

Accelerating Inspire

MAX LINDQVIST
Lund University, Sweden
September 15, 2017

Abstract

CERN has been involved in the dissemination of scientific results since its early days and has continuously updated the distribution channels. Currently, Inspire hosts a catalogues of articles, authors, institutions, conferences, jobs, experiments, journals and more. Successful orientation among this amount of data requires comprehensive linking between the content. Inspire has lacked a system for linking experiments and articles together based on which accelerator they were conducted at. The purpose of this project has been to create such a system. Records for 156 accelerators were created and all 2913 experiments on Inspire were given corresponding MARC tags. Records of 18404 accelerator physics related bibliographic entries were also tagged with corresponding accelerator tags. Finally, as a part of the endeavour to broaden CERN's presence on Wikipedia, existing Wikipedia articles of accelerators were updated with short descriptions and links to Inspire. In total, 86 Wikipedia articles were updated. This report provides a description of the work process which was conducted in the Scientific Information Service between 28 June 2017 and 15 September 2017 as a part of the CERN Summer Student Program.

1 Introduction

1.1 Background

Ever since the commencement of CERN in 1955, the field of High Energy Physics (HEP) and the advancements in related technology have brought together major collaborations between institutes and nations from all over the world. CERN has been involved in the open dissemination of scientific results since its early days and has continuously developed its distribution channels, including the birth of the World Wide Web and the currently used databases, CERN Document Server (CDS) and Inspire. Every year, thousands of articles and reports are uploaded to CDS and Inspire and it is of great importance that professionals in HEP, as well as the general public, have access to this information. To successfully orient oneself in this huge amount of data, the entries have to be linked together in efficient ways. Currently, Inspire provides author profiles and citation systems along with means of connecting experiments with associated articles. The database is currently lacking a system for linking experiments together which have been conducted at the same accelerator though. By implementing this functionality Inspire will increase its usability and enable discoverability of experiments which would otherwise be difficult to find.

1.2 Outline and purpose

This report aims to provide a description of the work I have conducted at the CERN Scientific Information System as a part of the CERN Summer Student Program. The main purpose of the project was to update the Inspire database by adding a collection of accelerator records and linking all existing experiments and accelerator physics related papers to them. This will allow for linking a large part of the literature and increase its discoverability. My

hope is that this report will also highlight some of the obstacles in the work process and methods for overcoming those for future Summer Students with similar tasks.

1.3 Inspire, CDS and MARC

- **Inspire** is an freely accessible digital library for HEP. It serves as the main information platform for high energy physics and is comprised of 7 interlinked databases for articles, authors, institutions, conferences, jobs, experiments and journals. INSPIRE is the successor of SPIRES (Stanford Physics Information Retrieval System), a bibliographic database which was founded in 1974 and was hosted by SLAC. Inspire is maintained and developed by as collaboration between CERN, DESY, Fermilab, IHEP and SLAC.^[1]
- The **CERN Document Server (CDS)** is CERN's institutional repository and hosts a collection of over 900000 records of CERN and non-CERN articles, preprints, reports and theses as well as internal and technical documentation, official CERN committee documents and multimedia objects.
- The set of digital formats used in the descriptions for cataloguing the entries on CDS and Inspire follow the **MARC (MACHine-Readable Cataloging)** standards.

2 Working with Inspire, CDS and Wikipedia

2.1 Experiment tagging and creating accelerator records

Currently, 2913 experiment records are available on Inspire. Most of these experiments are conducted at various accelerators, although about 472 experiments are so called Non-accelerator experiments which are usually on the topic of high energy astrophysics. The majority of work has gone into creating accelerator records and tagging these experiments with the corresponding accelerators.

In order to approach this task in a systematic way, the following workflow has been used.

- A research facility is chosen.
- Relevant sources such as home pages are found and a summary of the accelerators at the facility is created. The energy level and years of operation along with other useful information is noted down when available.
- An accelerator record for each facility is created on Inspire by using the experiment preset but adding the MARC field 980:ACCELERATOR. The accelerator record is marked with its unique name in the MARC field 119__b. In the cases where an accelerator has a general name such as "Main Ring" or "Proton Synchrotron" it is tagged as 119__b:affiliation-accelerator name, for example, 119__b:FNAL-MR is the Main Ring in Fermilab. Exceptions for this are the accelerators at CERN, which are all tagged by their accelerator names in 119__b, for example, 119__b:PS refers to the Proton Synchrotron at CERN while 119__b:KEK-PS refers to the Proton Synchrotron at KEK. The affiliation is tagged in the MARC field 119__u.
- The experiments are tagged with the corresponding accelerator name in the MARC field 119__b. Experiments which are not conducted at any accelerator facility are tagged as 119__b:NON.

This proved to be much more difficult than expected. While CERN provides information on precisely which accelerator has been used at every experiment in the **Greybook**, most research facilities do not provide this sort of information. The accelerator name is seldom apparent in the experiment name or the short descriptions. Therefore, the best way to identify the corresponding accelerator has been to read the associated papers of the

Table 1: Distribution of the number of associated experiments for each accelerator record.

Number of experiments	Accelerator
Over 200	NON, FNAL-MR, SPS, AGS
100-199	Tevatron, CEBAF, LANSCE, TRIUMF
50-99	U70, PS, Saturne, ZGS, KEK-PS, HIPA, HIPA, U10, SLC,
40-49	FNAL-MI
30-39	ISR
20-29	JPARC-MR, LENI-SC, Cooler ,LHC, NIMROD
10-19	Synchrophasotron, SPEAR, RCNP-Cyclotron, TRISTAN, LEAR, ATLAS, COSY, DAFNE, DORIS
5-9	RHIC, PEP, ISOLDE, LEP, FNAL-Booster, UNK, AD, LNS, SSC, ILC, CESR, PETRA, MAMI, ISIS, PBar-Accumulator, SC, VEPP-2
4	HERA, BEPC, MCDR, SmuS, SHB
3	SNS, SNS, ELSA, UNILAC, BATES, Bevalac
2	SuperKEKB, ESRF, JPARC-RCS, NUCLOTRON, PHASOTRON, BNL-LINAC, CDF, ILL
1	CLIC, J-PARC, PEP2, NICA, NSLS, NSCL, CELSIUS, SIS18, K500, RIBF, NICADD, SINQ, PBar-Debuncher, WGTS, BCD, HLMC
0	LCLS, HL-LHC, PF, NSLS2, Recycler, LINAC4, FCC, ALBA, SRS, VLHC, SLS, EMMA, CSNS-RCS, HIE-ISOLDE, CEPC, ALICE, Bevatron, LEIR, ELENA, ATF, PF-AR, AGOR, APS-LINAC, ADONE, VEPP-2M, IOTA, CLARA, IPNS, Super-FRS, AA, REX-ISOLDE, PSB, Super-ACO, TWAC, VELA, ACO, P-LINAC, LINAC2, Cosmotron, AdA, AC, U400, POPAE, TASCC, SPPC, U400M, LINAC3, VEP-1, DTL, ASPUN, LEUTL, DRIBs, CWDD, U200, TSR, SHE, RISING, RFQ, MT25, K150, ITEP-Booster, IC100, HEARTHFIRE, GEM, FRS

experiment. Most articles do not directly mention the specific accelerator where the experiment was conducted, but the information is usually given in an indirect wording such as “the 12 GeV beam” or “on the muon beamline” from which the correct accelerator can be deduced. Although, more often than not, there is either no mention of any specific beam at all or the description of the beam/accelerator is overly general. An example of this is the experiment [FNAL-E-0856](#), which has no description and only one associated article – a test beam request with beam requirement "Particle type: anything charged".[\[3\]](#) Many experiments do also have websites, with varying degree of usability, which in some cases provided relevant information, but generally the experiments which were conducted before the internet existed were also the ones which were the most difficult to categorise. To categorise those, references such as [\[2\]](#) were of use. Furthermore, e-mail contact with representatives from facilities as SLAC, FNAL, ANL, BNL and KEK has been of great value, they have provided historical timelines and details which have not otherwise been obvious.

All 2913 experiments have been tagged and Table 1 shows the distribution of the number of associated experiments for each accelerator record. The accelerator tag NON, with its 479 experiments, is the most common, followed by FNAL-MR (287 experiments), SPS (216 experiments) and AGS (203 experiments).

It is interesting to note that out of the 2913 experiments in Inspire, 897 do not have a single associated article. When categorising the experiments, it was clear that the experiments from some institutions had far fewer articles per experiments than other. Historical experiments is also lacking associated articles to a higher degree than newer ones. Figure 1 shows the distribution of associated articles per experiment for the 9 institutions which are linked to most experiments. Only experiments which have 30 or less associated articles are included to clearly show the distribution of the experiments with fewest associated articles, outlier experiments such as ATLAS which has 11762 associated articles on Inspire are not of any concern. The y-axis represents the number of experiments (frequency)

for a certain number of associated articles given on the x-axis.

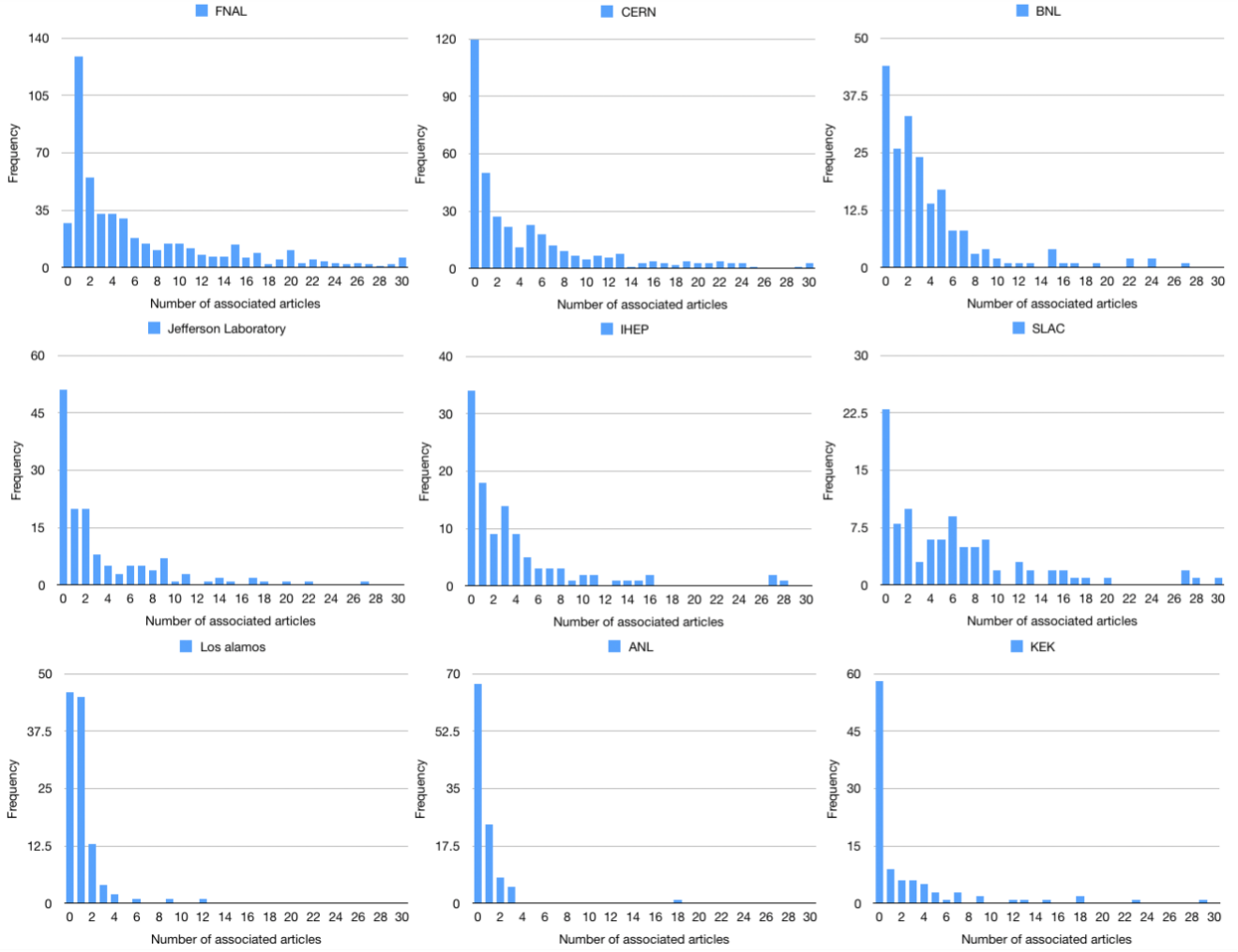


Figure 1: Distribution of number of associated articles for the 9 institutions with the most experiments on Inspire.

It is clear that all institutes, except Fermi National Laboratory (FNAL), have a rather large portion of experiments which are not associated with any articles. This is especially clear for the experiment records from Argonne National Laboratory (ANL) and the High Energy Accelerator Research Organization (KEK) for which the majority of experiments do not have any associated articles. Out of the experiments conducted at CERN, 29% of the experiments does not have any associated articles which can be compared with Fermilab which only has 5% without any associated articles. It was beyond the scope of this project to add experiment tags to reports which do not have any, this could be a task for a future Summer Student.

2.2 Accelerator reports

The HEP-collection on Inspire contains over 1.2 million records which are categorised by the field codes given in Table 2. Currently, 82185 of these records are on the topic of "Beams (accelerators)" and contain reports of accelerator physics. Some of these articles are directly connected to a specific accelerator and should thus contain the corresponding tag. This has been added to the MARC field 693__a. Many of these papers mention the accelerator of interest in its title and were therefore quickly identified with keyword searches and tagged with the

Inspire tool multiedit. Not all of them can be directly linked to an accelerator though, as some are general reports on accelerator physics. In total, 18,404 articles were tagged.

Table 2: Field codes for HEP articles

Field code	Subject
a	Astrophysics
b	Accelerators.
c	Computing
e	Experiment-HEP
g	Gravitation and Cosmology
i	Instrumentation
l	Lattice
m	Math and Math Physics
n	Theory-Nucl
o	Other
p	Phenomenology-HEP
q	General physics
t	Theory-HEP
x	Experiment-Nucl

After having exhausted most thinkable variations of keyword searches for the already existing accelerator records and in order to have an efficient method of identifying the remaining large groups of articles, phase frequency searches were conducted of all the titles and abstracts of the untagged articles. These searches quickly revealed relevant keywords.

2.3 Wikipedia updates

For a large portion of the people using the internet, Wikipedia is one of the major sources for information. Both laymen and professionals use the platform to form a basic understanding of a new subject, therefore CERN is engaging in dissemination information for the general public on this scene. As has been argued by Storehaug in [4], many of the CERN accelerators and historical experiments are either nonexistent or inadequate on Wikipedia. It is therefore in CERN's interest to broaden its prececence on Wikipedia. As a part of this endavour, existing Wikipedia articles of lists of experiments and accelerators in HEP were updated by providing short descriptions and links to Inspire as well as cleaning up old SPIRES links. In total, 77 Wikipedia articles were updated. Examples of such Wikipedia pages include [List of Proton Synchrotron experiments](#), [List of accelerators in particle physics](#) and [List of Super Proton Synchrotron experiments](#). The full list of edits is available [here](#).

2.4 Other tasks

In a further attempt to improve the cross linking between CDS and Inspire, all design reports related to accelerators on CDS have been linked to Inspire. The design reports which did not have a corresponding entry on Inspire were imported from CDS. General improvements, such as uploading of missing fulltexts to Inspire and replacement of broken or addition of missing links, were done continuously.

3 Conclusion

This final section will be used to summarise and conclude the report and also provide a personal reflection and some thoughts of the project. Working with INSPIRE provided an insight on the importance of collecting and effectively storing scientific information. Most of the working hours have been used to read abstracts of articles in HEP. Knowing which accelerator has been used during an experiment may be very obvious at the time, but over 30 years later, when the accelerator facility has completely changed and new machines are in use, it has proven to be rather difficult. In order to successfully tag the experiments, it has been of great value to have a background in physics as this has allowed for identification of experiments in certain areas which are unique to a certain accelerator in a given facility. Personally, having read hundreds of abstracts has given me a more enlightened view of HEP, I have gained an appreciation of what sorts of experiments have been conducted and an overview of all major accelerator facilities in the world. This will certainly be of good value in my future endeavours in physics.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor, Micha Moskovic, for the continuous support and for allowing me to spend my time as I saw fit.

Secondly, thank you Annette Holtkamp for always providing me with varying tasks and for never letting me get stuck with any given task for too long.

Finally, I would like to thank Ida Storehaug, who has taught me the historical context of all important experiments and discoveries, this has added an extra nuance to my work. You have also taught me to speak Norwegian, and for that I am forever grateful.

References

- [1] Annette Holtkamp, Salvatore Mele, Tibor Simko, and Tim Smith. INSPIRE: Realizing the dream of a global digital library in High-Energy Physics. (CERN-OPEN-2010-019):9 p, Jun 2010.
- [2] C P Horne, G P Yost, Alan Rittenberg, F E Armstrong, M S Hutchinson, D R Richards, T G Trippe, F Uchiyama, D M Chew, T A Coffeen, R L Crawford, J E Enstrom, R L Kelly, J J Kingston, Thomas A Lasinski, Michael J Losty, D L Pollard, Arthur Hinton Rosenfeld, V A White, C G Wohl, N Barash-Schmidt, G C Fox, P R Stevens, C S Cooper, F D Gault, B J Read, Y Oyanagi, R G Roberts, Louise Addis, and G M Row. *An indexed compilation of experimental high energy physics literature*. Lawrence Berkeley Nat. Lab., Berkeley, CA, 1978.
- [3] Sherwood Parker, Chris Kenney, Carl Haber, Nicolas Produit, Walter Snoeys, Chye Huat Aw, and James Plummer. Test beam request: An Integrated pixel detector. 1991.
- [4] Ida Storehaug. Improving CERN’s Presence on Wikipedia - Project Report, Summer Student Lecture Program 2017. Aug 2017.