# Information and Research Science connecting to Digital and Library Science

## Report on the 17th Italian Research Conference on Digital Libraries, IRCDL2021

Dennis Dosso
University of Padova
Padua, Italy
dennis.dosso@unipd.it

Stefano Ferilli
University of Bari
Bari, Italy
stefano.ferilli@uniba.it

Paolo Manghi
ISTI - Consiglio Nazionale
delle Ricerche
Pisa, Italy
manghi@isti.cnr.it

Antonella Poggi
"La Sapienza", University of Rome
Rome, Italy
antonella.poggi@uniroma1.it

Giuseppe Serra
University of Udine
Udine, Italy
giuseppe.serra@uniud.it

Gianmaria Silvello
University of Padova
Padua, Italy
gianmaria.silvello@unipd.it

## 1. INTRODUCTION

Since 2005 the Italian Research Conference on Digital Libraries is a yearly date for researchers on Digital Libraries and related topics, organized by the Italian Research Community. Over the years, IRCDL has become an essential national forum focused on digital libraries and associated technical, practical, and social issues. IRCDL encompasses the many meanings of the term *digital libraries*, including new forms of information institutions; operational information systems with all manner of digital content; new means of selecting, collecting, organizing, and distributing digital content; and theoretical models of information media, including document genres and electronic publishing.

Digital libraries may be viewed as a new form of information institution or an extension of the services libraries currently provide. Representatives from academia, government, industry, research communities, and research infrastructures, are usually invited to participate in this annual conference. The conference draws from a broad and multidisciplinary research area, including computer science, information science, librarianship, archival science and practice, museum studies and practice, technology, social sciences, cultural heritage and digital humanities.

This year the focus was bridging the broad field of Research and Information Science with the related field of Digital Libraries. Indeed, IRCDL2021 aimed at approaching "Digital libraries" by embracing the field at large, comprehending three critical areas of interest, that are *scholarly communication* (e.g. research data, research software, digital experiments, digital libraries), *e-science/computationally-intense research* (e.g. scientific workflows, Virtual Research Environments, reproducibility) and *library, archive and information science* (e.g. governance, policies, open access, open science). There were a total of 25 submissions, and 17 of these were accepted: 11 full papers, and 6 short papers. The overall aim of IRCDL is to help young researchers to move the first steps in the competitive research world. For this reason, weak papers are shepherded and a thorough feedback is provided to the authors before possibly leading to acceptance.

The conference has been organized in Padua (Italy) at the Department of Information Engineering of the University of Padua. 145 participants attended the conference, coming coming for the most part from Italy, and also from Switzerland, Sweden, Denmark, Russia, Brasil, Germany and the UK. Participants are mainly from academia, but also culture heritage related industry and institutions (e.g., libraries and archives). Due to the COVID-19 emergency, the conference has been entirely online on 18 and 19 February 2021. The proceedings are published in the CEUR-WS Vol-2816 `http://ceur-ws.org/Vol-2816/`. The papers are available in gold open access, and all the videos of the presentations are available on YouTube and accessible via the conference Website: `http://ircdl2021.dei.unipd.it/`

## 2. CONFERENCE CONTRIBUTIONS

All the contributions were peer-reviewed by three of the twenty-two members of the Program Committee, and eighteen were accepted, out of which seven were short papers. IRCDL 2021 [8] featured one invited speaker and six thematic sessions devoted to the presentation of the accepted research papers.

### Invited talk: The swings and roundabouts of a decade of fun and games with Research Objects

Prof. Carole Goble's keynote[1] focused on the concept of Research Objects[2][4]. We are transitioning toward the fourth paradigm of science, a concept based on the idea that computational science constitutes a new set of methods beyond empiricism, theory, and simulation. It requires the integration between tools, technologies, and platforms in shared methodologies and processes, also designing methods for researchers to share their data. It allows the integration of these tools and technologies, creating new research opportunities not available before [15]. In this paradigm research outcomes are more than publications and data, but include software, models, workflows, SOPs, lab protocols, etc. All these products of science should be considered as first class citizens of scholarship, and be treated following the FAIR and Reproducible principles (FAIR+R). The Research Object is a general framework, where research outcomes are related and bundled together, to form a single reusable and citable object in support of reproducibility. RO's metadata describes the attribution, reuse, relational properties of the bundled objects, to ensure their FAIR+Rness. These resources thus become shareable, citable, exchangeable across different platforms, present versioning and snapshots, can be identified with tools such as PIDs and can be registered and deposited on their own in services, thus to be later unpackaged, accessed, activated, and reproduced if appropriate.

### Data and Platforms

This section presented new platform to manage and share data, and shed light on important challenges faced by already existing and widely-used data infrastructures. Biasini et al. [7] presented FullBrain[3], a social e-learning platform where students can share and track their knowledge, helping each other in their learning process. Gargiulo et al. [13] noted

that the FAIR RDM initiatives in Italy are still based on communities of practice, thus investigate the perception of some leader initiative with respect to good practices, challenges and the strategic vision. Baglioni et al. [3] highlighted the "service misuses" suffered by the ORCID infrastructure, that put in jeopardy its very mission.

### Data Access and Monitoring

This section presented new systems and techniques based on data and how its correct use can have meaningful impacts on different aspects of people's everyday life, ranging from learning to security.

Zanichelli et al. [21] noted that digital games could be a valuable tool for enhancing learning and aim to understand how to improve the assessment of learning obtained through them. Avanzi et al. [2] presented *NoBis*, a new service for monitoring the crowding of indoor spaces during the COVID-19 era. Spadi et al. [19] presented the first results of the project RESTORE, whose main goal is the recovery, integration, and accessibility of data and digital objects produced in the last twenty years by the partners of the project, thus building a knowledge base on the history of the town of Prato (Tuscany, Italy).

### Information and Knowledge

This section presented new applications and measures to obtain and evaluate information. Irrera and Silvello [16] addressed the problem of finding context information for news articles by extracting entities and relationships from documents. The idea is to provide context to documents and improve learning to rank methods' effectiveness. Lancioni et al. [17] use Generative Adversarial Networks to predict keyphrases, i.e., short text sequences that convey the main semantic meaning of a document. Ferrante et al. [12] extend the AWARE measure with a set of supervised methods, by using TREC collections.

### Character Recognition

This section presented new approaches to address problems connected to the automatic recognition of handwritten text to uncover texts still unreadable and provide new and meaningful tools to the experts. Fante and Di Nunzio [10] proposed a mixed qualitative-quantitative approach to OCR error detection and correction to develop a methodology for compiling historical corpora. Heil et al. [14] introduced Handwritten Text Recognition and Document Image Analysis approaches to address the challenges inherent to the original drafts of Astring

---

[1] https://www.slideshare.net/carolegoble/the-swings-and-roundabouts-of-a-decade-of-fun-and-games-with-research-objects-242974091
[2] http://researchobject.org/
[3] https://fullbrain.org/

Lindren, a Swedish author whose original drafts and manuscripts are edited in the Melin system of short-hands, as of today considered undecipherable. Marinai et al. [18] described a system for the location of simple graphemes in medieval manuscripts based on the Mask R-CNN convolutional neural network. A first approach to provide paleographers with a tool to speed up and refine dating and determining the origin of manuscripts.

### Text Analysis

This section explores new potential applications connected to the analysis of documents, opening up new applications and research lines. The work proposed by Bernasconi et al. [6] explored the idea of navigating the semantic relationships among extracted entities to search a text corpus, and the evaluation carried with potential users has shown the feasibility and effectiveness of the approach. Ferilli [11] presented a paper which deals with architectural floorplans, proposing an approach based on formal representation and reasoning for their understanding and interpretation, opening up a viable and promising new line of research. Dosso and Silvello [9] proposed a new problem, called Data Credit Distribution, to annotate data in a database with a value, called credit, representing the impact and importance of this data in the scientific domain.

### DL Services

Bernasconi et al. [5] presented the Knowledge Graph they developed within the ERC NOTAE project[4] and showed how navigating such graph can help researchers at finding non trivial implications in data, providing them support in investigating graphic symbols dating back to the Roman State and Post-Roman Kingdoms. Svarre [20] studied labels used by Danish academic library websites that can support user interactions with library websites and their related content, shedding more light on the characteristics and variety of labels used across libraries. Almeida et al. [1] are developing the ROSSIO Thesaurus, whose objective is to support content discovery and management in the ROSSIO Portuguese research infrastructure for arts and humanities, and describe its modeling process and integration in the infrastructure, together with its publication as LOD.

## 3. CONCLUSION AND PROSPECT

The world of DLs is going through profound transformations, and the research activities and results presented at IRCDL2021 gives a clear indication of how multifaceted Digital Library research is.

A panel of experts was organized to continue the dialogue started in the last years to discuss DL's next future. The purpose is to enlarge DL's scope and open IRCDL to the challenges to come.

The panel opened with an introduction by Gianmaria Silvello, who reported the panel's outcomes at IRCDL 2020 about the future of IRCDL. In 2020 the panel concluded that IRCDL has a central role for the Italian DL community and is a reference point for the young researchers, who, thanks to this conference, can move the first steps in research in a friendly and open environment. Nevertheless, there is the need to update the topics of the conference and to open-up to new emerging fields while maintaining the roots: i) IRCDL is an Italian conference with an international view; thus, it can be organized by international researchers and in English, but it has to maintain its national connotation; ii) cultural heritage, the humanities and library science are essential topics for IRCDL that have to keep a central role in the conference; iii) IRCDL maintains its focus on students and young researchers.

This year's panel discussed the challenges for the DL community in an evolving world and how IRCDL can adapt and be a catalyst for these changes. There were five panelists. The first panelist was Nicola Ferro (University of Padua), who discussed the role of Computer Science and Engineering research in DL. The second was Julian Bogdani (University of Rome - Sapienza), who reported his experience working in the PAThs project [5], that is an ERC interdisciplinary project involving archaeology and computer science. Stefania Gialdroni (University of Roma3) further discussed the intersections between cultural heritage and computing, by reporting her experiences in the field of legal history and depicting the objectives of her future work within the ERC MICOLL project. Luigi Siciliano (Free University of Bozen-Bolzano) dug on digital libraries' role in the ever-changing world of information access; he analyzed prominent vendors and software houses and how they impact the management and access to knowledge. Finally, Donato Malerba (University of Bari) presented the many contact points between Data Science and Digital Libraries, opening up to future collaborations between these two fields.

The next edition of IRCDL will be held in 2022 in Italy. IRCDL2021 consolidated the Italian community presence and widened international participation opening up to authors from all around Europe.

---

Our goal for IRCDL2022 is to further open up to international involvement to shape DL's future.

## Acknowledgments

## 4. REFERENCES

[1] ALMEIDA, B., FREIRE, N., AND MONTEIRO, D. The development of the ROSSIO thesaurus: Supporting content discovery and management in a research infrastructure. In Dosso et al. [8], pp. 138–146.

[2] AVANZI, M., CONIGLIO, R., CISOTTO, G., GIORDANI, M., AND FERRO, N. Nobis: A crowd monitoring service against COVID-19. In Dosso et al. [8], pp. 126–137.

[3] BAGLIONI, M., MANNOCCI, A., MANGHI, P., ATZORI, C., BARDI, A., AND BRUZZO, S. L. Reflections on the misuses of ORCID ids. In Dosso et al. [8], pp. 117–125.

[4] BECHHOFER, S., BUCHAN, I., DE ROURE, D., MISSIER, P., AINSWORTH, J., BHAGAT, J., COUCH, P., CRUICKSHANK, D., DELDERFIELD, M., DUNLOP, I., ET AL. Why linked data is not enough for scientists. In *Future Generation Computer Systems* [8], pp. 599–611.

[5] BERNASCONI, E., BOCCUZZI, M., CATARCI, T., CERIANI, M., GHIGNOLI, A., LEOTTA, F., MECELLA, M., MONTE, A., SIETIS, N., VENERUSO, S., AND ZIRAN, Z. Exploring the historical context of graphic symbols: the NOTAE knowledge graph and its visual interface. In Dosso et al. [8], pp. 147–154.

[6] BERNASCONI, E., CERIANI, M., AND MECELLA, M. Exploring a text corpus via a knowledge graph. In Dosso et al. [8], pp. 91–102.

[7] BIASINI, M., CARMIGNANI, V., FERRO, N., FILIANOS, P., MAISTRO, M., AND NUNZIO, G. M. D. Fullbrain: a social e-learning platform. In Dosso et al. [8], pp. 25–41.

[8] DOSSO, D., FERILLI, S., MANGHI, P., POGGI, A., SERRA, G., AND SILVELLO, G., Eds. *Proceedings of the 17th Italian Research Conference on Digital Libraries, Padua, Italy (virtual event due to the Covid-19 pandemic), February 18-19, 2021* (2021), vol. 2816 of *CEUR Workshop Proceedings*, CEUR-WS.org.

[9] DOSSO, D., AND SILVELLO, G. Data credit distribution through lineage. In Dosso et al. [8], pp. 155–161.

[10] FANTE, D. D., AND DI NUNZIO, G. M. A quantitative/qualitative approach to OCR error detection and correction in old newspapers for corpus-assisted discourse studies. In Dosso et al. [8], pp. 53–63.

[11] FERILLI, S. Toward automatic floor plan interpretation. In Dosso et al. [8], pp. 13–24.

[12] FERRANTE, M., FERRO, N., AND PIAZZON, L. s-aware: Using crowd judgements in supervised measure-based methods for IR evaluation. In Dosso et al. [8], pp. 162–168.

[13] GARGIULO, P., GALIMBERTI, P., TAMMARO, A. M., AND ZANE, A. FAIR RDM (research data management): Italian initiatives towards EOSC implementation. In Dosso et al. [8], pp. 42–52.

[14] HEIL, R., NAUWERCK, M., AND HAST, A. Shorthand secrets: Deciphering astrid lindgren's stenographed drafts with HTR methods. In Dosso et al. [8], pp. 169–177.

[15] HEY, T., TANSLEY, S., AND TOLLE, K. M., Eds. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, 2009.

[16] IRRERA, O., AND SILVELLO, G. Background linking: Joining entity linking with learning to rank models. In Dosso et al. [8], pp. 64–77.

[17] LANCIONI, G., MOHAMED, S. S., PORTELLI, B., SERRA, G., AND TASSO, C. Efficient keyphrase generation with gans. In Dosso et al. [8], pp. 1–12.

[18] MARINAI, S., POMARO, G., RAFFAELLI, C., AND SCANDIFFIO, F. Location of simple graphemes in mediaeval manuscripts based on mask R-CNN. In Dosso et al. [8], pp. 103–116.

[19] SPADI, A., SPINELLI, F., AND DEGL'INNOCENTI, E. RESTORE: smart access to digital heritage and memory. In Dosso et al. [8], pp. 178–185.

[20] SVARRE, T. Labelling on academic library websites. In Dosso et al. [8], pp. 186–193.

[21] ZANICHELLI, F., ENCHEVA, M., THABET, B., TAMMARO, A. M., AND CONTI, G. Serious games for information literacy: Assessing learning in the NAVIGATE project. In Dosso et al. [8], pp. 78–90.

---