Representations of knowledge in living systematic reviews: ionization cross sections by electrons case study

Elisabetta Ronchieri

INFN CNAF

ISGC 2025
Taipei, 20 March

Living Systematic Reviews Introduction

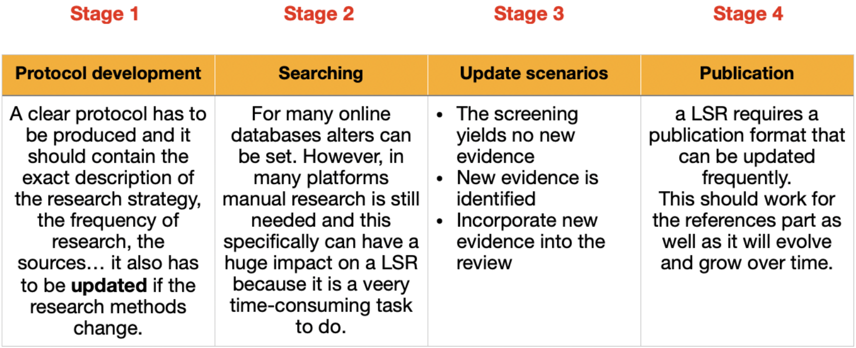
What is a Living Systematic Reviews (LSRs)?

A LSR is a continuously updated Systematic Review (SR) with a **priori commitment** of keeping the systematic review as current as possible.

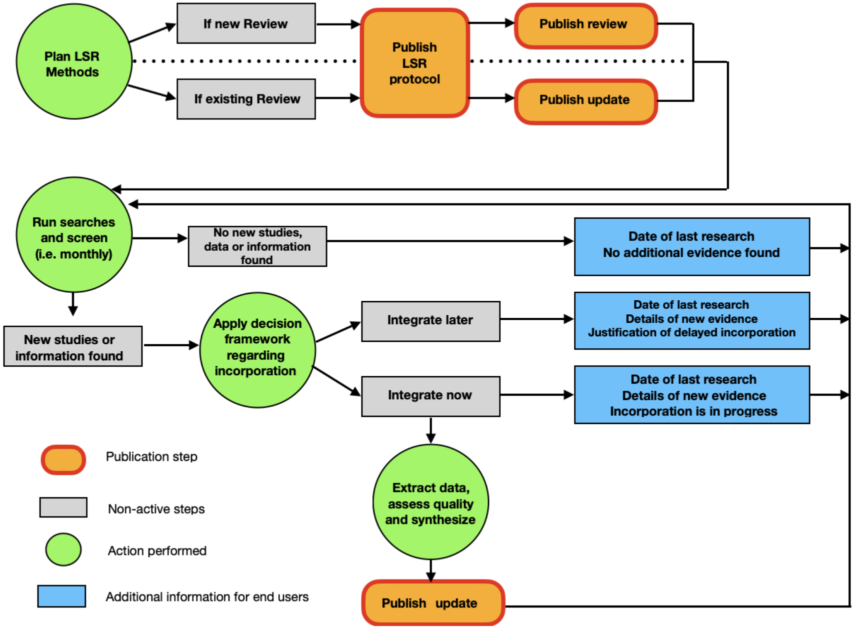
When is it appropriate?

- when the SR is a priority for decision making
- when certainty in the existing evidence is low or very low
- when new research evidence is likely to come up

LSR Stages



Stage 1	Stage 2	Stage 3	Stage 4
Protocol development	**Searching**	**Update scenarios**	**Publication**
A clear protocol has to be produced and it should contain the exact description of the research strategy, the frequency of research, the sources… it also has to be **updated** if the research methods change.	For many online databases alters can be set. However, in many platforms manual research is still needed and this specifically can have a huge impact on a LSR because it is a veery time-consuming task to do.	• The screening yields no new evidence • New evidence is identified • Incorporate new evidence into the review	a LSR requires a publication format that can be updated frequently. This should work for the references part as well as it will evolve and grow over time.



Plan LSR Methods

If new Review

If existing Review

Publish LSR protocol

Publish review

Publish update

Run searches and screen (i.e. monthly)

No new studies, data or information found

Date of last research
No additional evidence found

New studies or information found

Apply decision framework regarding incorporation

Integrate later

Date of last research
Details of new evidence
Justification of delayed incorporation

Integrate now

Date of last research
Details of new evidence
Incorporation is in progress

Extract data, assess quality and synthesize

Publish update

🟧 Publication step

▭ Non-active steps

🟢 Action performed

🟦 Additional information for end users

The updating process of a LSR

When should a systematic review be updated?

The updating approach should be individualized, depending on the author's resources and the editorial team

but also

- policy relevance
- the importance of the study
- founding

Studies Inclusion

"Identifying studies for inclusion is one of the most labor-intensive and time-consuming tasks." [1]

- Database searching
    - Automated searches are run regularly and reviewers are alerted
        - Not all databases support regular specific searches
        - Many databases do not offer API's to connect to
- Eligibility assessment
    - Machine learning techniques applied to titles and abstracts → probability score
    - Automation is not able to entirely perform eligibility assessment → Cochrane Crowd

*[1] Thomas, J., Noel-Storr, A., Marshall, I., et al. (2017). Living systematic reviews: 2. Combining human and machine effort. *Journal of Clinical Epidemiology*. doi:* `10.1016/j.jclinepi.2017.08.011`

BERT to extract papers

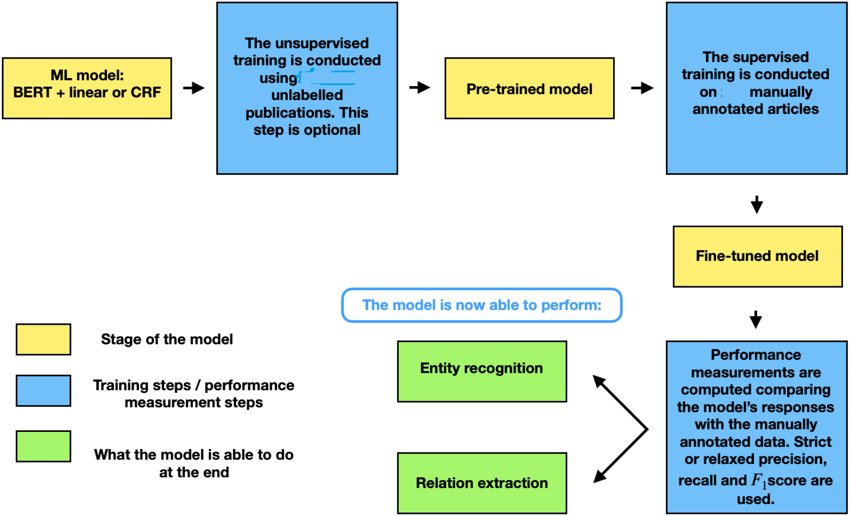
- The number of LSRs has grown significantly
- They rely on manual extraction making them prone to errors
- To automate data extraction→NLP and machine learning to find patterns in data

The state of the art of NLP: **BERT**

(Bidirectional Encoder Representations from Transformers)

- Pre-trained on general text, can learn the meaning of a word based on its surroundings
- Can also be pre-trained on specific texts and then fine tuned

How does BERT work?



ML model:
BERT + linear or CRF

The unsupervised training is conducted using [redacted] unlabelled publications. This step is optional

Pre-trained model

The supervised training is conducted on [redacted] manually annotated articles

Fine-tuned model

Performance measurements are computed comparing the model's responses with the manually annotated data. Strict or relaxed precision, recall and F_1 score are used.

The model is now able to perform:

Entity recognition

Relation extraction

🟨 **Stage of the model**

🟦 **Training steps / performance measurement steps**

🟩 **What the model is able to do at the end**

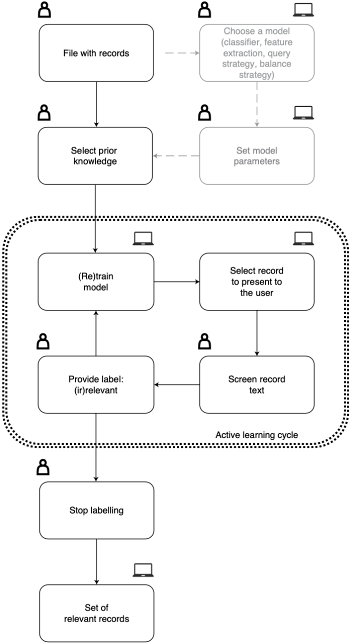
Active Learning SR pipeline

How does it work?

- the researcher downloads a file with records (titles and abstracts and other metadata) and uploads it
- prior knowledge is selected
- entering the active learning cycle until a stopping criterion is reached

↓

the output is a file with records labelled as relevant/irrelevant and unlabelled records ordered based on how likely is for them to be relevant, according to the model



Knowledge Representation

What's a Knowledge Graph (KG)?

"a KG is a directed labeled graph that describes the relationship between real-world entities and represents them in a network" [2]

A KG is composed by 3 objects:

- **A node**, a real world entity
- **An edge**, that captures the relationship between 2 nodes
- **A label**, that describes the meaning of the relationship

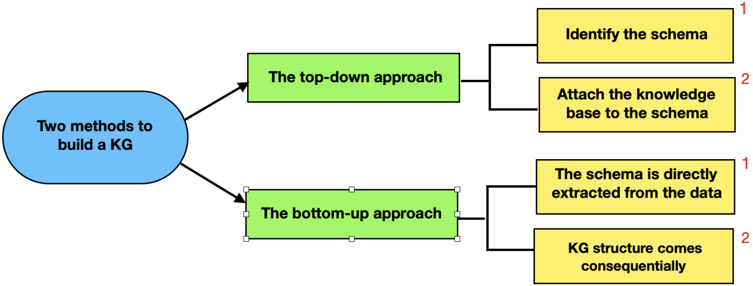
Two nodes connecting through an edge to create a triple

*[2] Akter, M. M., Rahoman, M. (2023). A systematic review on knowledge graphs classification and their various usages. *Artificial Intelligence Evolution, 4*(2), 187-192. doi: 10.37256/aie.4220233605*

KG implementation

When building a KG, two methods can be used:

- Top-down KG Construction Approach
- Bottom-up KG Construction Approach



KG construction approaches

The top-down KG approach

With this method the first step is to **identify** the **ontology** and the **schema** that will be used.

Successively the **knowledge base is attached** to the schema defined before.

→ This approach is mainly used when the data representation is **structured and controlled** and works better with domain-specific knowledge.

→ the top-down approach can be **disadvantageous** for rapidly evolving or dynamic fields.

The bottom-up KG approach

According to this other method, entities and relationships are extracted directly from the data and are **not predefined**.

After the extraction, these are used to **build** the KG structure.

→ This method is useful when dealing with **multiple sources** and is able to process a large amount of data and rapidly create a KG.

→ However, in many cases using this approach results in the need for a **specialist** to mark the actual relationships.

KG categorization

x Implementation

  - Resource Description Framework, Labeled Property Graphs

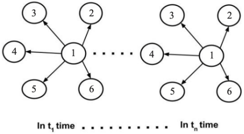
x Metadata

  - Text Knowledge Graphs, Visual Knowledge Graphs, Multi-modal Knowledge Graphs, constructed with both textual and visual data

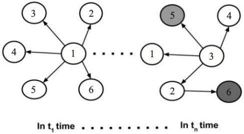
x Data

  - Public Knowledge Graphs, Private Knowledge Graphs

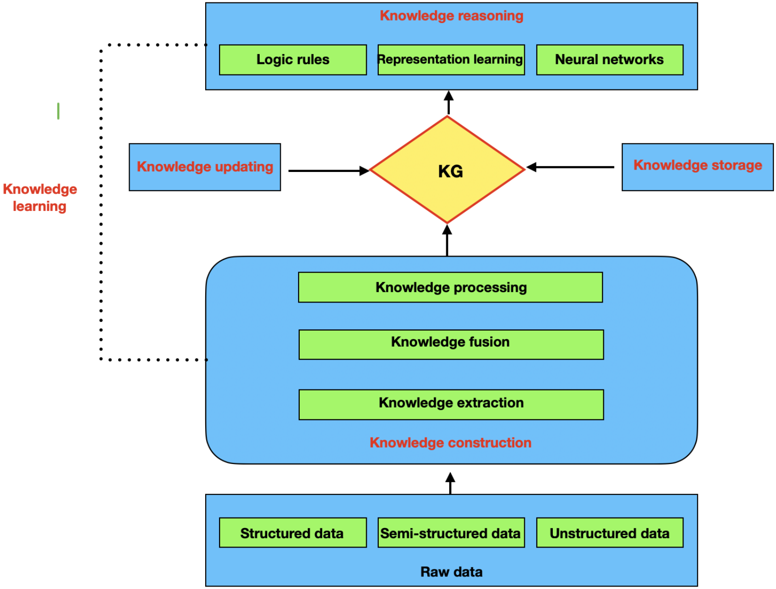
x Time



Static KG [2]



Dynamic KG [2]



KG components

Case Study

Can we represent knowledge about living systematic reviews on shell ionization cross sections data ?

I. Process
II. References
III. Data

IEEE TRANSACTIONS ON NUCLEAR SCIENCE, VOL. 65, NO. 8, AUGUST 2018 2279

Validation of Shell Ionization Cross Sections for Monte Carlo Electron Transport

Tullio Basaglia, Matteo Bonanomi, Federico Cattorini, Min Cheol Han, Gabriela Hoff, Chan Hyeong Kim, Sung Hun Kim, Matteo Marcoli, Maria Grazia Pia, and Paolo Saracco

***Abstract*—Theoretical and semi-empirical methods to calculate electron impact ionization cross sections for atomic shells are subject to validation tests with respect to a wide collection of experimental measurements to identify the state of the art for Monte Carlo particle transport. The validation process applies rigorous statistical analysis methods. Cross sections based on the EEDL Evaluated Electron Data Library, widely used by Monte Carlo codes, and on calculations by Bote and Salvat, used in the Penelope code, are generally equivalent in compatibility with experiment. Results are also reported for various formulations of the Binary-Encounter-Bethe and Deutsch-Märk models.**

***Index Terms*—Cross sections, Geant4, ionization, Monte Carlo simulation, software validation.**

I. INTRODUCTION

THE study reported in this paper complements a previous investigation [1] of ionization cross sections for electron transport with respect to experimental data: the previous publication examined total cross sections, with special emphasis on the low energy range up to a few keV, while the present study concerns the ionization of atomic inner shells by electron impact. Both studies aim to identify the state of the art for the calculation of electron ionization cross sections in Monte Carlo transport codes.

Modeling electron interactions with matter is a fundamental task of any particle transport code. The ability to calculate cross sections for the ionization of individual shells, along with the capability to simulate the subsequent atomic relaxation [2], [3], is required in a variety of experimental environments: in materials analysis performed by electron-probe microanalysis, in surface analysis performed through Auger electron spectroscopy and more generally in experimental scenarios where the simulation of characteristic X-ray or Auger electron emission is important.

Theoretical and semi-empirical models have been developed over several decades to calculate electron impact ionization cross sections for atomic shells; nevertheless, despite the experimental relevance of these cross sections, limited documentation is available in the literature about quantitative validation of their calculations. Comparisons with experimental data, such as those concerning the Deutsch-Märk model [4], often rest on the visual appraisal of plots only. A recent publication [5] illustrates comparisons between some theoretical calculations and experimental data published up to May 2013; however, it is limited to the domain of descriptive statistics, lacking statistical inference. Objective quantification is also missing in the assessment of the relative merits of the various calculation methods: their relative ability to reproduce experimental measurements has not been estimated with statistical methods yet.

This paper evaluates quantitatively and objectively the capabilities of several calculation methods of electron impact ionization cross sections that are relevant for general purpose Monte Carlo transport codes. The evaluation concerns K shell, L and M subshell ionization cross sections, for which experimental measurements are reported in the literature. Statistical inference is applied both to validate cross section calculations with respect to experimental measurements and to detect significant differences in the ability of the various calculations to reproduce experiment. The outcome of this process identifies the state of the art in modeling electron impact ionization cross sections for K, L and M shells in Monte Carlo particle transport codes.

II. ELECTRON IMPACT IONIZATION CROSS SECTIONS

The validation study reported in this paper addresses the calculation of electron impact ionization cross sections in a pragmatic way, i.e. considering calculation methods that are sustainable within the computational constraints of particle transport codes, either by implementing simple analytical formulations or by interpolating available tabulations of theoretical cross section calculations. Since the focus is on general-purpose Monte Carlo codes, only methods able to calculate electron impact ionization for any shell, and covering an extended electron energy range, are considered in the validation tests.

Manuscript received April 24, 2018; revised June 22, 2018; accepted June 27, 2018. Date of publication June 29, 2018; date of current version August 15, 2018.

T. Basaglia is with CERN, CH-1211 Geneva, Switzerland (e-mail: tullio.basaglia@cern.ch).

M. Bonanomi, F. Cattorini, and M. Marcoli are with the University of Milano-Bicocca, 20126 Milan, Italy (e-mail: m.bonanomi@campus.unimib.it; f.cattorini@campus.unimib.it; m.marcoli@campus.unimib.it).

M. C. Han, M. G. Pia, and P. Saracco are with the INFN Sezione di Genova, 16146 Genova, Italy (e-mail: mincheol.han@ge.infn.it; mariagraria.pia@ge.infn.it; paolo.saracco@ge.infn.it).

G. Hoff is with the University of Cagliari, Monserrato, Italy, and also with the INFN Sezione di Cagliari, 09042 Monserrato, Italy (e-mail: ghoff.gesic@gmail.com).

C. H. Kim and S. H. Kim are with the Department of Nuclear Engineering, Hanyang University, Seoul 133-791, South Korea (e-mail: chkim@hanyang.ac.kr; ksh4249@hanyang.ac.kr).

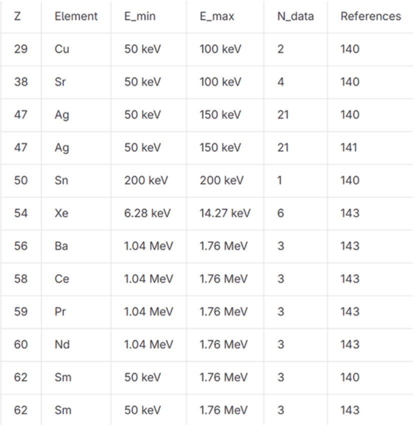
Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNS.2018.2851921

T. Basaglia, M. Bonanomi, et al, Validation of Shell Ionization Cross Sections for Monte Carlo Electron Transport, IEEE Transaction on Nuclear Science, vol. 65, n. 8, August 2018

What we have?

A set of tabular data



Z	Element	E_min	E_max	N_data	References
29	Cu	50 keV	100 keV	2	140
38	Sr	50 keV	100 keV	4	140
47	Ag	50 keV	150 keV	21	140
47	Ag	50 keV	150 keV	21	141
50	Sn	200 keV	200 keV	1	140
54	Xe	6.28 keV	14.27 keV	6	143
56	Ba	1.04 MeV	1.76 MeV	3	143
58	Ce	1.04 MeV	1.76 MeV	3	143
59	Pr	1.04 MeV	1.76 MeV	3	143
60	Nd	1.04 MeV	1.76 MeV	3	143
62	Sm	50 keV	1.76 MeV	3	140
62	Sm	50 keV	1.76 MeV	3	143

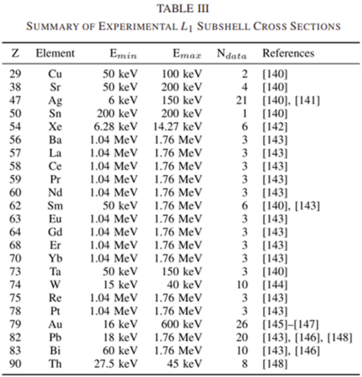


TABLE III

SUMMARY OF EXPERIMENTAL L_1 SUBSHELL CROSS SECTIONS

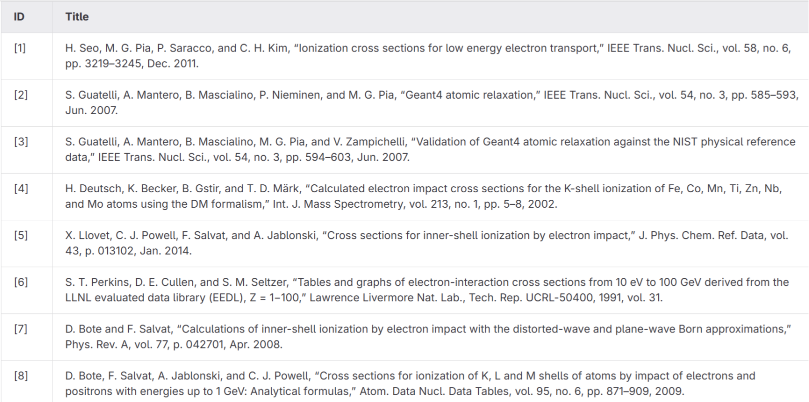
Z	Element	E_{min}	E_{max}	N_{data}	References
29	Cu	50 keV	100 keV	2	[140]
38	Sr	50 keV	200 keV	4	[140]
47	Ag	6 keV	150 keV	21	[140], [141]
50	Sn	200 keV	200 keV	1	[140]
54	Xe	6.28 keV	14.27 keV	6	[142]
56	Ba	1.04 MeV	1.76 MeV	3	[143]
57	La	1.04 MeV	1.76 MeV	3	[143]
58	Ce	1.04 MeV	1.76 MeV	3	[143]
59	Pr	1.04 MeV	1.76 MeV	3	[143]
60	Nd	1.04 MeV	1.76 MeV	3	[143]
62	Sm	50 keV	1.76 MeV	6	[140], [143]
63	Eu	1.04 MeV	1.76 MeV	3	[143]
64	Gd	1.04 MeV	1.76 MeV	3	[143]
68	Er	1.04 MeV	1.76 MeV	3	[143]
70	Yb	1.04 MeV	1.76 MeV	3	[143]
73	Ta	50 keV	150 keV	3	[140]
74	W	15 keV	40 keV	10	[144]
75	Re	1.04 MeV	1.76 MeV	3	[143]
78	Pt	1.04 MeV	1.76 MeV	3	[143]
79	Au	16 keV	600 keV	26	[145]–[147]
82	Pb	18 keV	1.76 MeV	20	[143], [146], [148]
83	Bi	60 keV	1.76 MeV	10	[143], [146]
90	Th	27.5 keV	45 keV	8	[148]

A subset of data about L1 SUBSHELL CS

Also available for K and M SHELLs CS, and for L2 and L3 SUBSHELLs CS

What we have?

A set of scientific papers



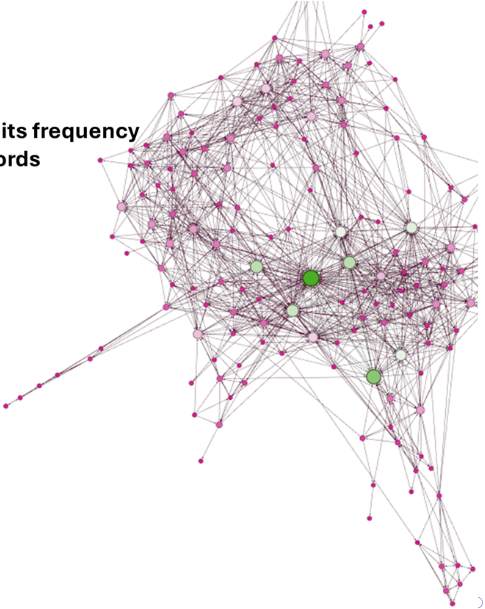
ID	Title
[1]	H. Seo, M. G. Pia, P. Saracco, and C. H. Kim, “Ionization cross sections for low energy electron transport,” IEEE Trans. Nucl. Sci., vol. 58, no. 6, pp. 3219–3245, Dec. 2011.
[2]	S. Guatelli, A. Mantero, B. Mascialino, P. Nieminen, and M. G. Pia, “Geant4 atomic relaxation,” IEEE Trans. Nucl. Sci., vol. 54, no. 3, pp. 585–593, Jun. 2007.
[3]	S. Guatelli, A. Mantero, B. Mascialino, M. G. Pia, and V. Zampichelli, “Validation of Geant4 atomic relaxation against the NIST physical reference data,” IEEE Trans. Nucl. Sci., vol. 54, no. 3, pp. 594–603, Jun. 2007.
[4]	H. Deutsch, K. Becker, B. Gstir, and T. D. Märk, “Calculated electron impact cross sections for the K-shell ionization of Fe, Co, Mn, Ti, Zn, Nb, and Mo atoms using the DM formalism,” Int. J. Mass Spectrometry, vol. 213, no. 1, pp. 5–8, 2002.
[5]	X. Llovet, C. J. Powell, F. Salvat, and A. Jablonski, “Cross sections for inner-shell ionization by electron impact,” J. Phys. Chem. Ref. Data, vol. 43, p. 013102, Jan. 2014.
[6]	S. T. Perkins, D. E. Cullen, and S. M. Seltzer, “Tables and graphs of electron-interaction cross sections from 10 eV to 100 GeV derived from the LLNL evaluated data library (EEDL), Z = 1–100,” Lawrence Livermore Nat. Lab., Tech. Rep. UCRL-50400, 1991, vol. 31.
[7]	D. Bote and F. Salvat, “Calculations of inner-shell ionization by electron impact with the distorted-wave and plane-wave Born approximations,” Phys. Rev. A, vol. 77, p. 042701, Apr. 2008.
[8]	D. Bote, F. Salvat, A. Jablonski, and C. J. Powell, “Cross sections for ionization of K, L and M shells of atoms by impact of electrons and positrons with energies up to 1 GeV: Analytical formulas,” Atom. Data Nucl. Data Tables, vol. 95, no. 6, pp. 871–909, 2009.

A subset of references over 142 selected papers

What have we done so far?

- Identified keywords for labeling each paper
  - e.g., K shell and L1 subshell
- Applied BERT-based models for classifications
  - using e.g. title, abstract, keywords
- Evaluating LLM models for classification
  - using full paper
- Evaluating KG models and tools
  - e.g., Gephi and WizMap

e.g. Gephi Representation



I. Nodes are keywords
II. Size of each node depends on its frequency
III. Connections are among keywords
IV. Keep trace of each paper

e.g. WizMap Representation

I. The graph has the shape of the embedding's projection in two dimensions

II. Each paper is represented by a point

III. Available information of the specific paper

Conclusions & Next Steps

Key Findings

- ✓ Implemented BERT for paper classification
- ✓ Tested different Knowledge Graph (KG) models
- ✓ Initial graphical representations created with Gephi and WizMap

Next Steps

- Optimizing KG graphical representation
- Exploring new text analysis methods with LLMs
- Integrating figures into models for better interpretability

Thanks

INFN CNAF Co-authors: Marco Canaparo
Master Student Co-author: Federica Frabetti
Acknowledgements: Maria Grazia Pia, Giovanni Zurlo, and Alessandro Costantini