description**
ONIX DOI Name Registration Formats	EDItEUR has developed a collection of DOI name registration formats allowing publishers and others to communicate the metadata required by an RA in order to record the assignment of a DOI name. The formats allow a DOI name to be assigned at "work" or "manifestation" level, to whole monographs, chapters or parts of monographs, serials, serial issues, or serial contributions.
MPEG-21 Rights Data Dictionary	The MPEG-21 Rights Data Dictionary (ISO/IEC 21000-6), often abbreviated "RDD", is maintained by the IDF. Any terms which are consistent with the RDD will be consistent with the DOI Data Model. Terms in the RDD are publicly available. It has two parts: terms that are defined in the base standard, and additional terms that are registered by anyone who finds it useful.
Connotea	Connotea, a free online reference management and social bookmarking service (see article at doi:10.1045/april2005-lund), recognises and stores DOI names, enabling bookmarking a DOI name directly. Connotea will convert the DOI name to a web link and retrieve the citation information for that DOI name from CrossRef. The Connotea bookmarklet also allows users to highlight and bookmark a DOI name from within an article web page.
CrossRef OpenURL Resolver	CrossRef's OpenURL resolver, functioning as a DOI name look-up resource for the public, accepts URLs sent to http://www.crossref.org/ structured according to the 0.1 or 1.0 NISO specification (and some common deviations) and offers users another way to be directed to publications identified by DOI names.
DOI® Resolver Dashboard Widget	**Dashboard Widget for Mac OS X Dashboard.** "Say you're reading a journal article and you want to read another paper that the article references. If the referenced paper's DOI is listed in its citation, simply type/paste it into the DOI Resolver widget, press Enter, and your default browser will take you to the web page currently associated with the DOI." (Independently Developed)
Google Toolbar Button	**Google™ ToolBar button that resolves a DOI name via the Proxy.** This tool, developed by Zac Hanley (http://www.ortholog.com) works with the latest version of Google Toolbar for Internet Explorer. After installation, selecting and right-clicking a DOI name in a page will show "Fetch Digital Object by DOI" on the context menu. Clicking there will direct you (via the Proxy) to the authoritative document for that DOI name. (Get more information.) Add to Toolbar
DOI® Button	**JavaScript™ that adds functionality to your Web Browser.** Add a link (or icon) to your browser's tool bar which opens up a dialog box in which you can paste or type a DOI name and then resolve it, or you can highlight a DOI name that appears on a web page, and resolve it without retyping it.
DOI System API	**The DOI System API is an application programming interface** that will help developers build applications for accessing and managing DOI names, Application Profiles, and Services. There are two versions of the DOI System API: a Java implementation and its accompanying Javadoc, and a C implementation. **[This software is currently under trial by IDF members only. Members please send a note to contact@doi.org if you are interested in using the API.]**
HDL Plug-in (Ver. 1.5) for Adobe Acrobat® and Acrobat Reader®	**The HDL Plug-in (Ver. 1.5) for Adobe Acrobat and Acrobat Reader is an extension to Acrobat and Acrobat Reader** that looks for an embedded handle (or DOI name) identifying a PDF file as that file is being opened, and, if it finds one, resolves it and uses the resulting information to customize icons, or pop-up windows, to offer services specific to the document being opened, e.g., the availability of a new version of the document or a way to negotiate rights for the document. The publisher data that enables this behavior is currently available for only a few test files but work is underway to add this additional information to existing DOI names. Please send comments and requests to the Handle System Administrator at hdladmin@cnri.reston.va.us.
CNRI Handle System	This **extension to Firefox** embeds a handle client in the browser, eliminating the need to rely on a Proxy Server System to redirect Firefox to a

DIGITAL OBJECT IDENTIFIER (DOI®) SYSTEM

Article in: Encyclopedia of Library and Information Sciences
(forthcoming) third edition (Taylor & Francis)

http://www.doi.org/overview/070710-Overview.pdf

DOI SYSTEM

Persistent Identifier eScience Seminar
Munich, March 27/28

International DOI Foundation