

30 Years of DL Research @ UNIPD

Maristella Agosti, Giorgio Maria Di Nunzio,
Nicola Ferro, Maria Maistro, Stefano Marchesin,
Nicola Orio*, Chiara Ponchia*, Gianmaria Silvello

Department of Information Engineering

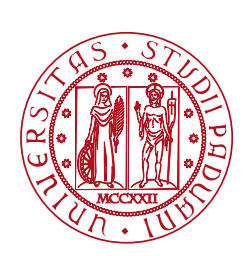
*Department of Cultural Heritage

University of Padua, Italy

14TH ITALIAN RESEARCH CONFERENCE ON DIGITAL LIBRARIES

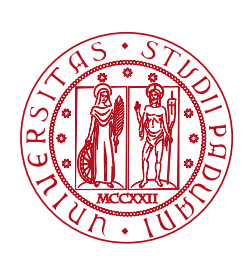
University of Udine

26 gennaio 2018

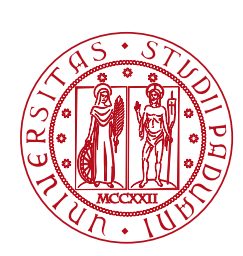


Outline

- DUO and the birth of the DL area
- Annotations
- CULTURA and Digital Archives
- Log Analysis
- Personalisation
- User-System Interaction
- Data Driven Digital Libraries



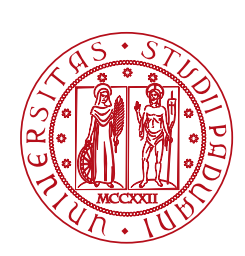
DUO: An Innovative OPAC



The Italian Library Automation Project and the OPAC



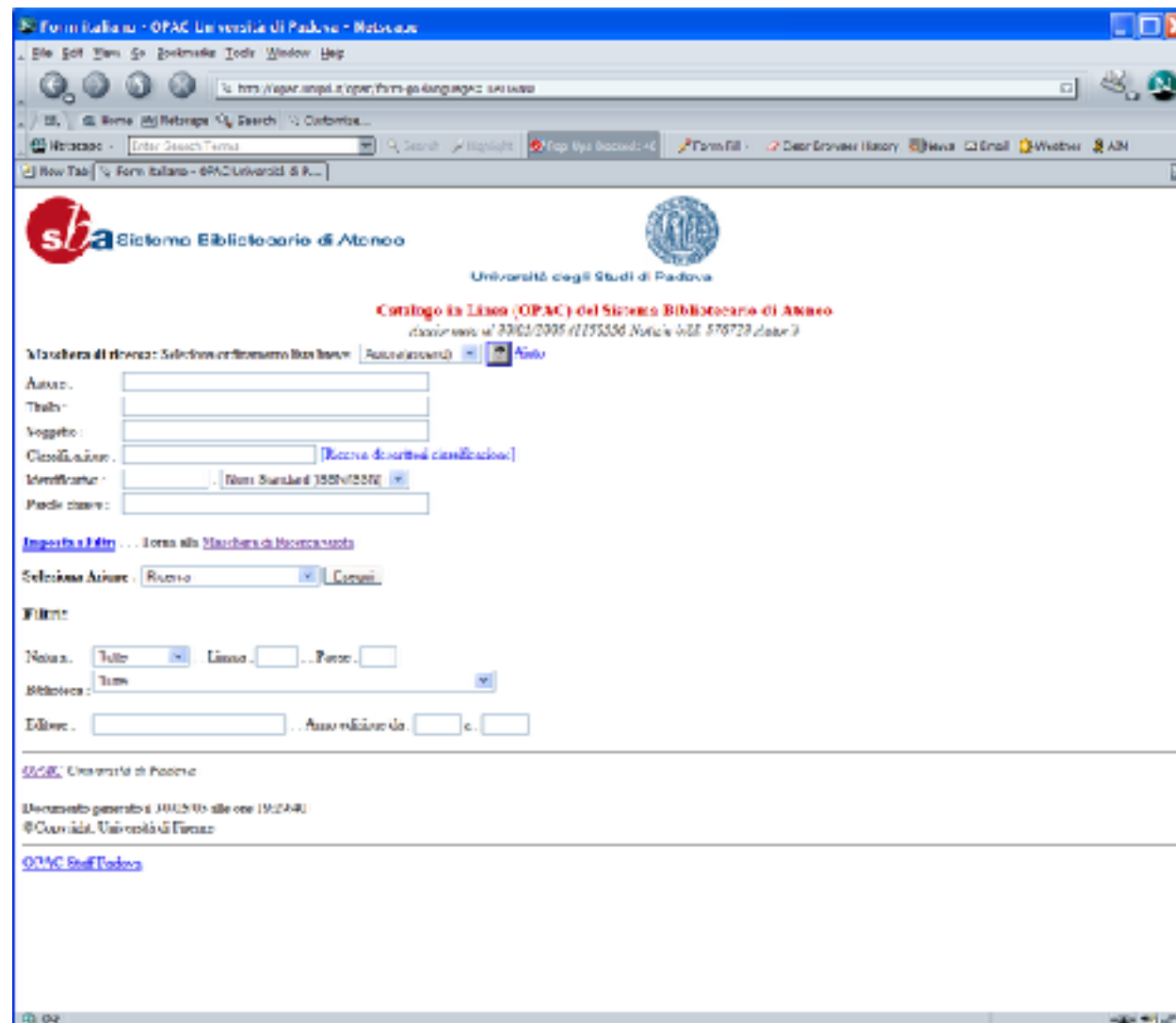
- The Italian national project of library automation, called SBN - Servizio Bibliotecario Nazionale, is an advanced library automation project started in 1970s
- Different library automation systems at national/regional/local level cooperating in a networked/hierarchical organisation
- Until late in the 1980s
 - The public online access to bibliographic data was not available, only traditional card catalogues were in use



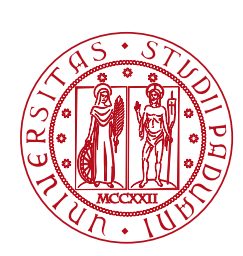
OPAC Access at the University of Padua



- The University of Padua became a node of the SBN project in the late 1980s
- At that time, there was much interest in OPAC
 - A first indication that information retrieval might start to interest the general public of libraries
- We launched a project for a third generation OPAC with advanced library catalogue and IR functions



- The time was not ripe for Web applications: the IR functions were lost



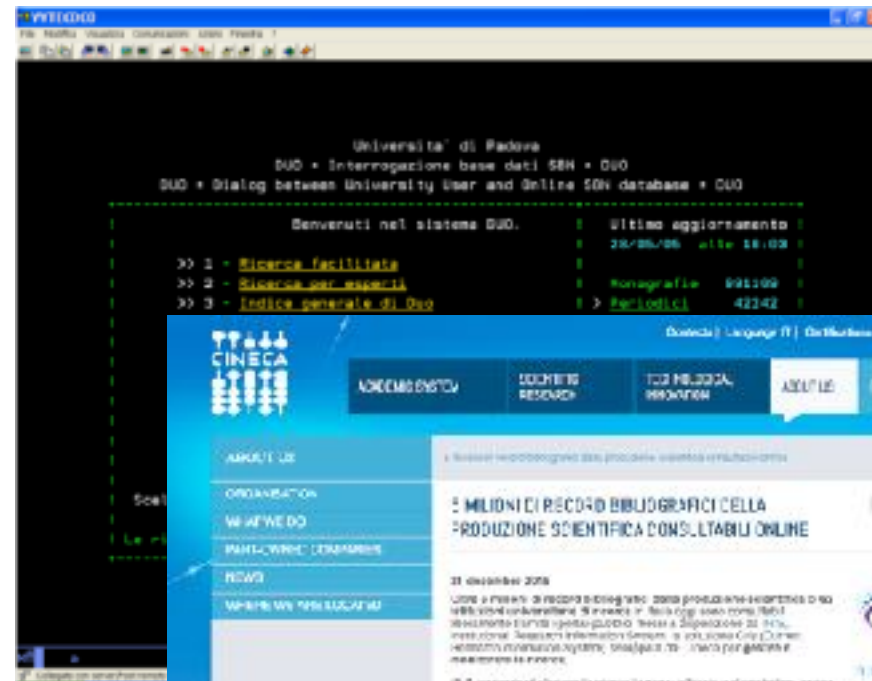
The Birth of the Digital Library Area



- The Library Automation community realises the lack of computer science and engineering knowledge
- The area of Digital Library starts in those years as a new scientific area
 - In USA - Digital Libraries Initiatives (DLI-1 and DLI-2) of the National Science Foundation (from late 1993)
 - In Europe - A group of projects supported by the European Commission under the 4th, 5th and 6th Framework Programme named DELOS Working Group 1996-99, first DELOS Network of Excellence 2000-2003, and DELOS Network of Excellence for Digital Libraries 2004-2007
- It is an area of confluence: library automation, database management, information retrieval, the Web, ...

DL a meeting point of disciplines and research fields:

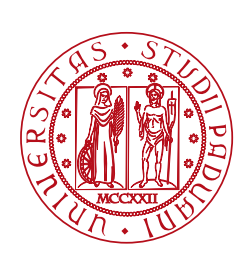
- DB management
- Information retrieval
- Library and information systems
- Document and information systems
- The Web
- Information visualization
- Artificial intelligence
- Human-computer interaction



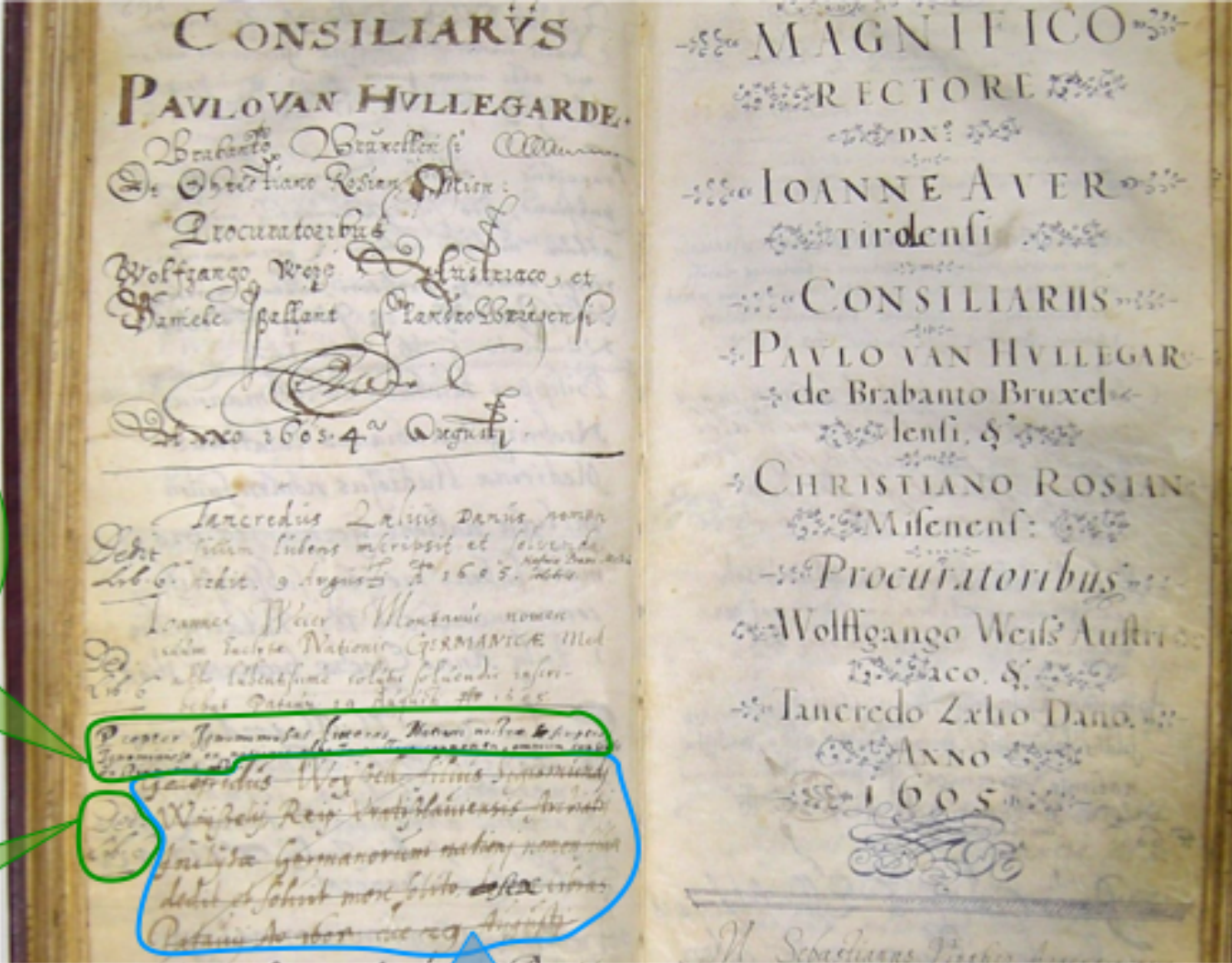


References

- M. Agosti, M. Masotti, A.M. Moressa. An Online Public Access Catalogue (OPAC) for University Library End-users Using TRS: project and prototype. Proc. Software AG's European Users' Conference, Hamburg, Germany, 1990, Vol.1, Paper N.52
- M. Agosti, M. Masotti. Design of an OPAC database to permit different subject searching accesses in a multi-disciplines universities library catalogue database. In: N. J. Belkin, P. Ingwersen, A. M. Pejtersen (Eds.). Proc. of the 15th Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval. Copenhagen, Denmark, ACM, 1992, 245-255
- S. Walker. Improving subject access painlessly: recent work on the Okapi online catalogue projects. Program, 1988, 22(1), 21-31
- S. Robertson, On the history of evaluation in IR. Journal of Information Science, 2008, 34(4), 439-456
- M. Agosti. Digital Libraries. Mondo Digitale, n. 43, settembre 2012, pp. 1-13



Annotations



a₃

Propter ignominiosas litteras Nationi nostrae scriptas ignominiose ex Nationis albo in publico conventu, omnium consensu, extirpatus est

[He was ignominiously expelled by the Council of Association, because he wrote ignominious letter to the Association]

Dedit libras 6

[He payed 6 liras]

a₂

Godefridus Woyssel, filius Sigismundi Woysseli reipublicae Vratislaviensis archiatri, inclitae Germanorum nationi nomen suum dedit et solvit more solito sex libras. Patavii anno 1605 die 29 augusti

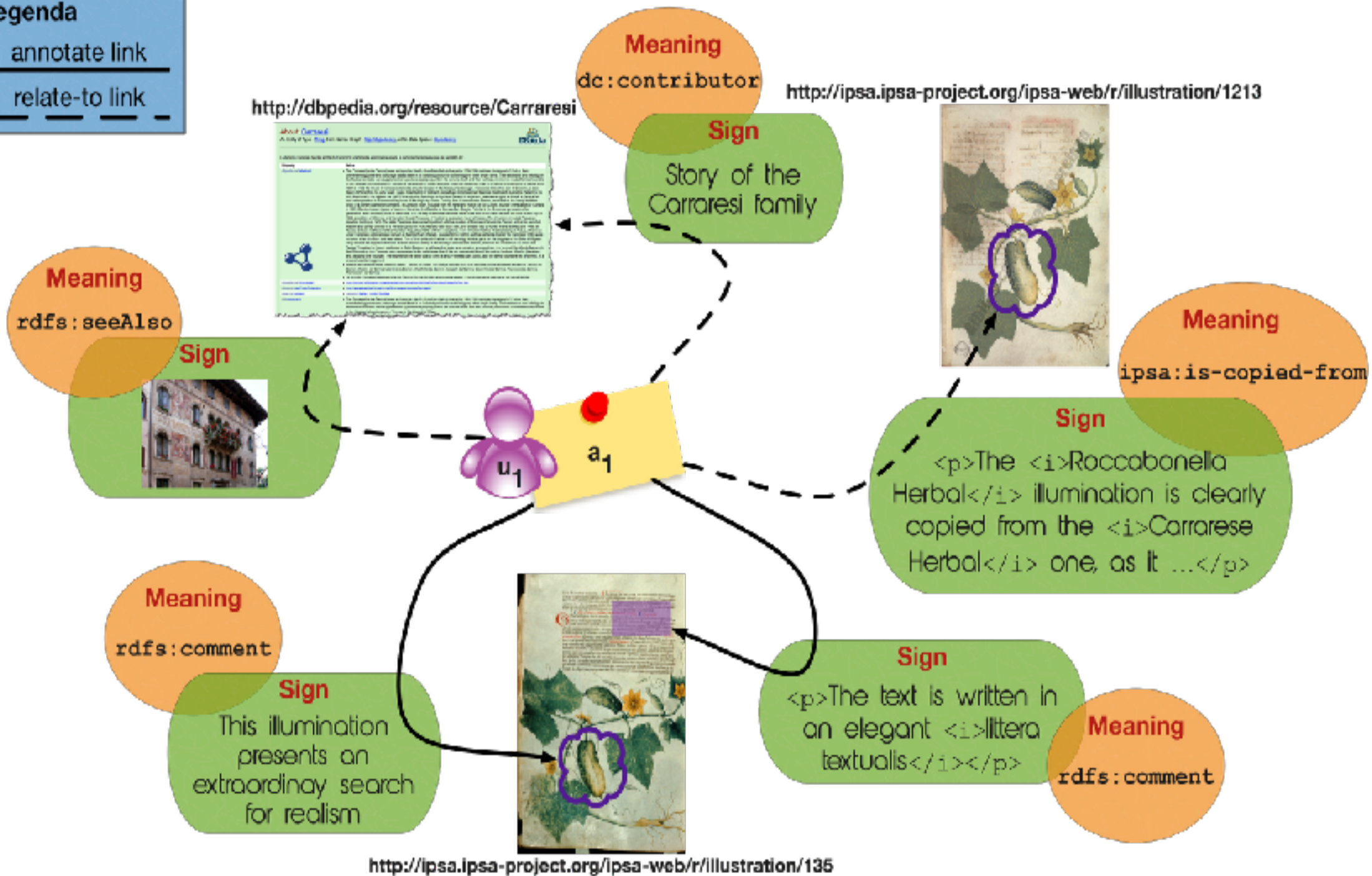
[Godfrey Woyssel, Sigmund's son, from Bratislava, enrolled and payed 6 liras]

a₁

- Italia, Padova, Archivio dell'Università di Padova, Archivio antico, Matricula Nationis Germanicae artistarum, reg. 465, c. 69v

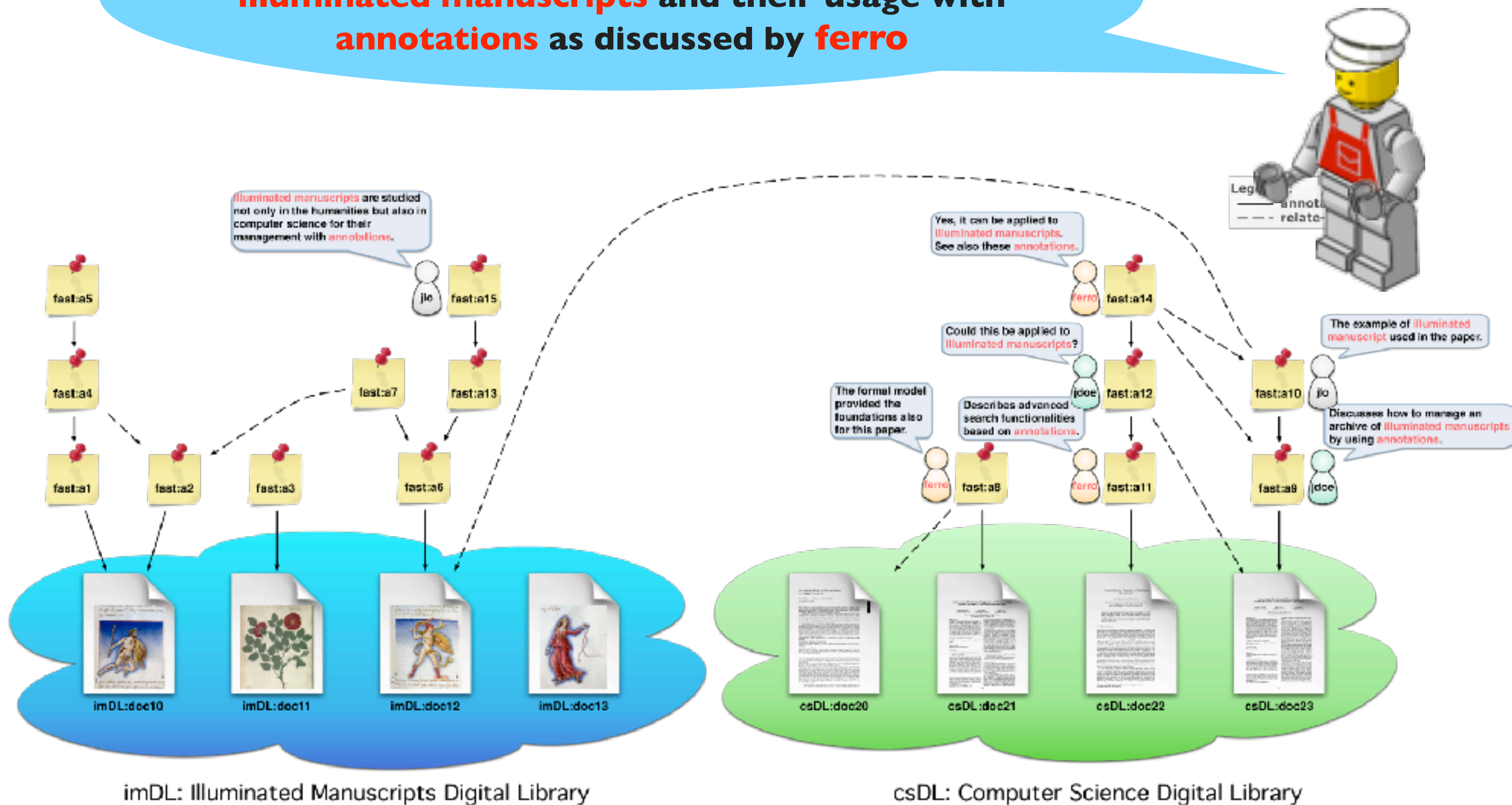
Annotation Model

Legenda
 annotate link
 relate-to link



Search by Using Annotations

I would like to find information about **illuminated manuscripts** and their usage with **annotations** as discussed by **ferro**

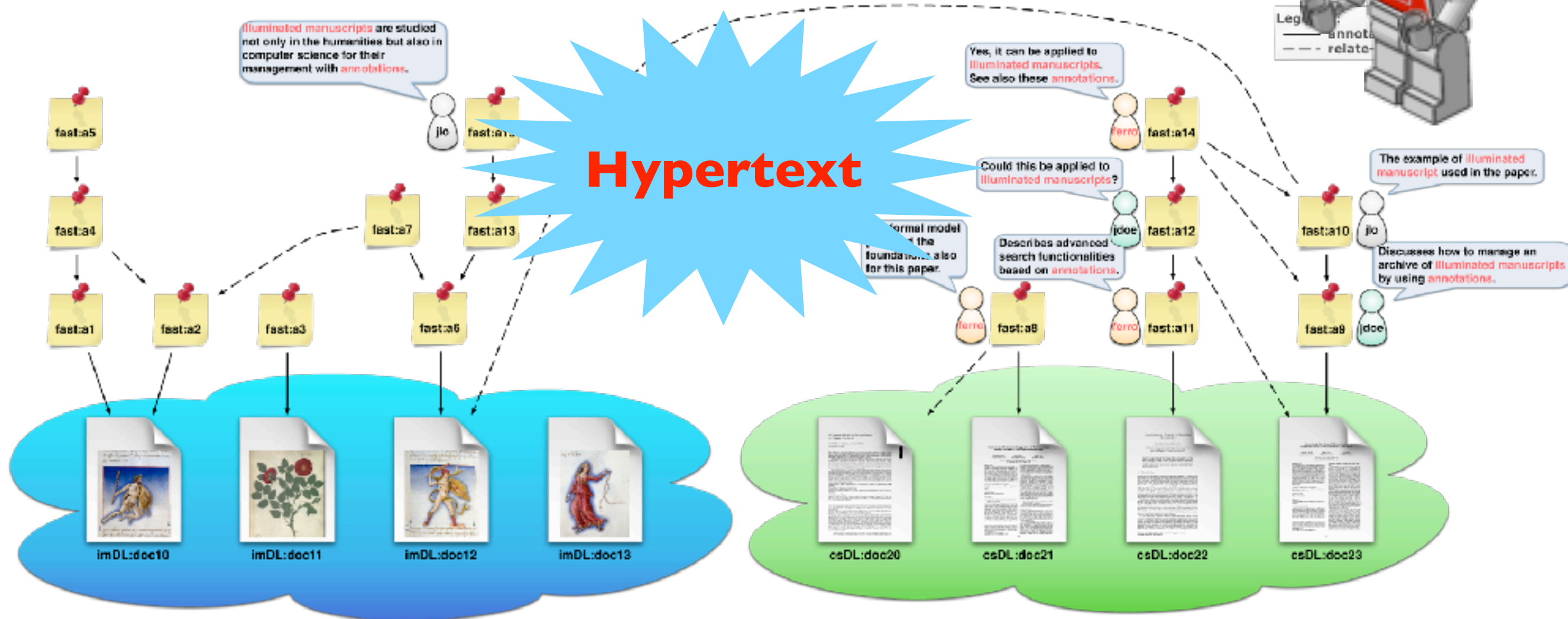


Search by Using Annotations

I would like to find information about
illuminated manuscripts and their usage with
annotations as discussed by **ferro**

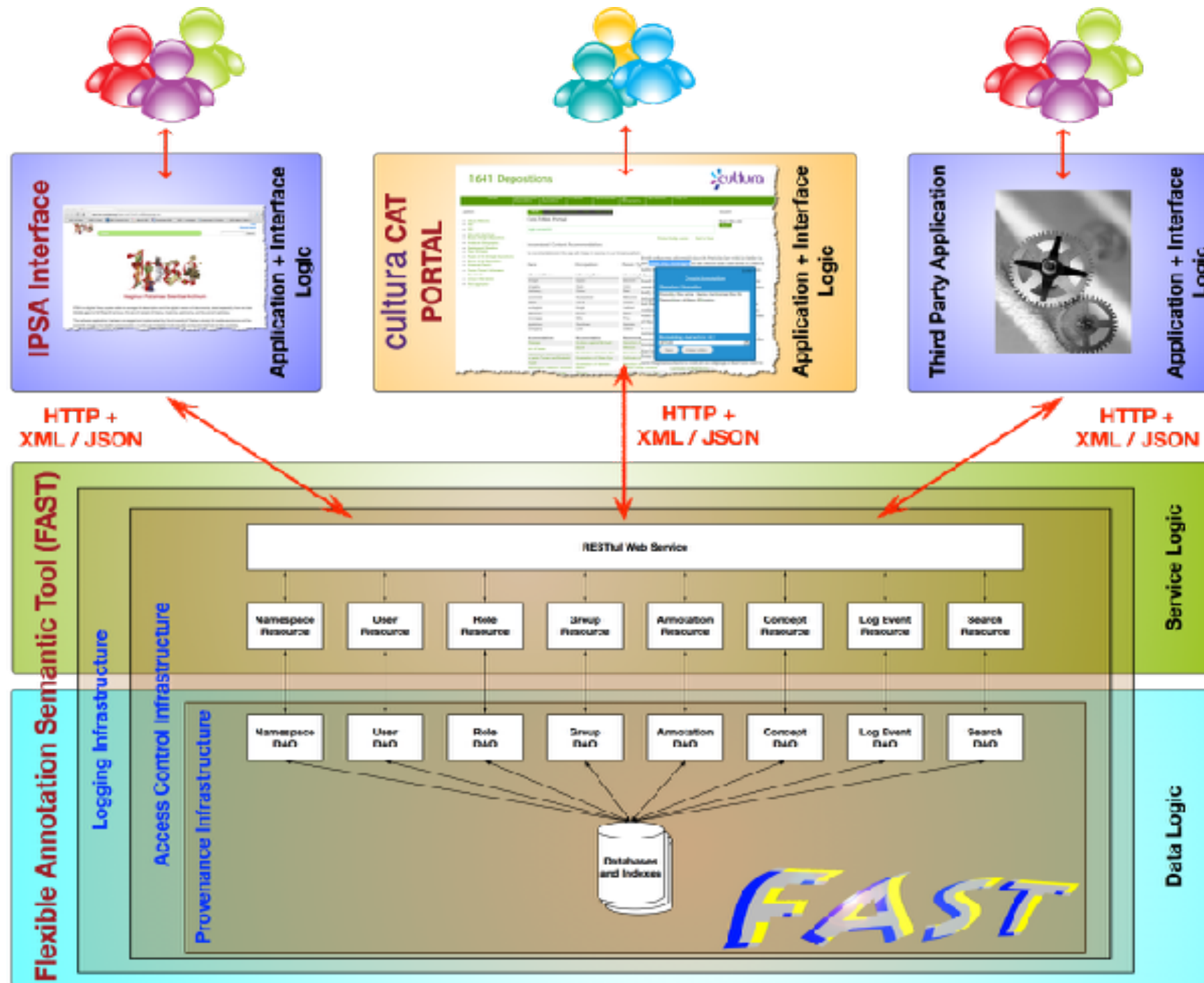
Best Match

Exact Match



imDL: Illuminated Manuscripts Digital Library

csDL: Computer Science Digital Library

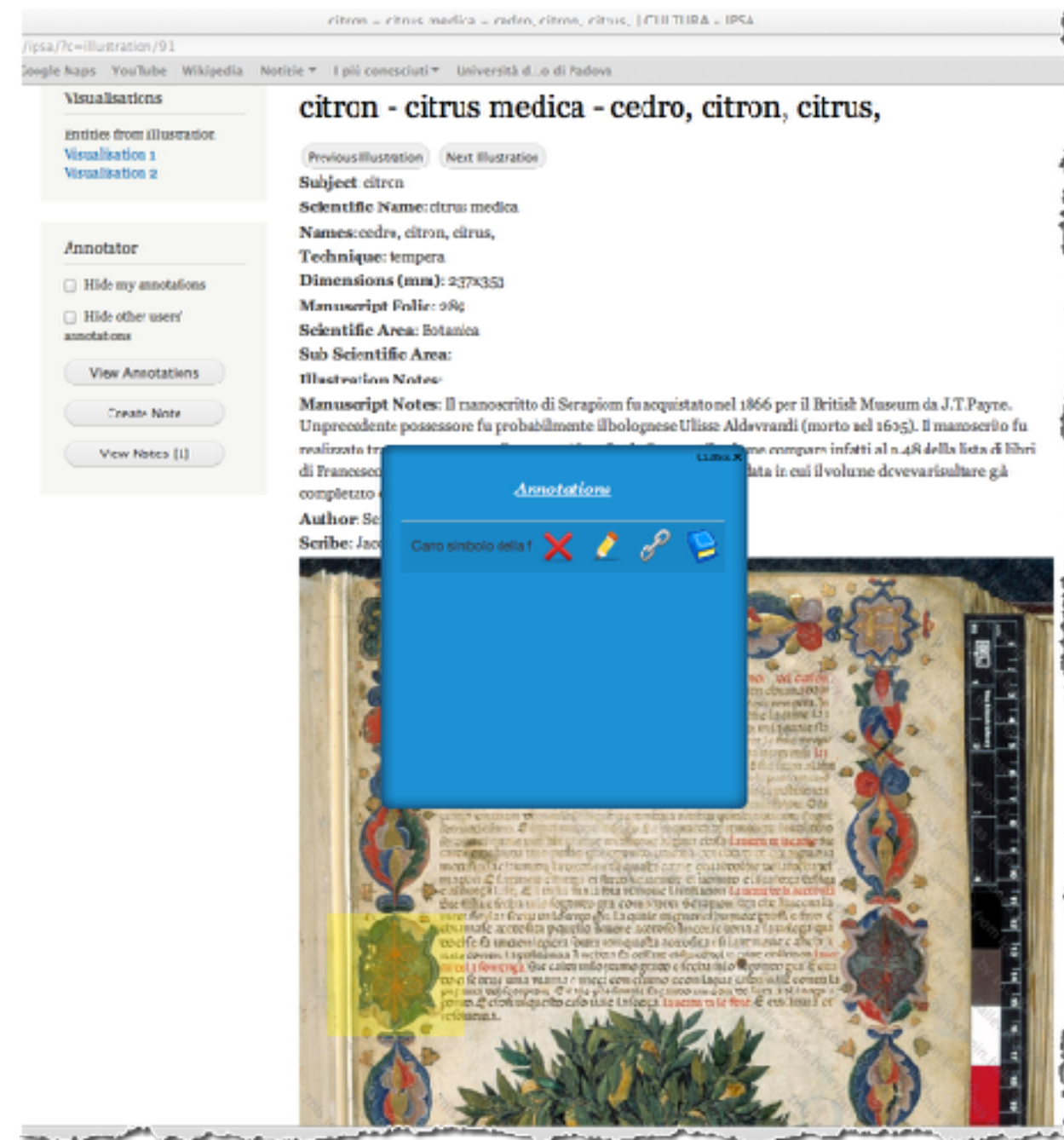


○ The CULTURA project

- innovative environment for users with a range of different expertise
- users can collaboratively explore, interrogate and interpret complex and diverse digital cultural heritage collections

○ Use cases

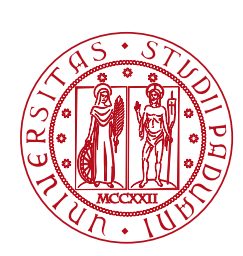
- IPSA: a digital archive of illuminated manuscripts produced in northern Italy during the 14th and 15th centuries
- The 1641 Depositions: the documents contain witness testimonies from men and women from all over Ireland and report on the rebellion of October 1641





References

- M. Agosti, N. Ferro. A Formal Model of Annotations of Digital Content. ACM Transactions on Information Systems (TOIS), 2008, 26(1):3:1-3:57
- M. Agosti, G. Bonfiglio-Dosio, N. Ferro. A Historical and Contemporary Study on Annotations to Derive Key Features for Systems Design. International Journal on Digital Libraries, 2007, 8(1):1-19
- M. Agosti, N. Ferro. Annotations as Context for Searching Documents. In: F. Crestani, I. Ruthven (Eds). Proc. of CoLIS 5, LNCS 3507, Springer, Heidelberg, Germany, 2005, 155-170
- N. Ferro. Annotation Search: The FAST Way. In: M. Agosti, J. Borbinha, S. Kapidakis, C. Papatheodorou, G. Tsakonas (Eds). Proc. 13th ECDL, LNCS 5714, Springer, Heidelberg, Germany, 2009, 15-26
- M. Agosti, O. Conlan, N. Ferro, C. Hampson, G. Munnely, G. Interacting with Digital Cultural Heritage Collections via Annotations: The CULTURA Approach. In S. Marinai, K. Marriot (Eds). Proc. 13th ACM DocEng, ACM Press, New York, USA, 2013, 13-22



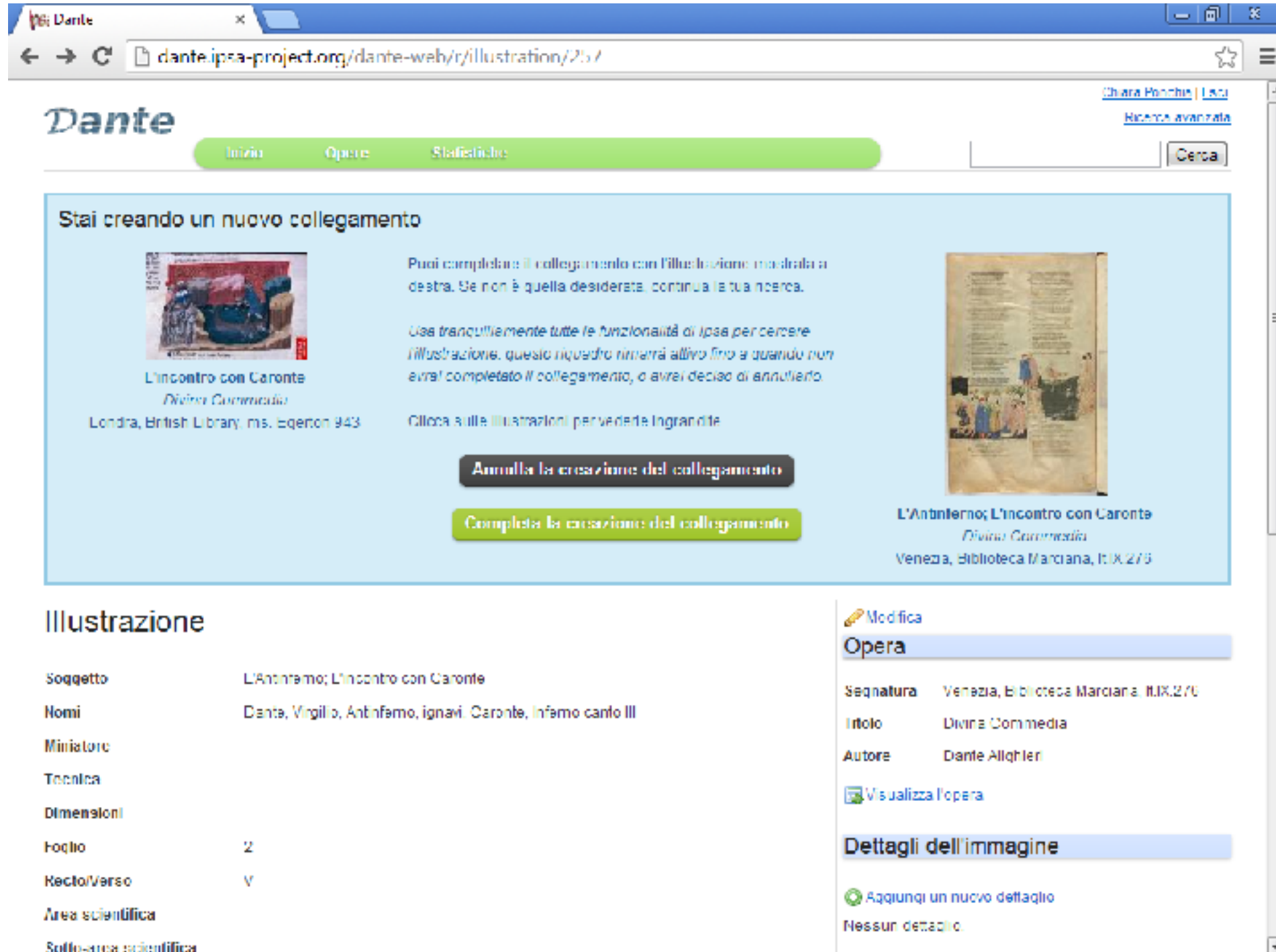
Digital Archives. IPSA and Dante

- Digital archive of illuminated manuscripts
- Result of a joint research project
 - Ex-Department of History of Visual Arts and Music
 - Department of Information Engineering
- Initially designed for a specialist public of scholars and researchers
- 56 manuscripts and more than 3000 images



DANTE.

A web-application for the History of Art



The screenshot shows a web browser window with the URL `dante.ipisa-project.org/dante-web/r/illustration/257`. The page title is "Dante". The navigation bar includes links for "Inizio", "Opere", and "Statistiche". A search bar is located on the right with the text "Cerca".

The main content area is titled "Stai creando un nuovo collegamento" (You are creating a new link). It features two illustrations side-by-side. The left illustration is titled "L'incontro con Caronte" (The encounter with Charon) and is identified as "Divina Commedia, Londra, British Library, ms. Egerton 943". The right illustration is titled "L'Antinferno; L'incontro con Caronte" (The Ante-Inferno; The encounter with Charon) and is identified as "Divina Commedia, Venezia, Biblioteca Marciana, It.IX.2/6".

Between the illustrations, there is a text block explaining the linking process: "Puoi completare il collegamento con l'illustrazione mostrata a destra. Se non è quella desiderata, continua la tua ricerca. Usa tranquillamente tutte le funzionalità di ipisa per cercare l'illustrazione: questo riquadro rimarrà attivo fino a quando non avrai completato il collegamento, o avrai deciso di annullarlo. Clicca sulle illustrazioni per vederle ingrandite." Below this text are two buttons: "Annulla la creazione del collegamento" (Cancel the creation of the link) and "Completa la creazione del collegamento" (Complete the creation of the link).

Below the main content area, there is a section titled "Illustrazione" (Illustration) with a table of metadata:

| | |
|------------------------|---|
| Soggetto | L'Antinferno; L'incontro con Caronte |
| Nomi | Dante, Virgilio, Antinferno, ignavi, Caronte, Inferno canto III |
| Miniature | |
| Tecnica | |
| Dimensioni | |
| Foglio | 2 |
| Recto/Verso | V |
| Area scientifica | |
| Sotto-area scientifica | |

On the right side of the page, there is a "Modifica" (Edit) button and a section titled "Opera" (Work) with the following details:

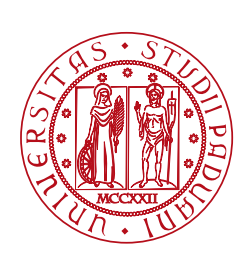
- Segnatura: Venezia, Biblioteca Marciana, It.IX.2/6
- Titolo: Divina Commedia
- Autore: Dante Alighieri

Below the "Opera" section, there is a "Visualizza l'opera" (View the work) button and a section titled "Dettagli dell'immagine" (Image details) with the text "Aggiungi un nuovo dettaglio" (Add a new detail) and "Nessun dettaglio" (No detail).



References

- Agosti, M., Benfante, L., Orio, N. IPSA: A Digital Archive of Herbals to Support Scientific Research. In: Proc. of the International Conference on Asian Digital Libraries (ICADL), Kuala Lumpur, MA, 2003, 253-264.
- Agosti, M., Ferro, F., Orio, N.: Annotating illuminated manuscripts: an effective tool for research and education. In: Marlino, M., Summer, T., Shipman, F., editors, Proc. 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2005), ACM Press, New York, USA, 2005, 121-130
- Agosti, M., Mariani Canova, G., Orio, N., Ponchia, C.: Methods of personalising a collection of images using linking annotations. In: Proceedings of the First Workshop on Personalised Multilingual Hypertext Retrieval (PMHR 2011), ACM, New York, NY, USA, 2011, 10-17.
- Ponchia, C. Dante. A web-application for the History of Art. In: Agosti, M., Tomasi, F., editors, Collaborative Research Practices and Shared Infrastructures for Humanities Computing, Padova 2014, 219-228.



CULTURA



- STREP project active from February 2011 to January 2014
- Engaging new user categories with digital cultural heritage collections
 - User centered-approach
- Exploring new ways to experience digital cultural heritage
 - Adaptivity
 - Interactive environment



Collections



- IPSA: a digital archive of illuminated manuscripts produced in northern Italy during the 14th and 15th centuries
- The 1641 Depositions: the documents contain witness testimonies from men and women from all over Ireland and report on the rebellion of October 1641

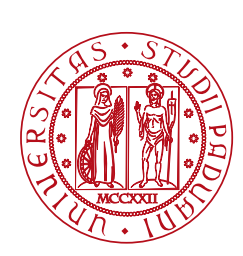


- Involving different user categories from the very beginning of the project
- Continuous Evaluation:
 - November and December 2011: baseline evaluations with professional researchers and students
 - April 2012: IPSA trials with postgraduate students
 - May 2012: interaction with professional researchers
 - September 2012: IPSA@CULTURA
 - December 2012 and January 2013: evaluation of IPSA@CULTURA with a new cohort of students and a new user category (Salvalarte Group)
 - October 2013 DILL Master evaluation
 - October 2013 High school students evaluation



References

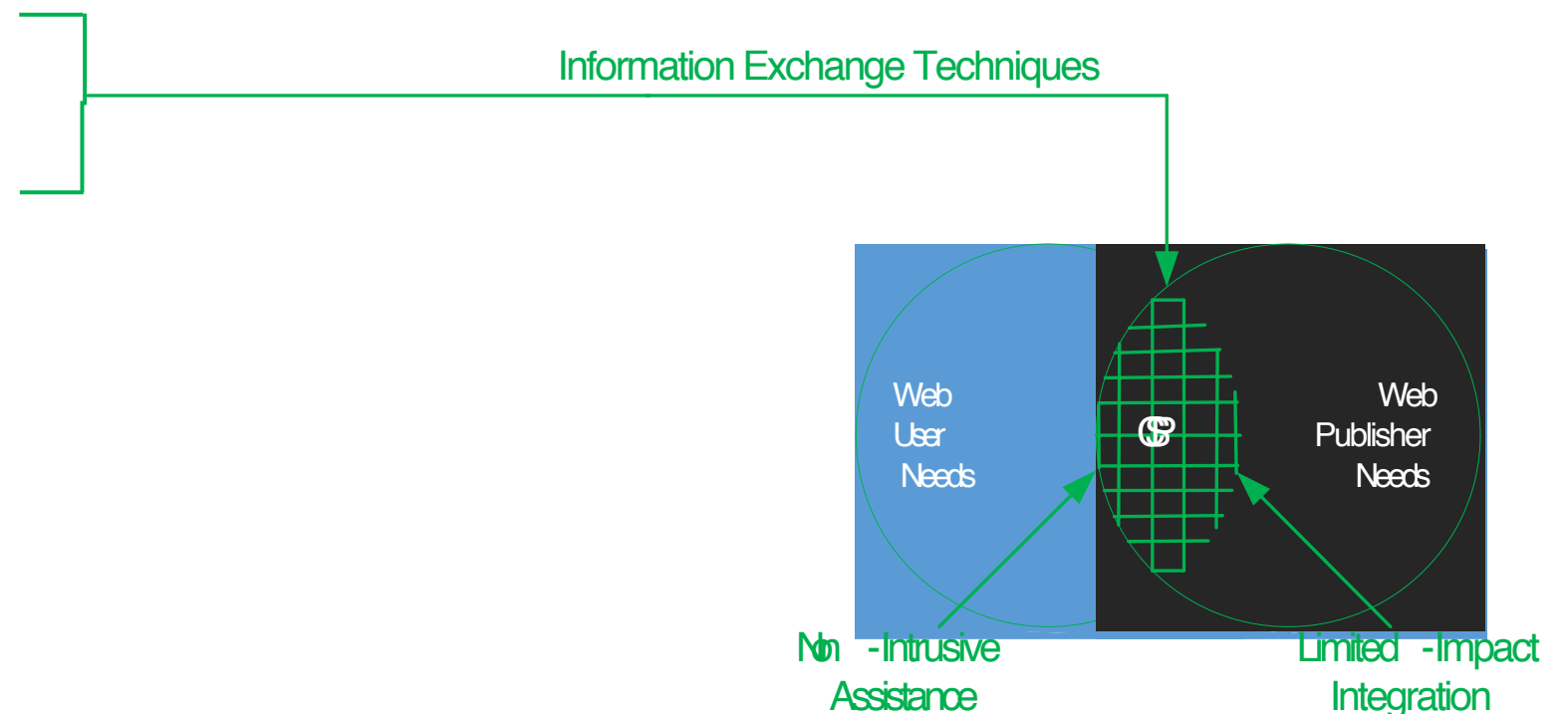
- Agosti, M., Ferro, N., Orio, N., Ponchia, C. CULTURA Outcomes for Improving the User's Engagement with Cultural Heritage Collections. In: "Procedia Computer Science", 38, 2014, Agosti, M., Catarci, T., Esposito, F. Proceedings of the 10th Italian Research Conference on Digital Libraries, IRCDL 2014, 34-39.
- Steiner, C.M., Agosti, M., Sweetnam, M.S., Hillemann, C., Orio, N., Ponchia, C., Hampson, C., Munnelly, G., Nussbaumer, A., Albert, D., Conlan, O. Evaluating a digital humanities research environment: the CULTURA approach, "International Journal on Digital Libraries", 15, 2014, 53-70.
- Agosti, M., Orio, N., Ponchia, C. Guided Tours Across a Collection of Historical Digital Images. In: Proceedings of the Third AIUCD Annual Conference on Humanities and Their Methods in the Digital Ecosystem (AIUCD 2014), New York 2015, 1-6.
- Agosti, M., Orio, N. The CULTURA project: CULTivating Understanding and Research through Adaptivity. In: Proceedings of the 7th Italian Research Conference on Digital Libraries, IRCDL 2011, Springer-Verlag, Berlin Heidelberg, 111-114.
- Agosti, M., Conlan, O., Ferro, N., Hampson, C., Munnelly, G., Ponchia, C., Silvello, G. Enriching Digital Cultural Heritage Collections via Annotations: The CULTURA approach. In: Greco, S., Picariello, A., Twentysecond Italian Symposium on Advanced Database Systems, SEBD 2014, 319-326.



Personalisation

Motivation

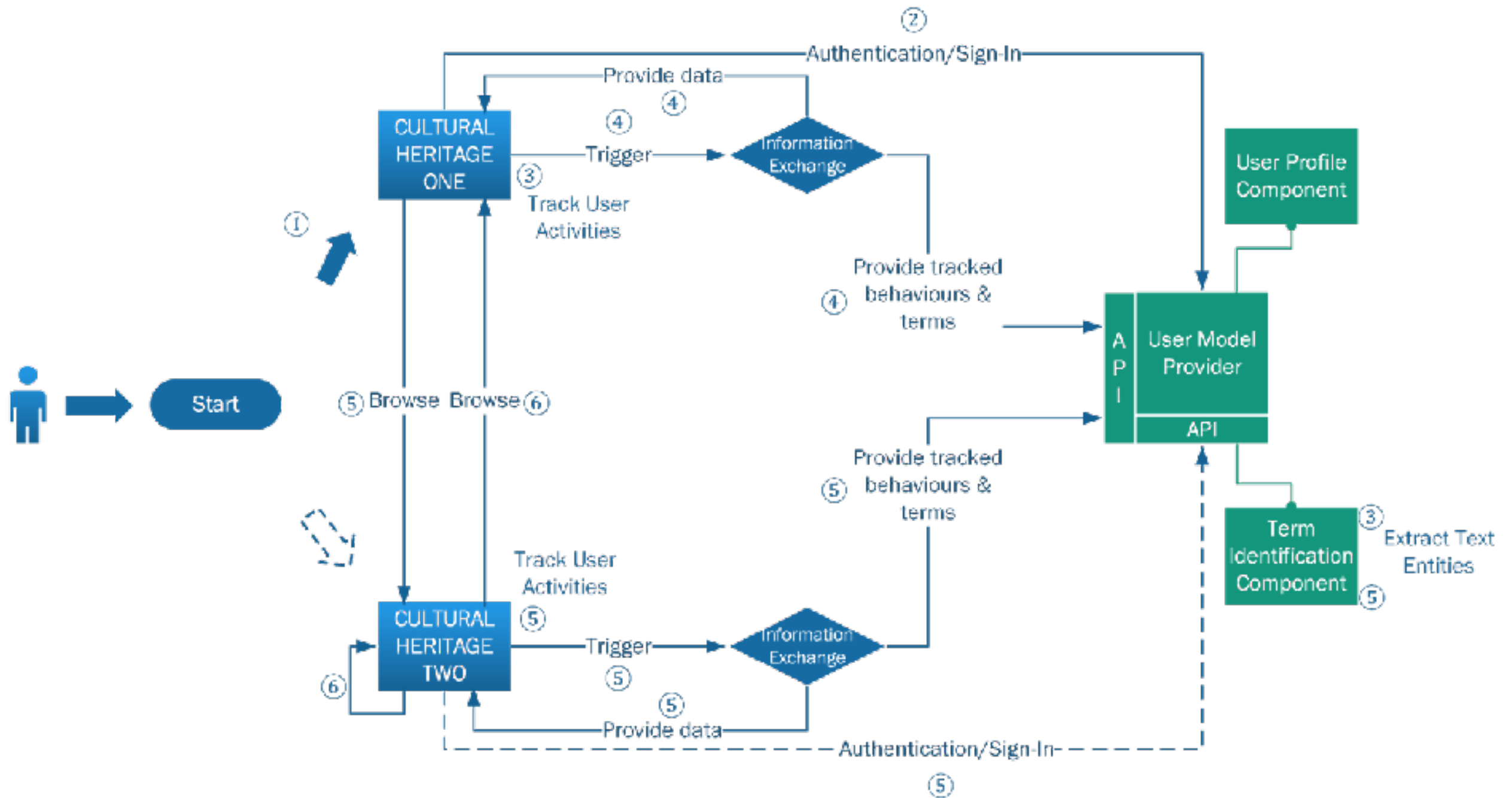
- Lack of a seamless browsing experience for users across (cultural) websites
- Most of personalisation approaches unable to assist users in information needs that span different (cultural) websites
- Need for a consistent Cross-Site Personalisation (CSP) support mechanism that provides:
 - Valuable exchange of information between a target website and the user modelling service
 - Users' privacy needs
- Use case:
 - The CULTURA project



Cross-Site Approach

- Complementary approach to the CULTURA Virtual Research Environment's demonstrator

- Designed as an open domain approach



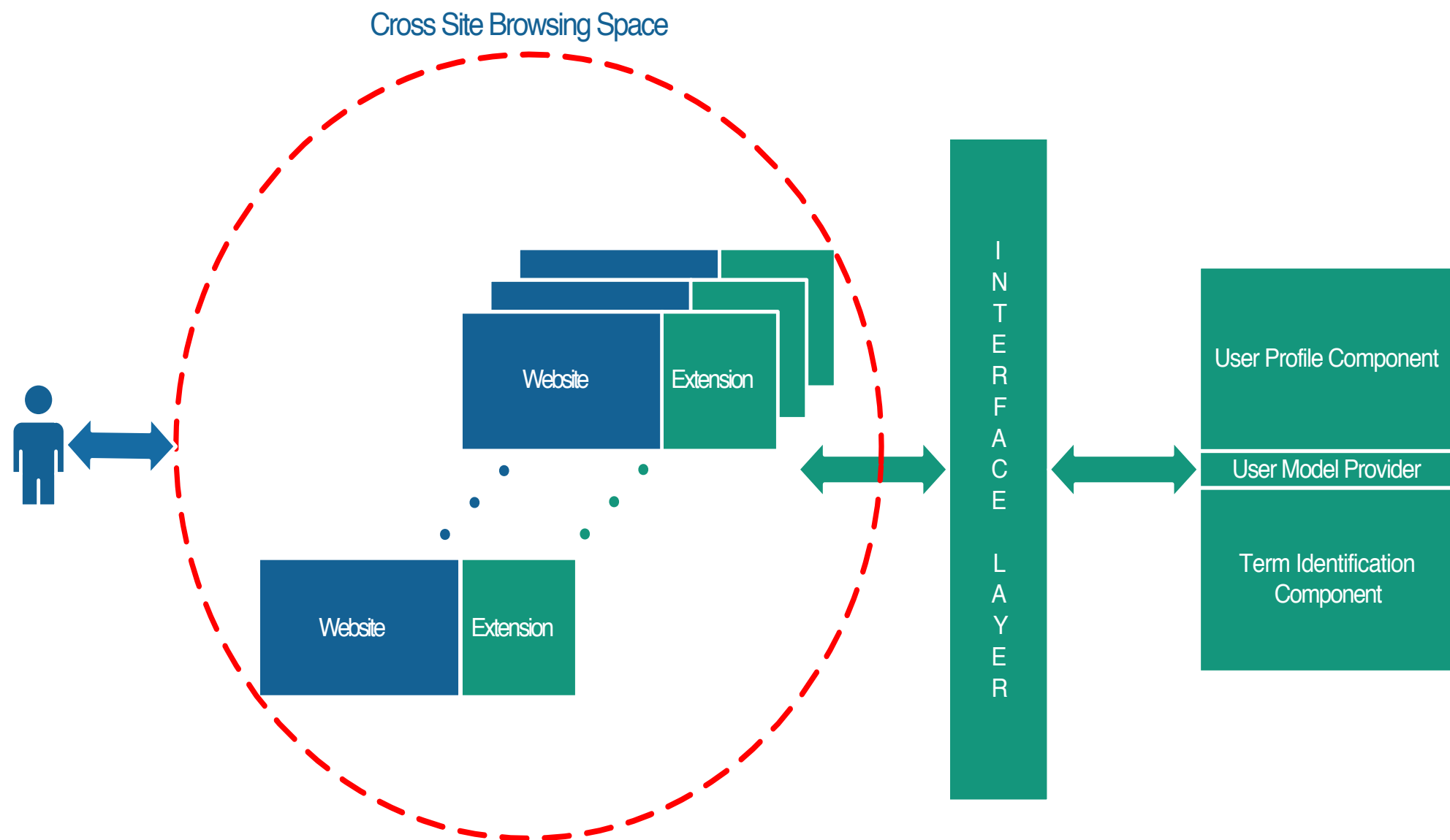
Architectural Design

Cross-Site Browsing Space

Web space introduced by websites that integrate the cross-site service through module extensions

Interface Layer

Facilitate communication between websites and the service



User Profile

Store and aggregate users activities and content related text entities

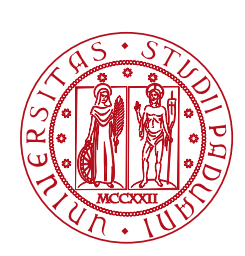
Term Identification

Extract relevant text entities from browsed content within websites



References

- Ahn, J.W., Brusilovsky, P., He, D., Grady, J., Li, Q.: Personalized web exploration with task models. In: Proceedings of the 17th International Conference on World Wide Web, pp. 1–10. ACM (2008)
- Berghel, H.: Cyberspace dealing with information overload. Commun. ACM 40(2), 19–24 (2000)
- Koidl, K., Conlan, O., Wade, V.: Cross-site personalization: assisting users in addressing information needs that span independently hosted websites. In: Proceedings of the 25th ACM Conference on Hypertext and Social Media, pp. 66–76. ACM (2014)
- Sweetnam, M., Siochru, M., Agosti, M., Manfioletti, M., Orio, N., Ponchia, C.: Stereotype or spectrum: designing for a user continuum. In: The Proceedings of the First Workshop on the Exploration, Navigation and Retrieval of Information in Cultural Heritage, ENRICH (2013)
- Sweetnam, M., Agosti, M., Orio, N., Ponchia, C., Steiner, C., Hillemann, E.-C., Ó Siochrú, M., Lawless, S.: User needs for enhanced engagement with cultural heritage collections. In: Zaphiris, P., Buchanan, G., Rasmussen, E., Loizides, F. (eds.) Theory and Practice of Digital Libraries, TPDL 2012. LNCS vol. 7489, pp. 64–75. Springer, Berlin (2012)



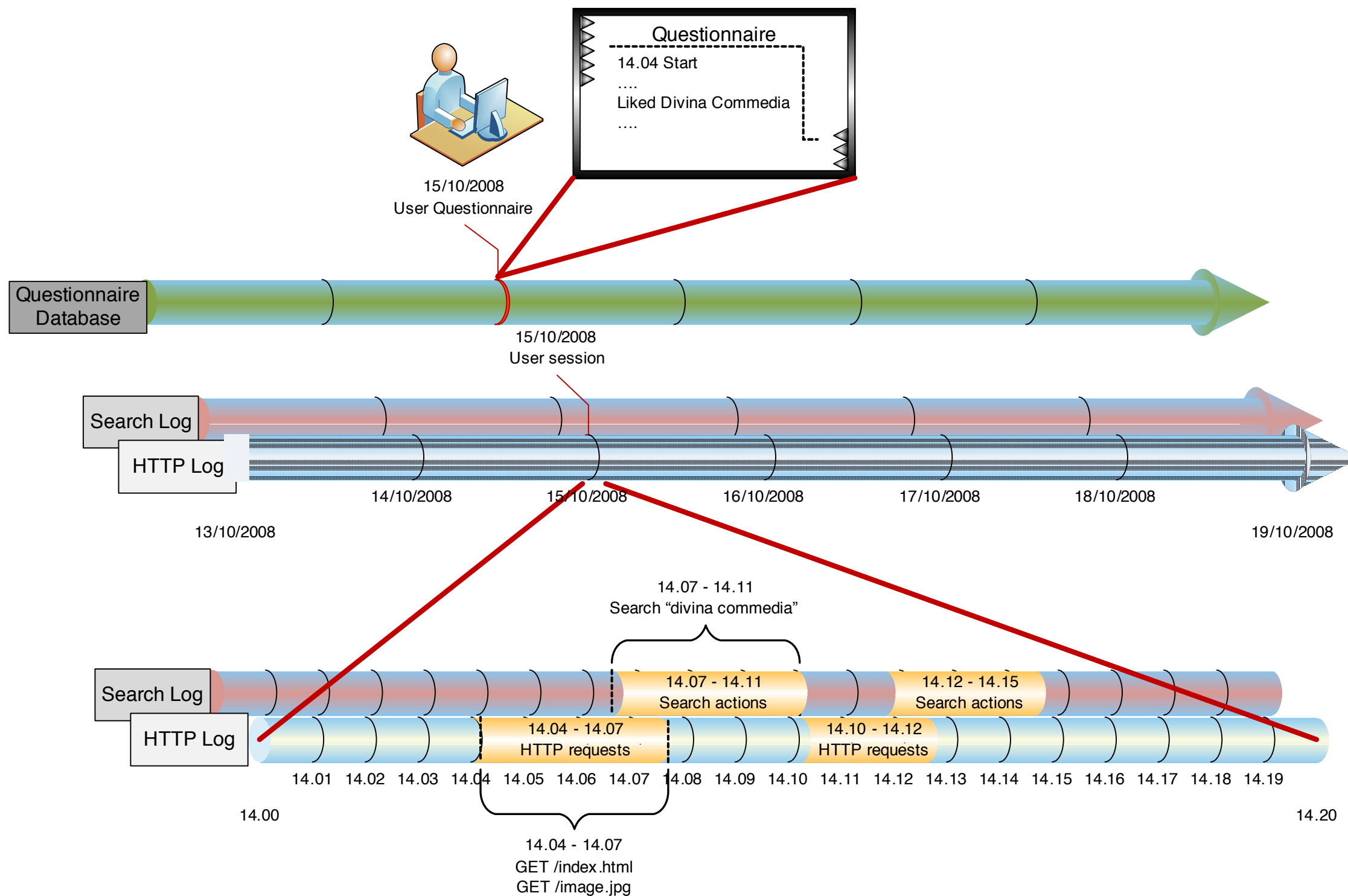
Log Analysis



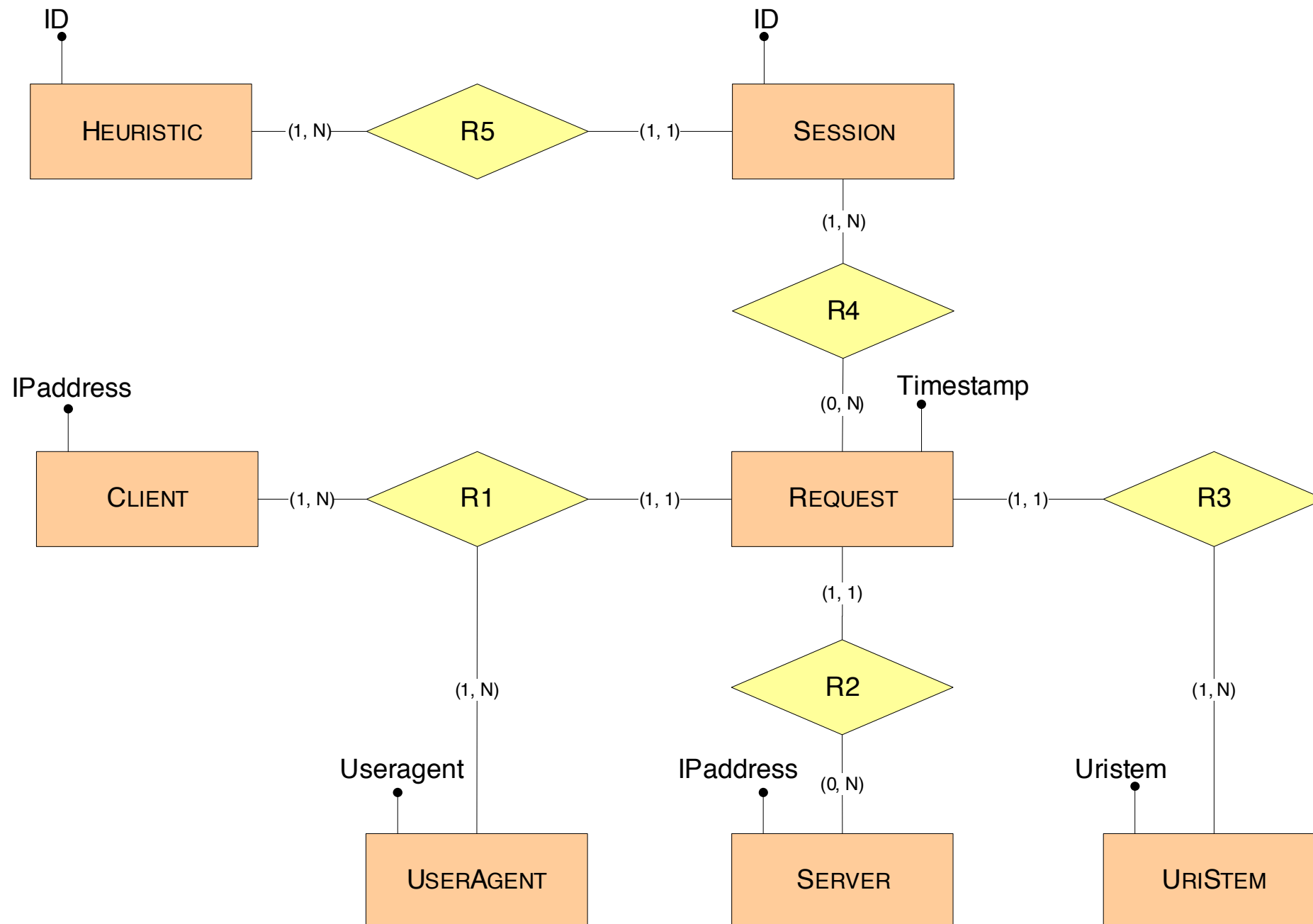
The European Library



- Case study: A portal born to offer access to the resources of 45 European National libraries.
- Analyse the data contained in the logs of their Web servers.
- Find a general methodology of HTTP Web Log analysis in order to
 - give advice about the security of the portal,
 - discover and study personalization issues.



TEL Logs





Multilingual LogCLEF

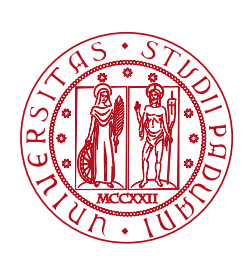


- Verifiability and repeatability of experiments
 - using the same period of time across different studies.
- From 2009 to 2011 LogCLEF Task
 - Language identification task
 - Query classification
 - Success of a query
 - Query refinement



References

- Maristella Agosti, Franco Crivellari, Giorgio Maria Di Nunzio: Web log analysis: a review of a decade of studies about information acquisition, inspection and interpretation of user interaction. *Data Min. Knowl. Discov.* 24(3): 663-696 (2012)
- Maristella Agosti, Franco Crivellari, Giorgio Maria Di Nunzio, Silvia Gabrielli: Understanding user requirements and preferences for a digital library Web portal. *Int. J. on Digital Libraries* 11(4): 225-238 (2010)
- Maristella Agosti, Franco Crivellari, Giorgio Maria Di Nunzio, Yannis E. Ioannidis, Elefterios Stamatogiannakis, Mei Li Triantafyllidi, Maria Vayanou: Searching and Browsing Digital Library Catalogues: A Combined Log Analysis for The European Library. *IRCDL 2009*: 120-135
- Maristella Agosti, Franco Crivellari, Giorgio Maria Di Nunzio: Evaluation of Digital Library Services Using Complementary Logs. *UIIR@SIGIR 2009*
- Giorgio Maria Di Nunzio, Johannes Leveling, Thomas Mandl: LogCLEF 2011 Multilingual Log File Analysis: Language Identification, Query Classification, and Success of a Query. *CLEF (Notebook Papers/Labs/Workshop) 2011*
- Giorgio Maria Di Nunzio, Johannes Leveling, Thomas Mandl: Multilingual Log Analysis: LogCLEF. *ECIR 2011*: 675-678
- Thomas Mandl, Giorgio Maria Di Nunzio, Julia Maria Schulz: LogCLEF 2010: the CLEF 2010 Multilingual Logfile Analysis Track Overview. *CLEF (Notebook Papers/LABs/Workshops) 2010*



User-System Interaction



Click Log Data



Query logs have proven to be a valuable and informative source of implicit user feedback:

- they can be easily collected by IR systems;
- they are available in real time;
- they represent personalized user preferences.

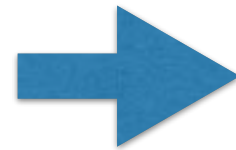
Query logs have proven to be a **valuable** alternative source of implicit user feedback.

- they can be used to **infer** user preferences.
- they are **anonymous**.
- they represent **personalized** user preferences.

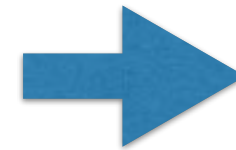


User Models

Different Types
of Users

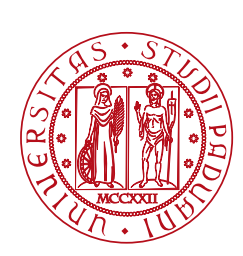


Different
Search



Different
Evaluation



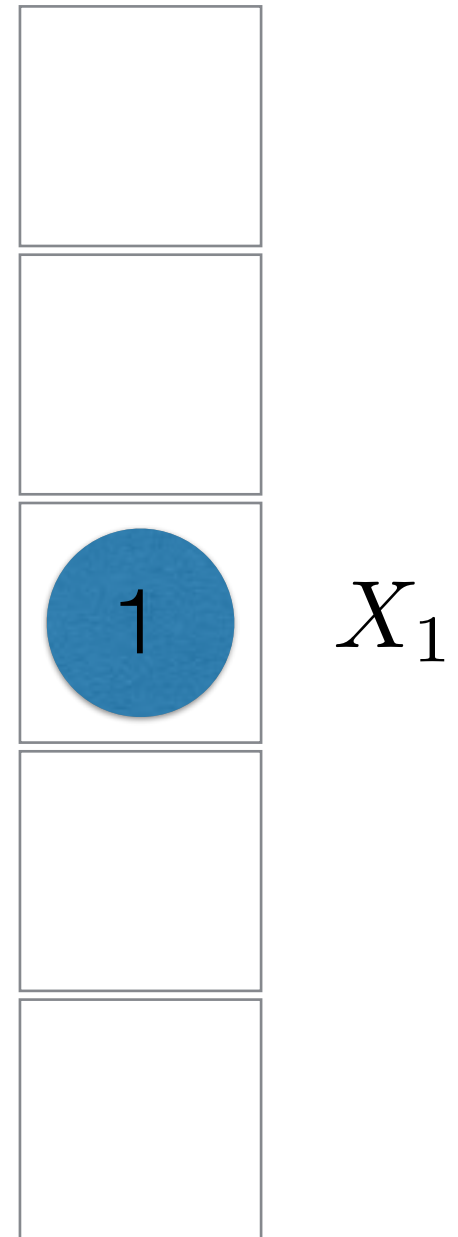


Markovian User Model

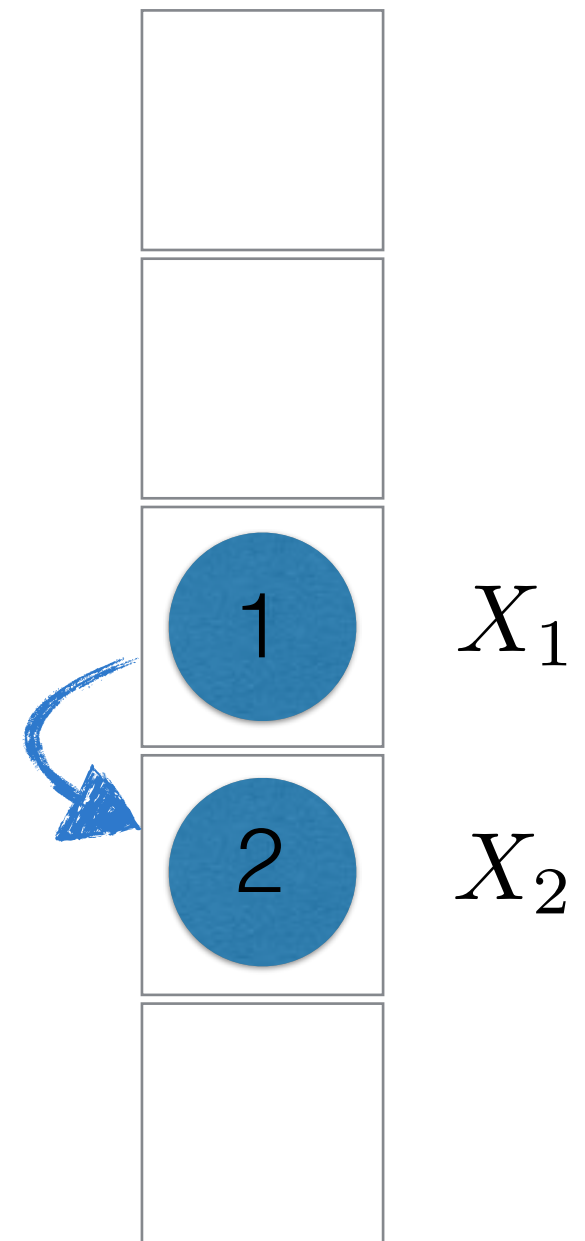
We assume that each user decides, independently from the random time spent in the first document, to move forward or backward to another document in the list.



We assume that each user decides, independently from the random time spent in the first document, to move forward or backward to another document in the list.



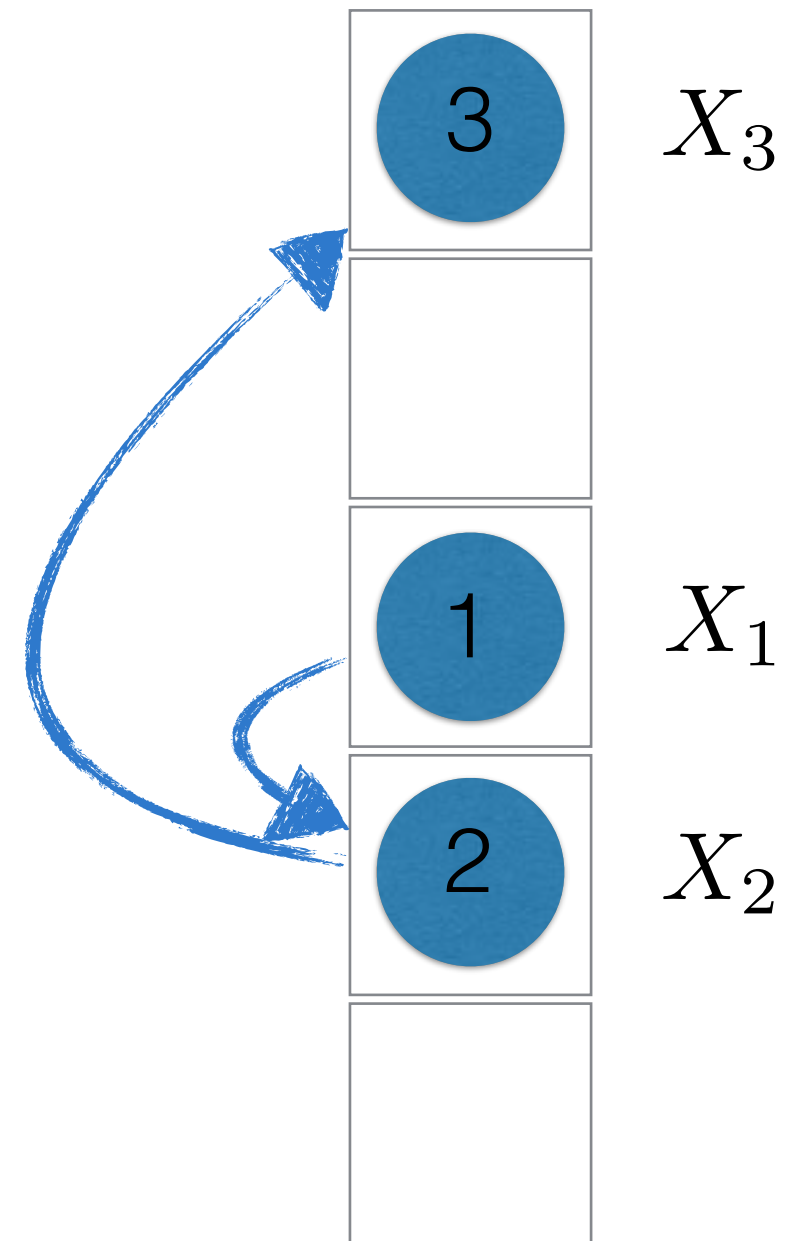
We assume that each user decides, independently from the random time spent in the first document, to move forward or backward to another document in the list.



We assume that each user decides, independently from the random time spent in the first document, to move forward or backward to another document in the list.

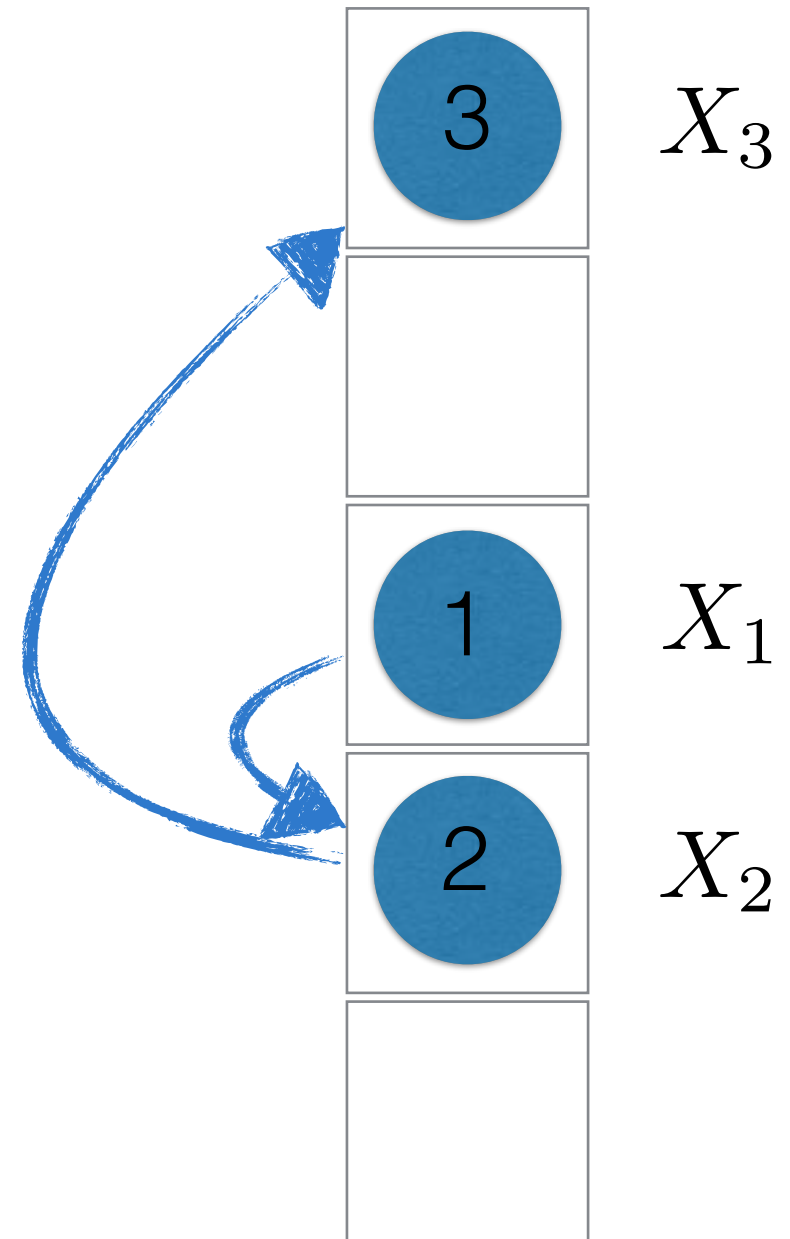
$$X_1, X_2, X_3, \dots \in \mathcal{R} = \{1, 2, \dots, R\}$$

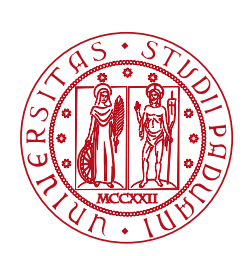
random sequence of document ranks visited by



22.6%

Backward Transitions





Application of the Markovian User Model

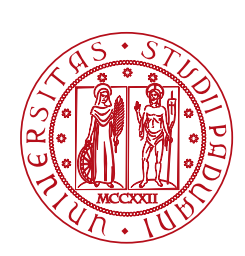


- New family of evaluation measures, called Markov Precision, which injects user models into precision
- Describe the user dynamic, which can be integrated in the Learning to Rank algorithm LambdaMart.



References

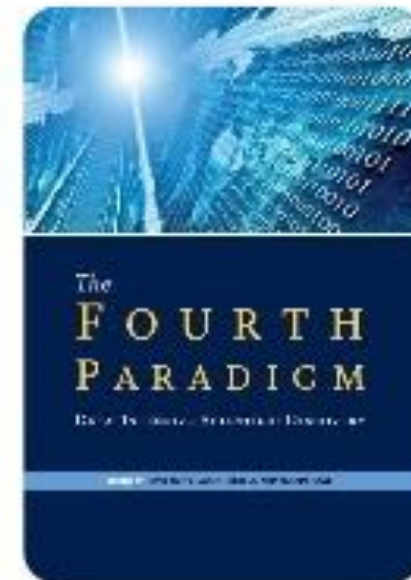
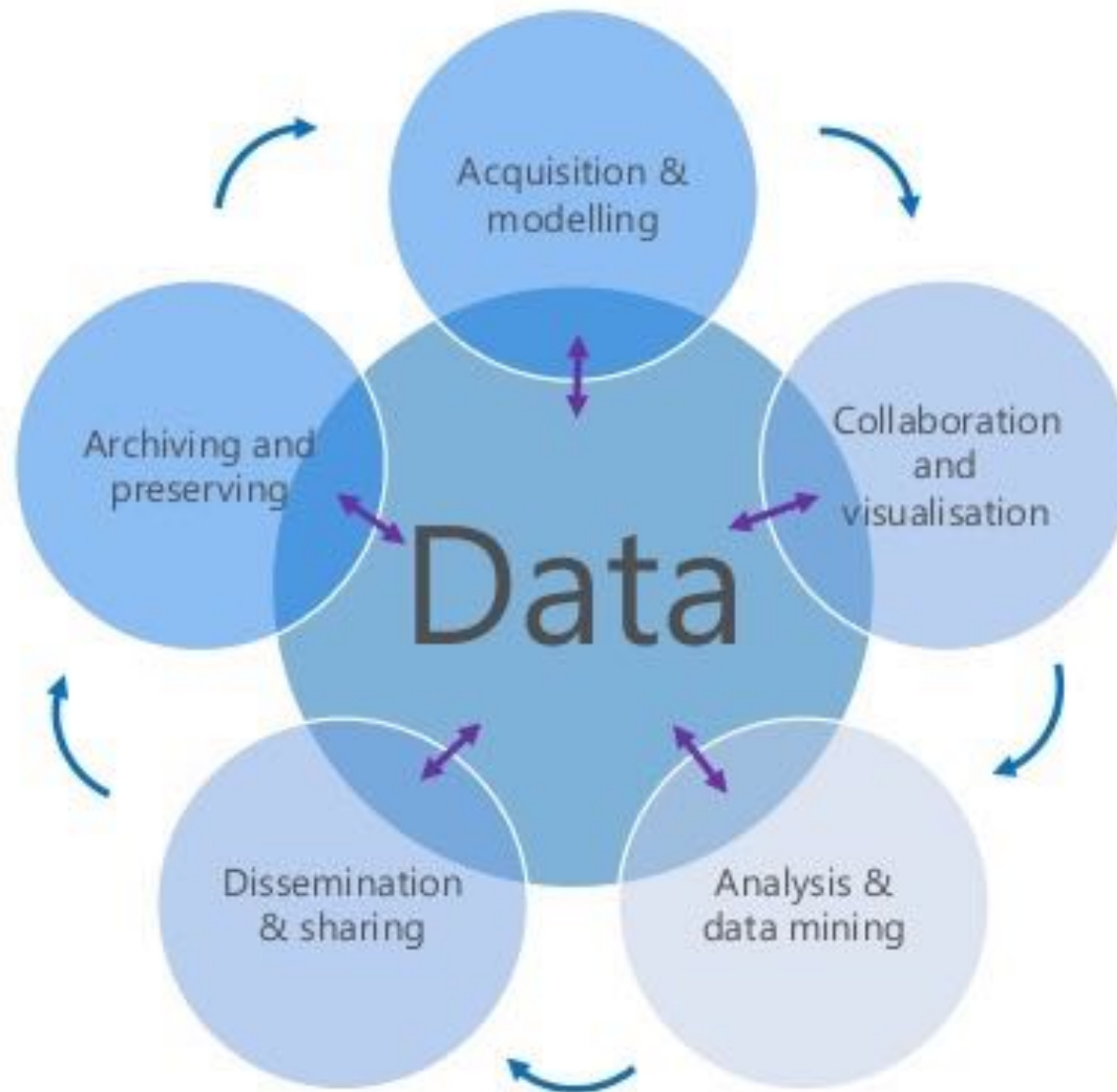
- Marco Ferrante, Nicola Ferro, and Maria Maistro. 2014. Injecting User Models and Time into Precision via Markov Chains. In Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '14). ACM, New York, NY, USA, 597-606.
- Nicola Ferro, Claudio Lucchese, Maria Maistro, and Raffaele Perego. 2017. On Including the User Dynamic in Learning to Rank. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). ACM, New York, NY, USA, 1041-1044.



Data Driven DL

Data-Driven Digital Libraries

Data-intensive Research



fourthparadigm.org





Reproducibility



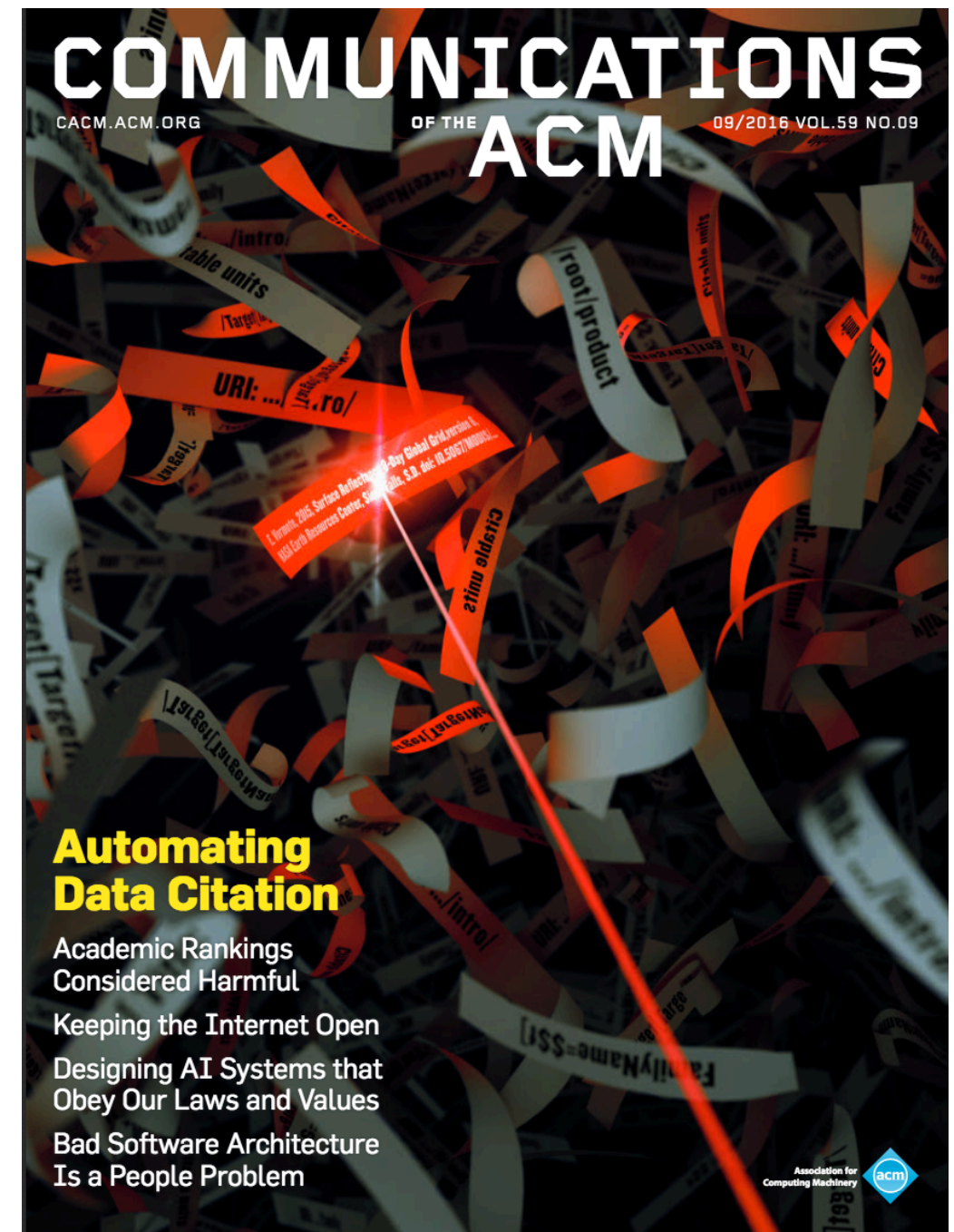
No research paper can ever be considered to be the final word, and the *replication* and corroboration of research results is key to the scientific process

[Nature, <http://www.nature.com/nature/focus/reproducibility/>]

An experimental result is not fully established unless it can be independently reproduced. Additional benefits ensue if the research artifacts are themselves made publically available so that any interested party may audit them. This also enables replication experiments to be performed.

[ACM Press, <https://www.acm.org/publications/policies/artifact-review-badging>]

- Principles and standards for data citation are unlikely to be used unless the process of extracting information is coupled with that of providing a citation for it.
- We need to automatically generate citations as the data is extracted.
- Data citation is a computational problem.



Buneman, Davidson, Frew:
Why data citation is a computational problem.
Commun. ACM 59(9): 50-57 (2016)



References

- G. Silvello (2018). Theory and Practice of Data Citation, Journal of the Association for Information Science and Technology (JASIST), 69 (1), pp. 6-20, 2018. DOI: 10.1002/asi.23917
- G. Silvello (2017). Learning to Cite Framework: How to Automatically Construct Citations for Hierarchical Data, Journal of the Association for Information Science and Technology (JASIST), 68 (6), pp. 1505–1524, June 2017. DOI: 10.1002/asi.23774
- S. B. Davidson, D. Deutch, T. Milo and G. Silvello (2017). A Model for Fine-Grained Data Citation. In Proc. of CIDR 2017, 8th Biennial Conference on Innovative Data Systems Research, 2017.
- S. B. Davidson, P. Buneman, D. Deutch, T. Milo and G. Silvello (2017). Data Citation: a Computational Challenge. In Proc. of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (PODS 2017), pages 1–4, ACM Press, New York, NY, USA, 2017.
- A. Alawini, L. Chen, S. B. Davidson, N. Portilho Da Silva and G. Silvello (2017). Automating data citation: the eagle-i experience. In Proc. of the 2017 ACM/IEEE Joint Conference on Digital Libraries, JCDL 2017, pages 169–178, IEEE Computer Society, 2017.
- G. Silvello and N. Ferro (2016). Data Citation is Coming. Introduction to the Special Issue on Data Citation, Bulletin of IEEE Technical Committee on Digital Libraries, Volume 12 Issue 1, pp. 1–5, May 2016.
- G. Silvello (2015). A Methodology for Citing Linked Open Data Subsets, D-Lib Magazine, 21(1/2). DOI: 10.1045/january2015-silvello
- P. Buneman and G. Silvello (2010). A Rule-Based Citation System for Structured and Evolving Datasets. Bulletin of the Technical Committee on Data Engineering Bulletin, September 2010, 3(3):33–41. IEEE Computer Society. ISSN 1053-1238.

Thanks

ANY
QUESTIONS?