Open Search Use Cases for Improving Information Discovery and Information Retrieval in Large and Highly Connected Organisations

DI Igor Jakovljevic, CERN & Graz University of Technology
Prof. Dr. Christian Guetl, Graz University of Technology
Dr. Andreas Wagner, CERN

Information Generation

- Humans generate 33 trillion gigabytes of information per minute

- User in an organisation sends and receives 128.8 emails per day, 24% contain additional information in the form of attachments

- Organisations use face-to-face meetings, broadcast media communications, mobile communication, electronic communication, and written communication as a means to share knowledge and to generate information

- Organisations process 500 TB of data per week

Information Generation in Large Organisations

- **Main issues with such large amount of information in organisations is how to effectively retrieve information and find information and messages when needed.**

- **Studies indicated that the wider the search for knowledge in an Organisations is, the higher the organisation's innovation.**

- **Information openness can stimulate innovative activities, creation of innovative approaches, and greater performance.**

- **Main disadvantages of an organisation's openness is the risk of losing competitive advantage and leaking private information.**

Focus of the Paper

The analysis of communication, information search, and retrieval habits in large and knowledge focused organisations, exemplary of CERN employees

With the goal

To determine possible improvements in their daily workflows, and to investigate how open search concepts can improve such settings.

Information Generation at CERN

Of the total 37057 users that are affiliated with CERN, 1232 users participated in the CERN survey which was used in this paper to explore the information use of CERN employees.

In the case of CERN, just the Large Hadron Collider generates 10 GB per second and the data produced by the collider is used for research, reports, visualisation, in communication and more

Information Generation at CERN

CERN employees also use various methods of communication and data sharing.

Method of Communication	User Count
Email	1216
Face-To-Face	1015
Chat	585
Phone	575
SMS	170
Social Media	111

Knowledge Sharing at CERN

Information is shared, transferred, and transformed with remarkable velocity nowadays. Sharing knowledge in organisations can help build a competitive advantage if the knowledge is managed adequately.

In recent years CERN has been working on the concept of open science, which uses the principles of open data and open search, to make the research done at CERN more visible and accessible to the general public.

Knowledge Sharing at CERN

Methods for sharing documentation	User Count
Via Email	899
Indico	479
Afs	455
CernBox	377
Dfs	318
Twiki	315
Other	237
Edms	196
Sharepoint	192
Cds	144
Eos	118
OnedriveSocialCernCh	44
No	25

Open Innovation at CERN

Properly documenting and sharing institutional knowledge throughout the organisation enables the organisation to tackle various problems and daily routines.

When a problem cannot be solved us- ing organisation's current routines, the organisation is forced to innovate by developing new knowledge.

The drive to use external sources of information for innovation and extension of knowledge, pushes organisations to be open and distribute information to facilitate innovation

Open Innovation

Open innovation can be defined as the use of purposive inflows and outflows of knowledge to stimulate internal innovation and to increase the demands for external use of innovation, respectively

Importance of Privacy in Open Data

The goal of open innovation and open information is to increase accountability, transparency, and to provide new and efficient services.

Before the information is made open, in general cases it is anonymised, with the intent to remove any personal identifiers from the data.

With a higher degree of anonymisation, the less useful the information is.

The survey shows that a large percentage of surveyed users do not prefer to use social media and avoid sharing personal information at the workspace.

User Analysis

- **Large and highly connected organisations, such as CERN, prefer the use of emails and face-to-face meetings as their main communication and knowledge sharing method.**

  - **The drawback of face- to-face meetings is that a part of the information produced in meetings can get lost due to inefficient documentation practices.**

  - **The documentation produced at face-to-face meetings is restricted to the group of individuals that attended the meeting.**

  - **This brings difficulties in sharing information with individuals outside of the work-group or meeting group and retrieving information.**

User Analysis

- **According to the CERN IT department study, most of the users use the CERN web-page and meetings for information discovery about CERN related information and events**

  - **This behaviour of searching for information within a specific group of users facilitates the creation of information bubbles.**

- **Besides the tools for information discovery of data in their departments, about CERN and events at CERN, users use software for technical documentation, version control software, blogging software, survey software, video software, project management software, wiki software, and multiple hardware components.**

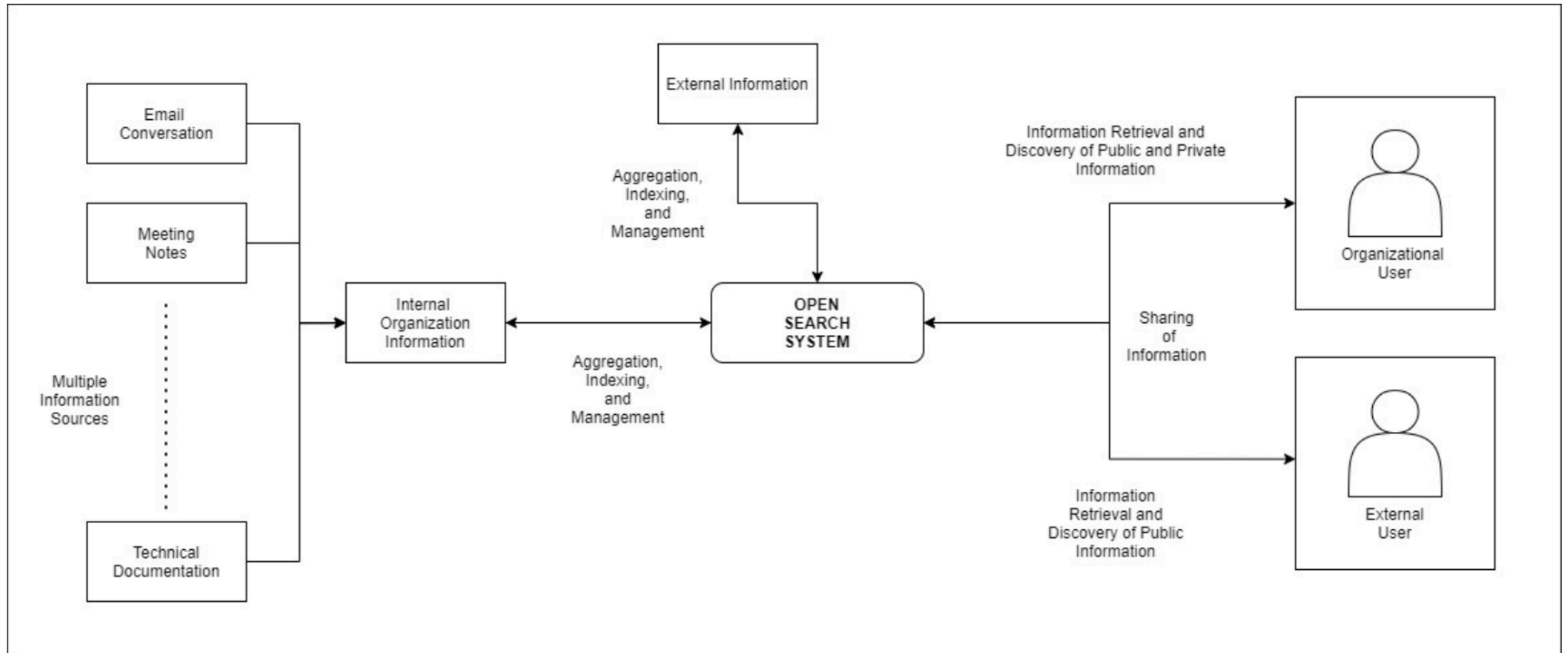# TOWARDS A CONSOLIDATED RETRIEVAL EXPERIENCE

We propose an open search solution that enables the search and sharing of private and public information within the organisation but also enables users outside of organisations access to organisations' public data.

TOWARDS A CONSOLIDATED RETRIEVAL EXPERIENCE

The concept system needs to meet the following requirements:

✦ Storage of large amounts of private and public data while taking into consideration privacy concerns

✦ Aggregation of different data sources (within an organisation or outside sources) and normalisation of data

✦ Implementation of the open data concept

✦ Intuitive access to information for internal and external

✦ users

✦ Restrictions of access to private information

✦ Possibility to share information within an organisation or outside of an organisation

✦ Ability to search for internal and external information

TOWARDS A CONSOLIDATED RETRIEVAL EXPERIENCE

TOWARDS A CONSOLIDATED RETRIEVAL EXPERIENCE

Enabling individuals inside and outside the organisation, groups, and organisations to consume and share the generated information is a basis point for enabling open innovation and increasing the value of the produced information.

home.cern