Added value for CRIS systems using visualizations

Ajay Sah, Bastian Melsheimer, Marcus Walther*

FAU Competence Center for Research Data and Information, Friedrich-Alexander-Universität Erlangen Nürnberg

* Corresponding author: marcus.walther@fau.de, Freyeslebenstr. 1, 91058 Erlangen

Introduction

Friedrich-Alexander Universität Erlangen-Nürnberg is a large research university in the German region of Frankonia, ranked the leading German university in innovation by the THE University Ranking. With over 40,000 students and employing more than 6,500 researchers and research supporting personnel in over 600 professorships, FAU has been using and improving their Converis based Current Research Information System (FAU CRIS) since 2015. FAU CRIS covers standard meta data models for research information, such as CERIF[i] and the German „Kerndatensatz Forschung“ (KDSF[ii]).

Through the long-standing operation of Current Research Information Systems (CRIS), a wide range of metadata on research outputs is collected in a structured way. The links between, for example, publications, research projects, project partners and other entity types produce a valuable, comprehensive dataset about the own institution. These details are not available elsewhere at this level of granularity, even not from commercial providers. Conversely, operating a research information system entails substantial costs, so ensuring the broadest possible use of the collected data must play an important role.

The use of research information by central bodies, such as the university leadership, is often ensured by the CRIS operating model via central units or offices. At FAU, this task is undertaken by the FAU Competence Center Research Data and Information (FAU CDI[iii]).

Re-use occurs in virtually all institutions that operate a research information system. This is usually limited to lists, the display of publications and grants on websites and sometimes to simple visualisations of connections. For the preparation of strategic decisions, bespoke reports tailored to the specific question are produced in close consultation with the relevant stakeholders.

Motivation & Data Overview

Research information allows for considerably more possibilities for re-use, however. It is also crucial for researchers’ acceptance that they themselves benefit from the data collected in the CRIS and can access analyses without major hurdles. At FAU, list-format reports have been standard practice for several years.

In addition, more attractive visual presentations are increasingly in demand. Some CRIS implementations (like Pure RIMS[iv]) provide out-of-the-box network graphs between people and world maps.

At FAU we go beyond these standard representations. In response to demand from across the university, we have developed the following graphical representations: (1) a map showing collaborations for the whole university; (2) maps based on the research information of individual units, such as departments and chairs; (3) collaboration networks of research groups; (4) visualisations of relationships with statistical attributes, such as disciplines.

It is important that the visualisations can be produced very easily while meeting a basic standard of visual design. Beyond use in presentations and funding applications, the ability to embed them in websites is necessary. The visualisations should also be exploratory in nature, allowing website visitors to dive into the data and interactively request further information.

All visualisations are based on author linkages, including affiliations on publications, and, where applicable, additional collaboration partners and partners in research projects. The data model in FAU CRIS includes both direct links between publications or projects and collaboration partners, and links via the „business“ card model (going beyond the CERIF standard, see figure 1 below). This model also allows relations to be represented at the person level.

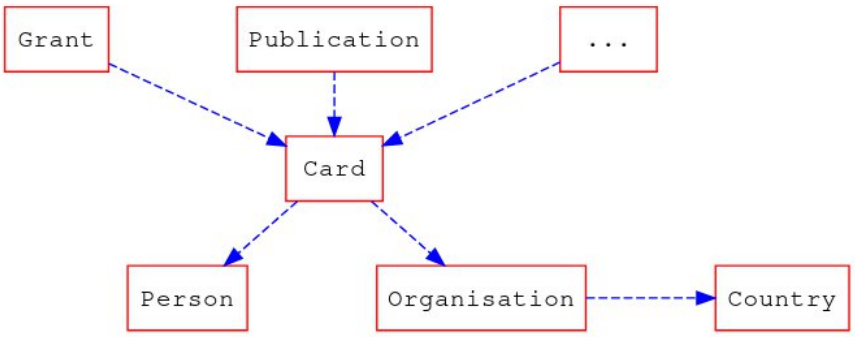


Figure 1: "Business" Card data model in FAU CRIS

Implementation

A very important measure is ensuring adequate data quality and managing duplicates. Problems in this area distort the visualisations and endanger trust in the research information system.

This issue is addressed by a broad use of persistent identifiers (PIDs) in the CRIS. Those identifiers include, among others, ROR ID[v], ORCID[vi], DOI[vii] and internal identifiers for persons (via identity management) and organizational subunits (from FAU’s central directory). However, this does not resolve all challenges; manual corrections are often required. While this increases the operating costs of the research information system, it does greatly improve its acceptance within the university.

The map presentation for the whole university is integrated into the public research portal[viii] (an example is given in Figure 2). The map tiles are based on data from OpenStreetMap[ix]. To reduce the load on their servers, FAU operates its own map server[x]. This also fulfils data-protection requirements — for example, no IP addresses are transmitted to external services or international content networks. Within the map, users can choose between data from the last five years and all available metadata. Visitors can use up to 18 zoom levels and thus virtually visit partner institutions on site. Navigation to neighbouring partners is provided, including a link back to the details in the research portal.

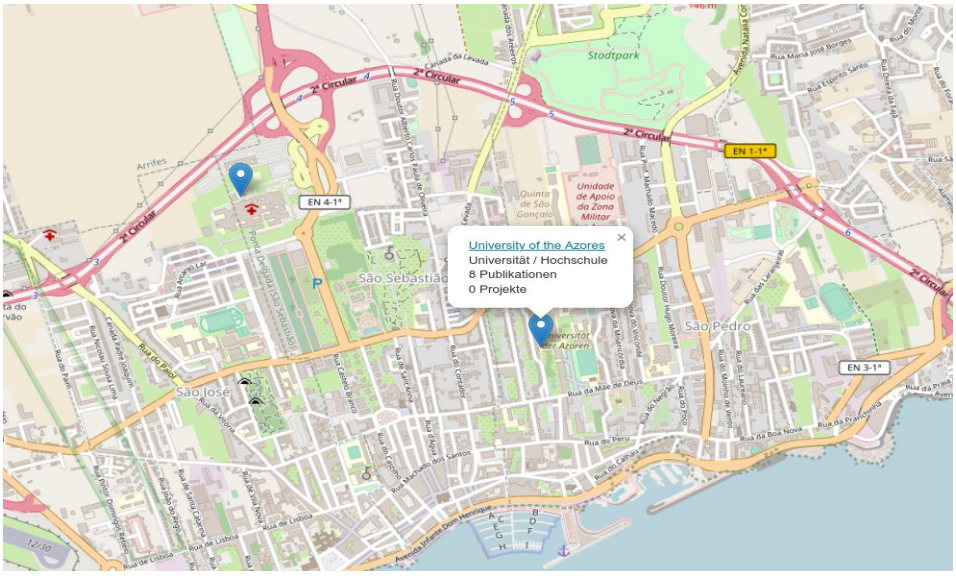


Figure 2: Map view in FAU's research portal (Ponta Delgado, Azores).

Maps are inherently not fully accessible for people with visual disabilities. In the current implementation, however, it is possible to increase contrast to make the presentation more accessible.

Our implementation uses the open-source framework Leaflet[xi], which can directly read and display data from CSV files. This enables a very simple technical integration of research information data and makes it independent of the underlying platform. By using ROR IDs for organizations, geographic coordinates can be retrieved via the ROR API; these originally come from Wikidata[xii]. For organisations not listed in ROR, manual entry in the CRIS has been implemented.

Unlike the university-wide presentation, data preparation for an internal subunit, such as a department or chair, is more complex. For each recorded research output, the intersection between the partner organisation and the subunit must be determined. The configuration of this visualisation can be performed directly in the CRIS by the respective subunit heads and their CRIS administrators (see Figure 3). Converis, the platform used at FAU, permits this flexible setup. Individual FAU subunits use WordPress as their web platform. The map is embedded via a shortcode, making it very easy to enhance a subunit's web presence with data from the CRIS.

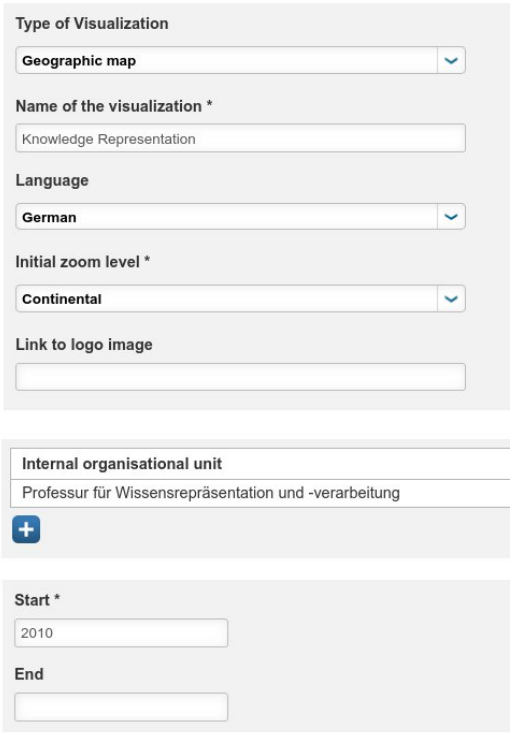


Figure 3: Self-Service configuration of map visualization

The third type of visualisation covers collaborations between individuals within a working group. A working group can have any background — for example joint research or shared use of a laboratory — and is handled very flexibly. Researchers at FAU can easily search for people respective their „business cards" in the CRIS — they are already present because their research outputs are recorded — and configure the displays. In the interactive visualisation based on the open-source framework vis.js[xiii], the depth of collaboration is indicated

by the distance between project partners: the smaller the distance, the stronger the collaboration (more joint publications and research projects). This allows a very quick assessment of a working group's network. In addition, a list of joint publications can be generated, including the number of group members who co-authored each work. These details are available at the click of a button and are mainly used in funding extension applications to document achieved outputs. Strategically, the network visualisation is particularly used in larger collaborative projects (e.g. DFG's funding lines Collaborative Research Centres (CRC) and Research Training Groups (RTG)) to analyse the involvement of project partners.

Another form of representation assigns research outputs to statistic values or classifications. FAU CRIS currently supports assignment to the subject classification of the German Research Foundation (DFG) and to the Scopus Subject Areas. Here too, collaborations are the focus of the visualisation, i.e. connections between subjects or research fields are shown (see Figure 4). Depending on the strength of the connection, interdisciplinary focus areas can be readily identified.

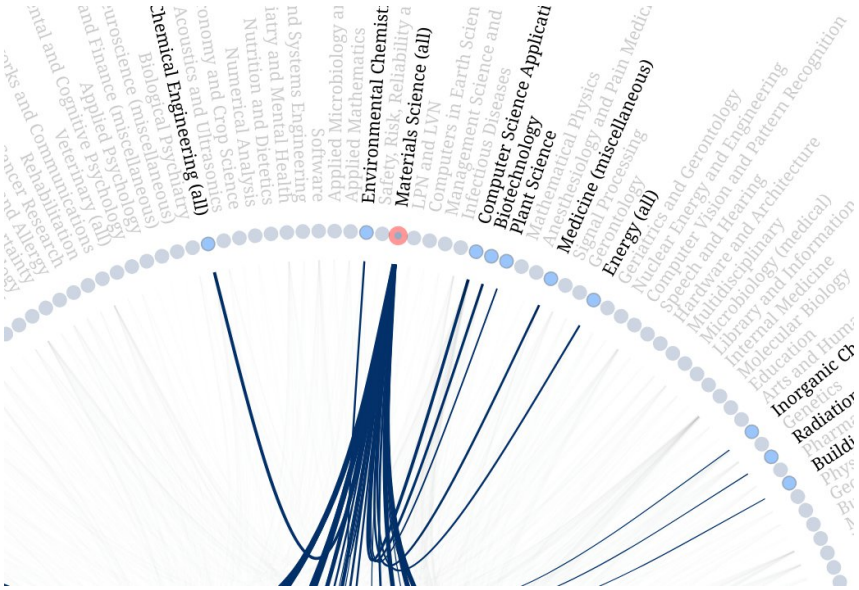


Figure 4: Collaborations indicated by Scopus Subject Areas

Conclusion

At FAU, visualisations of research information are used primarily to depict collaborations. This allows a rapid assessment without having to analyse extensive lists of e.g. publications first. The representation as a world map is available both university-wide and for individual subunits, and the latter can configure it directly in the CRIS.

Making the collected research information accessible in this way increases the CRIS's acceptance within the university, since analyses are available to everyone and not only to the university management.

A particular challenge lies in ensuring the necessary data quality with respect to completeness, accuracy and duplicate management. While the consistent use of PIDs already makes most entity types manageable, there is still room for optimisation when matching people with identical names at the same institution to reduce manual effort. This issue affects the network visualisation in particular: where no PID is available, it may be impossible to determine whether "Michael Smith" at Harvard University refers to one person or several. One possible solution would be to disambiguate by using the discipline or subject area derived from publications.

i Main features of CERIF. euroCRIS. https://eurocris.org/services/main-features-cerif
ii KDSF-Standard für Forschungsinformationen in Deutschland. https://kerndatensatz-forschung.de/
iii FAU Competence Center for Research Data and Information. FAU CDI. https://cdi.fau.de/
iv Pure RIMS: Research Information Management System. https://www.elsevier.com/de-de/products/pure
v Research Organization Registry (ROR). https://ror.org/
vi ORCID. https://orcid.org/
vii Digital Objects Identifier. https://doi.org/
viii The Current Research Information System of Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). https://cris.fau.de/
ix OpenStreetMap. https://www.openstreetmap.org/
x Kartendienst der FAU. https://karte.fau.de/
xi An open-source JavaScript library for interactive maps. Leaflet. https://leafletjs.com/
xii Wikidata. https://www.wikidata.org/
xiii Vis.js Community Edition. https://visjs.org/