

## Press release: Paving the way for machine learning with medical data

### Data science specialist CASUS and the PIONEER research consortium ally in the fight against prostate cancer

The Center for Advanced Systems Understanding (CASUS) at the Helmholtz-Zentrum Dresden-Rossendorf (HZDR) has joined PIONEER, a € 12.8m euro project funded by the public-private partnership Innovative Medicines Initiative 2 (IMI2). The HZDR is PIONEER's 36th member. PIONEER aims to transform the field of prostate cancer care by unlocking the potential of big data and big data analytics. Spread all across Europe, databases from clinical studies, public registries and electronic health records contain clinical data from thousands of prostate cancer patients. PIONEER collects, anonymizes and assembles these diverse data sets. CASUS takes over the task of providing a new centralised data and analytics platform for PIONEER. The cloud-based platform will provide data access and machine learning analytics capabilities for both academia and industry researchers. PIONEER operates both a central and federated model of data sharing. For the federated model, CASUS will take on the challenge of establishing a federated analytics network. The use of both data sharing models has allowed PIONEER to maximise both data protection and data utilisation.

Prof. James N'Dow, Academic lead for PIONEER welcomes new partner HZDR to the consortium: "The expertise of HZDR's CASUS in large-scale data management will provide a secure, scalable and sustainable infrastructure to host the PIONEER Prostate Cancer Big Data Platform. We are excited to embark on this next stage of PIONEER with CASUS".

Besides providing the PIONEER Big Data Platform cloud infrastructure, CASUS will also set up and support federated data analysis for all members of the consortium. For Dr. Michael Bussmann, Scientific Head of the Görlitz (Germany)-based research center, this aspect is of paramount importance: "By developing advanced machine learning algorithms, we expect to come up with better predictive models of patient outcomes and disease progression. The focus is on established and new clinical and biological indicators, so-called biomarkers. We will try to find out if and how recording such biomarkers improves predictions throughout a prostate cancer patient's care pathway".

### Safeguarding data access and data protection

Many stakeholders throughout the healthcare system collect medical data. Data-driven machine learning is considered a powerful tool for analyzing these data. To achieve this, however, data must all be available in a common format and all data protection concerns must be adequately addressed.

PIONEER operates with two data access models – a central and federated model. In the central model, a copy of the data is transferred to PIONEER, converted and stored in a central data warehouse for research. In the federated model, data owners standardize their own data sets and set up analytical tools within their own data environment supplying PIONEER with aggregated results from requested analytic tasks. In this data access model, the data does not leave its original site. Data from a variety of sources are effectively temporarily "linked" in order to address

specific remote queries with PIONEER bring the analysis to the data. Within PIONEER, CASUS will be responsible for the coordination and management of both data utilisation models.

Within PIONEER the data has been redacted to ensure sufficient anonymity such that the identification of the person to whom the data relates is virtually impossible. As a result of this anonymization process the data within PIONEER's Big Data Platform is not classified as personal data and as such the use of the data complies with all applicable data protection laws at the EU level. These data fall outside the scope of the EU's General Data Protection Regulation while maintaining their clinical relevance.

### **Open questions in prostate cancer research**

In general, PIONEER aims to both identify and close knowledge gaps in prostate cancer research. Among the most pressing open questions determined so far are: What are the relevant tumor-specific and patient-specific variables that affect prognosis of prostate cancer patients suitable for active surveillance? What is the natural history of prostate cancer patients undergoing conservative management (i.e., watchful waiting) and what is the impact of comorbidities and life expectancy on long-term outcomes? By scrutinizing data from diverse populations of prostate cancer patients across different stages of the disease (and from different European countries) PIONEER is expected to provide evidence-based answers to these questions to facilitate improved shared-decision making between physicians and patients. The final goal is to not only improve prostate-cancer related outcomes but also to increase healthcare system efficiency and the overall quality of health and social care.

At present the PIONEER platform, both central and federated, consists of a network of 29 data sets from consortium partners, industry and associated data partners. Of these, eleven datasets have been mapped to the European Common Data Model (CDM) of the Observational Medical Outcomes Partnership (OMOP), with mapping ongoing or about to start for an additional eight datasets. Once complete, the PIONEER Big Data platform will cover 1.8 million prostate cancer patients.

By combining complex systems research with cutting-edge digital methods from data and computational science, PIONEER aims to play a pivotal role in Europe's research landscape. The collaboration between PIONEER and CASUS was initiated by PIONEER member Prof. Manfred Wirth, Senior Professor at Dresden University of Technology and former Director of the Department of Urology at the University Hospital Carl Gustav Carus in Dresden (Germany). PIONEER is looking forward to a fruitful collaboration with CASUS as we enter this exciting phase of the project.

*PIONEER (Prostate Cancer DiagnOsis and TreatmeNt Enhancement through the Power of Big Data in EuRope) is a European project focused on using big data to improve the clinical understanding and inform the diagnosis and treatment of prostate cancer, funded through IMI2 JU, and is listed under grant agreement No. 777492. [www.prostate-pioneer.eu](http://www.prostate-pioneer.eu)*

*CASUS was founded 2019 in Görlitz/Germany and pursues data-intensive interdisciplinary systems research in such diverse disciplines as earth system research, systems biology or materials research. The goal of CASUS is to create digital images of complex systems of unprecedented fidelity to reality with innovative methods from mathematics, theoretical systems research, simulations as well as data and computer science to give answers to urgent societal questions. Partners are the Helmholtz-Zentrum Dresden-Rossendorf (HZDR), the Helmholtz Centre for Environmental Research in Leipzig (UFZ), the Max Planck Institute for Molecular Cell Biology and Genetics in Dresden (MPI-CBG), the Technical University of Dresden (TUD) and the University of Wrocław. CASUS is funded by the Federal Ministry of Education and Research (BMBF) and the Saxon State Ministry for Science, Culture and Tourism. [www.casus.science](http://www.casus.science)*

**Additional information:**

*Dr. Michael Bussmann | Scientific Head*

*Center for Advanced Systems Understanding (CASUS) at HZDR*

*phone: +49 3581 375 2311, +49 351 260 2616 | e-mail: m.bussmann@hzdr.de*

*Dr. Martin Laqua | Officer Communications, Press and Public Relations*

*Center for Advanced Systems Understanding (CASUS) at HZDR*

*phone: +49 3581 375 23 24 | e-mail: m.laqua@hzdr.de*

*Emma Jane Smith | PIONEER Communications*

*European Association of Urology*

*phone: +31 26 389 0680 | e-mail: e.smith@uroweb.org*