Questions and answers are in data – the development of eInfrastructure and Digital Health technologies

Cezary Mazurek

Poznan Supercomputing and Networking Center
affiliated to the Institute of Bioorganic Chemistry,
Polish Academy of Sciences

EU Digital Health Tech
& Innovation Conference
Brussels, 27th September 2022

Institute of Bioorganic Chemistry Polish Academy of Sciences

- The Institute of Bioorganic Chemistry of the Polish Academy of Sciences (IBCh PAS) is a unique research unit in Europe, which carries out interdisciplinary research in chemistry, biology, bioinformatics and computer science
- Together with its affiliated Poznań Supercomputing and Networking Center (PSNC) it is one of the largest institutes of Polish Academy of Sciences, employing in total over 750 people
- The core mission of PSNC is to foster scientific excellence by providing reliable and cutting-edge e-Infrastructure such as communication networks, data and supercomputing systems, as well as highly-specialized laboratories

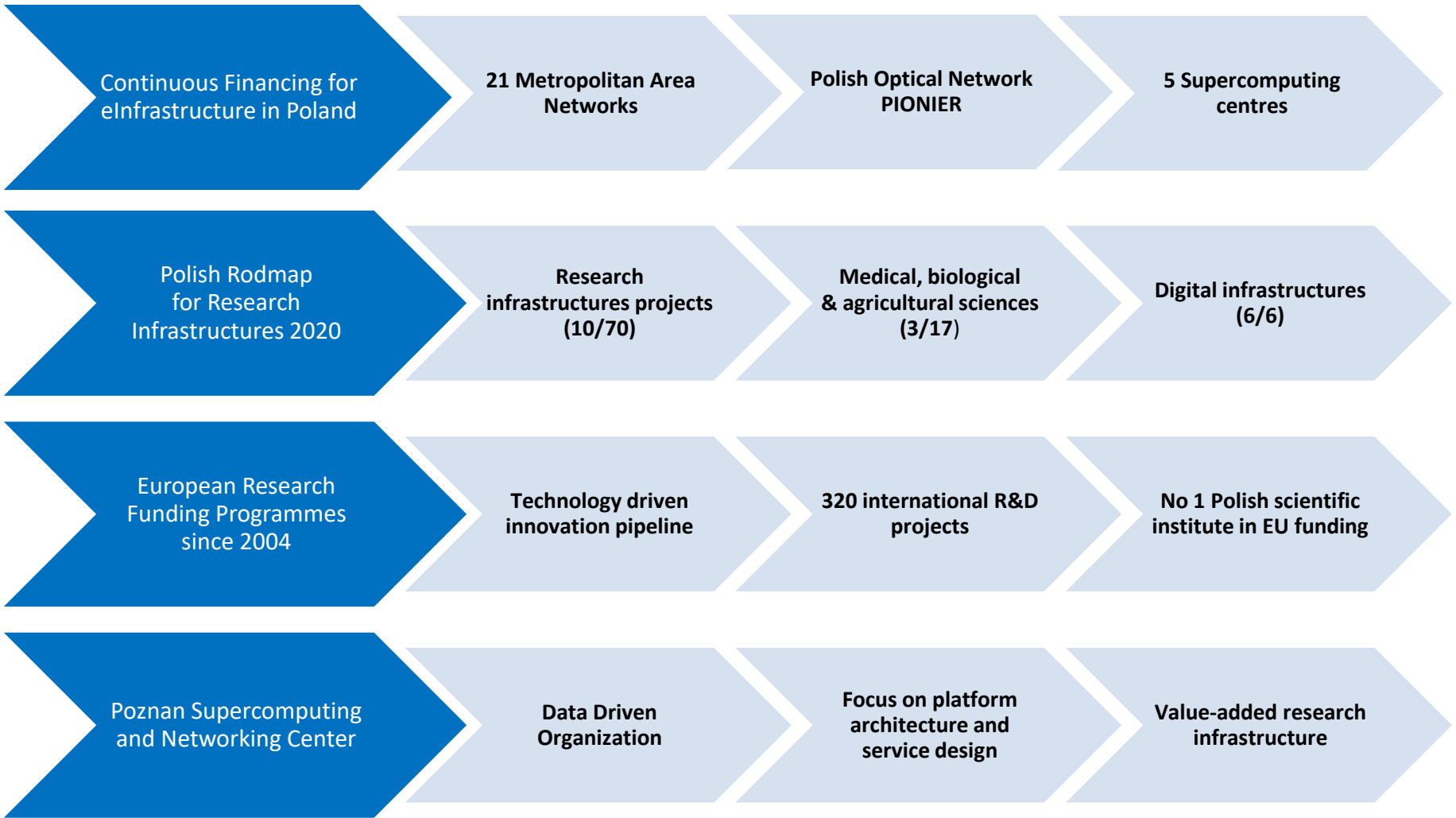
PSNC Mission

- The core mission of PSNC is to foster scientific excellence by providing reliable and cutting-edge e-Infrastructure such as communication networks, data and supercomputing systems, as well as highly-specialized laboratories
- We achieve this goal particularly through extensive R&D activities related to information and communication technologies and their innovative applications.
- In collaborative project teams we draw from the synergy of science, technology, business, and education to tackle the greatest challenges for our future.
- It is our highest objective to ensure steady growth of the competencies of our people. They are the pillar for the successful exploration of new technologies and the development of scientific and engineering innovations, and ultimately the winning deployment of our services.

Locations

Focus on Digital Transformation based on advanced einfrastrucutre

- Continuous R&D activities related to information and communication technologies and their innovative applications



28
years of activity

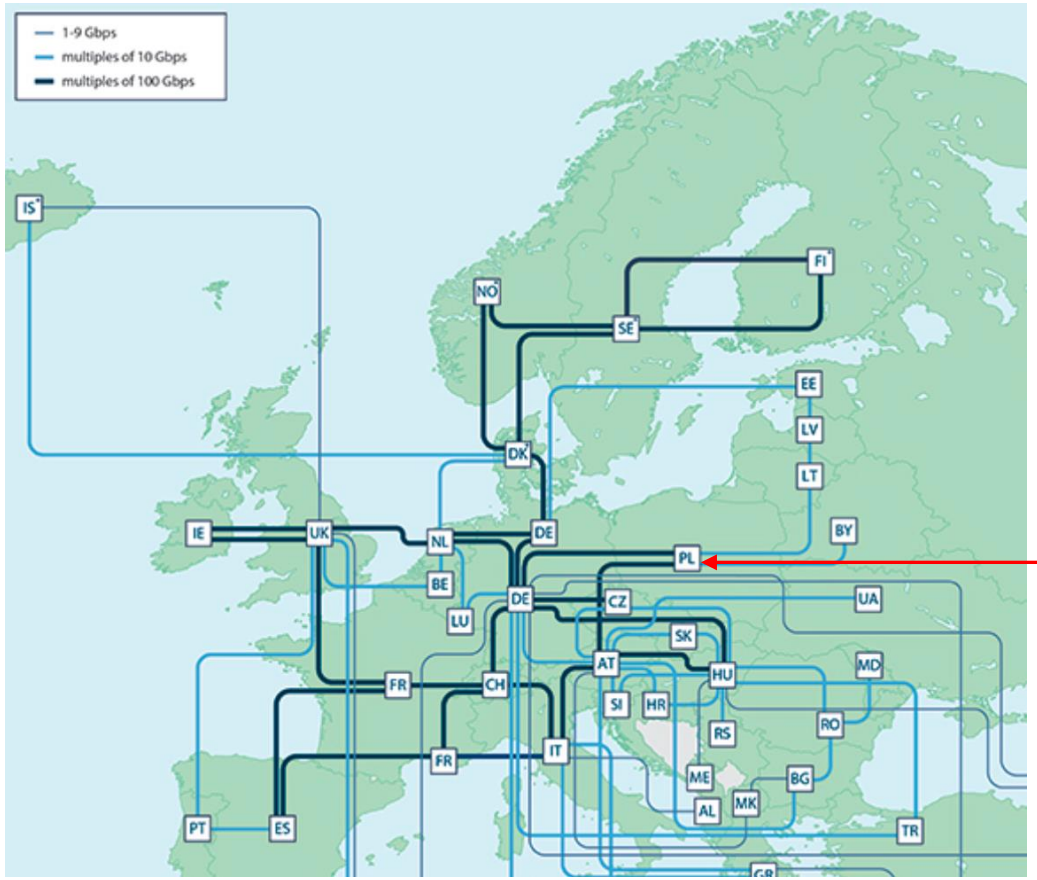
1.34M*
scientific users
* all users served for 21 MANs

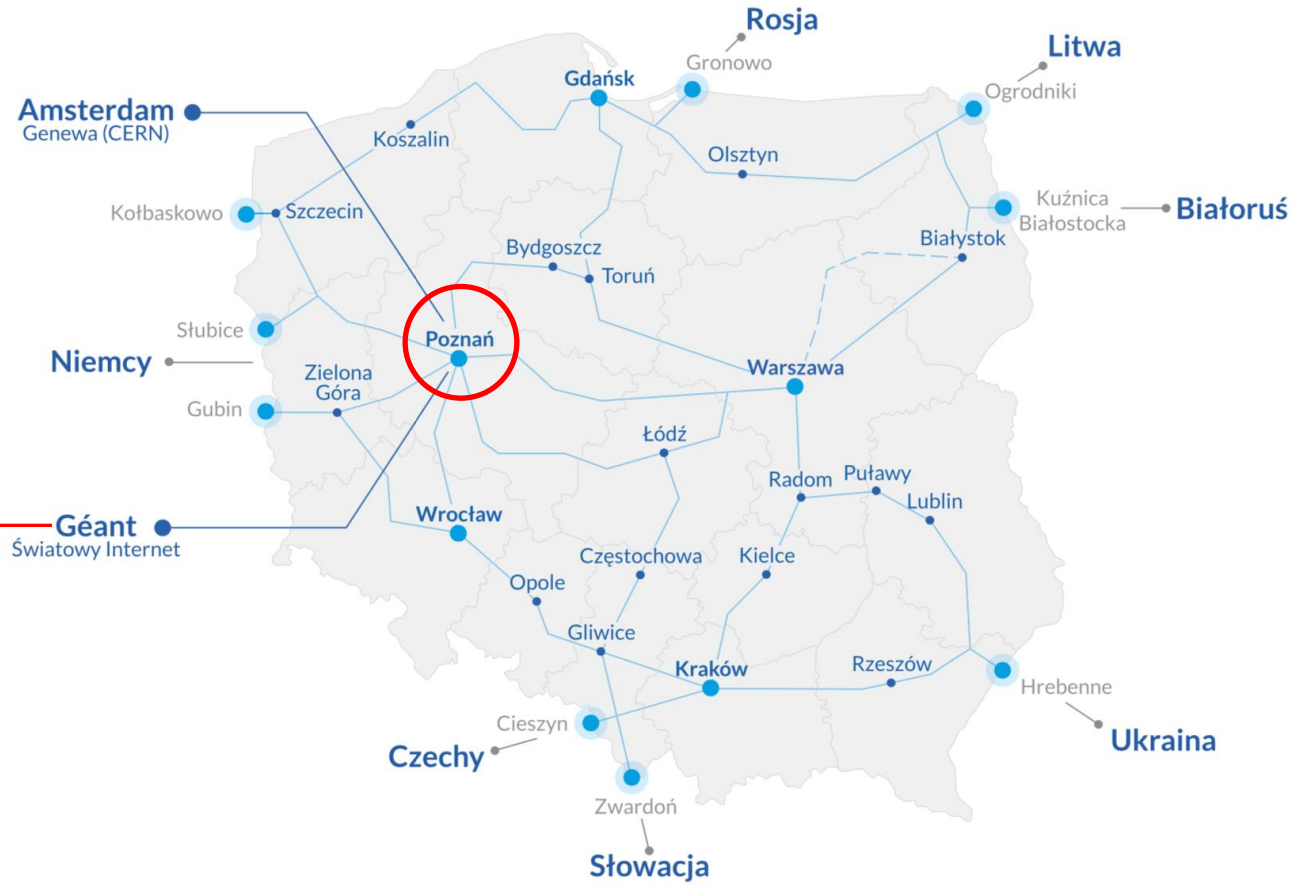
1680*
customers
* since 2016

432
employees

301
projects

Networking: GEANT and PIONIER (NREN in Poland operated by PSNC)

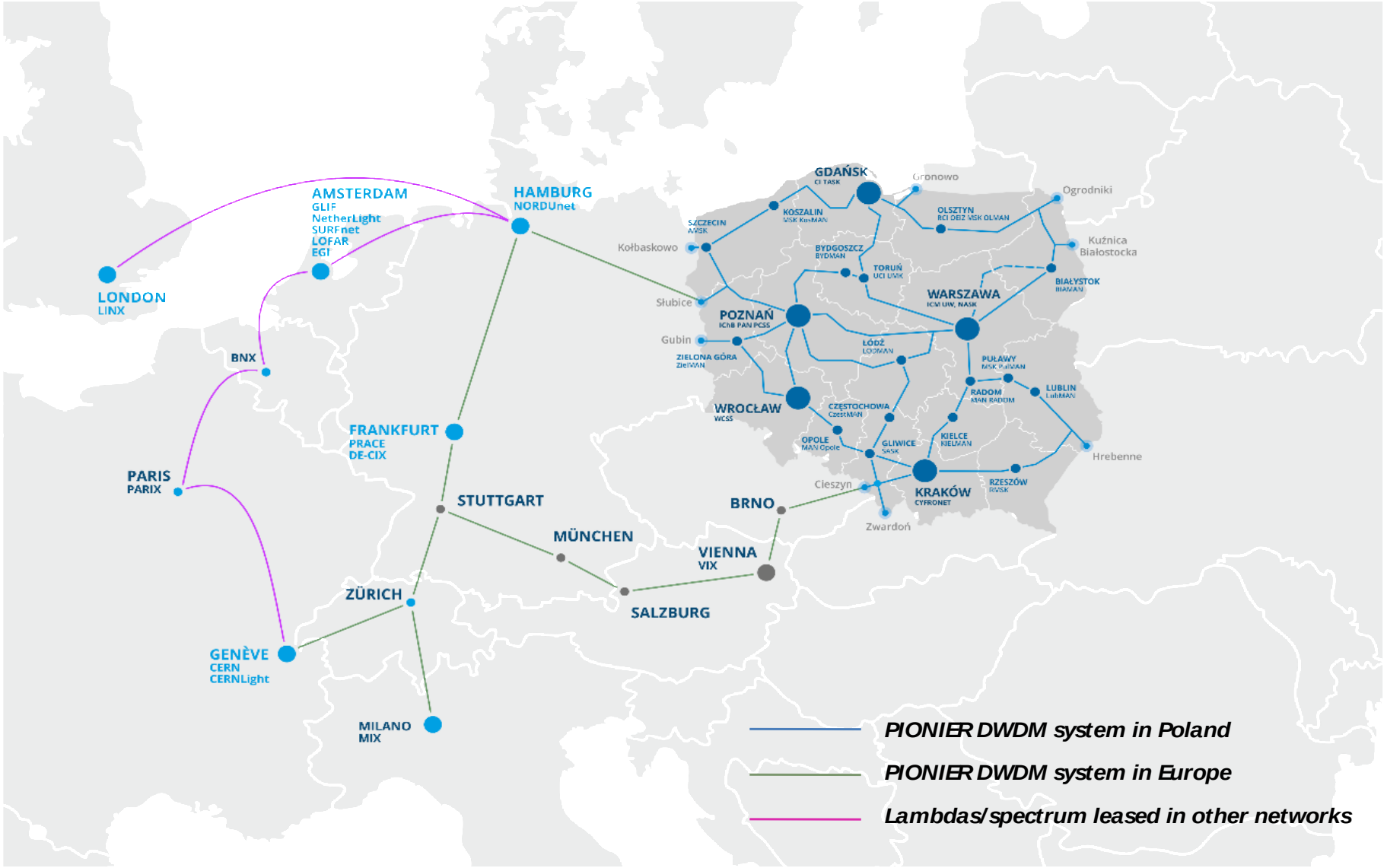




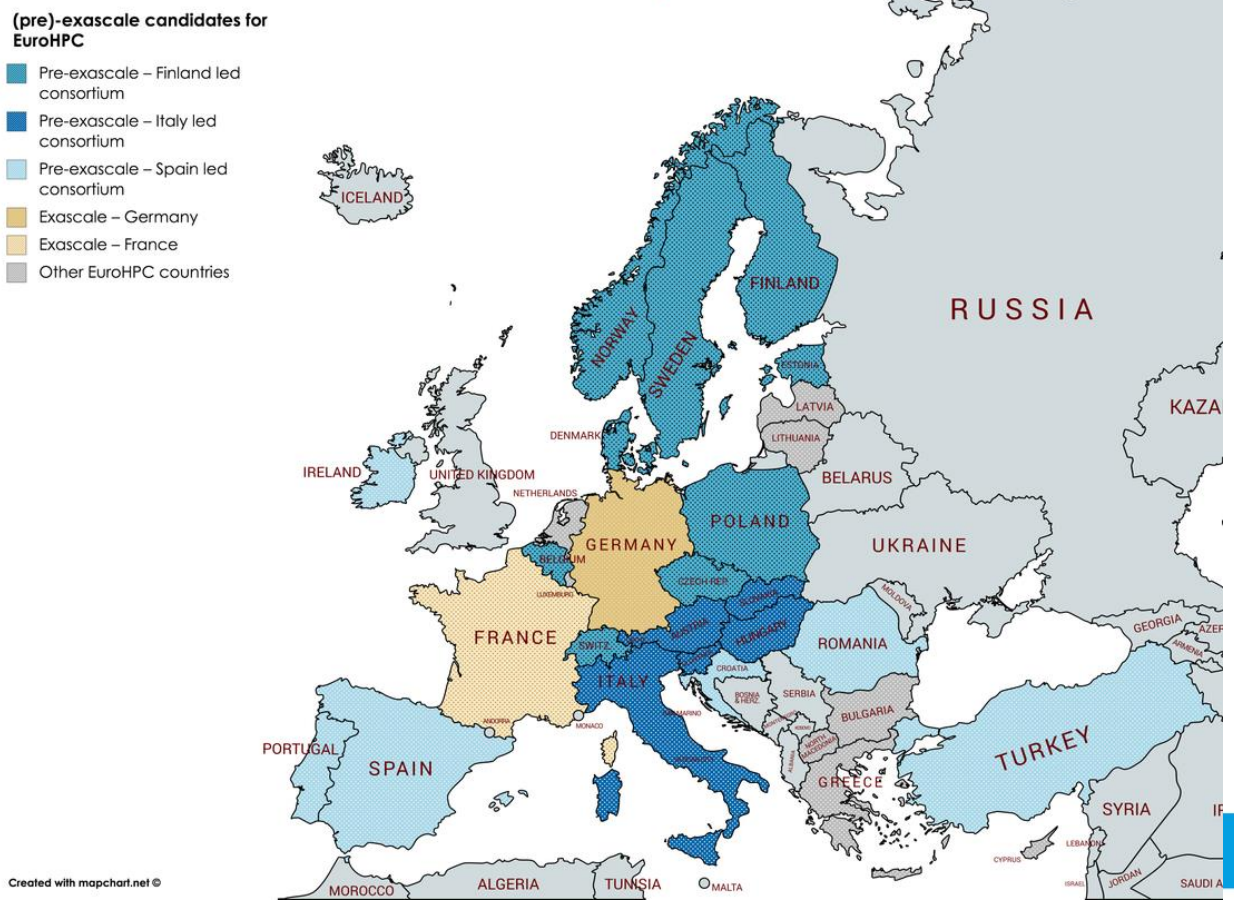
GÉANT
www.geant.net

PSNC

PIONIER – Polish Optical Internet



High Performance Computing: PRACE/EuroHPC and PRACE-LAB I/II in 2021



PRACE-LAB:

Instytut Chemii Bioorganicznej PAN, PCSS – Lider

Akademickie Centrum Komputerowe CYFRONET AGH

Politechnika Białostocka

Politechnika Częstochowska

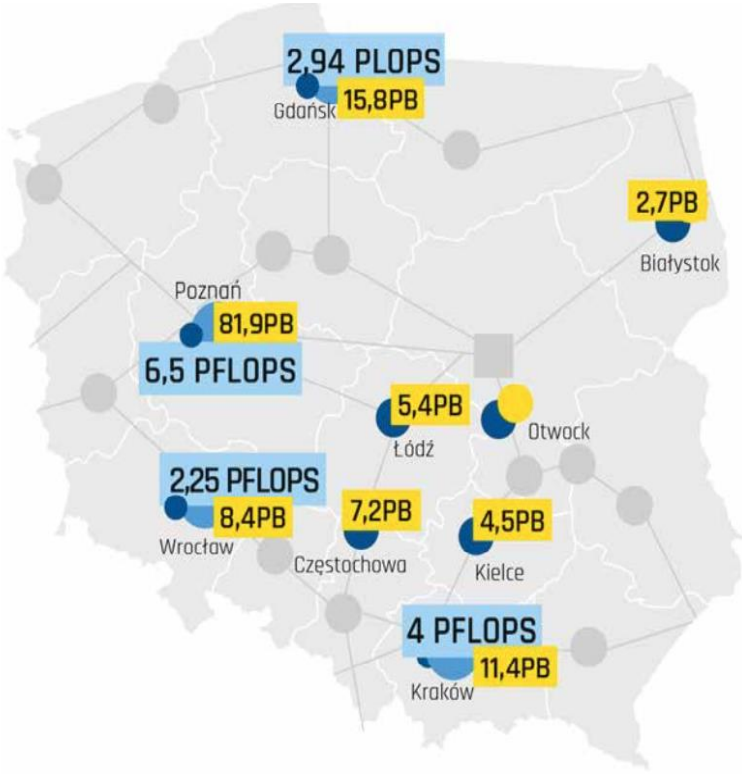
Politechnika Gdańska CI TASK

Politechnika Łódzka

Politechnika Świętokrzyska

Politechnika Wrocławska – WCSS

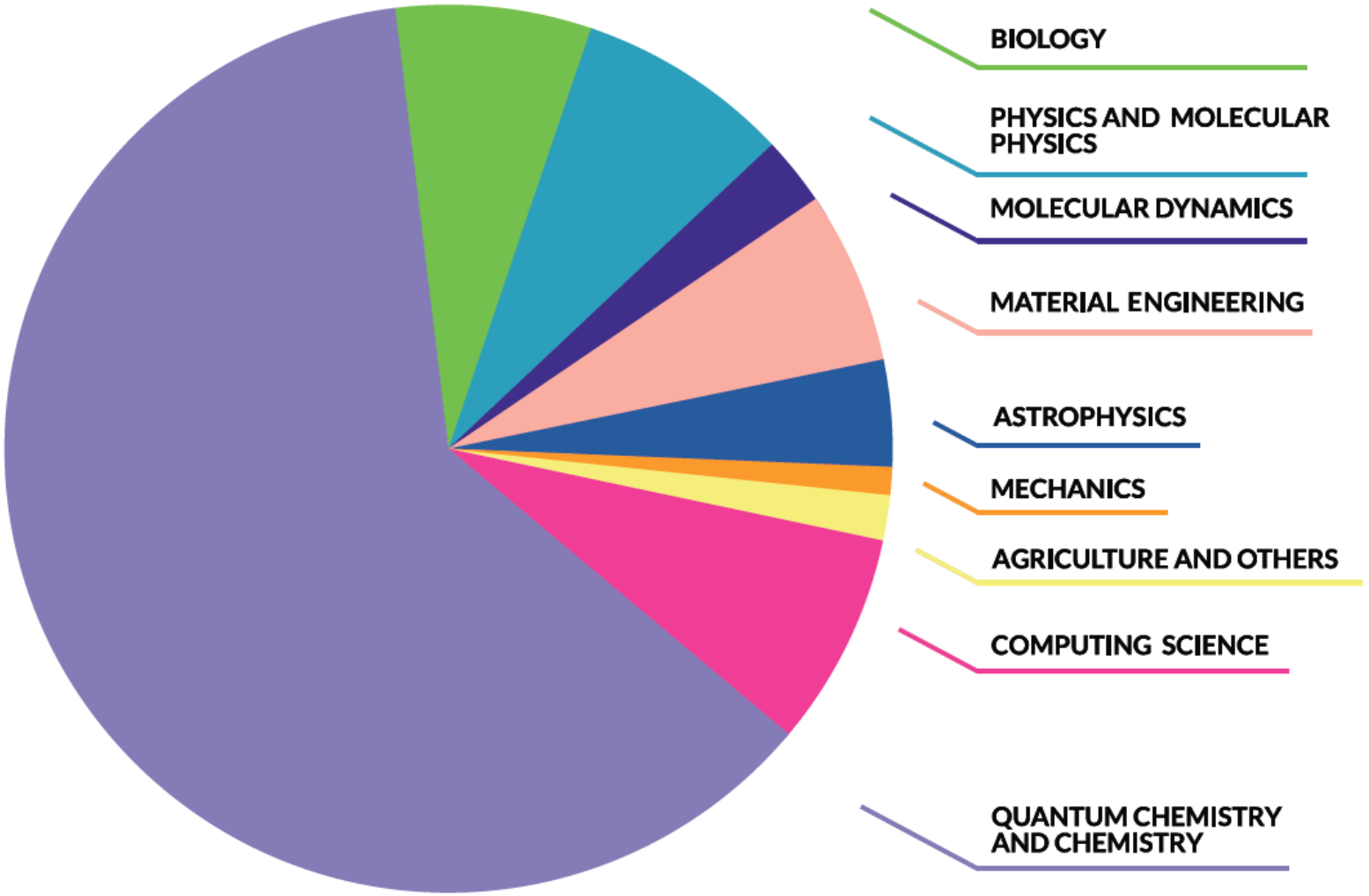
prace-lab.pl



High Performance Computing @PSNC in 2021

- All major HPC services are provided by two systems: Eagle (since 2016) and Altair (since 2021), which deliver conventional power and a specialized accelerating environment GPGPU for AI and big data analysis.
- The remaining cloud services are delivered for specific requirements at universities (dedicated resources, higher availability and critical services), industry and administration entities.
- PSNC computing systems have defined 200 grants and 5,000 computing accounts in both HPC and clouds.
- Access is provided for national and international users based on the contracts signed for the exchange of computing power under:
    - Partnership for Advanced Computing in Europe (PRACE),
    - nuclear physics—Worldwide LHC Grid Computing (WLCG),
    - Low-frequency Array for radio astronomy (LOFAR),
    - number of international flagship ESFRI projects.

High Performance Computing @PSNC in 2021 – core-hours/disciplines

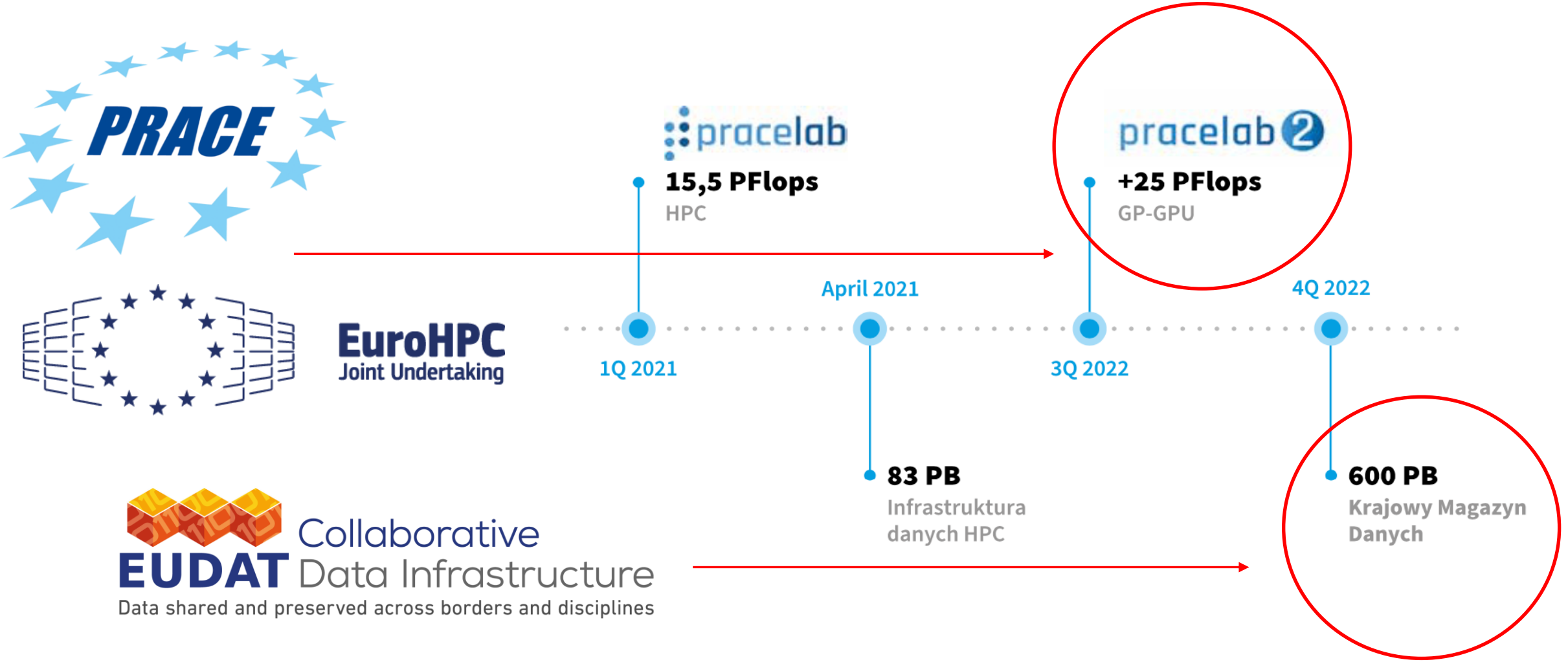


PRACE (Partnership for Advanced Computing in Europe) grant:

Prof. Agnieszka Kaczor
(Medical University of Lublin)

"Molecular simulations of drug targets involved in the central nervous system diseases (NERVOMOLSIM)"

National Data Storage and HPC Infrastructures in Poland in 2021 and 2022



Polish Rodmap for Research Infrastructures 2020 – PSNC projects

01 – 01 PIONIER-LAB National Platform for Integrating Research Infrastructures with Ecosystems of Innovation

02 – 02 National Laboratory for Photonics and Quantum Technologies - NLPQT

03 – 02 PRACE - Cooperation in the field of advanced computing in Europe

04 – 03 PRACE-LAB2 - Cooperation in the field of advanced computing in Europe

05 – 03 NEBI - National Centre for Advanced Analysis of Biological and Biomedical Imaging

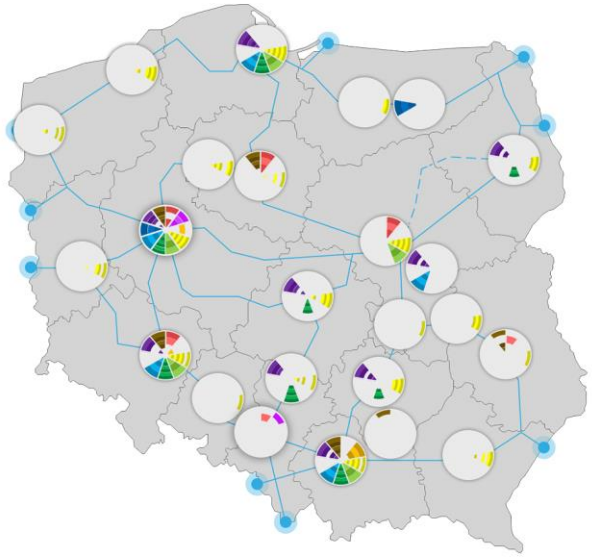
06 – 04 Digital Research Infrastructure for Arts and Humanities DARIAH-PL

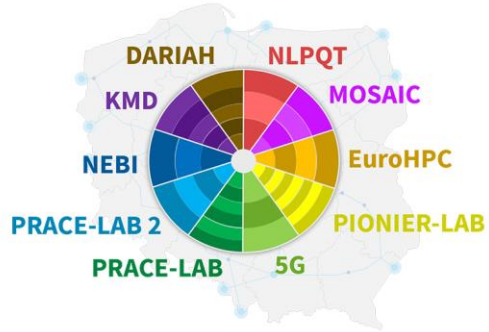
07 – 04 National Laboratory for Advanced 5G

08 – 04 National Data Storage

09 – 04 National Supercomputing Infrastructure for EURO HPC PL

10 – 04 ECBIG - The European Center for Bioinformatics and Genomics – MOSAIC





R&D Activities, Projects, People

Healthcare – our emphasis on data management and integration towards better diagnosis and treatment

- ICT
  - ADMIRE - Adaptive multi-tier intelligent data manager for Exascale (Brain Super-resolution imaging)
  - INSENSION - Personalized intelligent platform enabling interaction with digital services to individual with PMLD (Profound and multiple learning disabilities)
  - GlaucomAI - Multimodal Advanced Glaucoma Diagnosis Model
  - medVC - Real-time audio-video collaboration for doctors
- AAL
  - PELOSHA: Personalizable services for supporting healthy ageing
  - Fit4Work: Self-management of physical and mental fitness of older workers
- eInfrastrucutre
  - NEBI – National Imaging Centre for biological and biomedical sciences
  - MOSAIC – AI Platform to integrate and analyze multiomics and clinical data for new insights and tools for broadly accessible, personalized prevention, diagnosis and medical therapy

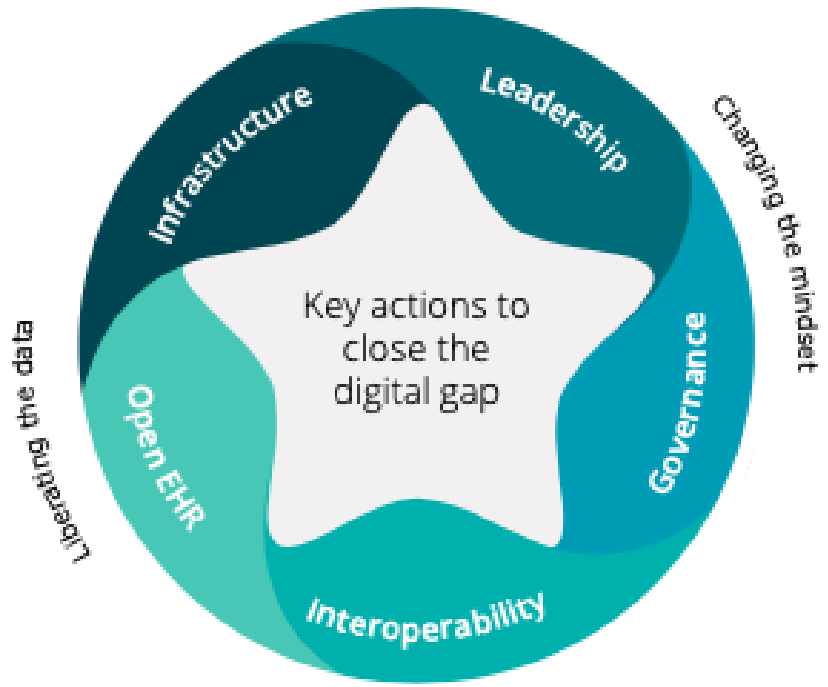
Digital Transformation: Shaping the future of European healthcare – Deloitte report (2020)

Top challenges facing HC organizations in implementing digital technologies

Europe

1. **Bureaucracy in healthcare (57.4%)**
2. **Cost of technology (50.3%)**
3. **Finding the right technologies (49.0%)**
4. Training staff to adequately use technology (35.8%)
5. Complexity of technology (28.9%)
6. Challenges in sharing patient data (27.7%)
7. Convincing staff of the benefits of technology (20.5%)
8. Scaling up the use cases of technologies (16.1%)
9. Existence of evidence of outcomes (12.5%)

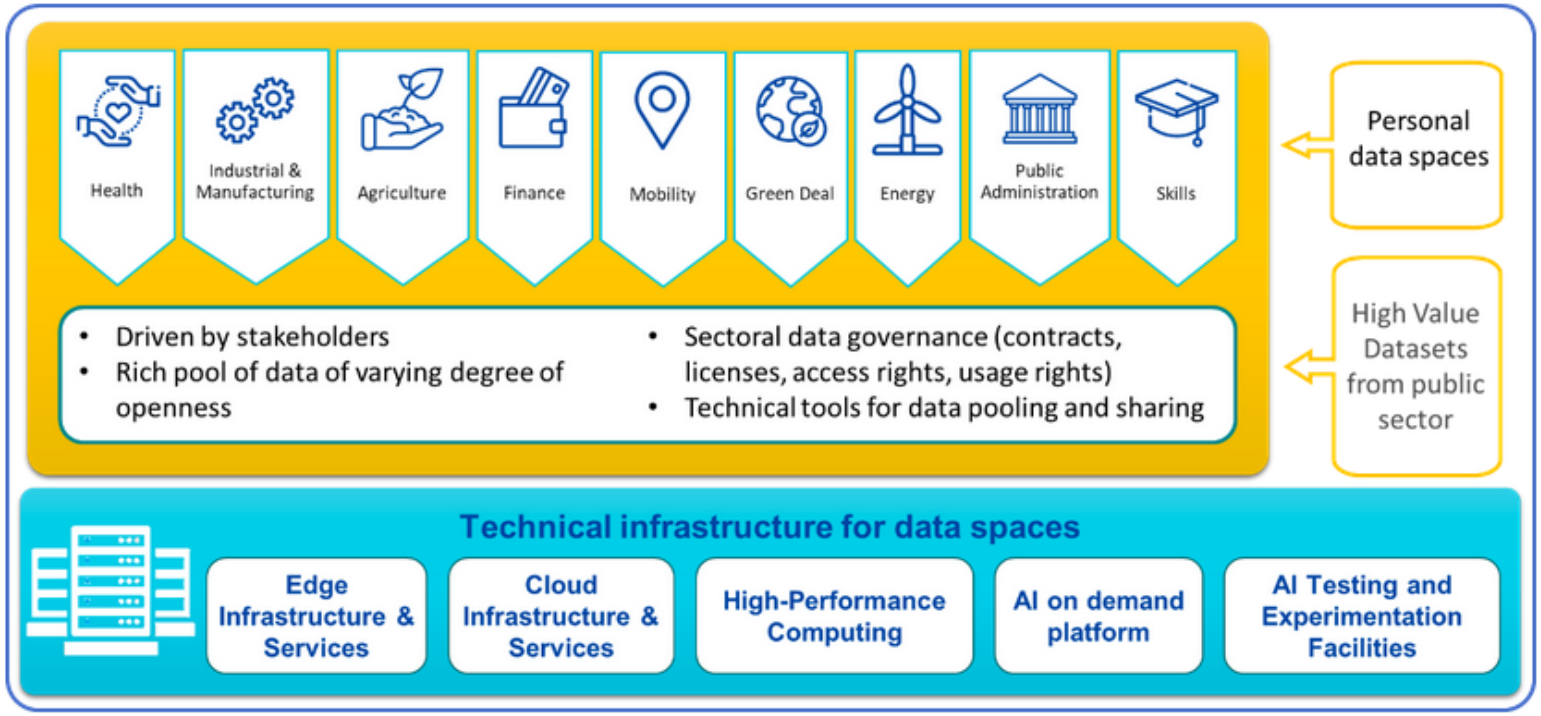
Figure 12. Key actions to close the digital gap

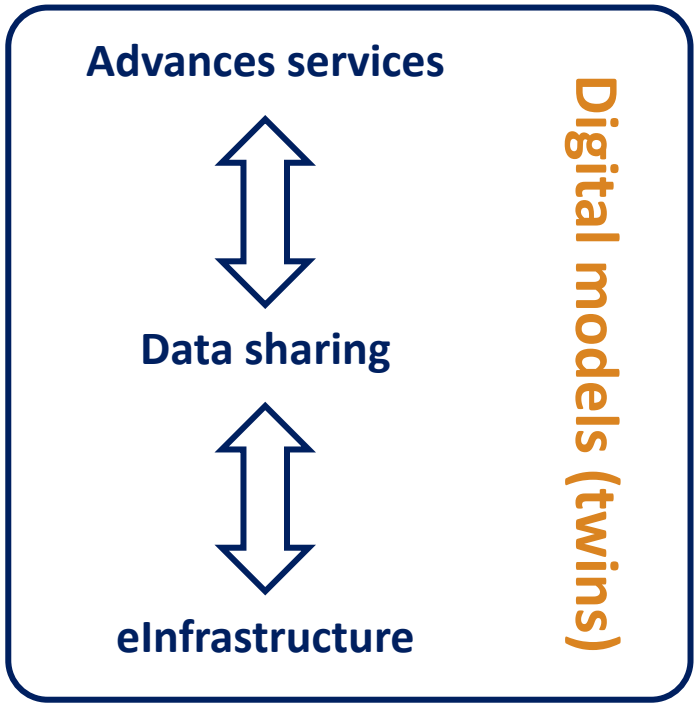


Source: Deloitte research and analysis, 2020

Common European Data Spaces

- The medicine of today is increasingly employing IT technologies to better understand the observed processes and support diagnostic and individual therapeutic procedures based on digital models





https://digital-strategy.ec.europa.eu/en/library/building-data-economy-brochure

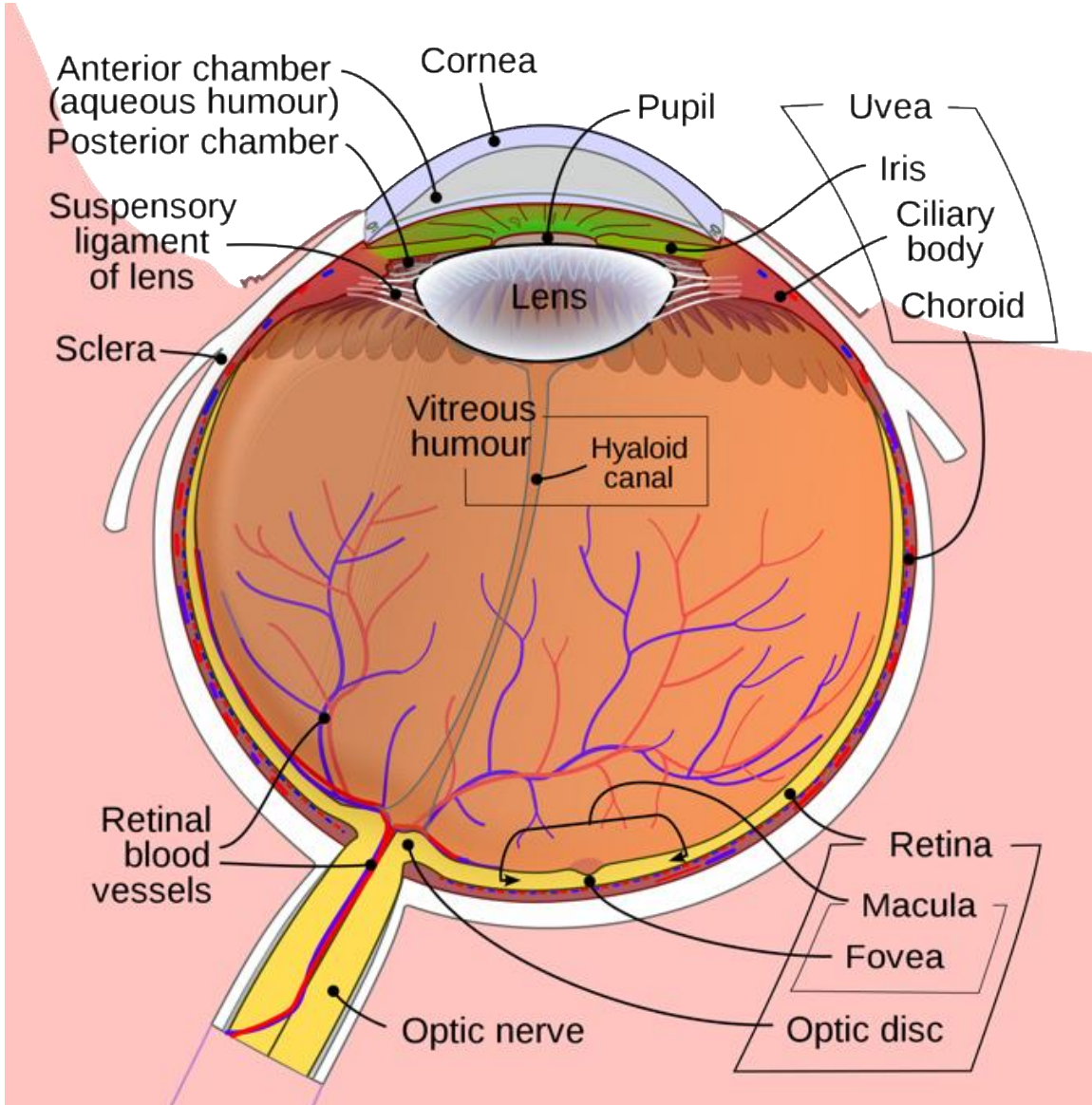
GlaucomAi - Multimodal Advanced Glaucoma Diagnosis

Glaucoma is a progressive optic neuropathy which is the most common cause of irreversible vision loss in the world.

Its pathomechanism is associated with a change in lamina cribrosa phenotype under the influence of excessive mechanical forces, which leads to neurotrophic deprivation and subsequent accelerated process of retinal ganglion cell apoptosis.

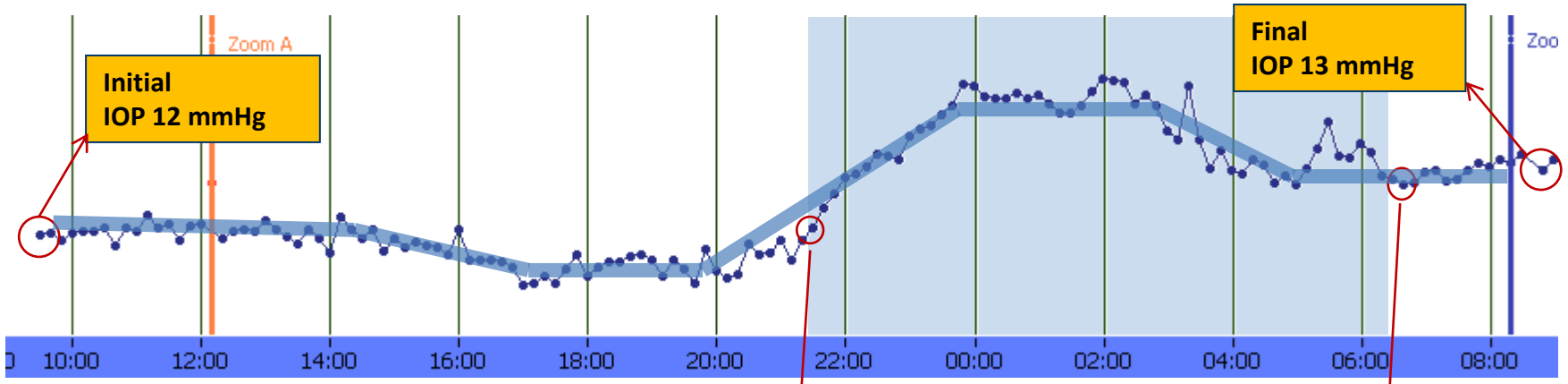
Loss of retinal ganglion cells leads to disruption of the functional visual pathway continuity which leads to the development of specific (depending on the architecture of the retina and optic nerve) defects in the visual field.

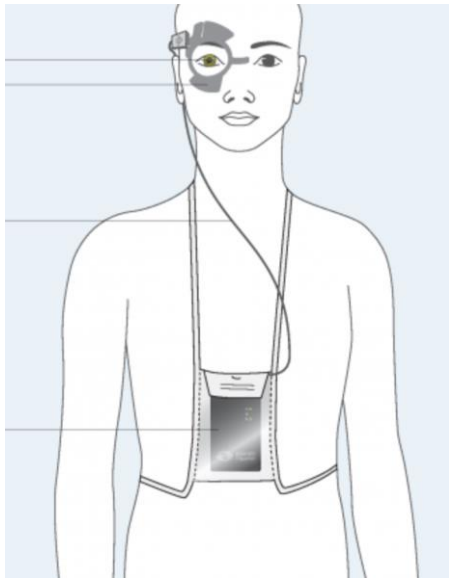
Those defects appear clinically when at least 30 to 50% of the ganglion cells of a given area of the retina is lost.

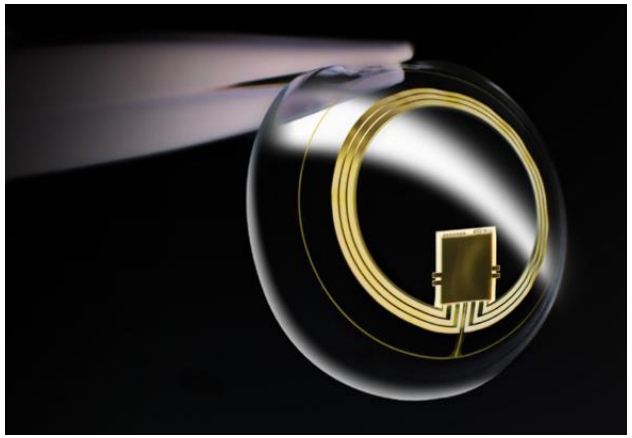


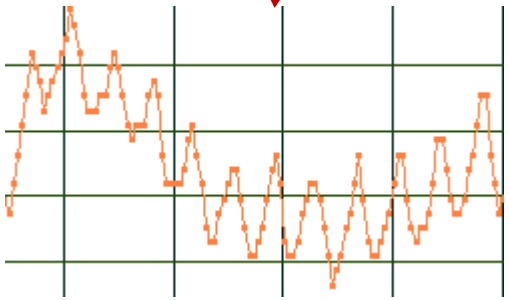
Multimodal Advanced Glaucoma Diagnosis

Data collected from SENSIMED Triggerfish®

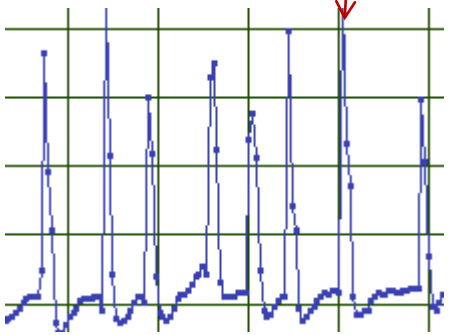








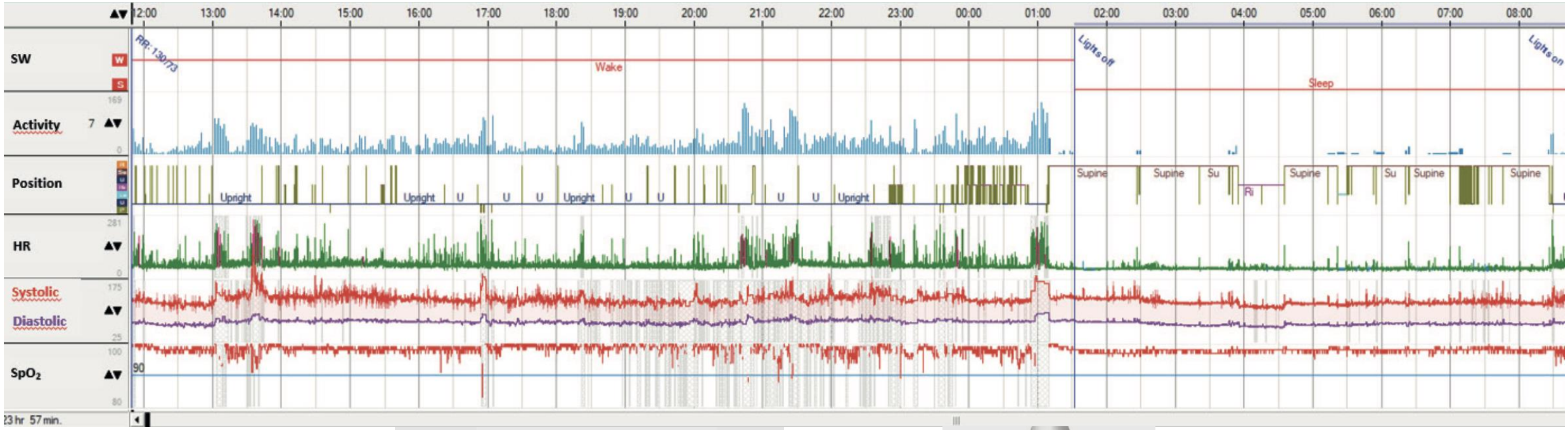
Ocular Pulse amplitude- during sleep

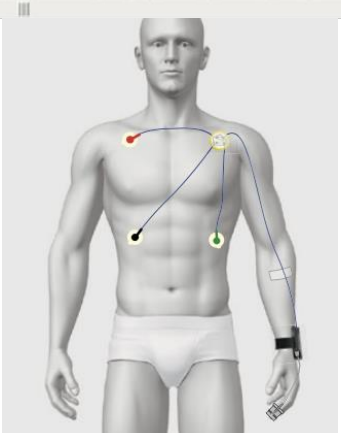


Eye blinking – high peaks

Multimodal Advanced Glaucoma Diagnosis

Data collected from SOMNOtouchTM NIBP



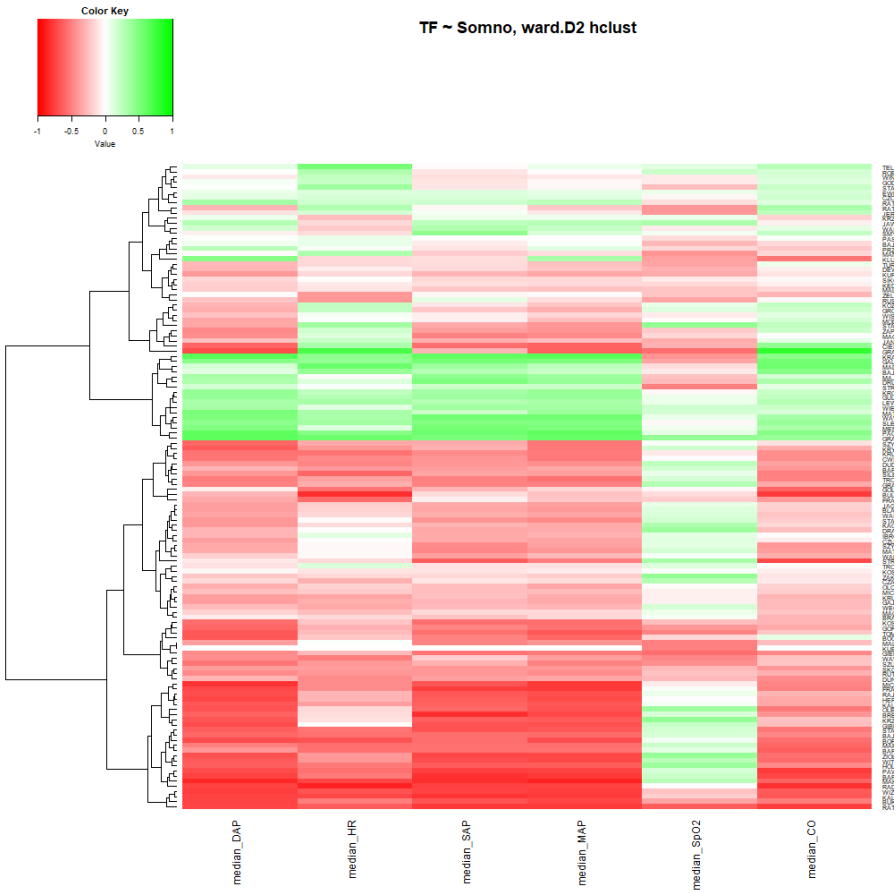


Atributes

Feature engineering during reference periods:

- Sum of areas under the TF curve
- The slope angle of the linear regression straight line for TF measures
- The total pace of change in TF
- Power and frequency of the main FFT components of the TF signal
- Correlations of TF and cardiovascular sensor signal (SAP, DAP, HR, SpO_2)
- Discrete convolution product of TF with HR, SAP for short periods after sleep (1-1.5h)

- Additional attributes: clinical metadata describing eyeball biomechanical properties (CH, CRF),

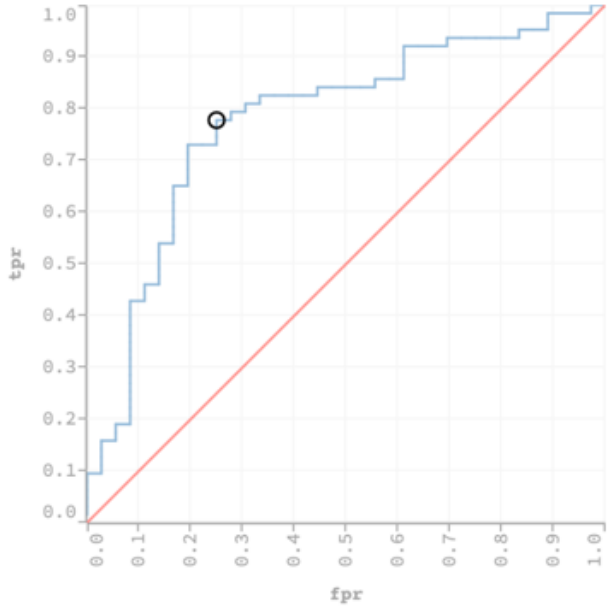


Multimodal Advanced Glaucoma Diagnosis

I 2021: 100 subjects, 8-fold cross-validation

Model #3: CH i sensor attribiutes, {NORM, POAG|NTG} (*GLM*)

▾ ROC CURVE - CROSS VALIDATION METRICS , AUC = 0.779101



Threshold: Choose...

Criterion: max accuracy

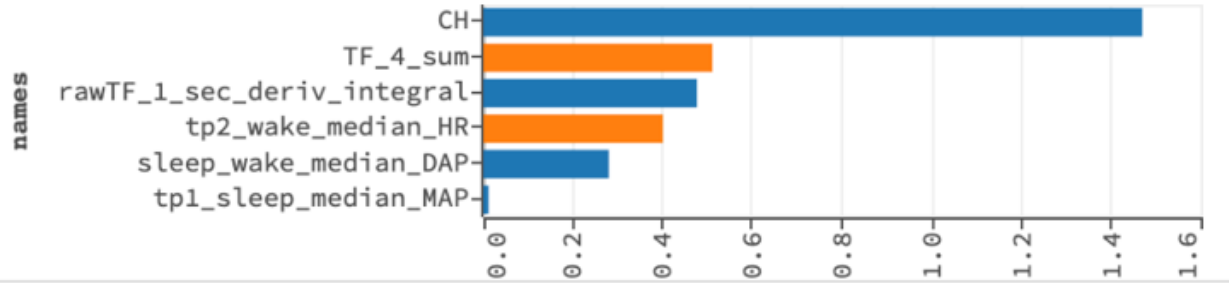
Selected mark(s):

threshold	0.6144
f1	0.8099
f2	0.7903
f0point5	0.8305
accuracy	0.7677
precision	0.8448
recall	0.7778
specificity	0.7500
absolute_mcc	0.5154
min_per_class_accuracy	0.7500
mean_per_class_accuracy	0.7639
tns	27
fns	14
fps	9
tps	49
tnr	0.7500
fnr	0.2222
fpr	0.2500
tpr	0.7778
idx	57

CM

Actual/Predicted	NORM	NOT_NORMAL	Error	Rate
NORM	27	9	0.2500	9 / 36
NOT_NORMAL	14	49	0.2222	14 / 63
Total	41	58	0.2323	23 / 99

▾ STANDARDIZED COEFFICIENT MAGNITUDES



Multimodal Advanced Glaucoma Diagnosis

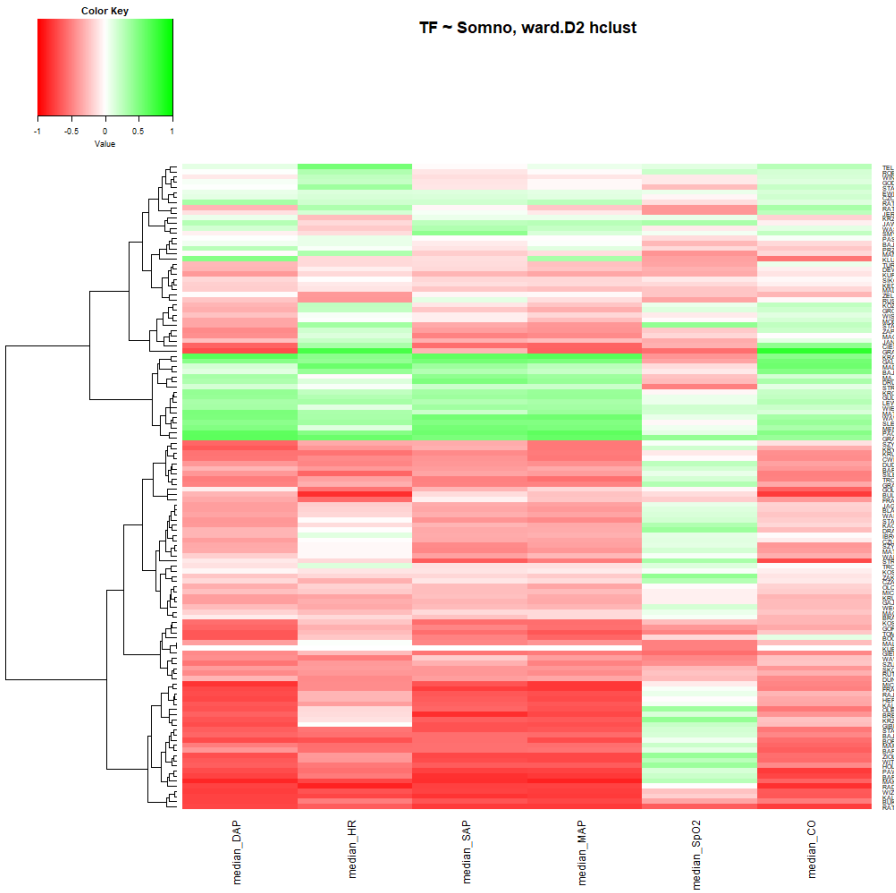
Table 3 Key diagnostic metrics (specificity, sensitivity, PPV, F-measure, AUC) as a result of using the various individual and combined sources

		Screening setpoint				Optimal ROC knee setpoint				
		Specificity	Sensitivity	PPV	F-measure	Specificity	Sensitivity	PPV	F-measure	AUC
Single source	Genome information (G)	0.850	0.543	0.143	0.227	0.744	0.760	0.122	0.183	0.810
	Retinal fundus image features (I)	0.850	0.420	0.115	0.180	0.720	0.662	0.101	0.166	0.722
	Personal data (P)	0.850	0.197	0.057	0.089	0.596	0.554	0.060	0.108	0.561
Combined source	Genome information + retinal image features (G+I)	0.850	0.647	0.166	0.264	0.764	0.812	0.139	0.192	0.856
	Personal data + genome information (P+G)	0.850	0.639	0.164	0.261	0.750	0.824	0.134	0.186	0.853
	Personal data + retinal image features (P+I)	0.850	0.454	0.123	0.193	0.714	0.718	0.107	0.171	0.753
	AGLAIA-MII: Personal data + genome information + retinal fundus image features (P+G+I)	*0.850*	*0.671*	*0.171*	*0.273*	*0.786*	*0.816*	*0.153*	*0.201*	*0.866*
Current glaucoma screening using IOP		0.850	0.304	0.086	0.134	0.614	0.560	0.064	0.118	0.604

For specificity, sensitivity and PPV, two conditions are considered: screening setpoint by setting specificity constant at 0.85, and at the optimal ROC knee point. The best results are highlighted in italics.
AGLAIA-MII, automatic glaucoma diagnosis through medical imaging informatics; AUC, area under the (ROC) curve; IOP, intraocular pressure; PPV, positive predictive value; ROC, receiver operating characteristic.

Challenges and future plans

- Number of cases to analyze - the disproportion between the parameters and the size of the set to analyze
- Data mining to determine the parameters of the case description - strictly controlled by the physician's expertise
- Feedback effect - model building based on medical knowledge and interpretation of physiology based on work with data and model
- Assessment of model reliability and validation of results
- Transition from the level of the diagnostic model to the proposal of therapeutic intervention for doctors



Multimodal Advanced Glaucoma Diagnosis

Result: Data Analytics App - patient classification

Glaucoma

Medical Diagnostic Hypothesis Refinement Through Data Analysis

Predict diagnosis for u_125

Describe this experiment...

Data analysis method: ○ Time series summary ○ Data distribution summary ○ Gradient Boosting Machine model ◉ Diagnosis assessment

Enter time (hh:mm)

TF start: 11:33 | Sleep from: 23:05 | Wake at: 07:18

Enter CH (numeric)

CH: 10.6

Select raw Triggerfish CSV file for patient

File selected 3-u_1...

Run conversion of the .CSV file

Select cardio record in zip file for patient

File selected 3-u_1...

Run conversion of the .zip file

Select classification model

◉ basic TF/Somno model

Zip file conversion has been successful. Now you can check classification.

Check classification

Predicted Diagnosis: NORM (*Confidence*: 1.60%)

Conlusions – Challenges for Digital Health solutions

- A multidisciplinary approach in medicine no longer means only medical teams working together
- For data-driven solutions, it is crucial to have access to sufficient number of records to avoid imbalances between numer of parameters and the size of the set to analyze
- Data mining leading to determination of the crucial parameters which characterize the case and disease should be strictly supervised by the physician's expertise
- Feedback loop for improved understanding – modeling based on medical knowledge and physiology cross-referencing while working with data and model
- Advancing beyond diagnosis support towards specific therapeutic interventions suggested for physicians

Thank you!

Contact:

Cezary Mazurek
mazurek@man.poznan.pl

EU Digital Health Tech
& Innovation Conference
Brussels, 27th September 2022