Information and Research Science connecting to Digital and Library Science

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

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

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

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

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

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

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

1 INTRODUCTION

...

2 PAPERS

Invited talk: The swings and roundabouts of a decade of fun and games with Research Objects. Prof. Carole Goble discussed her work on the concept of Research Objects[1][4] in her keynote[2]. We are today transitioning toward the fourth paradigm of science, where the research outcomes are more than just publications and data, but also 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. Their metadata describe their content and context, enriching them with additional information required to make them FAIR+R. These resources thus become shareable, citable, exchangeable, present versioning and snapshots, can be identified with tools such as PIDs and can be registered and deposited on their own in services like Zenodo, 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 famous 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. [12] note 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] highlight 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 the everyday life of people in different aspects of their life, ranging fro, learning to security. Zanichelli et al. [19] note that digital games can be a valuable tool for enhancing learning, and aims 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. [17] 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 to build a knowledge base on the history of the town of Prato.

Information and Knowledge. This section presents new applications and measures to obtain and evaluate information. Irrera and Silvello [14] addresses the problem of finding context information for news articles by extracting entities and relationships from documents that can provide context to documents, and use these entities to improve the effectiveness of ranking methods. Lancioni et al. [15] use Generative Adversarial Networks to predict keyphrases, i.e. short text sequences that convey the main semantic meaning of a document. Ferrante et al. [11] extend the AWARE measure with a set of supervised methods by using several TREC collections.

Character Recognition. This section presented new approaches to address problem 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 [9] propose a mixed qualitative-quantitative approach to OCR error detection and correction to develop a methodology for compiling historical corpora. Heil et al. [13] introduces Handwritten Text Recognition and Document Image Analysis approaches to address the challenges inherent to the original drafts of Astring Lindren, a Swedish author whose original drafts and manuscripts are edited in the Melin system of shorthands, as of today considered undecipherable. Marinai et al. [16] describe a system for the location of simple graphemes in mediaeval manuscripts based on the Mask R-CNN convolutional neural network, a first approach to provide paleographers with a tool to speed up and refine the process of dating and determining the origin of manuscripts.

---

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

Text Analysis. In this section it is possible to find works that explore new potential applications connected to the analysis of documents, opening up new applications and research lines. The work proposed by Bernasconi et al. [6] explores the idea of navigating the semantic relationships among extracted entities as a way to search a text corpus, and the evaluation carried with potential users has shown the feasibility and effectiveness of the approach. Ferilli [10] presents 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 [8] 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] work in the context of the NOTAE project, which investigates graphic symbols, visual units in written text, dating back to the Roman State and Post-Roman Kingdoms. The authors argue that a different, knowledge graph-based, approach to the task can ease the work of researchers at finding non trivial implications in the data. Svarre [18] studies labels used by Danish academic library websites, which 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 infrasctcture, together with its publication as Linked Open Data.

3 PANEL

...

ACKNOWLEDGMENTS

...

REFERENCES

[1] B. Almeida, N. Freire, and D. Monteiro. 2021. The development of the ROSSIO Thesaurus: Supporting Content Discovery and Management in a Research infrastructure. In *Proceedings of the 17th Italian Research Conference on Digital Libraries (CEUR Workshop Proceedings, Vol. 2816)*. CEUR-WS.org, 138–146. http://ceur-ws.org/Vol-2816/short1.pdf

[2] M. Avanzi, R. Coniglio, G. Cisotto, M. Giordani, and N. Ferro. 2021. NoBis: A Crowd Monitoring Service Against COVID-19. In *Proceedings of the 17th Italian Research Conference on Digital Libraries (CEUR Workshop Proceedings, Vol. 2816)*. CEUR-WS.org, 126–137. http://ceur-ws.org/Vol-2816/paper11.pdf

[3] M. Baglioni, A. Mannocci, P. Manghi, C. Atzori, A. Bardi, and S. La Bruzzo. 2021. Reflections on the Misuses of ORCID iDs. In *Proceedings of the 17th Italian Research Conference on Digital Libraries (CEUR Workshop Proceedings, Vol. 2816)*. CEUR-WS.org, 117–125. http://ceur-ws.org/Vol-2816/paper10.pdf

[4] S. Bechhofer, I. Buchan, D. De Roure, P. Missier, J. Ainsworth, J. Bhagat, P. Couch, D. Cruickshank, M. Delderfield, I. Dunlop, et al. 2013. Why linked data is not enough for scientists. *Future Generation Computer Systems* 29, 2 (2013), 599–611.

[5] E. Bernasconi, M. Boccuzzi, T. Catarci, M. Ceriani, A. Ghignoli, F. Leotta, M. Mecella, A. Monte, N. Sietis, S. Veneruso, and Z. Ziran. 2021. Exploring the Historical Context of Graphic Symbols: the NOTAE Knowledge Graph and its Visual Interface. In *Proceedings of the 17th Italian Research Conference on Digital Libraries (CEUR Workshop Proceedings, Vol. 2816)*. CEUR-WS.org, 147–154. http://ceur-ws.org/Vol-2816/short2.pdf

[6] E. Bernasconi, M. Ceriani, and M. Mecella. 2021. Exploring a Text Corpus via a Knowledge Graph. In *Proceedings of the 17th Italian Research Conference on Digital Libraries, Padua, Italy (virtual event due to the Covid-19 pandemic), February 18-19, 2021 (CEUR Workshop Proceedings, Vol. 2816)*. CEUR-WS.org, 91–102. http://ceur-ws.org/Vol-2816/paper8.pdf

[7] M. Biasini, V. Carmignani, N. Ferro, P. Filianos, M. Maistro, and G. M. Di Nunzio. 2021. FullBrain: a Social E-learning Platform. In *Proceedings of the 17th Italian Research Conference on Digital Libraries (CEUR Workshop Proceedings, Vol. 2816)*. CEUR-WS.org, 25–41. http://ceur-ws.org/Vol-2816/paper3.pdf

[8] D. Dosso and G. Silvello. 2021. Data Credit Distribution through Lineage. In *Proceedings of the 17th Italian Research Conference on Digital Libraries (CEUR Workshop Proceedings, Vol. 2816)*. CEUR-WS.org, 155–161. http://ceur-ws.org/Vol-2816/short3.pdf

[9] D. Del Fante and G. M. di Nunzio. 2021. A Quantitative/Qualitative Approach to OCR Error Detection and Correction in Old Newspapers for Corpus-assisted Discourse Studies. In *Proceedings of the 17th Italian Research Conference on Digital Libraries (CEUR Workshop Proceedings, Vol. 2816)*. CEUR-WS.org, 53–63. http://ceur-ws.org/Vol-2816/paper5.pdf

[10] S. Ferilli. 2021. Toward Automatic Floor Plan Interpretation. In *Proceedings of the 17th Italian Research Conference on Digital Libraries (CEUR Workshop Proceedings, Vol. 2816)*. CEUR-WS.org, 13–24. http://ceur-ws.org/Vol-2816/paper2.pdf

[11] M. Ferrante, N. Ferro, and L. Piazzon. 2021. s-AWARE: Using Crowd Judgements in Supervised Measure-Based Methods for IR Evaluation. In *Proceedings of the 17th Italian Research Conference on Digital Libraries (CEUR Workshop Proceedings, Vol. 2816)*. CEUR-WS.org, 162–168. http://ceur-ws.org/Vol-2816/paper4.pdf

[12] P. Gargiulo, P. Galimberti, A. M. Tammaro, and A. Zane. 2021. FAIR RDM (Research Data Management): Italian Initiatives Towards EOSC Implementation. In *Proceedings of the 17th Italian Research Conference on Digital Libraries (CEUR Workshop Proceedings, Vol. 2816)*. CEUR-WS.org, 42–52. http://ceur-ws.org/Vol-2816/paper4.pdf

[13] R. Heil, M. Nauwerck, and A. Hast. 2021. Shorthand Secrets: Deciphering Astrid Lindgren's Stenographed Drafts with HTR Methods. In *Proceedings of the 17th Italian Research Conference on Digital Libraries (CEUR Workshop Proceedings, Vol. 2816)*. CEUR-WS.org, 169–177. http://ceur-ws.org/Vol-2816/short5.pdf

[14] O. Irrera and G. Silvello. 2021. Background Linking: Joining Entity Linking with Learning to Rank Models. In *Proceedings of the 17th Italian Research Conference on Digital Libraries (CEUR Workshop Proceedings, Vol. 2816)*. CEUR-WS.org, 64–77. http://ceur-ws.org/Vol-2816/paper6.pdf

[15] G. Lancioni, S. S. Mohamed, B. Portelli, G. Serra, and C. Tasso. 2021. Efficient Keyphrase Generation with GANs. In *Proceedings of the 17th Italian Research Conference on Digital Libraries (CEUR Workshop Proceedings, Vol. 2816)*. CEUR-WS.org, 1–12. http://ceur-ws.org/Vol-2816/paper1.pdf

[16] S. Marinai, G. Pomaro, C. Raffaelli, and F. Scandiffio. 2021. Location of Simple Graphemes in Mediaeval Manuscripts based on Mask R-CNN. In *Proceedings of the 17th Italian Research Conference on Digital Libraries (CEUR Workshop Proceedings, Vol. 2816)*. CEUR-WS.org, 103–116. http://ceur-ws.org/Vol-2816/paper9.pdf

[17] A. Spadi, F. Spinelli, and E. Degl'Innocenti. 2021. RESTORE: Smart Access to Digital Heritage and Memory. In *Proceedings of the 17th Italian Research Conference on Digital Libraries (CEUR Workshop Proceedings, Vol. 2816)*. CEUR-WS.org, 178–185. http://ceur-ws.org/Vol-2816/short6.pdf

[18] T. Svarre. 2021. Labelling on Academic library Websites. In *Proceedings of the 17th Italian Research Conference on Digital Libraries (CEUR Workshop Proceedings, Vol. 2816)*. CEUR-WS.org, 186–193. http://ceur-ws.org/Vol-2816/short7.pdf

[19] F. Zanichelli, M. Encheva, B. Thabet, A. M. Tammaro, and G. Conti. 2021. Serious Games for Information Literacy: Assessing Learning in the NAVIGATE Project. In *Proceedings of the 17th Italian Research Conference on Digital Libraries (CEUR Workshop Proceedings, Vol. 2816)*. CEUR-WS.org, 78–90. http://ceur-ws.org/Vol-2816/paper7.pdf