

INSPIRE VRE

INSPIRE is the next-generation High Energy Physics (HEP) information system, which empowers scientists with innovative tools to promote successful research. It is based on Invenio digital library technology developed at CERN, and provides access to almost one million records.

INSPIRE uses advanced computing services to provide enhanced functionality to the worldwide community of over 30,000 scientists. D4Science helps INSPIRE in providing faster and more complex functionality than earlier HEP repositories which had technical, architectural and computational limitations.

OCR

Many digital libraries contain large numbers of scanned documents where textual information is not available. Optical Character Recognition (OCR) enables textual analysis and indexing of these documents. Traditionally the OCR process has been carried out using commercial tools or services where libraries have to pay for the OCR process, but INSPIRE now provides tools to enable free OCR on the D4Science infrastructure using open-source software.

The INSPIRE team has implemented a procedure and packaged the software so that the OCR of scanned documents can be executed as a grid job in the D4Science Process Execution Engine.



