Dr. Tobias Weigel
11 Oct 2019 .

This year's Open Science Fair in Porto highlighted recent developments and future opportunities in many areas of interest to stakeholders in the earth sciences, particularly as regards e-infrastructure and service provision. The following commentary has been provided to us by FAIRsFAIR Champion and earth science data specialist **Dr. Tobias Weigel**.

The Challenge of Data Reproducibility

In his welcome address, Yannis Ioannidis of OpenAIRE noted that the trend toward using artificial intelligence and machine learning to improve large scale data analysis poses a challenge for Open Science as regards ensuring openness and reproducibility.

The Earth System sciences community is a case in point. While the uptake of the new data analysis methodologies is gradually increasing in this field, questions on their reproducibility and transparency remain open.

This is understandable in the light of the existing challenges for transparency and reproducibility of Earth System science workflows in general. My hope is that these issues will receive more attention, possibly also as part of future Open Science Fairs. The challenges cannot be ignored but nor should new methodologies be avoided out of fear.

Findability vs Discovery

The discussion in the discovery workshop around the fine distinction between findability, discovery and search is worth mentioning. Data may have become more findable due to improved processes for gathering metadata, but to enable discoverability in practice, one needs to shape infrastructure adequately to the requirements of different discovery scenarios.

This is not necessarily and widely the case today. Typically, the focus is on tuning metadata for better findability, and that does not fully reflect the user perspective. Expert users of climate data for example, are quite happy with faceted search interfaces as long as they provide the means to filter precisely along well-curated metadata categories. However, designing for different user needs is critically important, as is designing with human factors in mind. Making the search process fun is an absolute pre-requisite for encouraging data reuse on the widest scale possible.

Text Mining

In the session on open science graphs, text mining was identified as an intelligent technique which is already being explored. Also discussed was the potential of AI to further enhance text mining capabilities by evolving patterns in data across published articles, graphics, and other media with the aim of creating

complete knowledge graphs comprised of both the underlying data and software.

As graphs become too large for human users to easily consume, even with more sophisticated visual interfaces, one might consider integrating recommender systems into the catalogs and discovery services associated with graphs. Doing so could also reinvigorate semantic matching and reasoning efforts. The result would be to provide users with a better interaction experience. The reuse of research artifacts from other domains or groups would also be facilitated.

I believe the Open Science community should engage the AI community on these topics. The FAIRsFAIR project could take a leading role in stimulating activities. I was stimulated by what I saw and heard at this event. In terms of the topics mentioned above – and many others not discussed here – there is enormous scope for future activity and I look forward to my participation in unfolding events.

About Tobias Weigel

Dr. Weigel has a background in geoinformatics and computer science and works at the German Climate Computing Center (DKRZ) in the area of e-infrastructures at European and international level. An advocate of persistent identifiers, machine-interpretable metadata and reusable software components for data management, Dr. Weigel has a strong interest in building innovative solutions across domains, and harnessing computer science concepts to make work with research data more productive and open.

- Findability
- Reproducibility
- FAIRsFAIR Champion

3,238 Read

```php
<?php// print render($content['links']); ?>
```