

Tools for NextGen Scholarly Communication:

Work underway at the University of Minho

29/Jan/2007

Ana Alice Baptista

<http://www.dsi.uminho.pt/~analice>



Contents



- Scope:
 - UMinho and DSI
 - SDUM and RepositóriUM
 - Odisseia
- CRiB
- Tea for Two
- Other Projects
- Future work

Contents



- Scope:

- UMinho and DSI
- SDUM and RepositóriUM
- Odisseia
- CRiB
- Tea for Two
- Other Projects
- Future work





The University of Minho

<http://www.uminho.pt>



- Since 1974
- 2 campi
- 11 schools/ institutes (faculties)
- 30 research centres
- 13500 undergrad
- 1500 grad (MSc + doctorate)
- 1116 FTE



Information Systems Department

<http://www.dsi.uminho.pt>



- School of Engineering
- One of the two computer departments: **technology in context**
- Main interface areas: management and information science
- 2 graduation programs, 3 MSc programs and 1 doctorate program
- 37 teachers and 12 hired researchers
- 114 MSc students - 43 doing thesis
- 42 PhD students





SDUM

<http://www.sdum.uminho.pt>

- Documentation Services
- 4 main libraries
- + 21 other
- Several digital services (catalog, search portal, digital publications, repositóriUM,...)

RepositóriUM

<http://repositorium.sdum.uminho.pt>



- Developed in the scope of the e-UM Virtual Campus project;
- Implemented in DSpace
- Start date: April 2003
- Open to the public: November 2003
- Provided a framework for R&D activities:
 - 2 + 4 add-ons for DSpace (2 included in version 1.4)





RepositóriUM

<http://repositorium.sdum.uminho.pt>

Consultas e downloads - Total

Consultas	Downloads
-----------	-----------

470184	531118
--------	--------

Since
01/Jan/06

Médias por documento

Docs	Consultas	Downloads
------	-----------	-----------

5151	91.3	103.1
------	------	-------

RepositóriUM

<http://repositorium.sdum.uminho.pt>

Por Origem

Origem	Consultas	Downloads
 Portugal	194413	250505
 Brazil	103878	133603
 Universidade do Minho	70553	32484
 United States	50456	18026
 India	3165	8954
 N/A	6086	8802
 China	3156	8208
 Spain	4026	5682

Since
01/Jan/06



Odisseia

<http://odisseia.dsi.uminho.pt>

- Research group on Scholarly Communication
- Scientific sub-area: Information Society
- Started in 2004 with some projects, including the “DSpace Dev@University of Minho”



Odisseia

<http://odisseia.dsi.uminho.pt>

- Currently:
 - CRiB - digital preservation
 - Tea for Two - informal communication + social networks
 - plus some other satellite projects

Contents



- Scope:
 - UMinho and DSI
 - SDUM and RepositóriUM
 - Odisseia
- **CRiB**
- Tea for Two
- Other Projects
- Future work



CRiB

- Basically a PhD and satellite projects
- Start date: Jan/2005, End date: Dec/2007
- Problem: Repositories are
 - excellent at archiving and disseminating materials
 - poor at preserving those materials in the long run

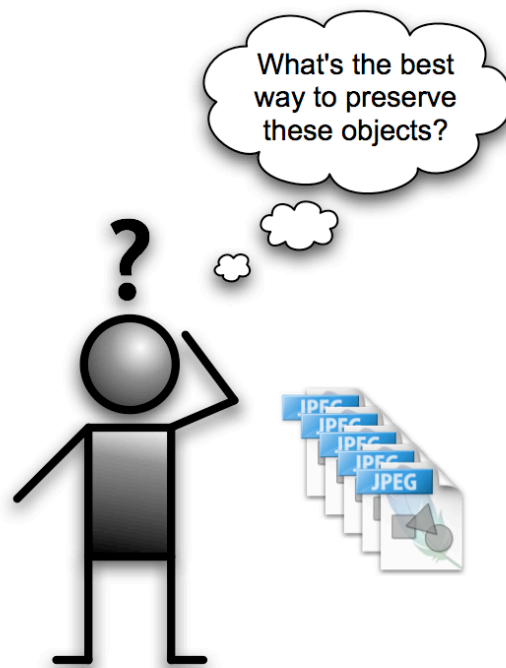


CRiB

- Service Oriented Architecture (SOA)
- Preservation approach: migration
- Recommendation service - Recommends an optimal migration strategy taking into account
 - Requirements of each client institution
 - Behavior/quality of each migration service
- Migration services
 - Service composition
- Evaluates the outcome of each migration
 - Performance, data loss, format characteristics
 - Produces an evaluation report (authenticity)

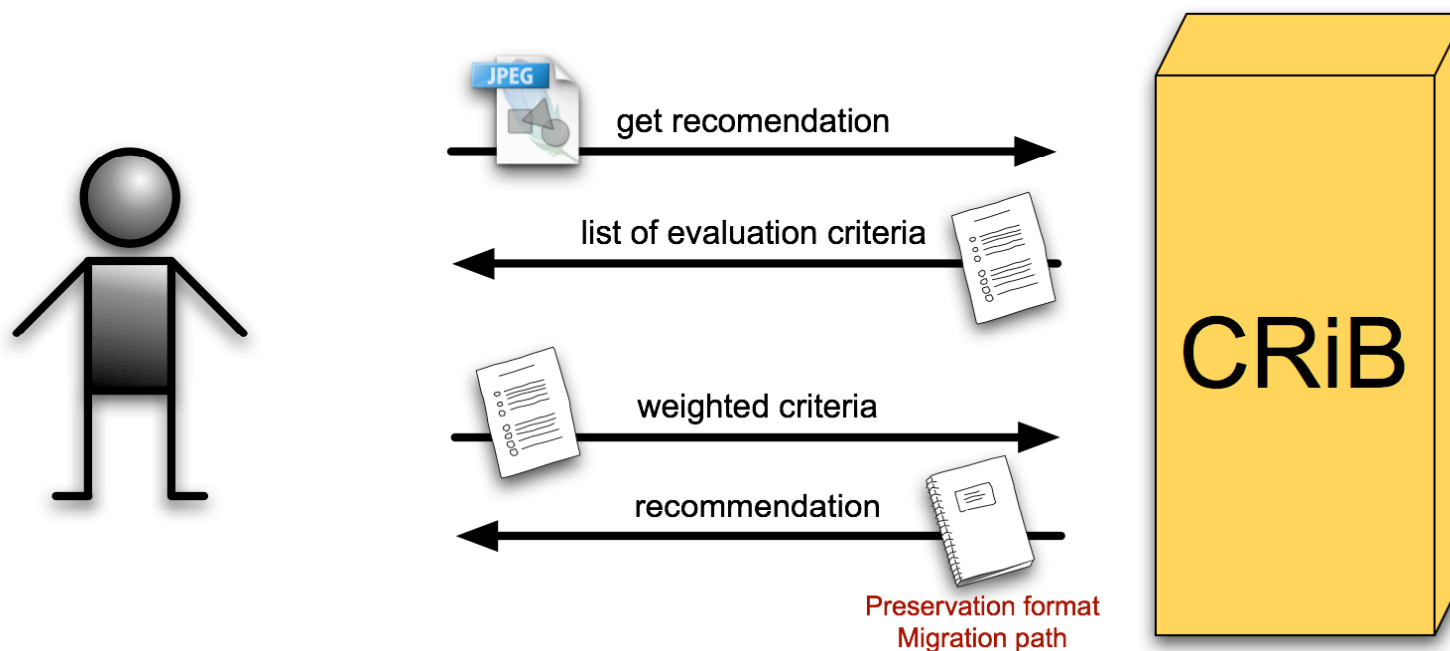
CRiB - Scenario

- A collection of digital objects of a certain format
 - e.g. JPEG files collected from a digital camera
 - e.g. A collection of text documents



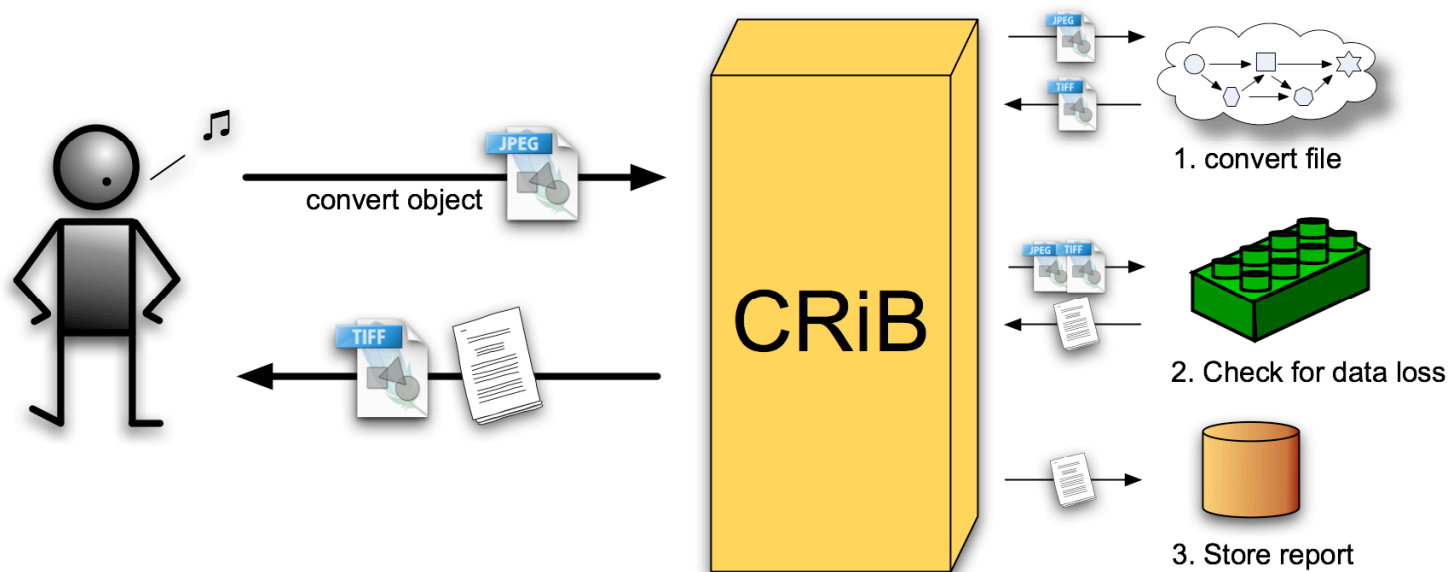
CRiB - Scenario

- Using the recommendation service
 - Preservation format (i.e. The target format)
 - Migration service (or combination of services)



CRiB - Scenario

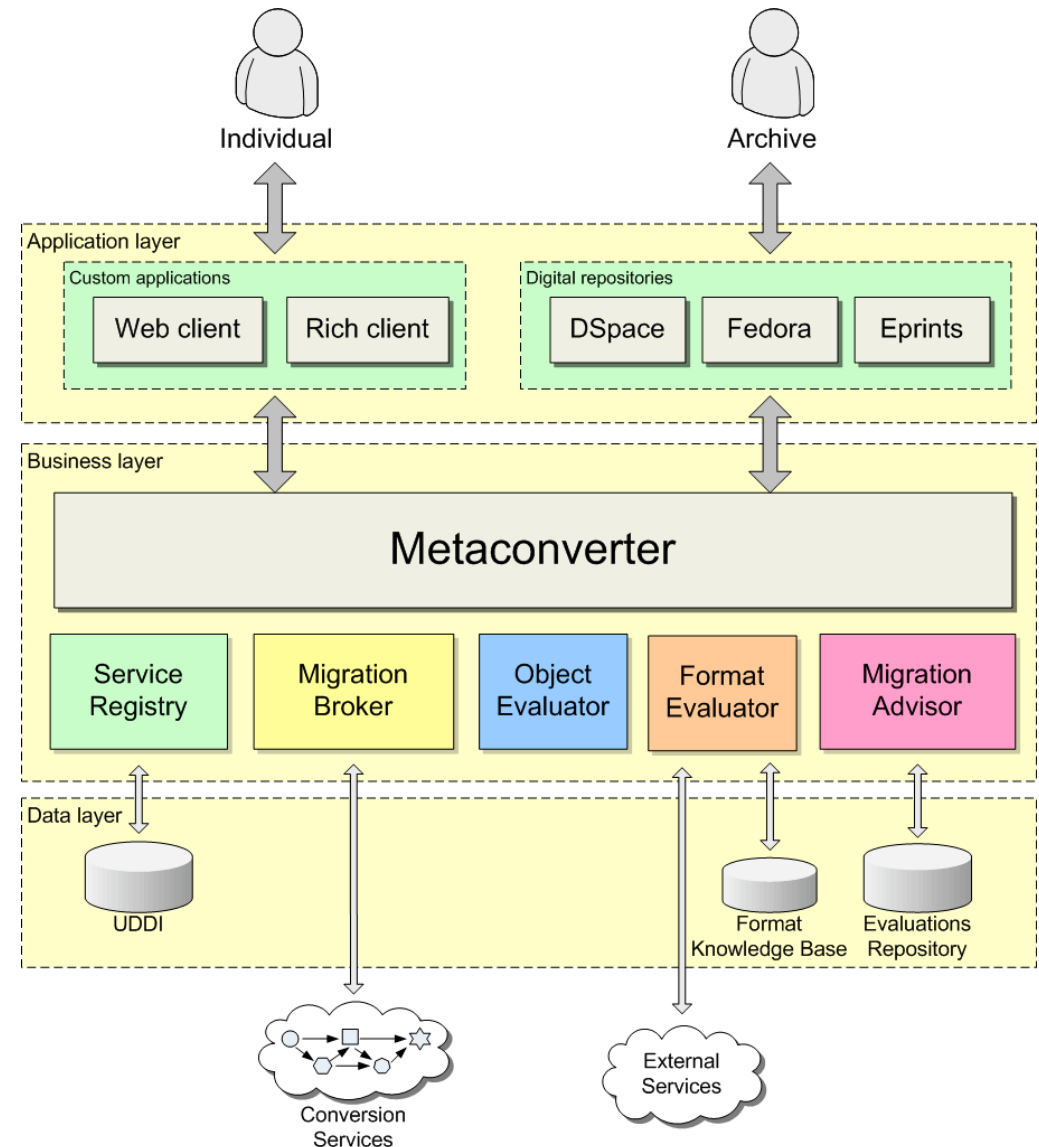
- Using the conversion services
 - Check for data loss and generate a migration report
 - Store the report
 - Return the converted file and the report back to the





CRiB

General Architecture





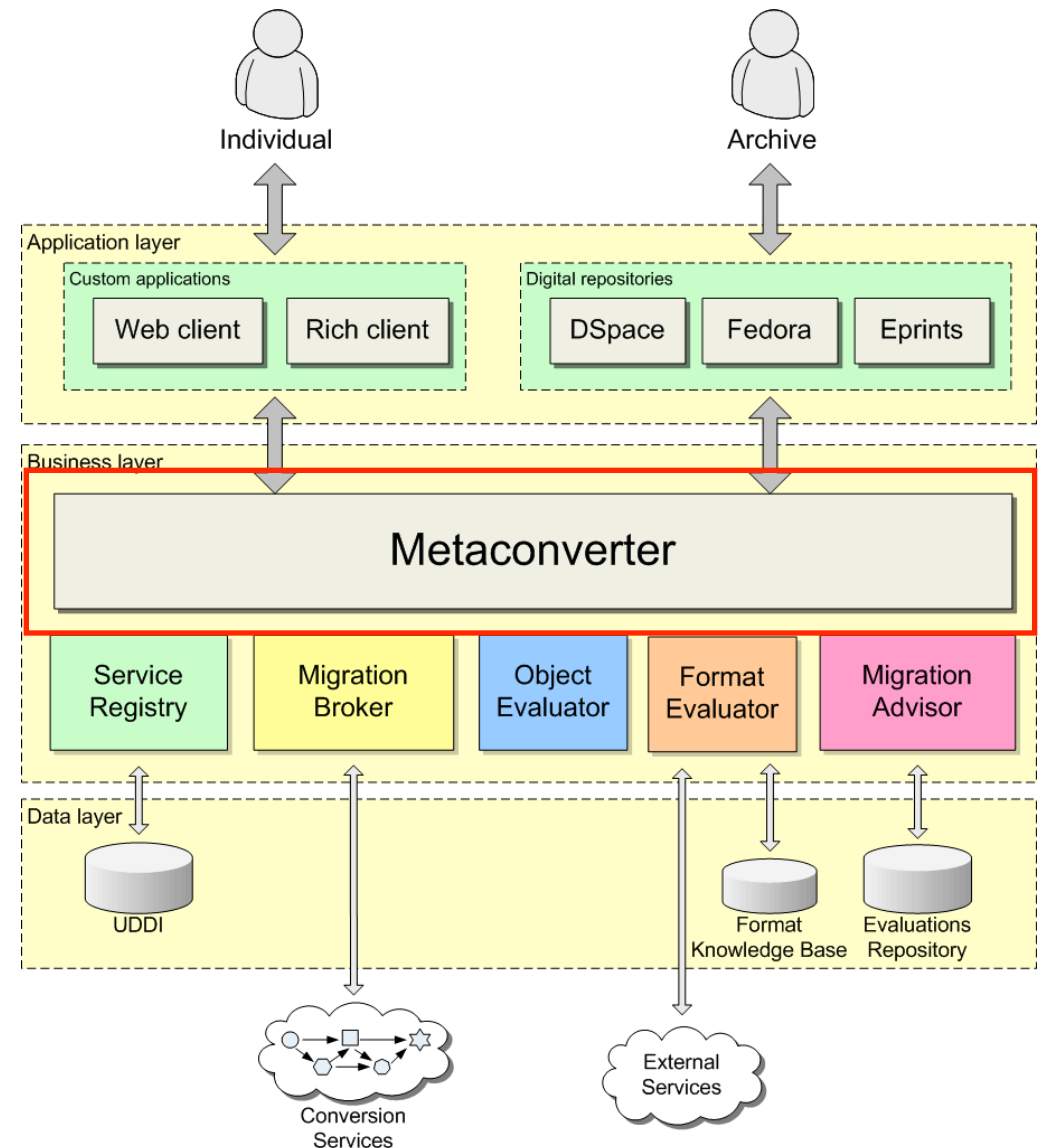
CRiB

General Architecture

Handles all communication between the client and the CRiB system

Its a web service

Orchestrates the communication within the system





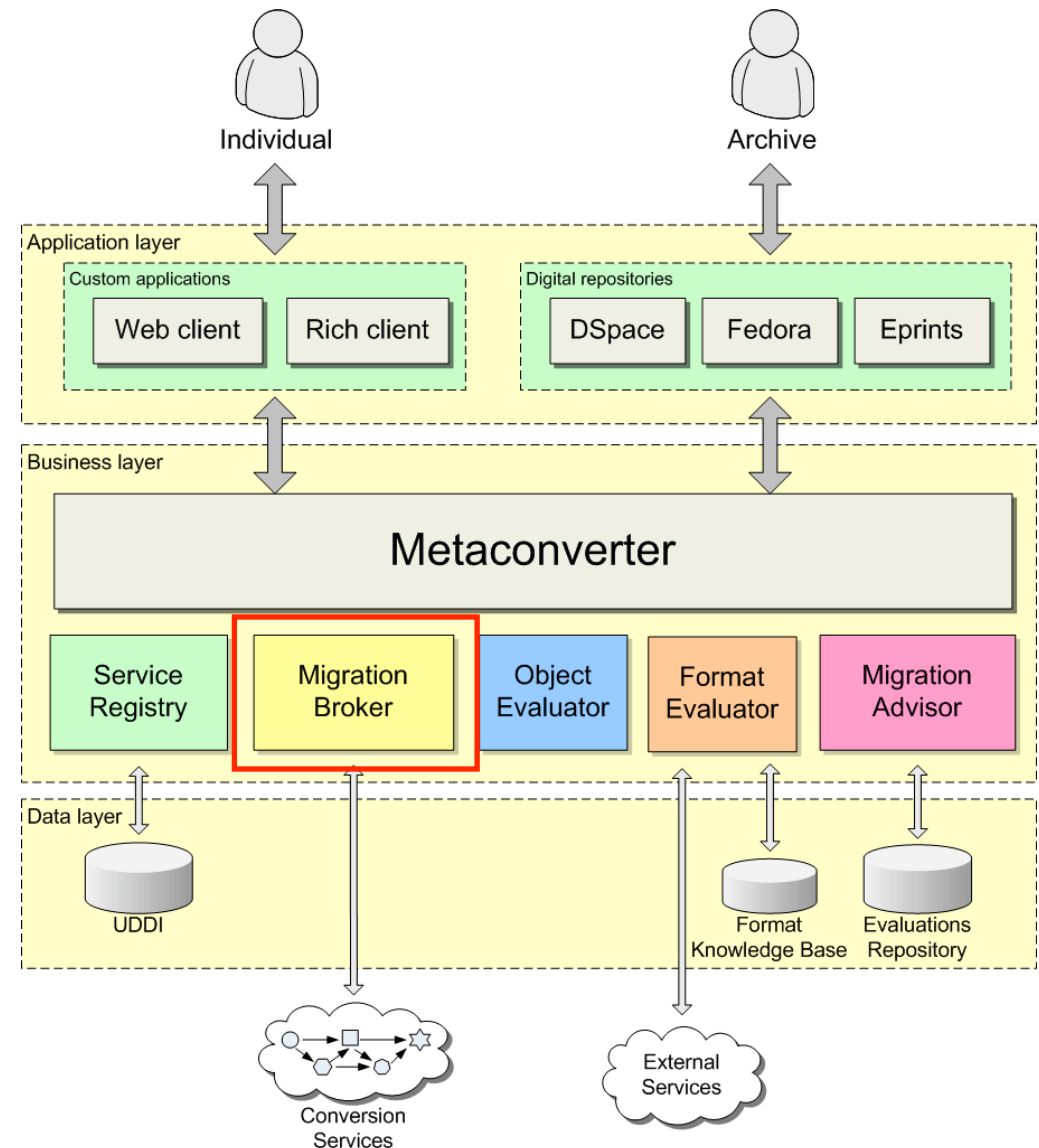
CRiB

General Architecture

Carries out format migrations

Invokes all the necessary conversion services

Measures the performance of the conversion process





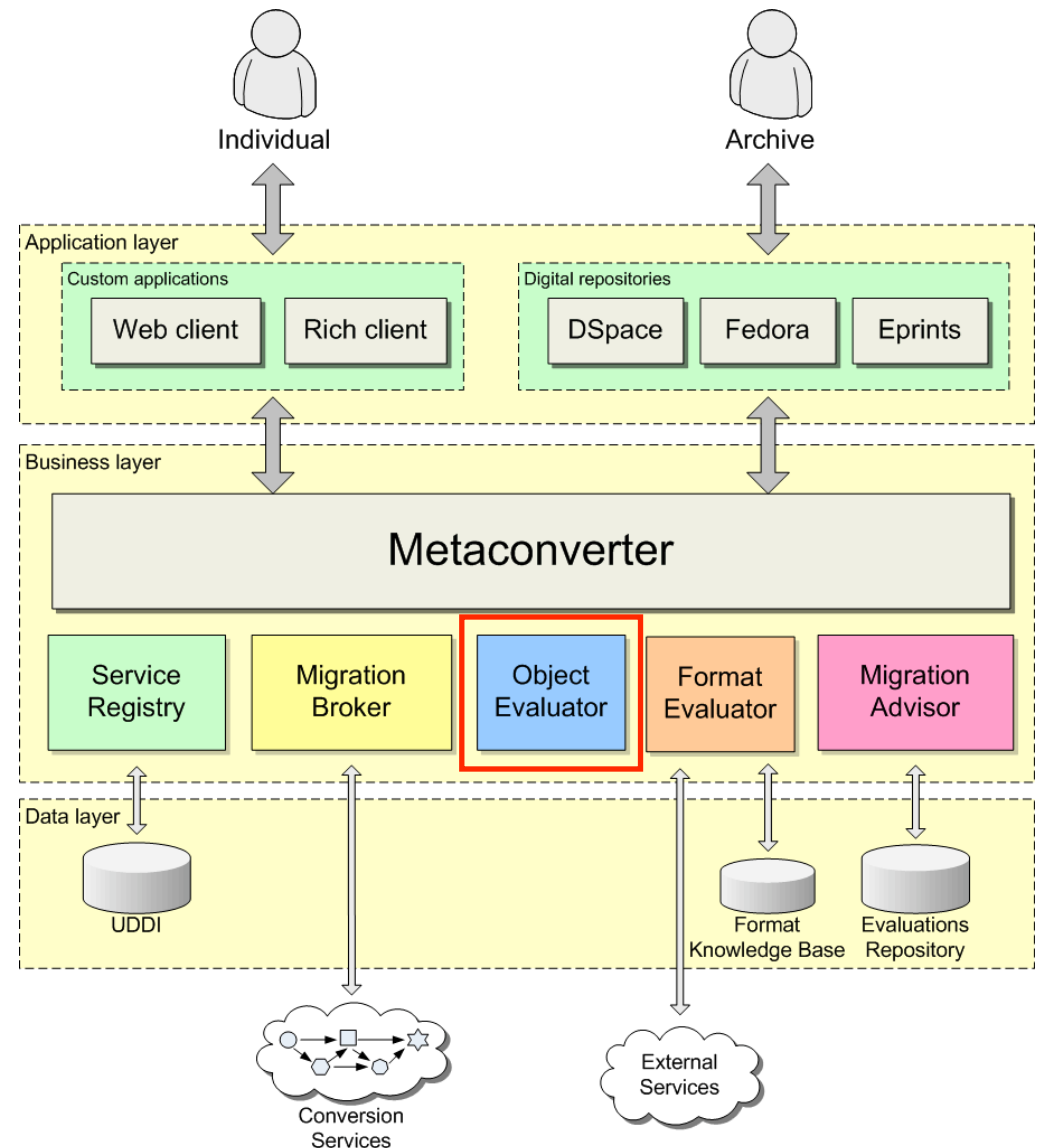
CRiB

General Architecture

Determines the amount of data loss involved in the migration

Detects the similarity of each significant property of the digital object

Produces evaluation reports in PREMIS format (eventOutcomeDetail)





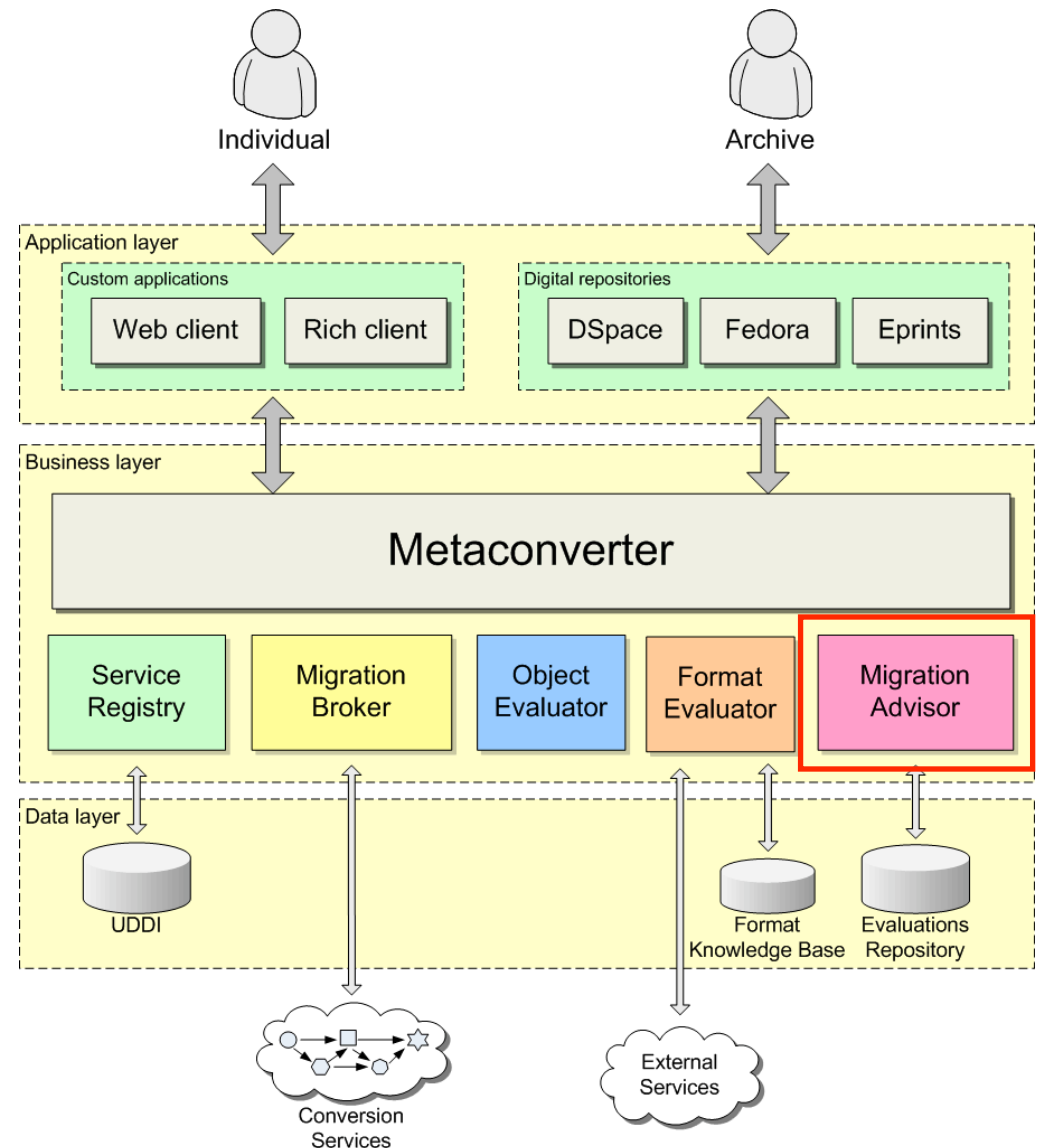
CRiB

General Architecture

Generates recommendations of optimal migration choices

Uses information provided by the client to determine the best available option

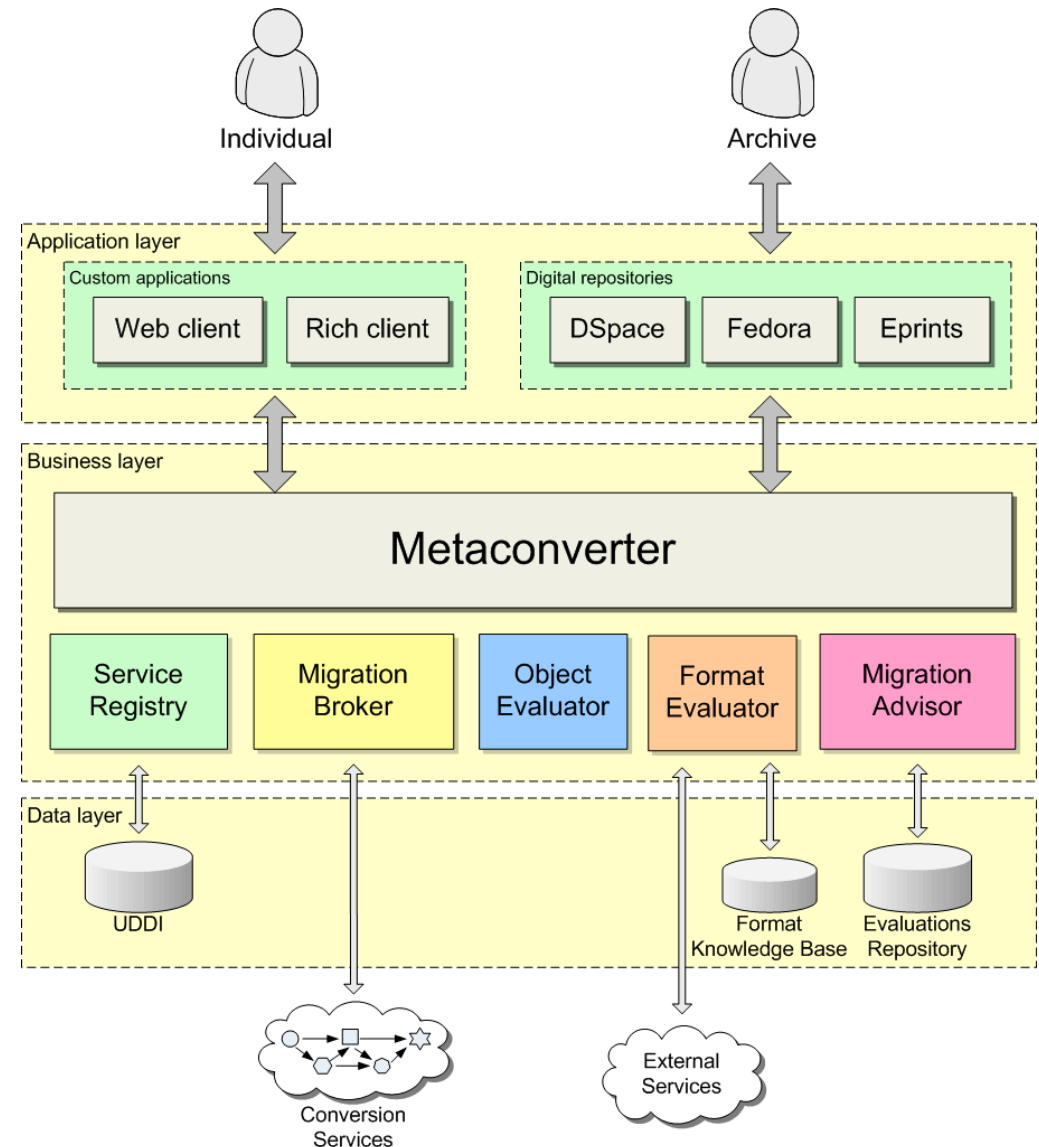
Confronts those requirements with the accumulated knowledge about the behavior of each conversion service





CRiB

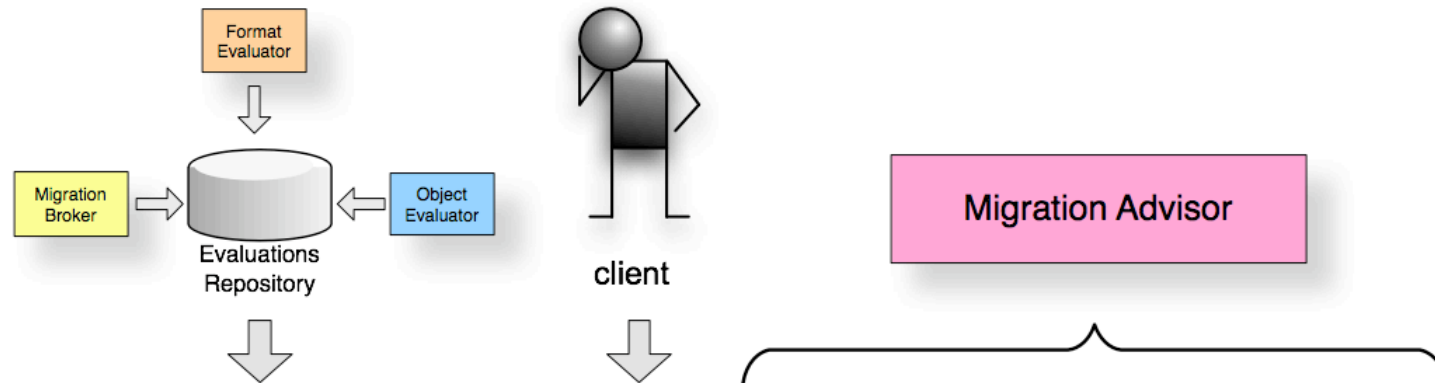
General Architecture





CRiB

Recommendation Engine



Evaluation criteria		WS1	WS2	Weights	WS1	WS2	Weights	WS1	WS2
Process	Availability	95%	90%	40%	5	5	0.4	2	2
	Stability	80%	90%	30%	5	5	0.3	1.5	1.5
	Scalability	1	1.5	0%	1	5	0	0	0
	Throughput	1 Kb/s	2 Kb/s	20%	1	5	0.2	0.2	1
	Cost	0.01 €	0.05 €	10%	5	1	0.1	0.5	0.1
Object	Number of chars	70%	95%	40%	4	5	0.4	1.6	2
	Ordering of chars	60%	85%	40%	4	5	0.4	1.6	2
	Page layout	80%	20%	20%	5	2	0.2	1	0.4
Format	Market share	30%	70%	30%	2	4	0.3	0.6	1.2
	Support level	5	4	30%	5	4	0.3	1.5	1.2
	Open standard	yes	no	40%	5	1	0.4	2	0.4
								12.5	11.8

(1)
Average evaluations per service

(2)
User assigned weights of importance

(3)
Normalised evaluations [1-5]

(4)
Ranking process

(5)
Overall rating

CRiB - Who



Miguel Ferreira
mferreira@dsi.uminho.pt



Ana A. Baptista
analice@dsi.uminho.pt

Other people in a non-regular basis.

+

PhD co-supervisor

Contents



- Scope:
 - UMinho and DSI
 - SDUM and RepositóriUM
 - Odisseia
- CRiB
- [Tea for Two](#)
- Other Projects
- Future work

DSpace Dev@University of Minho

<http://dspace-dev.dsi.uminho.pt>

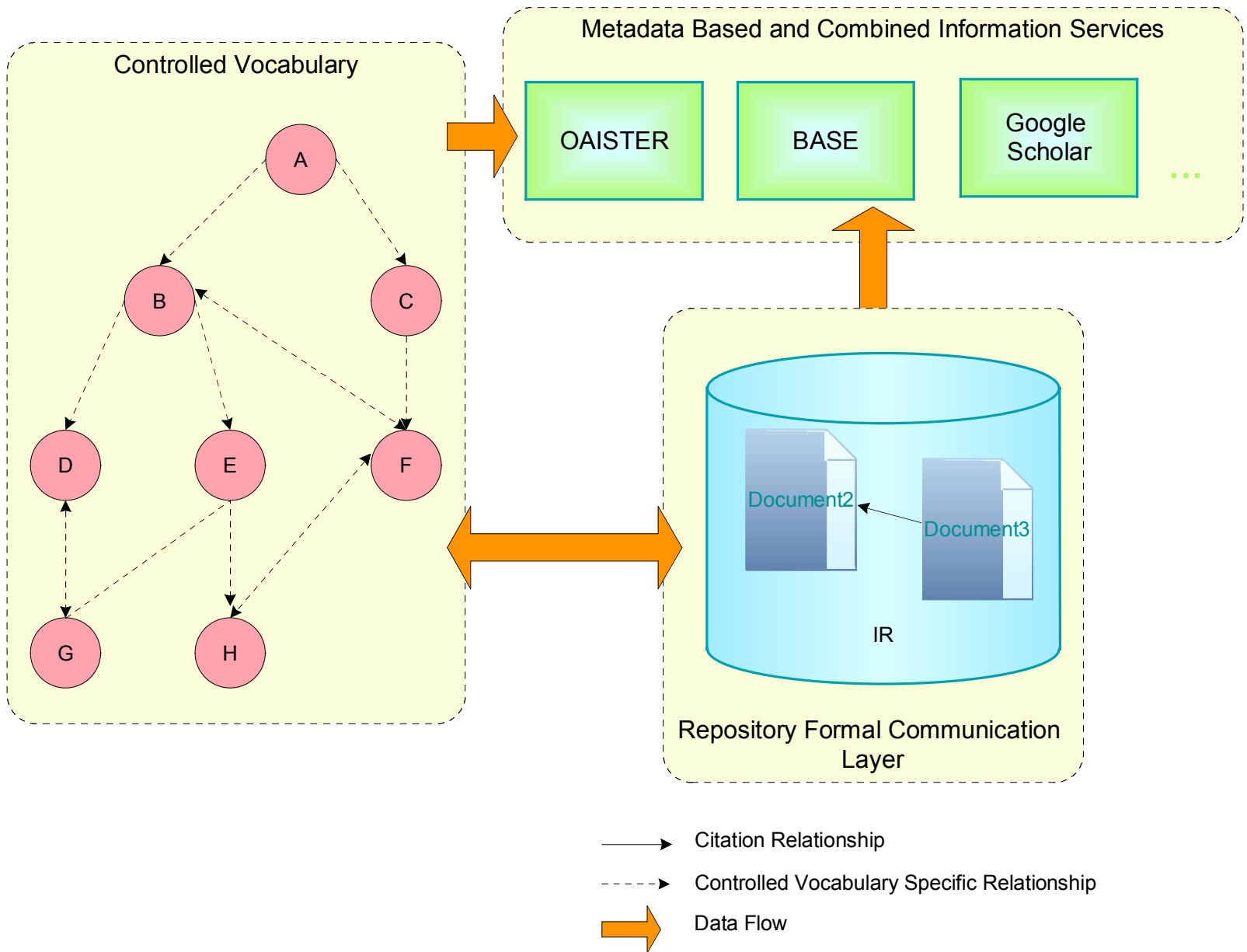


- Started and Finished in 2004
- Implementation of some “old” ideas
- Participants: Ana Alice Baptista and Miguel Ferreira
- Add-ons developed:
 - Controlled Vocabulary
 - Commenting
 - Web of Communication
 - Recommendation

Controlled Vocabularies add-on



- Goal: to facilitate and improve resource discovery by means of controlled vocabularies
- First Implementation: papadocs with ACM CCS
- Additional deliverables: Specific XML Schema; Official XML version of the ACM CCS
- Integrated in version 1.4 of DSpace
- Being upgraded in a joint project with FAO (Food and Agriculture Organization - United Nations)





Seleccione os conceitos a pesquisar

Filtro:

Filtrar árvore...

- ACMCCS98
 - General Literature
 - Hardware
 - Computer Systems Organization
 - Software
 - Data
 - Theory of Computation
 - Mathematics of Computing
 - Information Systems
 - GENERAL
 - MODELS AND PRINCIPLES
 - DATABASE MANAGEMENT
 - INFORMATION STORAGE AND RETRIEVAL
 - INFORMATION SYSTEMS APPLICATIONS

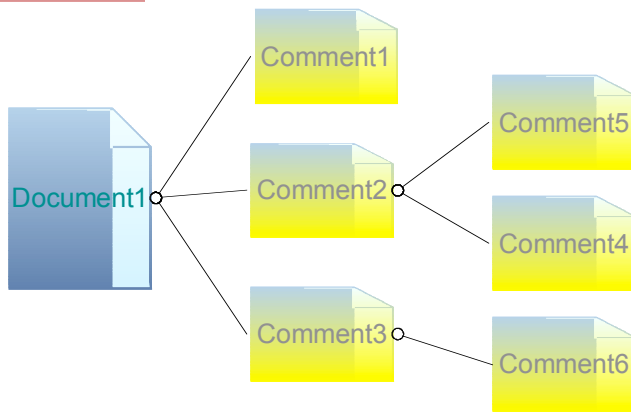


Commenting add-on

- Goal: to bring informal communication to DSpace in a way that it would be directly related to formal communication.
- First Implementation: papadocs
- Specific issue: graph (and not tree) based threads

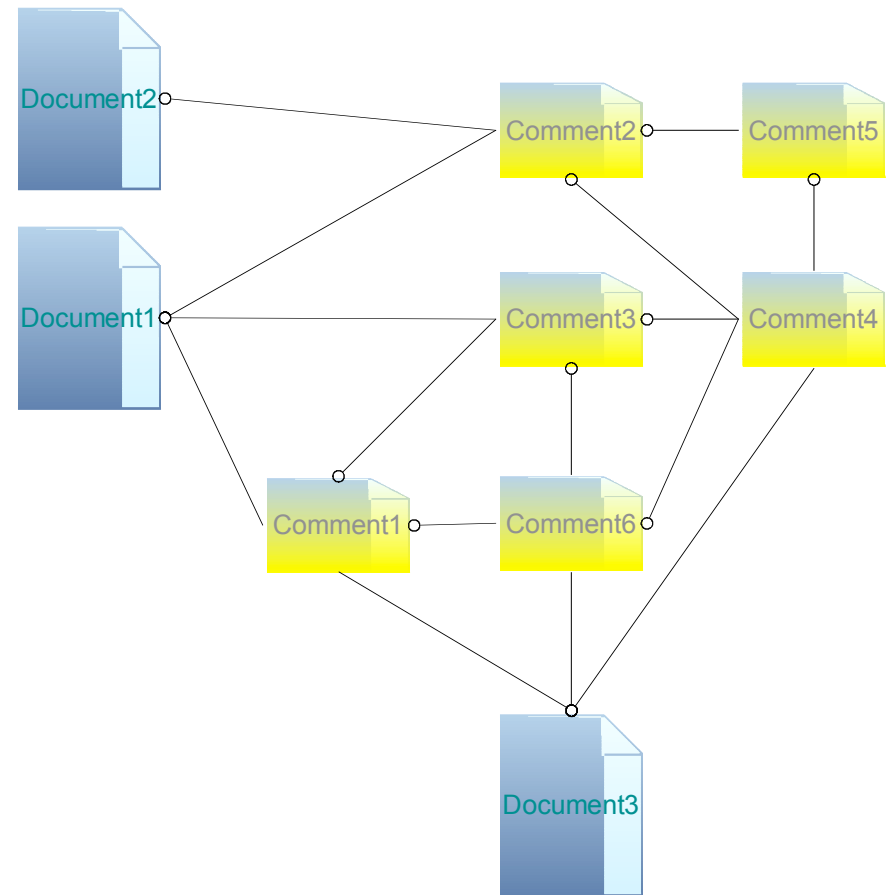


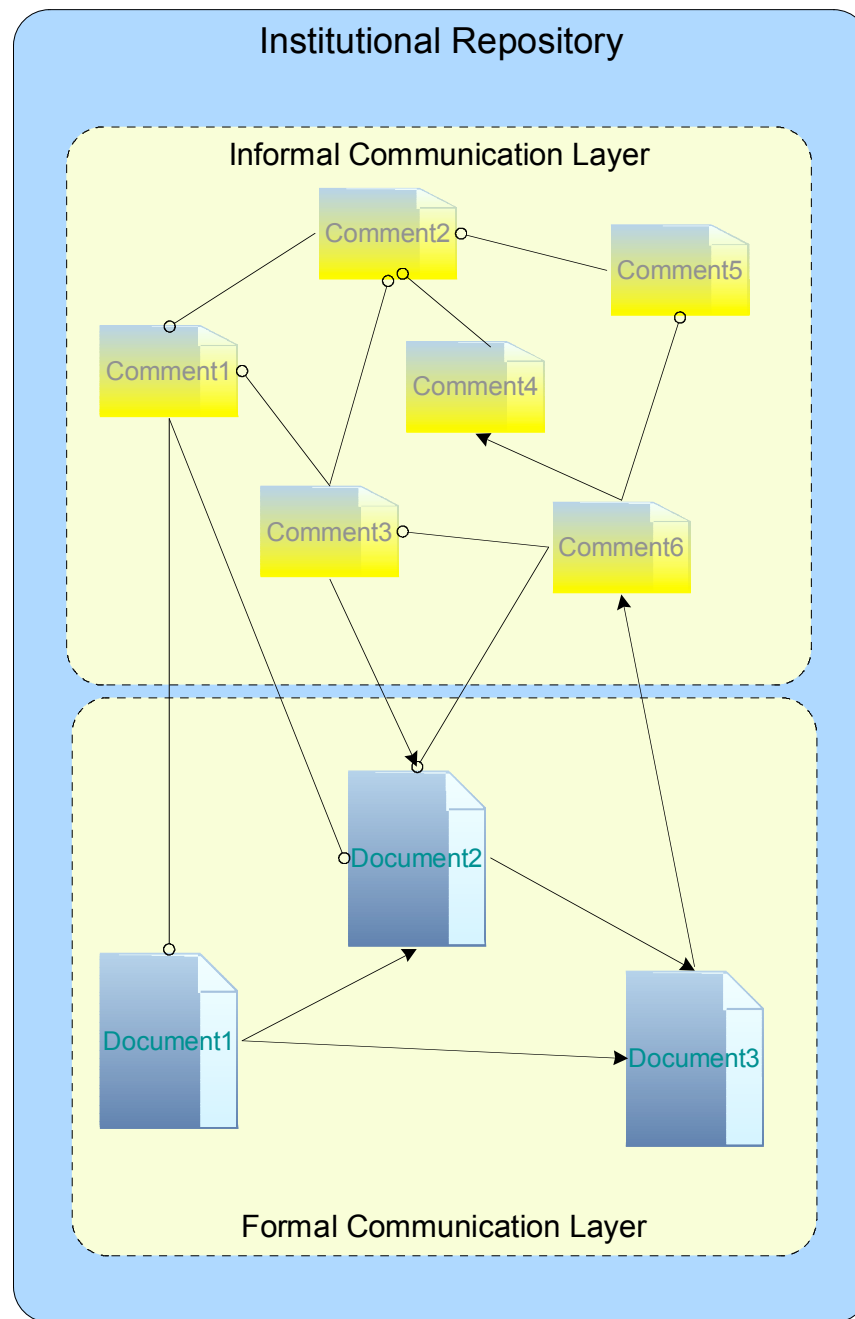
Tree-based comments' network



—○ Comment Relationship

Graph-based comments' network





- Citation Relationship
- Comment Relationship



- [Por Data](#)
- [Participantes](#)

Entrar:

- [Subscrever](#)
- [Meu Repositório](#)
utilizadores autorizados
- [Editar Conta](#)

- [Ajuda](#)
- [F.A.Q.](#)
- [Acerca do Repositório](#)

Ficheiros deste Item:

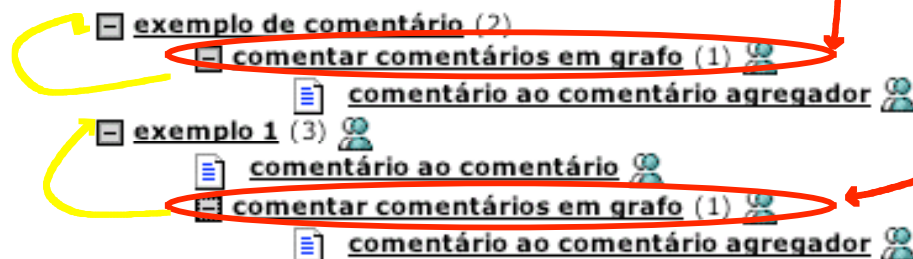
Ficheiro	Descrição	Tamanho	Formato	
mascote_papadocs.jpg		133Kb	JPEG	Ver/Abriu
mascote_papadocs-withoutsign.psd		8626Kb	Photoshop	Ver/Abriu
mascote_papadocs-withsign.psd		9076Kb	Photoshop	Ver/Abriu

[Mostrar registo em formato completo](#)

[\[Comentar\]](#) [\[Marcar\]](#) [\[Teia de Recursos\]](#)

Same Comment

Comentários:

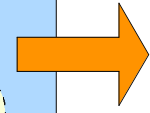
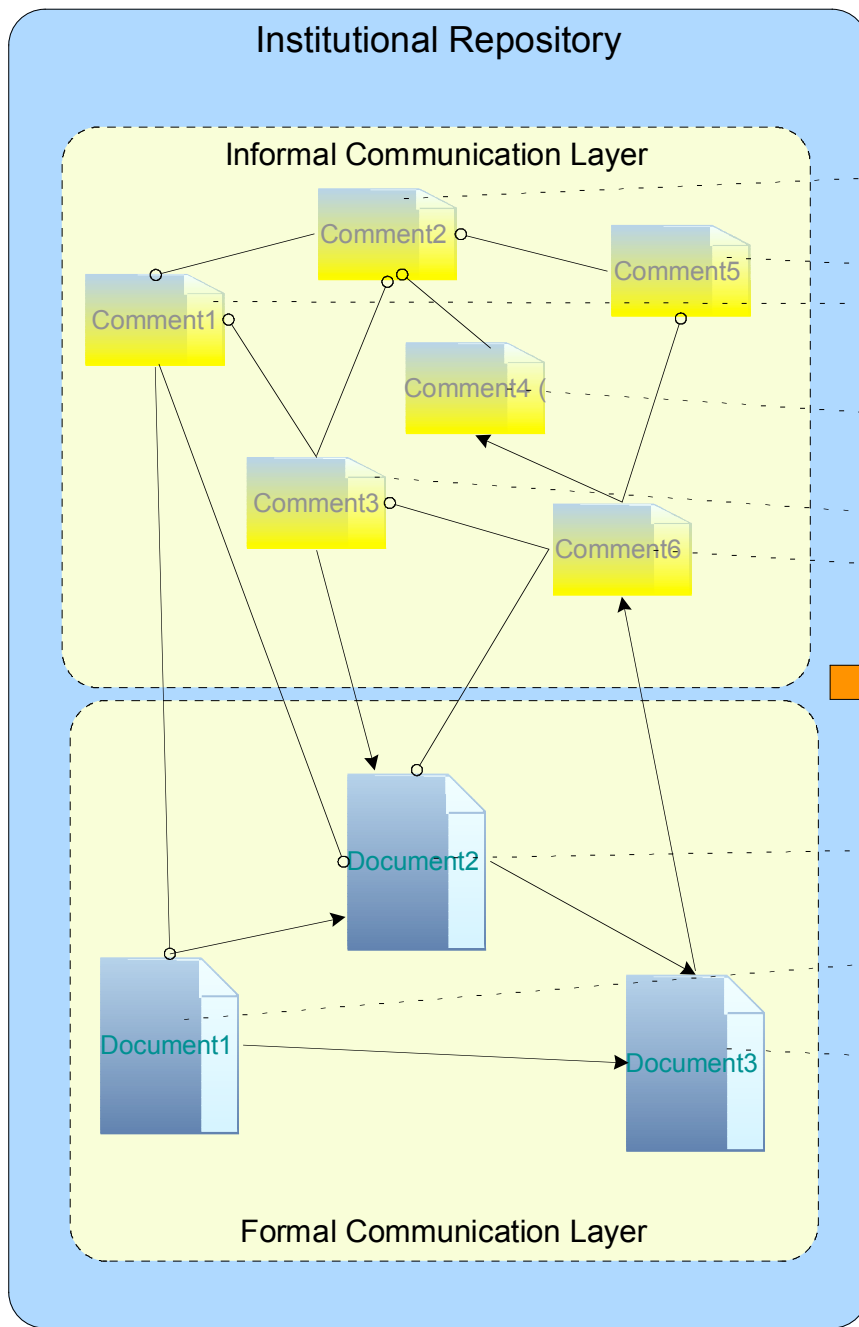


Todos os registos no repositório estão protegidos por leis de copyright, com todos os direitos reservados.

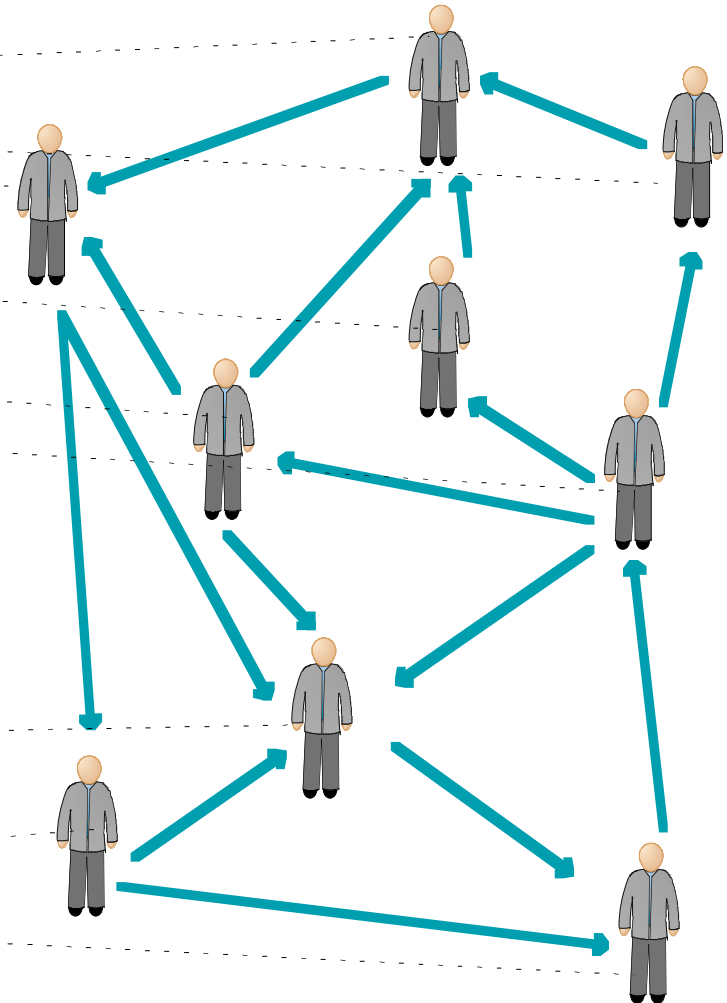
WoC add-on



- Goal: to provide a way to determine and depict the existing relationships between scholars that would otherwise be invisible.
- A Web of Communication is a network of weighted communication relationships between persons.
- First implementation: papadocs.

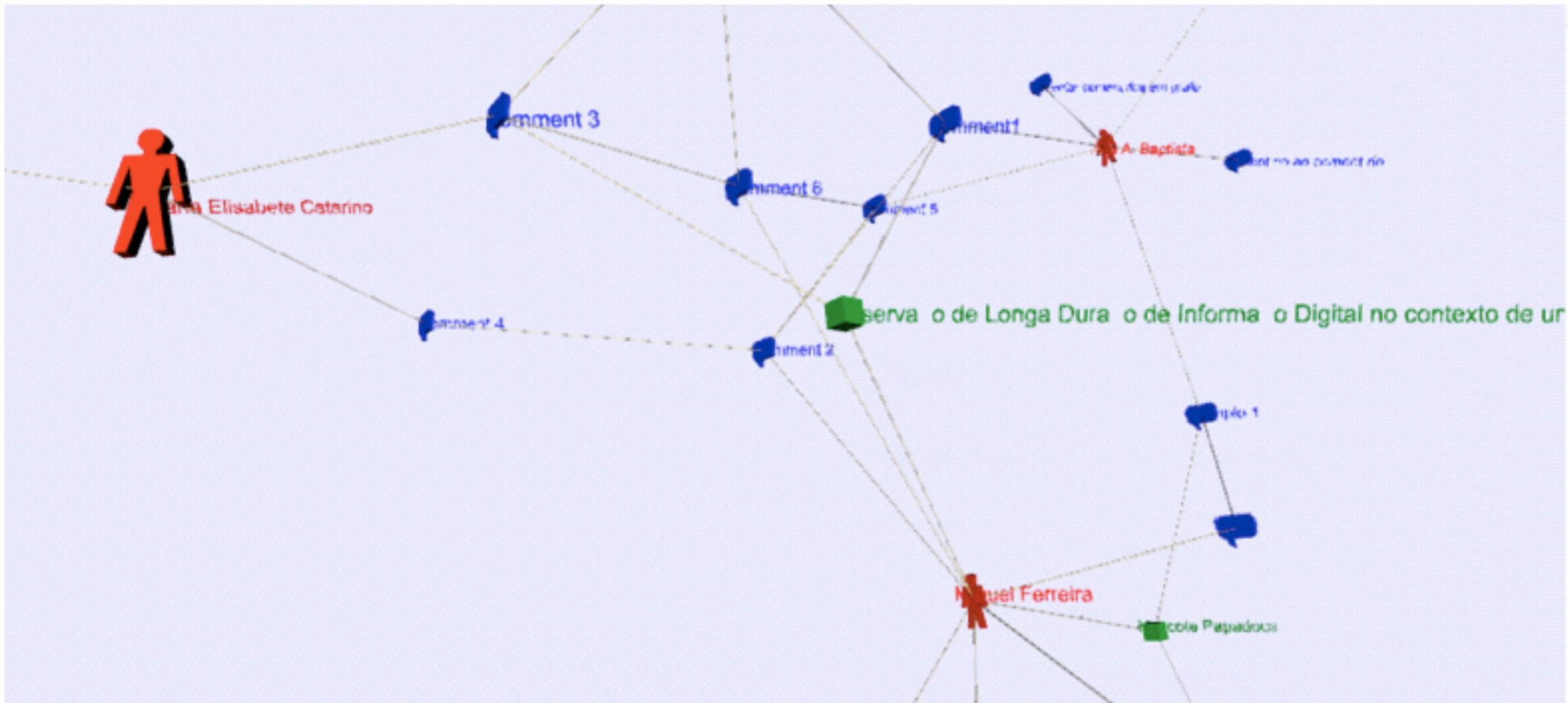


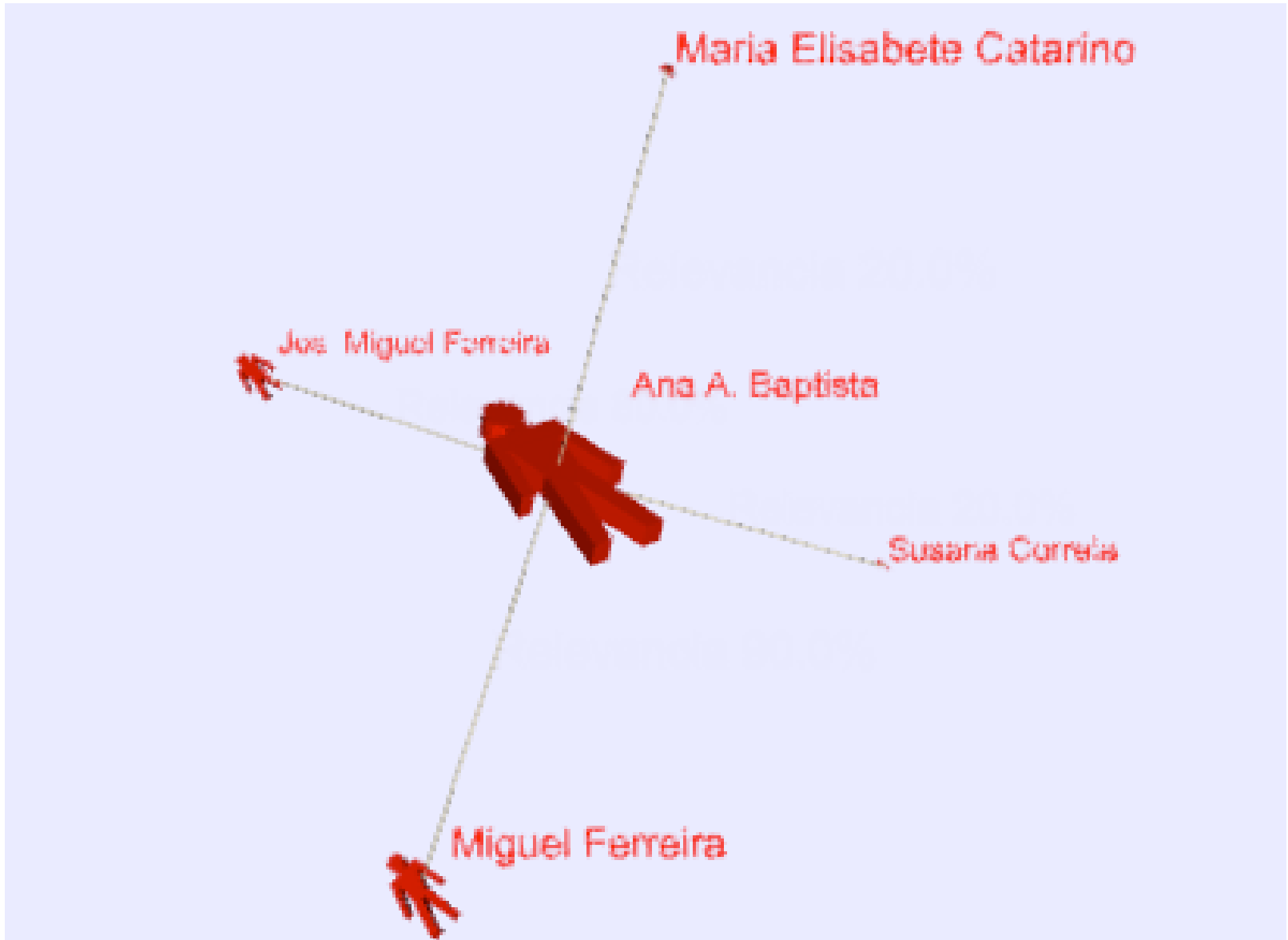
Web of Communication



- Citation Relationship
- Comment Relationship

- Communication Relationship
- Data Flow





Recommendation add- on



- A kind of WoC in text mode
- Visible relationships:
 - related people
 - related formal documents
 - related comments



http://papadocs.dsi.uminho.pt

Página Principal | Dep. Sistemas de Informação | Escola de Engenharia | RepositóriUM | Universidade do Minho | 14 visitantes em linha

Ligado como analice@dsi.uminh... (Sair)

Pesquisar Repositório:

Pesquisa Avançada
Pesquisar Conceitos

Percorrer:

- [Cursos & Disciplinas](#)
- [Títulos](#)
- [Autores](#)
- [Por Data](#)
- [Participantes](#)

Entrar:

- [Subscrever](#)
- [Meu Repositório](#)
utilizadores autorizados
- [Editar Conta](#)
- [Ajuda](#)

[Papadocs](#) >
[Comissão Científica](#) >
[Propostas de Doutoramento](#) >

Título: Preservação de Longa Duração de Informação Digital no contexto de um Arquivo Histórico

Autor(es): Ferreira, Miguel (n/a)

Orientador(es): Baptista, Ana Alice
Ramalho, José Carlos

Submetido por: [Miguel Ferreira](#)

Data submissão: sem data

Resumo: Certo tipo de documentação produzida por determinados ramos da Administração Pública (e.g. Registo Civil, Cartórios Notariais e Tribunais) é obrigatoriamente incorporada e preservada em Arquivos Históricos. Acontece que, uma grande parte dessa documentação é, actualmente, produzida em formato digital, o que levanta sérios problemas no que diz respeito à sua preservação. Os documentos digitais estão intrinsecamente relacionados com o software que os produziu, no entanto, o software possui um tempo de vida limitado sendo a sua descontinuidade um facto incontornável. O futuro previsível resume-se a um número variável de documentos prisioneiros do seu próprio formato de codificação. Diversas estratégias têm sido apontadas no sentido de solucionar o problema, no entanto, todas elas apresentam lacunas variadas e não abordam o problema

5 pessoas relacionadas

- [Miguel Ferreira](#)
- [Ana A. Baptista](#)
- [Maria Elisabete Catarino](#)
- [Susana Correia](#)
- [José Miguel Ferreira](#)

5 itens relacionados

- [Web Semântica - XML & RDF](#)
- [Mascote Papadocs](#)
- [Projectos Digitarq & CRAV](#)
- [Word to XML to Latex to PDF](#)
- [RODA - Repositório de Objectos Digitais Autênticos](#)

5 comentários relacionados

- [Comment 1](#)
- [comment 3](#)
- [Comment 6](#)
- [Comment 2](#)
- [Comment 5](#)

Tea for Two - Rationale



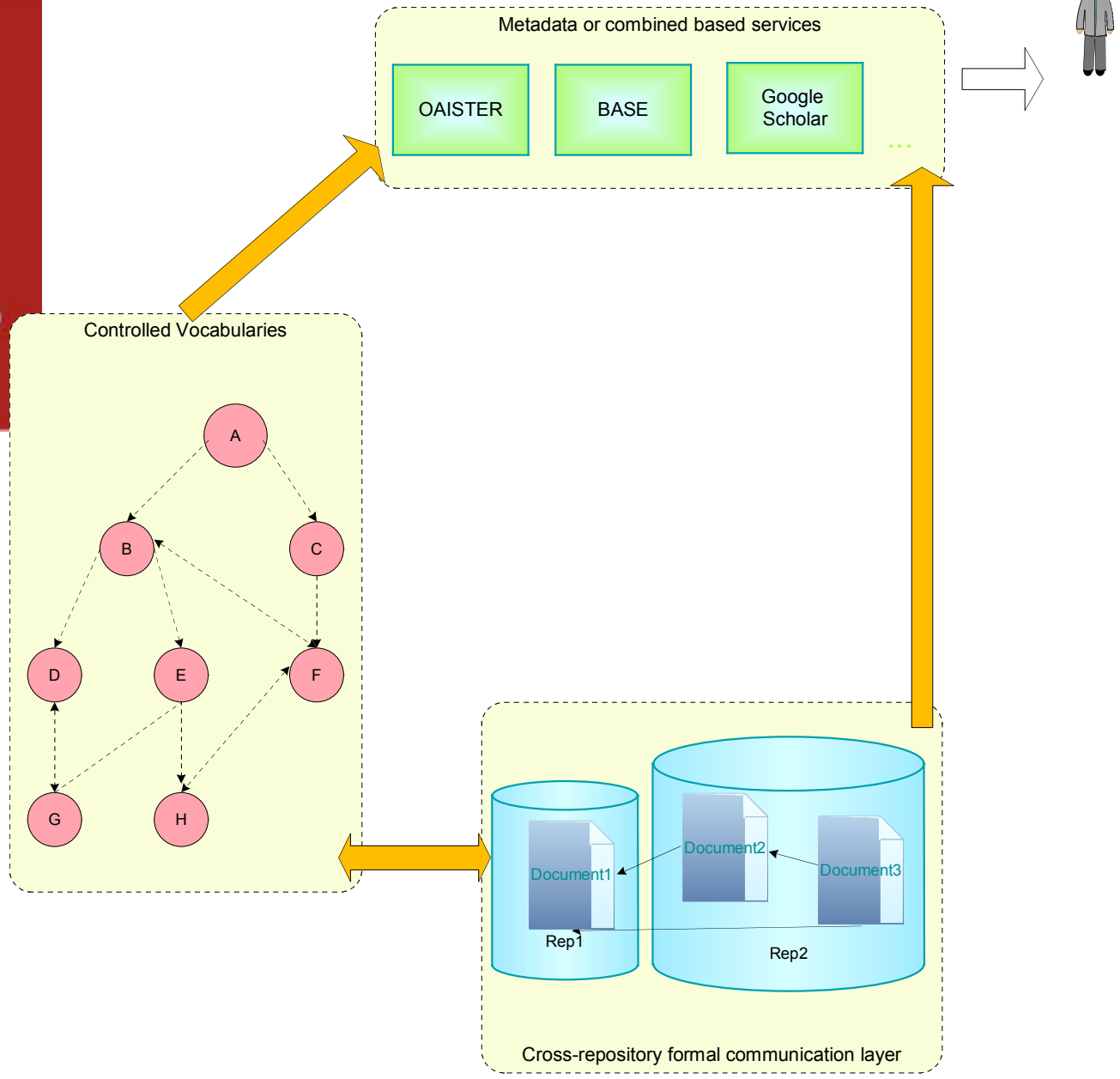
- Knowledge communities are built having formal documents as a basis
- People belong to one or more subject related communities
- There are persons especially gifted to perform the role of information mediators within and across communities
- The knowledge these persons share is not properly stored and accessible
- If it was, it could help to discover the “undiscovered public knowledge (Swanson, 1986)

Tea for Two - Aims

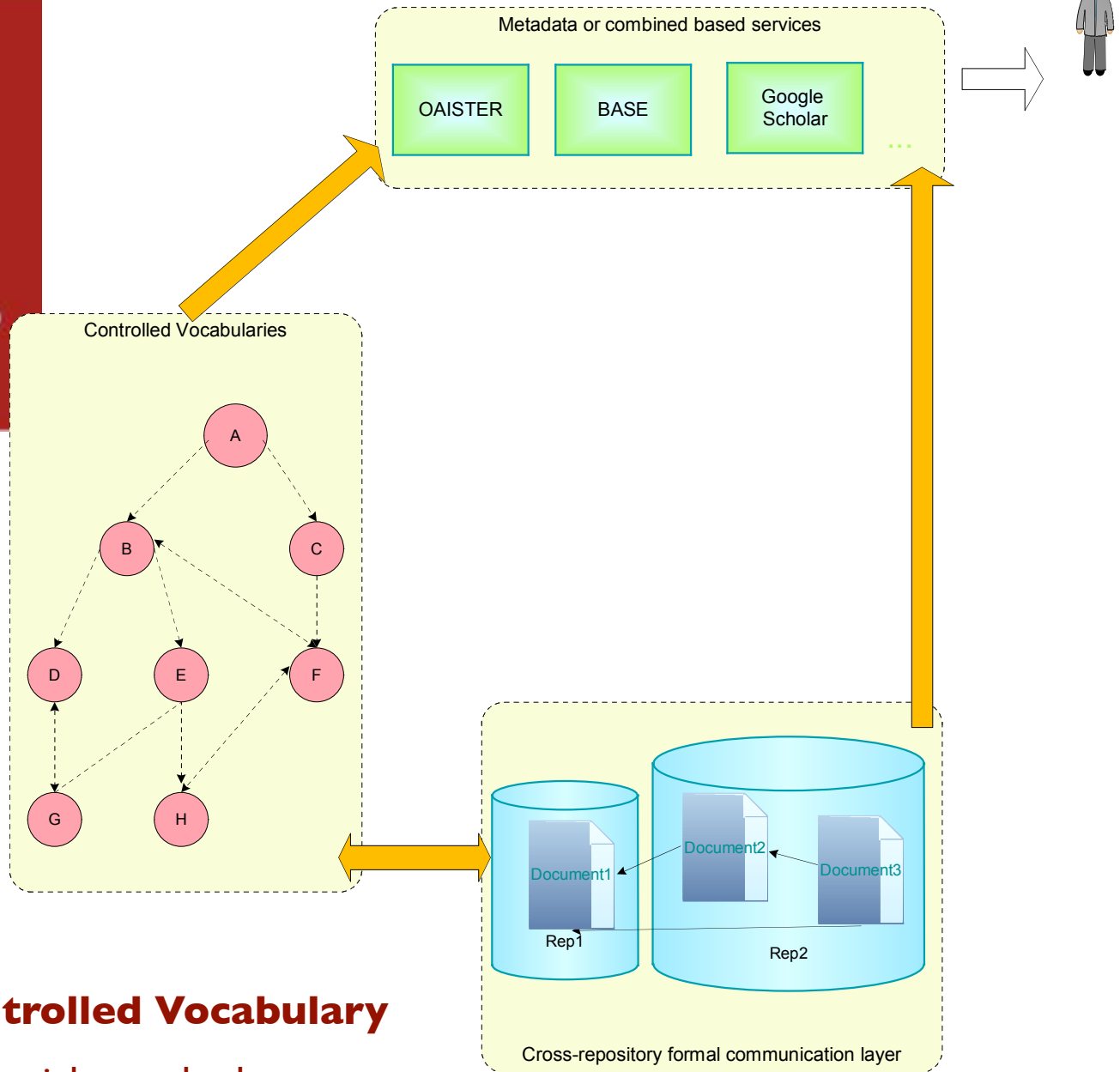


- To provide more favorable means of **uncovering undiscovered public knowledge**. Creating the necessary conditions for the increase in informal interactions, storing them, and making them more visible, thus facilitating the access to them.
- To create an **overall view of the communication relationships** between the scholars, i.e., a cross repository, using the (formal and informal) communication interactions.





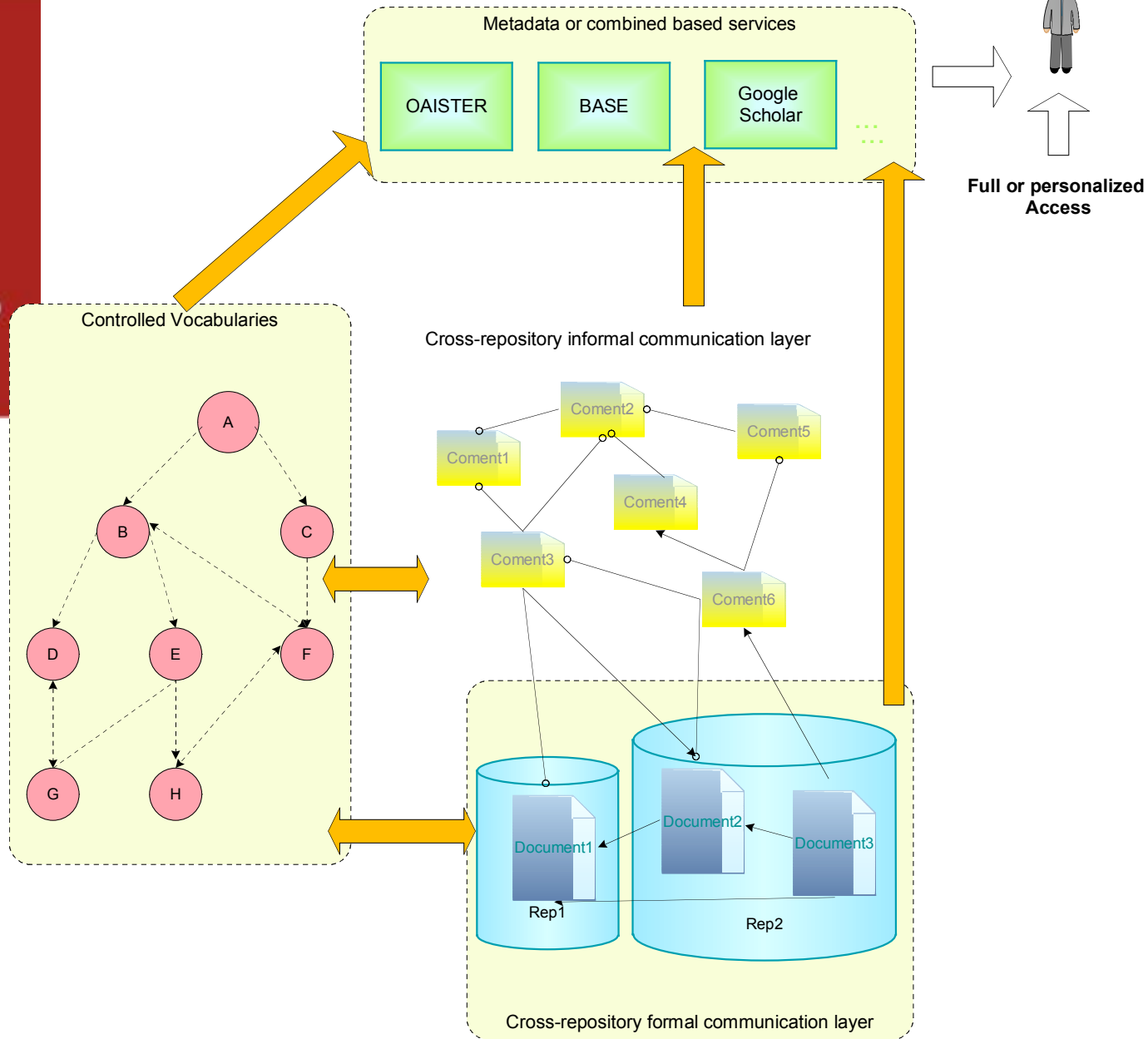
- Citation relationship
- > Controlled vocabulary specific relationship
- ➔ Data flow



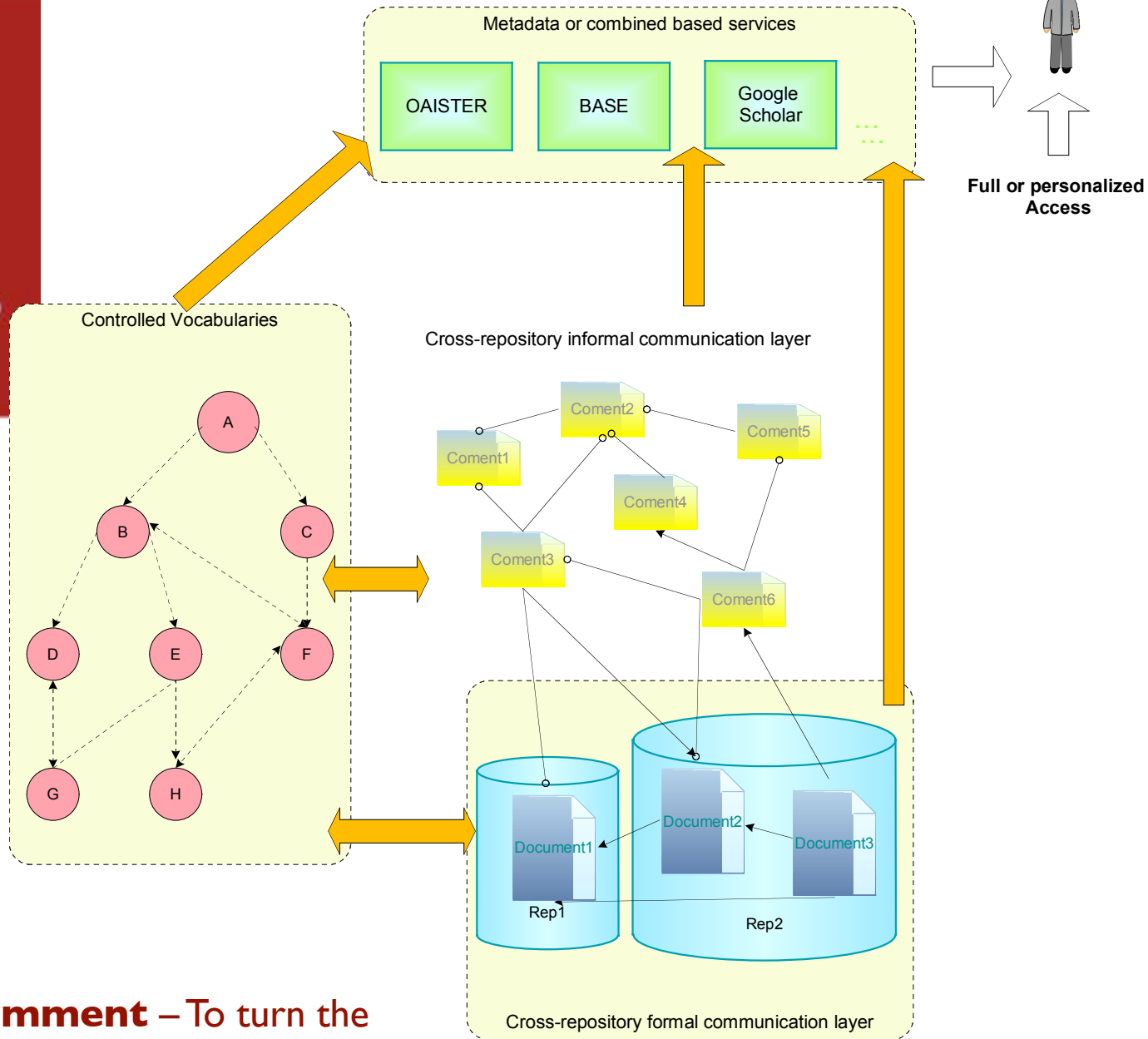
I - Controlled Vocabulary

- To use a rich standard description language to express controlled vocabularies.

- Citation relationship
- - - -> Controlled vocabulary specific relationship
- Data flow

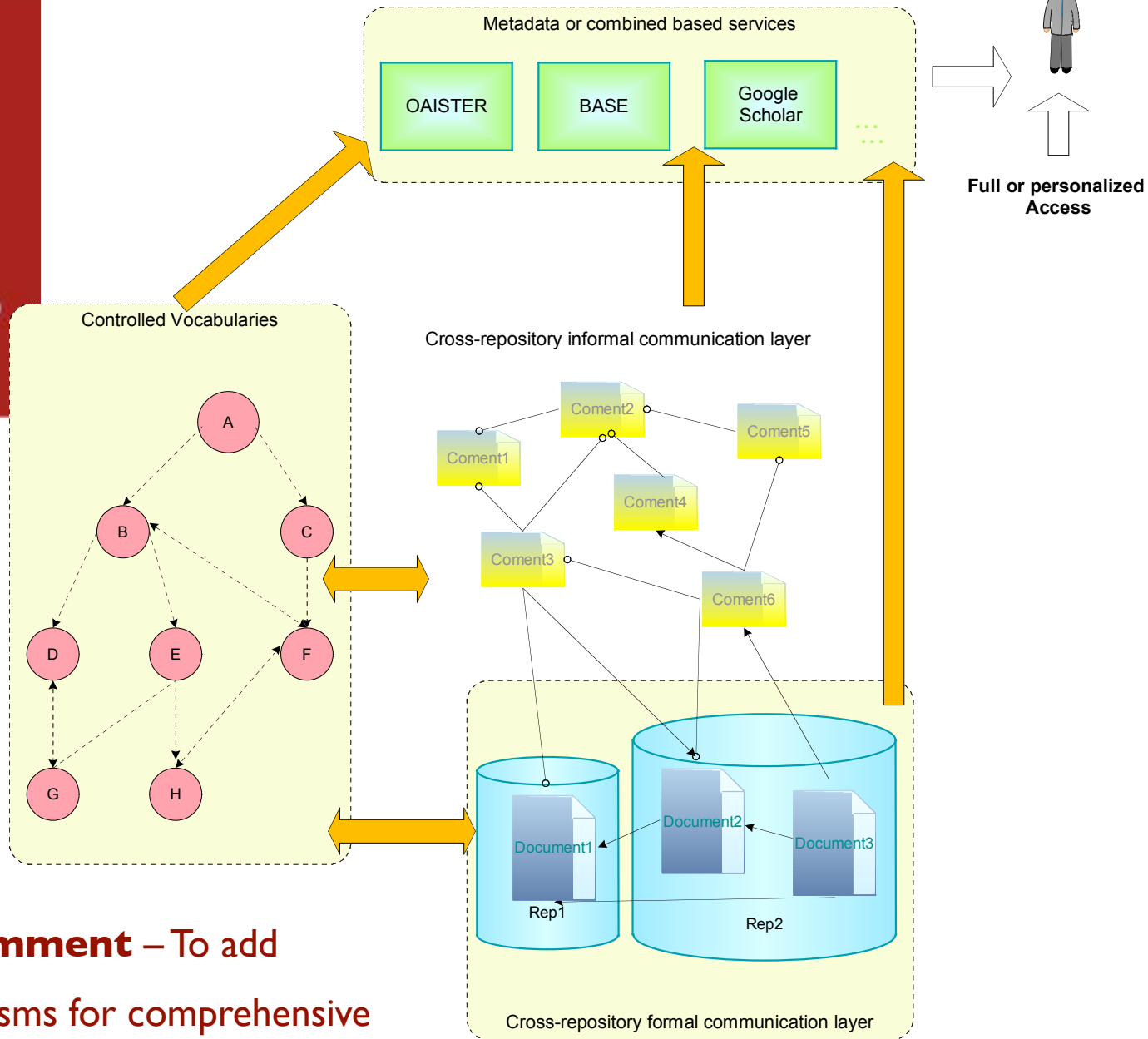


- Citation relationship
- Comment relationship
- - -→ Controlled vocabulary specific relationship
- ➔ Data flow



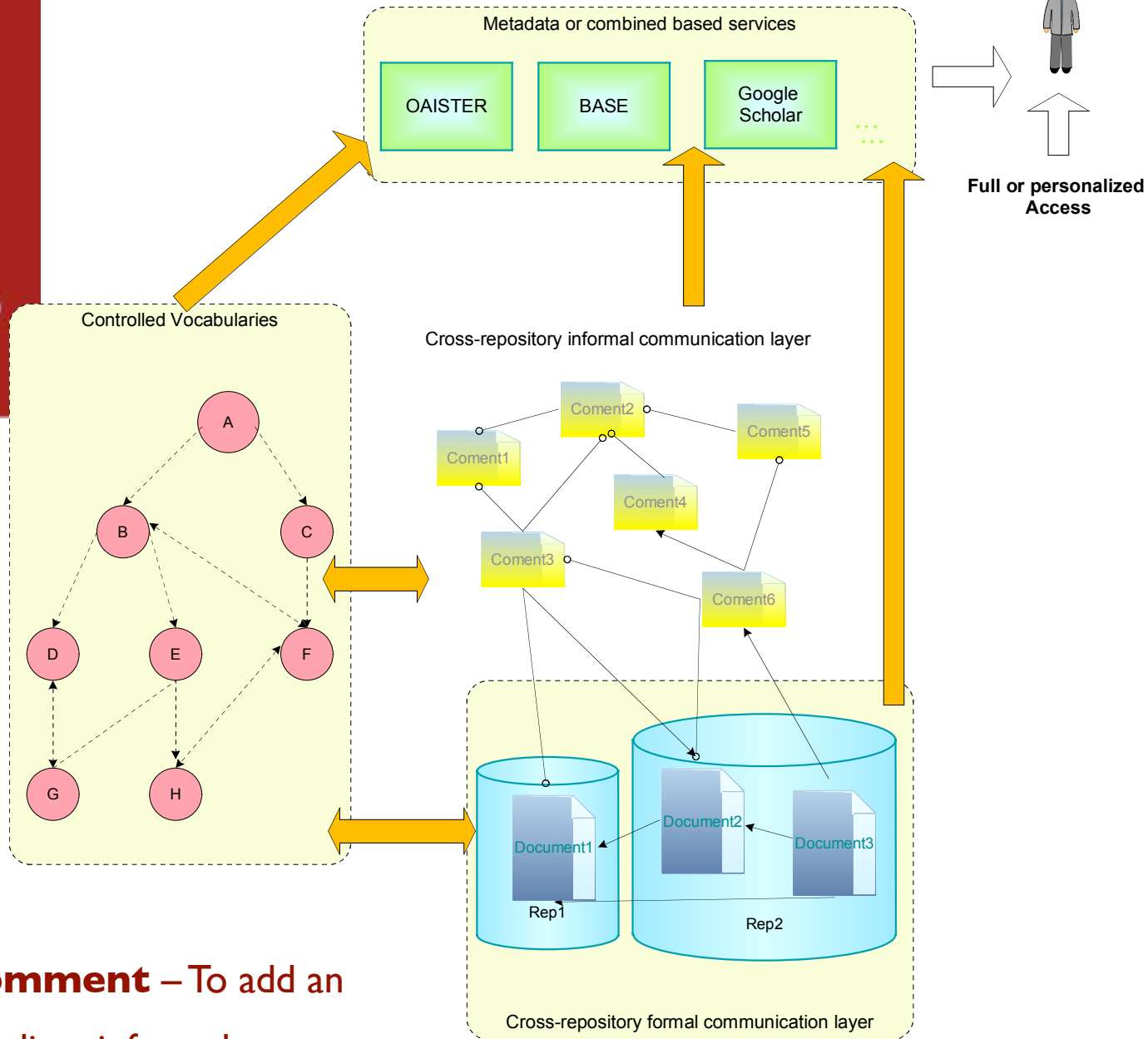
2 - Comment – To turn the add-on into a cross-repository service.

- > Citation relationship
- Comment relationship
- - -> Controlled vocabulary specific relationship
- ➡ Data flow



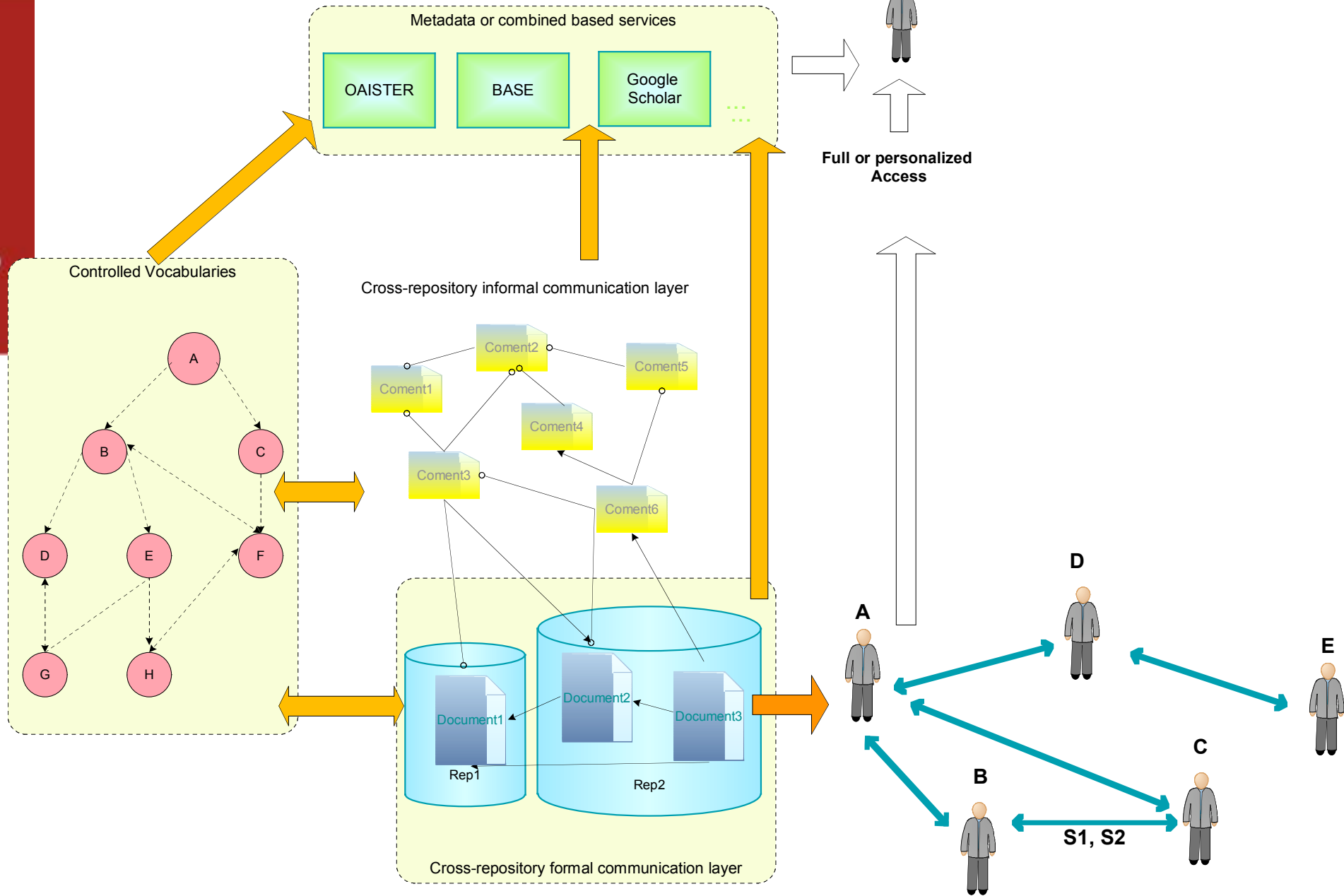
3 - Comment – To add mechanisms for comprehensive metadata description of informal documents, including subject, author and type information;

- Citation relationship
- Comment relationship
- Controlled vocabulary specific relationship
- ➔ Data flow

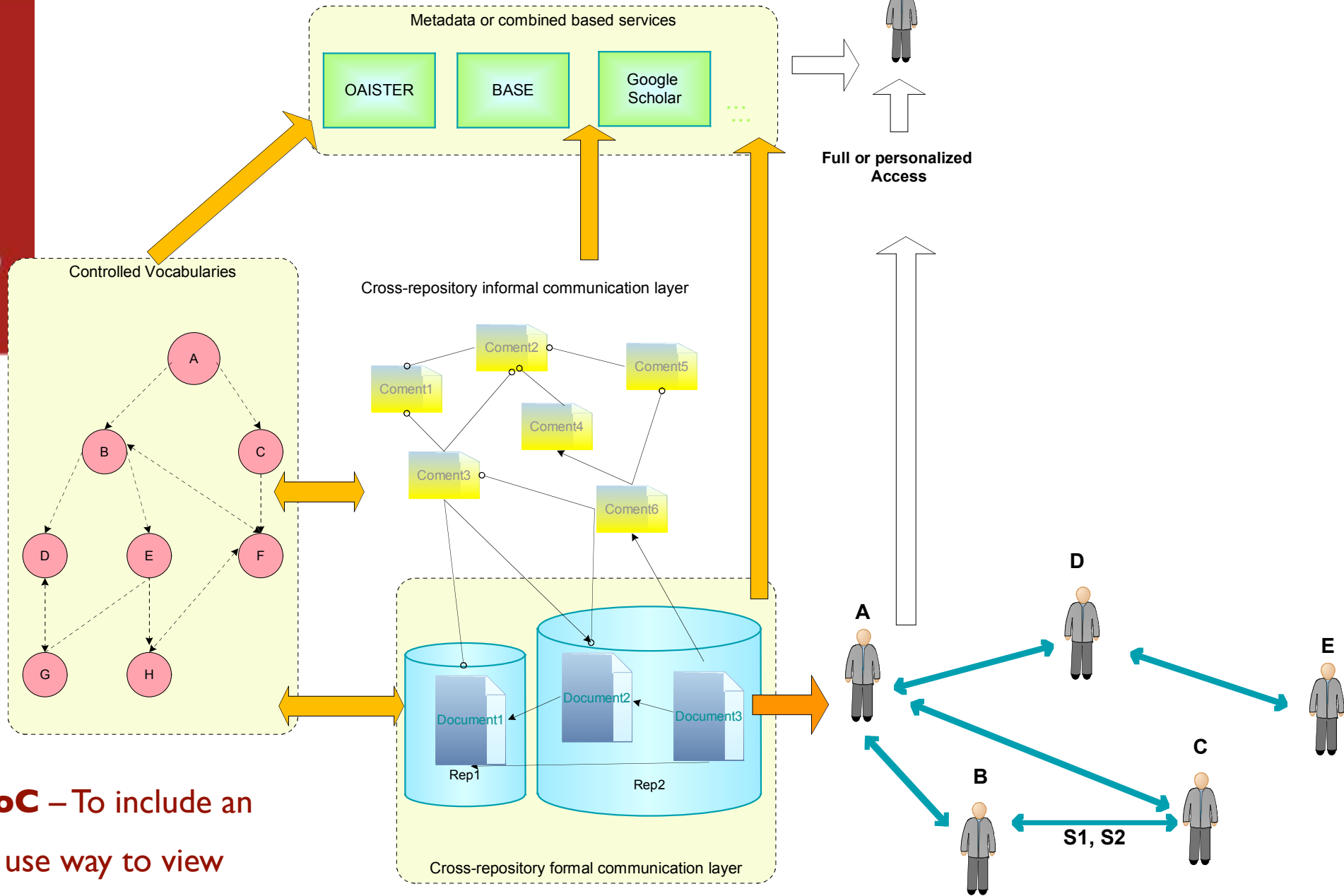


4 - Comment – To add an intermediate informal communication layer between documents and comments

- Citation relationship
- Comment relationship
- > Controlled vocabulary specific relationship
- ➔ Data flow

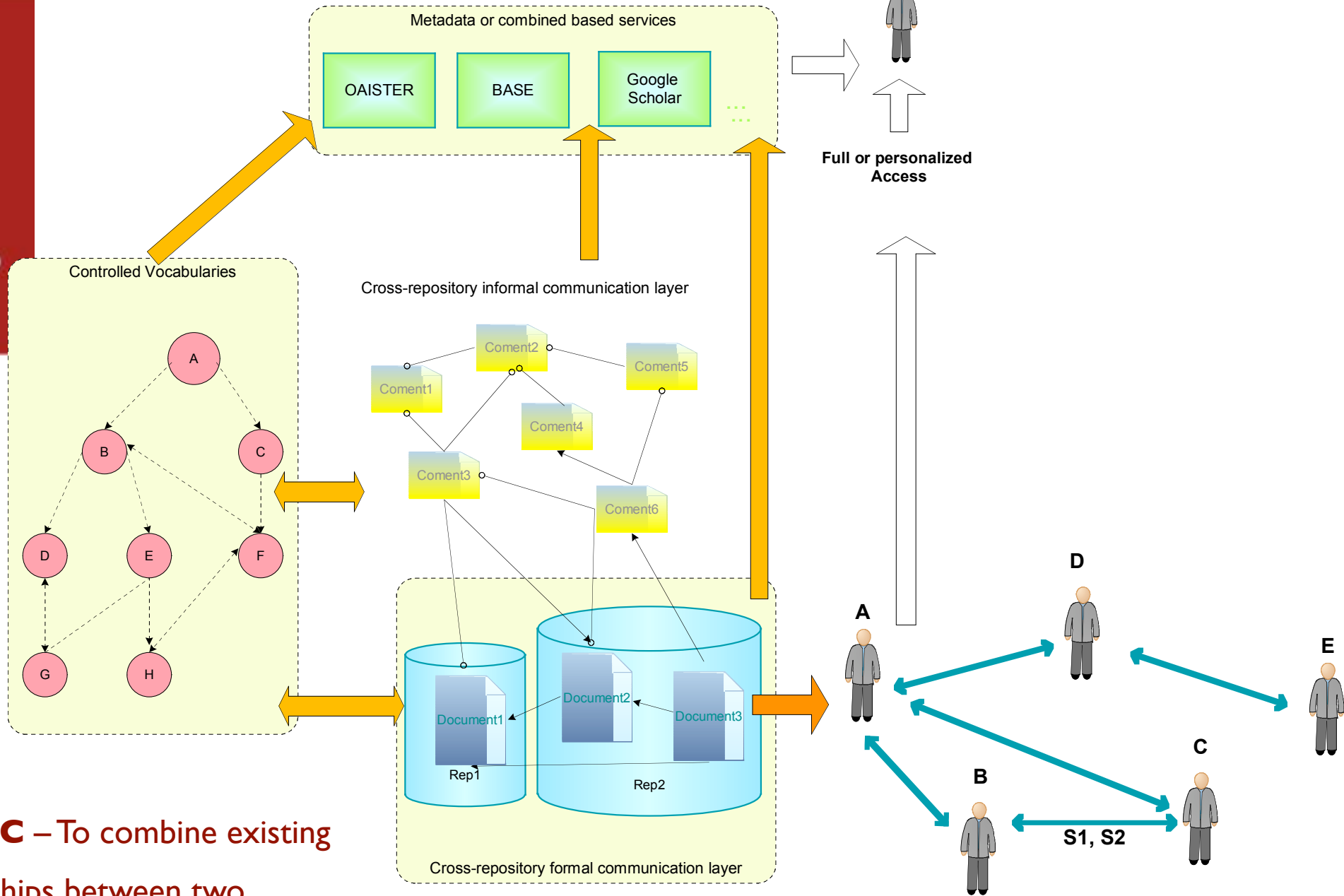


- Citation relationship
- Comment relationship
- > Controlled vocabulary specific relationship
- Data flow
- ↔ Communication Relationship

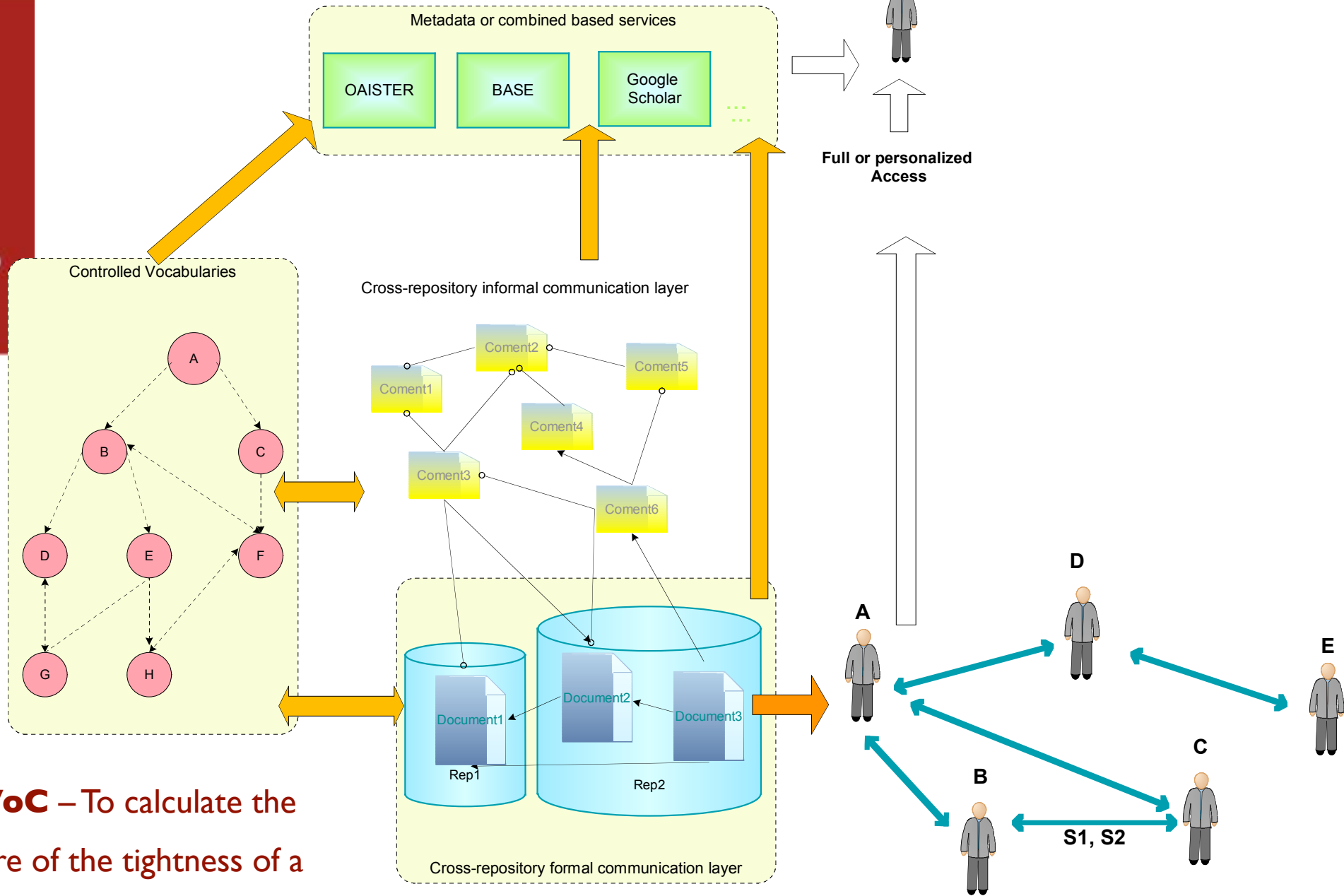


5 - WoC – To include an easy to use way to view subject information in a relationship between two persons;

- Citation relationship
- Comment relationship
- > Controlled vocabulary specific relationship
- ➡ Data flow
- ↔ Communication Relationship



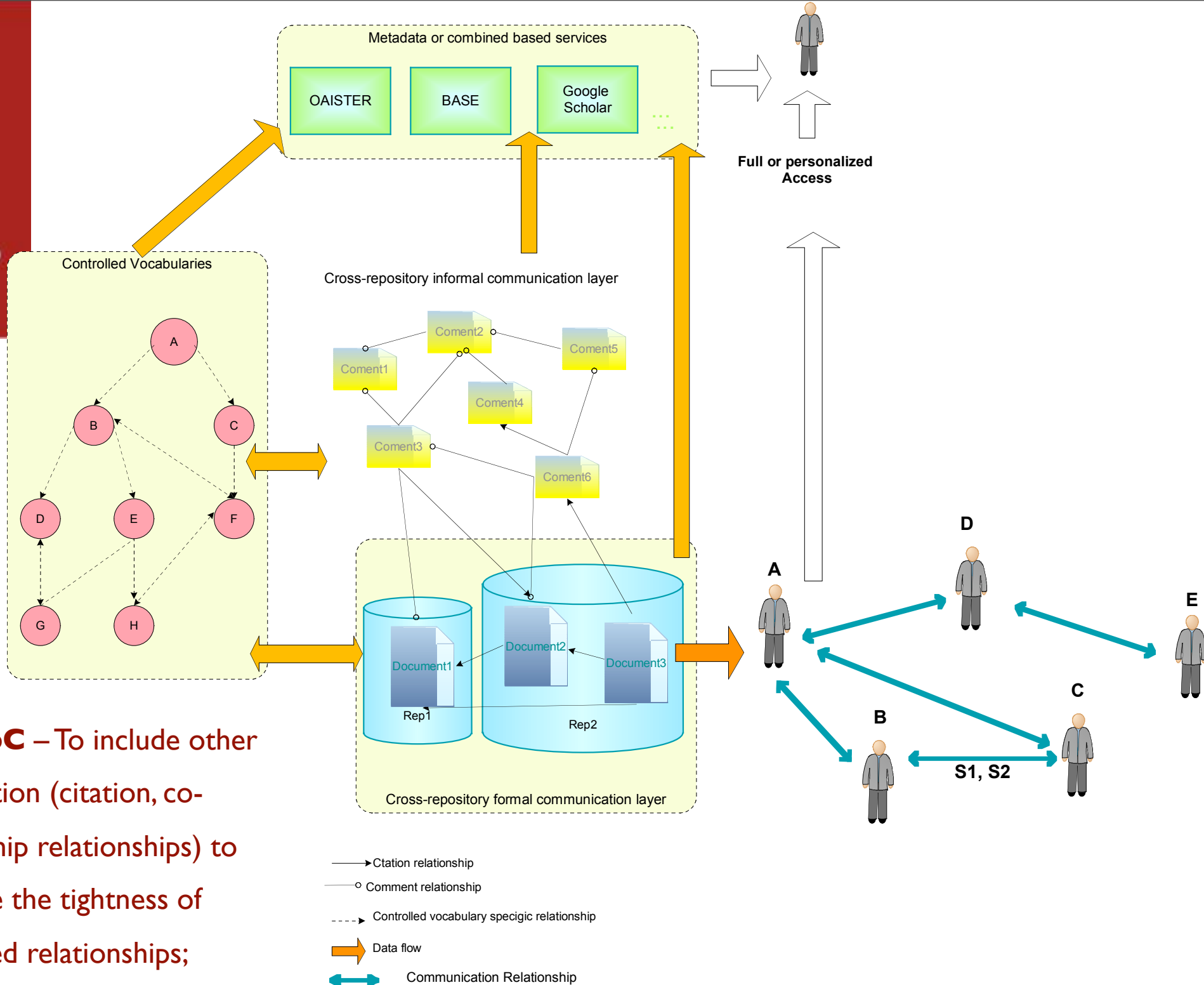
6 - WoC – To combine existing relationships between two persons in a single one and depict it;

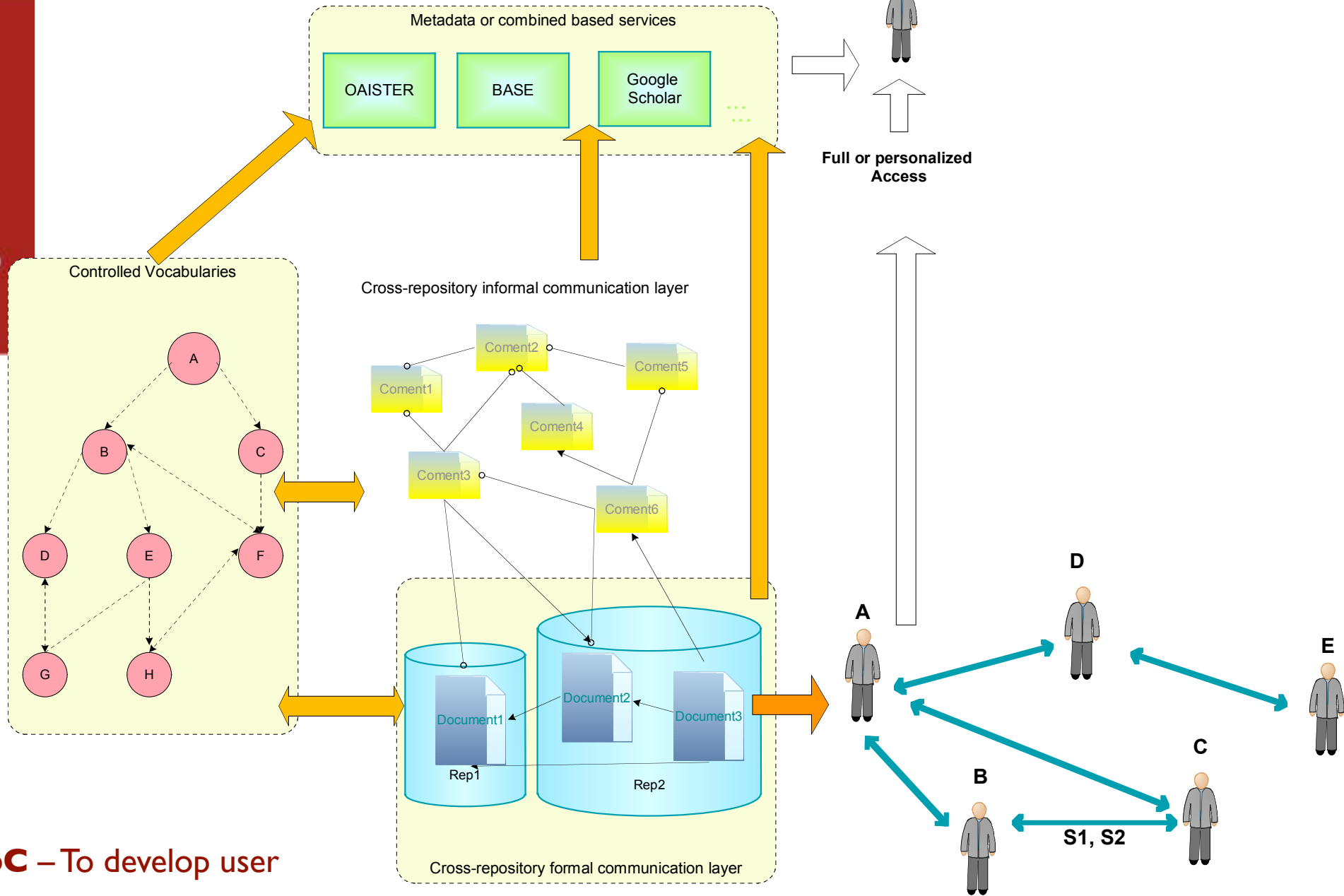


7 - WoC – To calculate the measure of the tightness of a combined relationship between two persons and find a way to depict it;

- Citation relationship
- Comment relationship
- > Controlled vocabulary specific relationship
- ➔ Data flow
- ↔ Communication Relationship

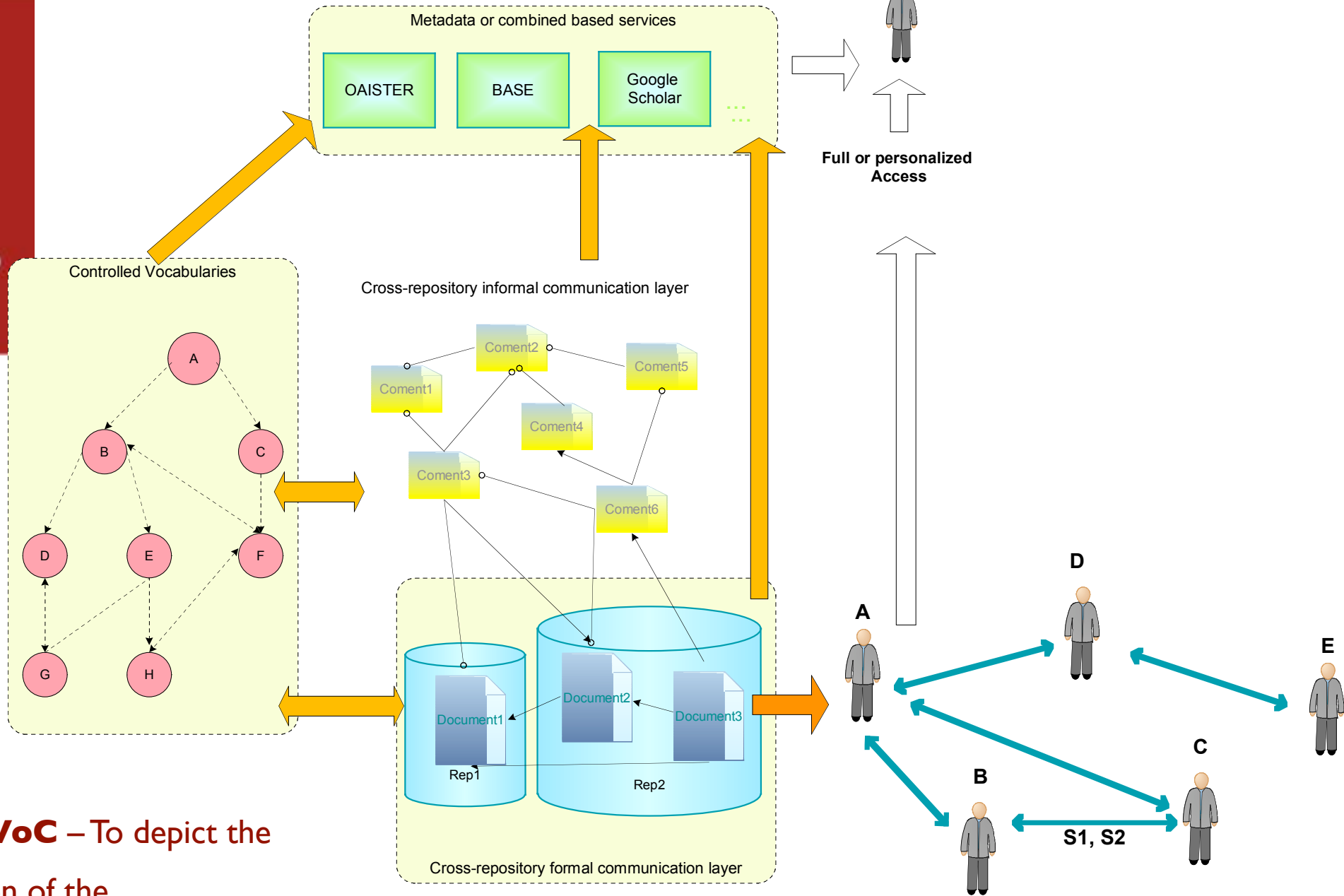
8 - WoC – To include other information (citation, co-authorship relationships) to calculate the tightness of combined relationships;



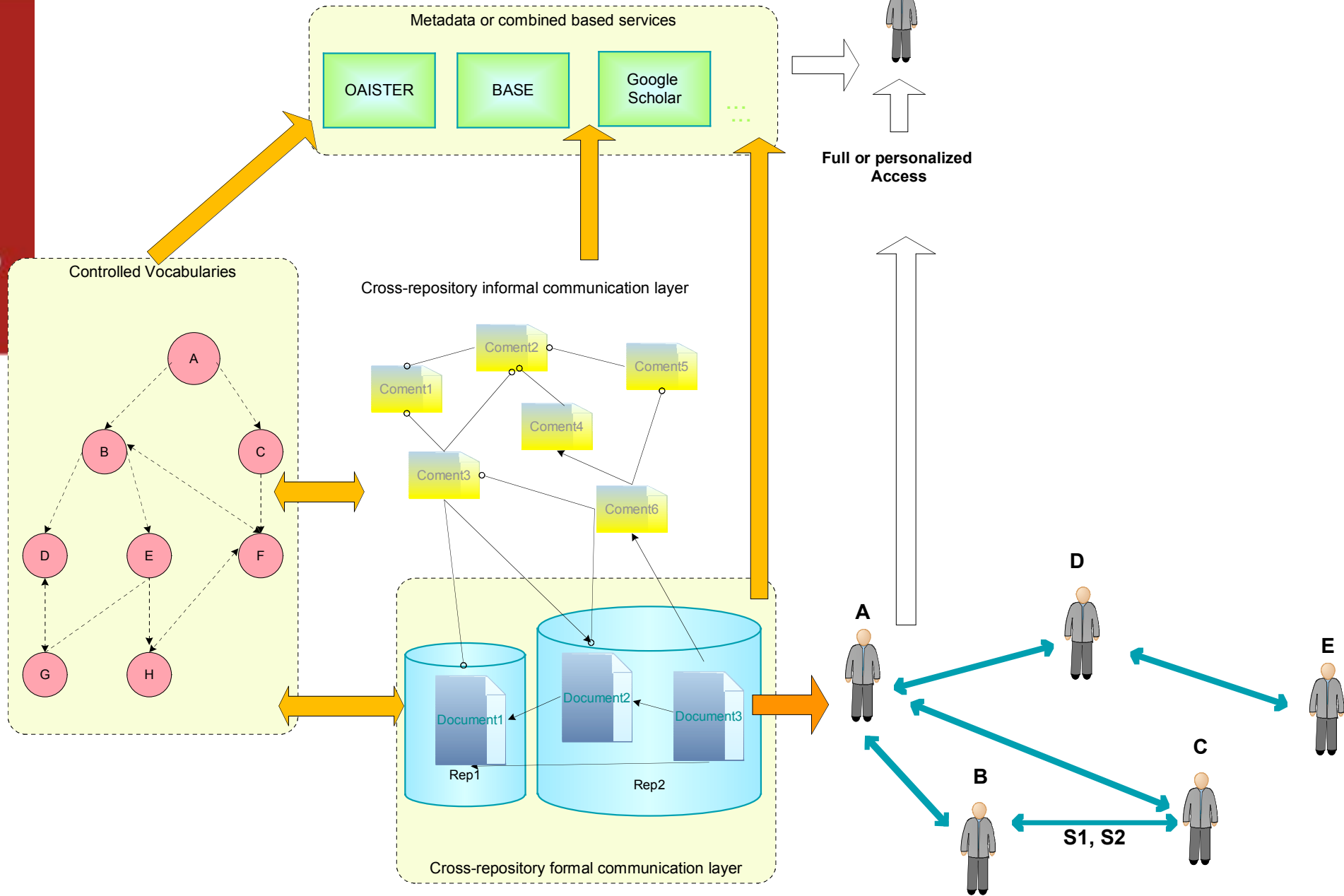


9 - WoC – To develop user configurable filters for the WoC depiction;

- Citation relationship
- Comment relationship
- > Controlled vocabulary specific relationship
- ➡ Data flow
- ↔ Communication Relationship



IO - WoC – To depict the direction of the communication between two authors.



- Citation relationship
- Comment relationship
- > Controlled vocabulary specific relationship
- Data flow
- ↔ Communication Relationship



- 1 - Controlled Vocabulary** – To use a rich standard description language to express controlled vocabularies.
- 2 - Comment** – To turn the add-on into a cross-repository service.
- 3 - Comment** – To add mechanisms for comprehensive metadata description of informal documents, including subject, author and type information;
- 4 - Comment** – To add an intermediate informal communication layer between documents and comments
- 5 - WoC** – To include an easy to use way to view subject information in a relationship between two persons;
- 6 - WoC** – To combine existing relationships between two persons in a single one and depict it;
- 7 - WoC** – To calculate the measure of the tightness of a combined relationship between two persons and find a way to depict it;
- 8 - WoC** – To include other information (citation, co-authorship relationships) to calculate the tightness of combined relationships;
- 9 - WoC** – To develop user configurable filters for the WoC depiction;
- 10 - WoC** – To depict the direction of the communication between two authors.

Tea for Two - Who



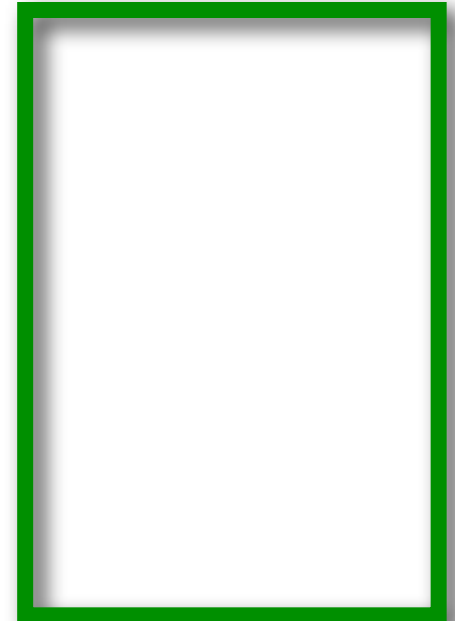
Ana A. Baptista
analice@dsi.uminho.pt



Miguel Ferreira
mferreira@dsi.uminho.pt



Eva Oliveira
eva.oliveira@gmail.com



Simone Costa
scosta@reitoria.uminho.pt

Contents



- Scope:
 - UMinho and DSI
 - SDUM and RepositóriUM
 - Odisseia
- CRiB
- Tea for Two
- [Other Projects](#)
- Future work

Other Projects



- Educational Contents Repository
- User Studies (waiting for approval)
- Social Tagging
- Informattica Online (standby)



Other Projects - Who

(Educational Contents Repository)



Ana Dias

anadias@tecminho.uminho.pt



Ana A. Baptista

analice@dsi.uminho.pt



José Carvalho

jcarvalho@tecminho.uminho.pt



Elisabete Cardoso

elisabete@dsi.uminho.pt

Other Projects - Who (Social Tagging)



Ana A. Baptista

analice@dsi.uminho.pt



Elisabete Catarino

ecatarino@dsi.uminho.pt



Anabela Novais

anabalanovais@gmail.com



Susana Pinheiro

susana-pinheiro@hotmail.com

Contents



- Scope:
 - UMinho and DSI
 - SDUM and RepositóriUM
 - Odisseia
- CRiB
- Tea for Two
- Other Projects
- Future work



Future Work (CRiB)

- to change the way the overall rating (recom. engine) is calculated
- to finish the integration of evaluators
- to validate the prototype
- “Real” test as a service for digital preservation



Future work (T4.2)

- to do more research on social networks
- to get together with other teams working on same/similar subjects
- to develop and test the new prototype
- to perform user studies in order to evaluate its achievements - validate the system

Future work (Odisseia)



- Pursue the “informal way” to “open” channels for scholarly communication to occur and
- to share knowledge and
- to allow information mediators to do their job so that they can
- help to discover the “undiscovered public knowledge”.

Contents



- Scope:
 - UMinho and DSI
 - SDUM and RepositóriUM
 - Odisseia
- CRiB
- Tea for Two
- Other Projects
- Future work

End...



- ... so far.

THANK YOU!

analice@dsi.uminho.pt