

Shaping the Czech National Infrastructure for Research Data through Community Consensus

by Martin Dvořák, Pavlína Špringerová and Matej Antol (Masaryk University)

Four years ago, the Czech Republic laid out its plan for a National Data Infrastructure for FAIR (Findable, Accessible, Interoperable, Reusable) research data under the Czech chapter of the European Open Science Cloud initiative (EOSC CZ). After years of planning and development, the first comprehensive national survey of researcher practices delivers insight into its progress: although two-thirds of Czech researchers declare some familiarity with the Open Science principles in research data management, more than half still store research data primarily on portable devices, and a similar proportion have never heard of EOSC. These findings are directly influencing the initiative's approach to working with the scientific community and its deployment of infrastructure, tools, and services.

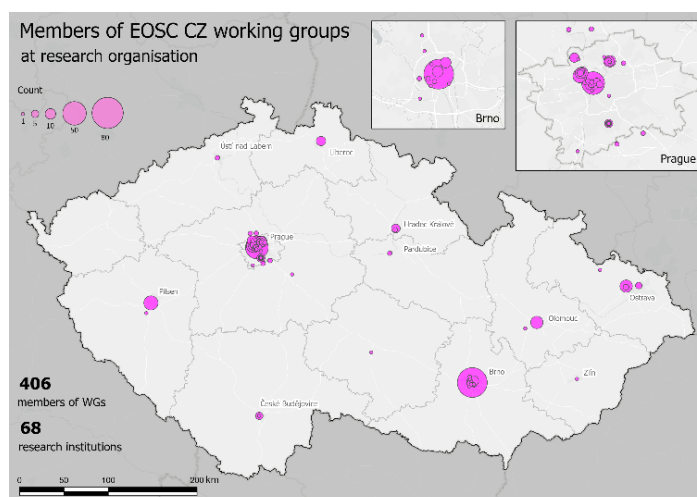


Figure 1: Members of the EOSC CZ Working Groups.

Since 2022, 12 working groups (WGs) have been established to identify the most pressing needs of Czech researchers regarding FAIR data management. More than 400 WG members drafted a comprehensive plan that encompasses hardware infrastructure, repository systems, data repositories, accompanying services and systemic support for data management training. The EOSC CZ initiative has received funding from the Ministry of Education, Youth and Sports (MEYS) through a series of closely interconnected projects. The whole effort is coordinated by Masaryk University and currently encompasses over 20 institutional partners. Its ambition is clear: to provide every Czech researcher with the necessary infrastructure, tools, and skills to manage data responsibly and in accordance with FAIR principles.

Build-up of the infrastructure

The national data infrastructure for FAIR research data is steadily taking shape. Its repository-as-a-service platform, including repo-systems CESNET Invenio, CLARIN-DSpace, ASEP/ARL, and Islandora, is in pilot production and already provides approximately 10% of its 50 PB target storage capacity. It currently hosts a generalist repository [L1] and domain-specific repositories for biodiversity, molecular biophysics, and archaeological data. These repositories are further complemented with services for data processing, authentication and authorisation infrastructure, data management planning tools, metadata models and more. Interoperability is ensured by the newly introduced Czech Core Metadata Model (CCMM)[L2].

To build the human capacity needed alongside the technical infrastructure, the EOSC CZ Training Centre has trained more than 3,000 researchers, data stewards, curators, and enthusiasts, who, in turn, bring new sets of requirements for the infrastructure and services. Alongside the EOSC CZ Working Groups, a national data steward community has been established and has grown to over 100 professionals [L3] actively improving data management processes and standards within research institutions. Moreover, a new research assessment methodology provided by MEYS now recognises datasets as a standalone result type, providing a systemic incentive for awareness of the importance of research data.

Reflecting on the current progress

While the infrastructure is being built, the EOSC CZ initiative conducted the most comprehensive assessment of Open Science and FAIR data practices in the Czech Republic to date, surveying 1,121 principal investigators, with a 29.5% response rate [2]. The results reveal a persistent gap between awareness and practice, reflecting slow adaptation of research workflows to FAIR principles.

About 58% of researchers report familiarity with FAIR data principles, yet two-thirds (66%) still store research data primarily on personal computers and 53% use portable devices (USB drives, CD/DVD, or external hard drives), exposing them to risks of loss, damage, or theft. National e-infrastructure remains underutilised, with only 14.5% adopting it during their projects. Repository adoption varies sharply by discipline: 64% of life science scientists upload data to repositories, compared to 39% in the humanities. Data Management Plans split the community – 44% see them as useful, 54% as a bureaucratic burden – though researchers with access to data stewards are significantly more likely to embrace FAIR practices. Notably, 62% of researchers are unfamiliar with EOSC CZ, the very initiative that is building the infrastructure for them.

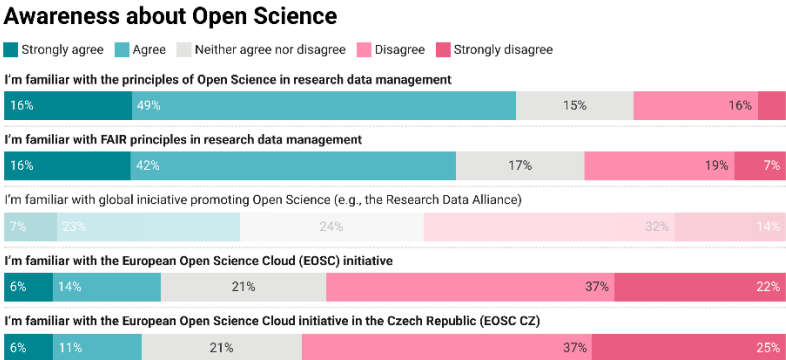


Figure 2: Awareness about Open Science.

The perception of bureaucracy also extends beyond data management planning. When asked what prevents researchers from sharing their data, researchers primarily point to administrative and legal concerns – copyright and licensing issues top the list of obstacles, flagged by 63% of respondents. Over 40% of respondents simply do not know which repository to use, directly mirroring the survey's finding that 62% are unfamiliar with EOSC CZ itself. Fears of misinterpretation and misuse of data (around 50% and 40%, respectively) suggest that reluctance to share is not merely administrative but rooted in a broader lack of trust and guidance.

What kind of obstacles do you see in sharing the research data?

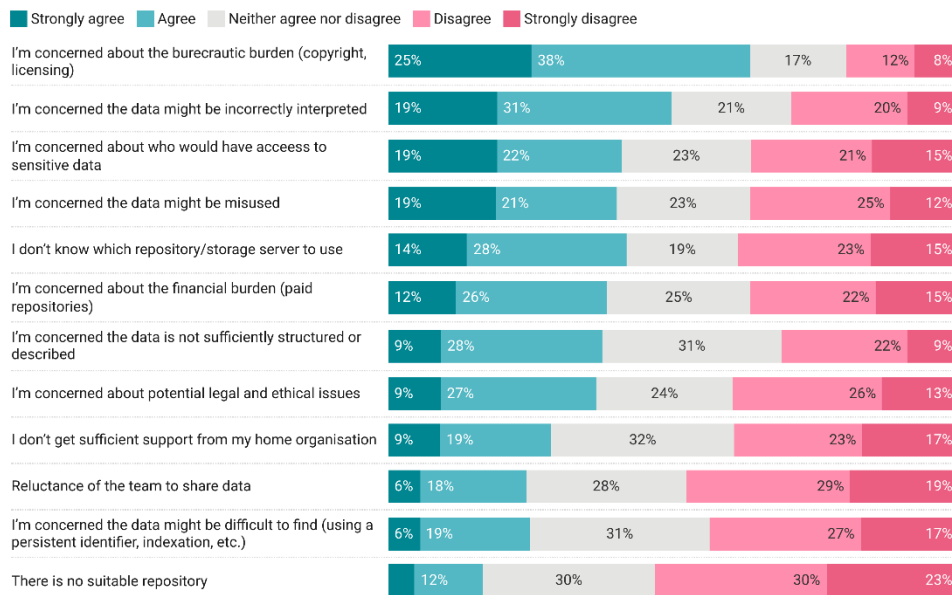


Figure 3: Perceived obstacles in sharing the research data.

The imminent future

The survey findings make clear that building technical infrastructure is not enough. Researchers' reluctance to share data stems not only from practical gaps – not knowing which repository to use or how to structure data – but from deeper concerns about trust, misinterpretation, and administrative burden. Shifting this mindset is as important as deploying the tools themselves, and the infrastructure must be designed to actively address these concerns rather than simply meet technical standards.

The community has been established, and the infrastructure is entering its operational phase. Its real value, however, will materialise only if the barriers and needs identified through the surveys are systematically addressed, and the heterogeneity of scientific domains and their specific requirements is fully acknowledged. In the end, building a national infrastructure for research data requires continuous alignment between researchers' evolving needs and the initiative's design, services, and priorities.

References:

- [1] M. Antol et al., "EOSC CZ: Towards the development of Czech national ecosystem for FAIR research data," 2024, <https://arxiv.org/abs/2402.13343>.
- [2] M. Vávra et al., "Průzkum správy výzkumných dat v České republice 2025," CSDA, 2026. <https://doi.org/10.14473/CSDA/OJTMJF>.

Links:

- [L1]: Generalist Repository, available at <https://datarepo.eosc.cz>
- [L2]: Czech Core Metadata Model, available at <https://www.cmm.cz/en/>
- [L3]: Map of Data Stewards in Czech Republic, available at <https://arcg.is/0HGKPr3>

Figures:

- Figure 1: Members of the EOSC CZ Working Groups
- Figure 2: Awareness about Open Science
- Figure 3: Perceived obstacles in sharing the research data

Please contact:

Martin Dvorak

Masaryk University, Czech Republic

dvorak@ics.muni.cz