

Citizen Science in Data-Intensive Physics: PUNCH4NFDI Perspective

V. Tokareva, M. Kramer, A. Haungs, R. Karuppusamy

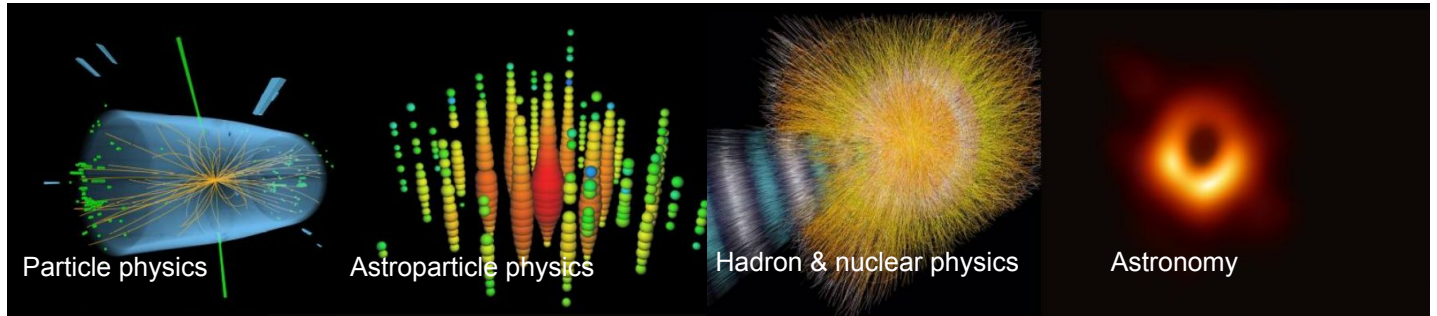


Max-Planck-Institut
für Radioastronomie

Particles,
Universe,
Nu-
Clei and
Hadrons
4 (for)
Nationale
Forschungs-
Daten
Infrastruktur



is the consortium of **particle, astroparticle, hadron & nuclear physics and astronomy** (as well “PUNCH Communities”), representing ~10000 scientists in Germany
Data-Intensive Physics: Big Data, “heavy” complex computations



Citizen Science in the context of PUNCH4NFDI TA7

WP1. Training

For foster expertise and career prospects

WP2. Education

Educational and data resources for university lecturers and students

WP3. Outreach

Fostering communication of Scientists with schools, media, public

WP4. Citizen Science

Involve public into scientific research, democratise science

Objective: "Educate professionals and non-professionals in data science technologies"

WP4 goals:

- Actively engage the public in citizen science projects such as Einstein@Home
- Provide incentives and access to data infrastructure and methods to involve the public in ongoing research

WP4 tasks:

- Map out potential research applications in citizen science
- Prepare datasets and soft- and hardware infrastructure for Citizen Science projects
- Run Citizen Science projects, engage schools, universities and general public, involve the physics community

Motivation for Citizen Science

Citizen science is a collaborative approach to scientific research in which members of the public, who may not have formal scientific training, actively participate in various stages of the scientific process.

Citizen science includes:

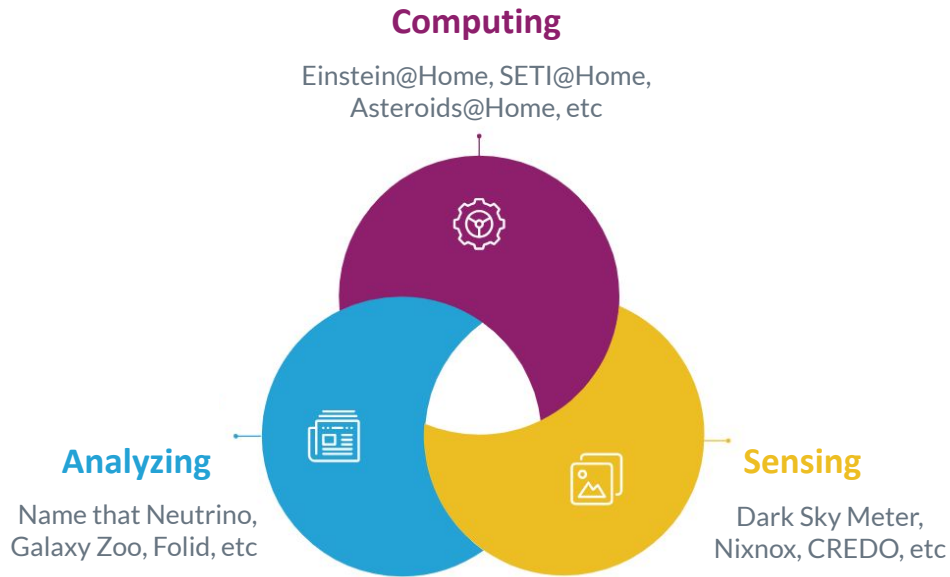
- **Data Collection** (*CREDO*)
- **Data Analysis** (*Zooniverse - Galaxy Zoo, Name that neutrino, etc.; Kaggle - IceCube - Neutrinos in Deep Ice, Higgs Boson Machine Learning Challenge, etc.*)
- **Calculations** (*SETI@Home, Einstein@Home*)
- **Problem Solving and Innovation** (*Foldit*)
- **Public Engagement and Education** (*iNaturalist*)
- **Community-Based Research** (*Flint Water Study*)
- **Scientific Research and Discovery** (*Einsten@Home - pulsar discoveries*)

Citizen Science Landscape

World (?)	Citizen Science Association (CSA)
Europe	European Citizen Science Association (ECSA)
	Research Infrastructures FOR Citizens in Europe (REINFORCE)
DACH	Citizen Science DACH AG
Germany	Bürger schaffen Wissen (GEWISS)
	Wissenschaft im Dialog (WiD)
Projects	Einstein@Home, CREDO, Radio Galaxy Zoo, etc.

Other national level organisations: Citizen Science Network Austria (CSNA), Australian Citizen Science Association (ACSA)

Classification of Citizen Science activities



A popular classification of activities



Our classification of activities done in
CitSci Projects

Audience of Citizen Science*

Computing

- Monthly audience, Jun 2021: 71k
- 10% contributing 71%
- > 60% are in the field of science or IT
- SETI - the biggest project: >14mio people total
- Demographics (SETI):
 - 90% male
 - med age 34
- BOINC: 15 PFLOPS/24h -> “comparable to fastest supercomputers” (Horeka: 17 PFLOPS - 15th in Europe)

Analyzing

- Monthly audience, Jun 2021: 23k
- 10% contributing 79%
- Gender: 68% male (Galaxy Zoo)
- Age :
 - projects, such as Galaxy Zoo - many school-age participants (35%)
 - Gamified projects (FoldIt): 65% of the participants are < 30 y.o. (median age =25)
- Occupation (FoldIt): 80% in science or IT
- Background (FoldIt): 90% in science or IT

Sensing

- Monthly audience, Jun 2021: 665k
- 10% contributing 69%
- Gender:
 - iNaturalist: 56% male
 - eBird: 51% male
- Highest participation (iNaturalist): Europe, USA and countries with rich biodiversity - Costa Rica, Panama, Taiwan, etc.
- Only biological projects

*According to [1] Strasser BJ, Tancoigne E, Baudry J, Pigué S, Spiers H, Luis-Fernandez Marquez J, et al. (2023) Quantifying online citizen science: Dynamics and demographics of public participation in science. PLoS ONE 18(11): e0293289. <https://doi.org/10.1371/journal.pone.0293289>

[2] Füchslin, Tobias; Schäfer, Mike S; Metag, Julia (2019). Who wants to be a citizen scientist? Identifying the potential of citizen science and target segments in Switzerland. Public Understanding of Science, 28(6):652-668. DOI: <https://doi.org/10.1177/0963662519852020>

Data collection for individual projects - in progress...

What? Data about existing CitSci projects within “PUNCH Sciences” (particle, astroparticle, astro-, nuclear physics and LQCD are under consideration)

Where?

- CSA, "Platforms for hosting participatory science projects" by CSA [24],
- Projects, recommended by ECSA [4]

What else?

- Kaggle competitions by PUNCH sciences' communities
- Aggregated statistics from [1], Quantifying Online Citizen Science [5]

[3] Citizen Science Association (CSA). Platforms for hosting participatory science projects. <https://participatorysciences.org/resources/platforms-for-hosting-participatory-science-projects/>. Accessed: 2023-10-01

[4] European Citizen Science Platform, Project search: <https://eu-citizen.science/projects?country=DE>

[5] Strasser Group, Quantifying Online Citizen Science, <https://doi.org/10.26037/yareta:lkrlrxq3njdh4hdzt6nrwdz4>

CitSci platforms summary:

- 11 platforms/projects: Anecdata, CitSci.Org, FieldScope, GLOBE Observer, iNaturalist, Zooniverse, SciStarter, BOINC, Bürger schaffen Wissen, EU-Citizen.Science, Kaggle[^]
- Platforms with presence of CitSci projects in physics (8):

Platform Name	Anecdata	CitSci.Org	EU-Citizen.Science	Bürger schaffen Wissen	Zooniverse	BOINC-based	SciStarter	Kaggle
NumAllProj	48	1142	321	168	97	29	1393	165 ("Research")
NumActPhysicsSpaceProj	2	2	16	2	25	6	101	12
NumActPUNCH-rel	0	0	5	2	23	6	≈65	7

[^] Kaggle is AI and ML platform. It's main goal is not CitSci

Status for datasets and infrastructure

- Initial approach (by J.T.): Conversion of the Effelsberg old raw data to 'MBFITS' format for broader accessibility. This was found to be error-prone for a significant fraction of data files
- Conversion script has been upgraded:
 - Initial tests show successful conversions after script modifications
 - Ongoing checks to ensure errors from previous conversions are resolved
- Currently, the efforts are focusing on primary data conversion and management, as well as exploring opportunities to enhance accessibility beyond the FITS format for easier visualization and interpretation by outsiders in Radio Astronomy
- Significant portion of data expected to be converted into FITS by Q1/2024
- Discussions are planned with Effelsberg operator for optimal utilization of these data products for visualization and analysis by the wider community
- Further discussions are needed for concrete plans on data utilization and accessibility enhancements

Conclusion

- Literature review was performed in order to study definitions, modern landscape, audience and methodologies, used in modern Citizen Science projects
- Data on Citizen Science projects in the field of data intensive physics are being collected
- Datasets of Effelsberg provider are being converted for the future Citizen Science project

Acknowledgements: Jacob Tremblay (MPIfR/Uni Bonn, Oct 2022 - Oct 2023), P. Jütte (Aug 2023), Katia Kornetzki (Jan 2024)

Future plans

- Citizen Science Projects overview:
 - Collect more data
 - Find similarities for popular projects
 - Find, what is missing for projects in data-intensive physics
 - Define other data insights
 - Review motivation and involvement component in CitSci projects
- => **Define recommendations for development CitSci projects in PUNCH Sciences**
- Employing Effelsberg provider data for Citizen Science
- Use of MeerKAT data set of globular clusters (together with AEI - Hannover)

Acknowledgements: J. Tremblay (MPIfR/Uni Bonn, Oct 2022 - Oct 2023), P. Jütte (Aug 2023), K. Kornetzki (Jan 2024)

Future plans

- Citizen Science Projects overview:
 - Collect more data
 - Find similarities for popular projects
 - Find, what is missing for projects in data-intensive physics
 - Define other data insights
 - Review motivation and involvement component in CitSci projects
- => **Define recommendations for development CitSci projects in PUNCH Sciences**
- Employing Effelsberg provider data for Citizen Science
 - Use of MeerKAT data set of globular clusters (together with AEI - Hannover)

Acknowledgements: J. Tremblay (MPIfR/Uni Bonn, Oct 2022 - Oct 2023), P. Jütte (Aug 2023), K. Kornetzki (Jan 2024)

Thank you for your attention! Questions?

Contact me: victoria.tokareva@kit.edu