



Shaping the Future of Digital Infrastructures for Data-Intensive Science in High-Energy Physics and Radio Astronomy

Sergio Andreozzi

Head of Strategy, Innovation and Communications, EGI Foundation

SPECTRUM Project Director

Advanced Computing User Day
12 Dec 2024



Funded by
the European Union

SPECTRUM is funded by the European Union - Grant Agreement Number 101131550



Outline

- **Challenges**
- **Project Overview**
- **Main Expected Results**
- **Timeline**
- **Current outputs**
- **What's Next**

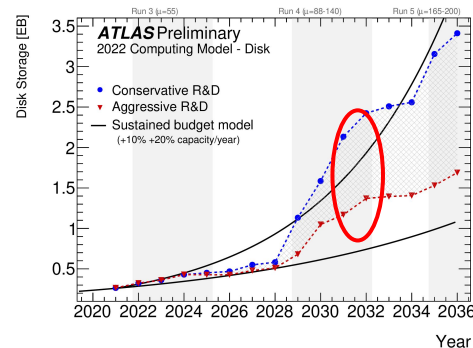
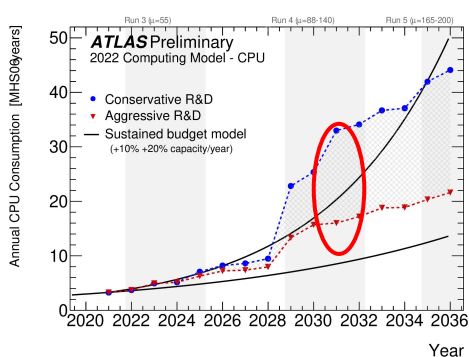
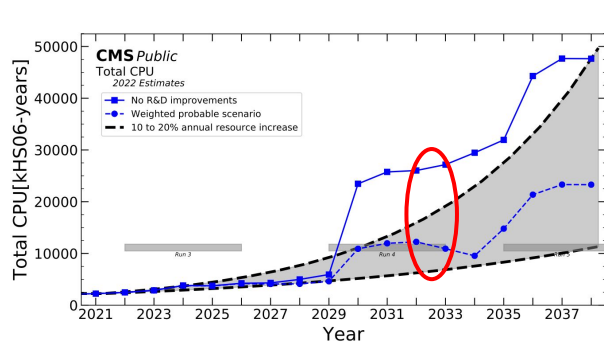
Challenge

The computing for the next generation of scientific instruments in High-Energy Physics and Radio Astronomy is an unsolved problem

- There are different domain-specific initiatives
- Can we join forces and have a more organic approach to solve this challenge?

HEP: LHC Experiments at CERN

- HL-LHC is expected to generate more than **one exabyte of new data for each year** in the period ~2029–2040
- This data must be exported in ~real time from CERN to compute sites

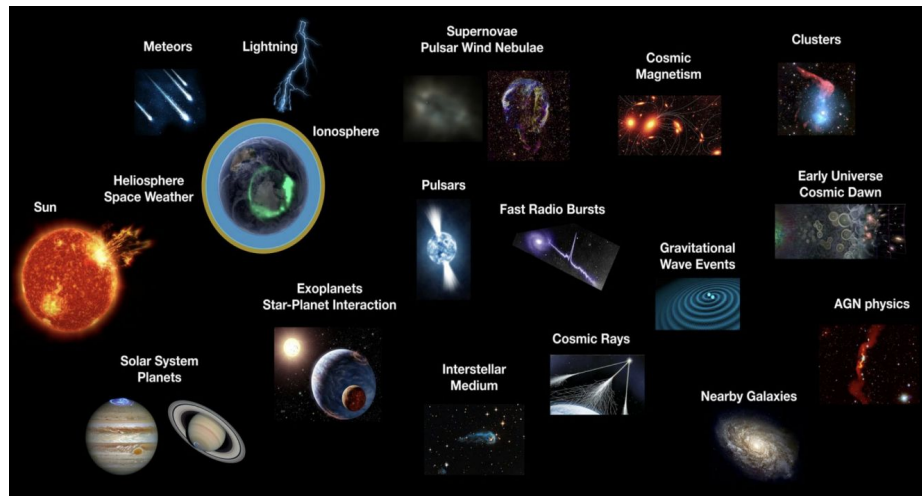


CMS and ATLAS Compute Resource projections

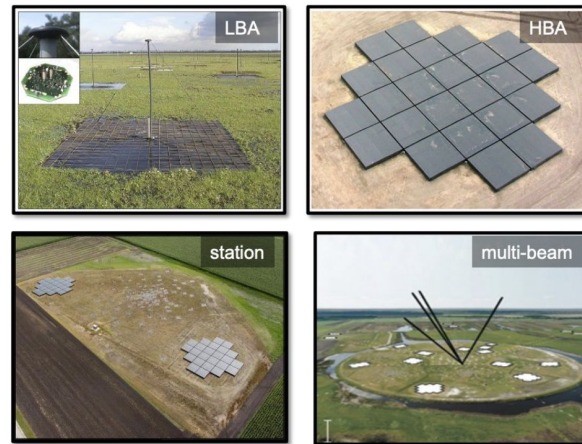


RA: The LOw Frequency ARray (LOFAR)

Key LOFAR Science



LOFAR Antennas (52 Stations across Europe)

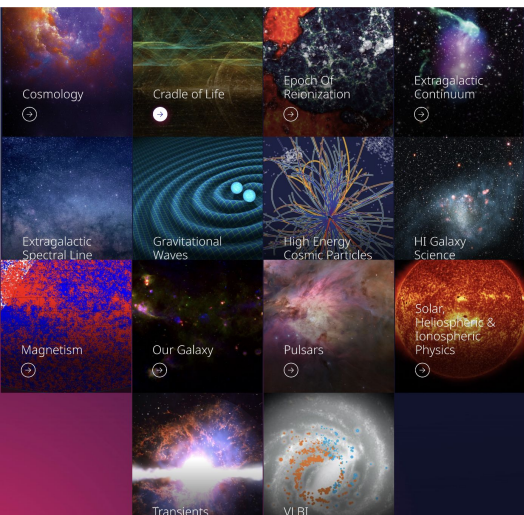


- *Already 60 PB under management: the world's largest astronomical archive.*
- Upgrade to LOFAR2.0 underway: on-sky in 2026
 - >1 billion core hours required to process first five years of data
 - >100 PB of legacy data archived

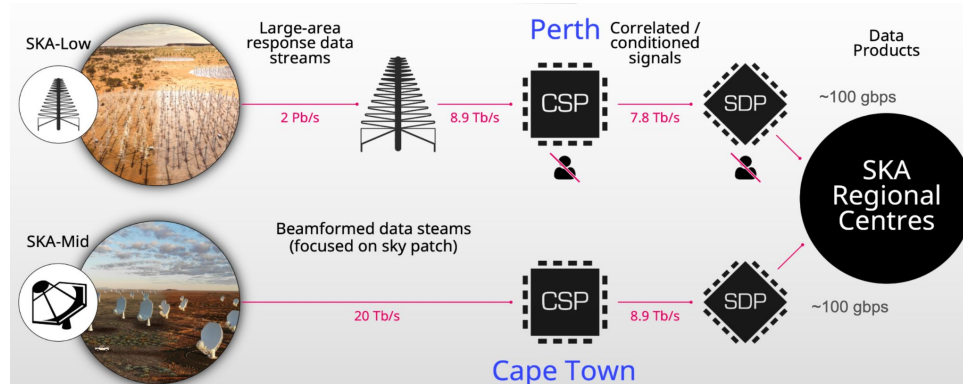
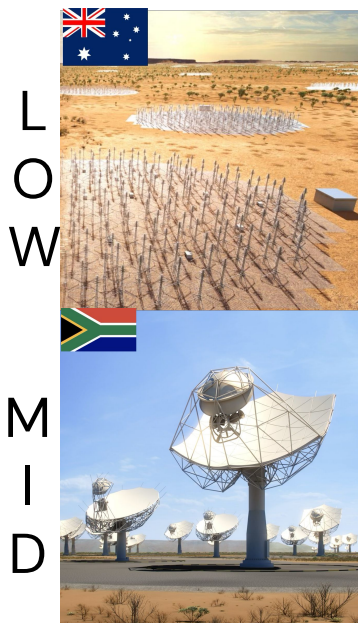
SKA: The Square Kilometer Array (SKA)

The largest radio telescope in the world

A general purpose observatory addressing 14 science themes



12 December 2024 | ACUD 2024



- Dedicated processing within the Observatory generates science-ready products
- Globally-distributed Regional Centre Network provides archiving and analysis facilities
- SRC Network accumulates data at around **1 EB/year from 2028 onwards**



Project Vision and Overall Objective

VISION

Data-intensive scientific collaborations have access to a European exabyte-scale research data federation and compute continuum

PROJECT OVERALL OBJECTIVE

*Deliver a **Strategic Research, Innovation and Deployment Agenda (SRIDA)** which defines the vision, overall goals, main technical and non-technical priorities, investment areas and a research, innovation and deployment roadmap for data-intensive science and infrastructures*



Project Key Data

Fact sheet: [See page on CORDIS](#)

Duration: 30 months – **Start date:** 1 Jan 2024 – **End date:** 30 June 2026

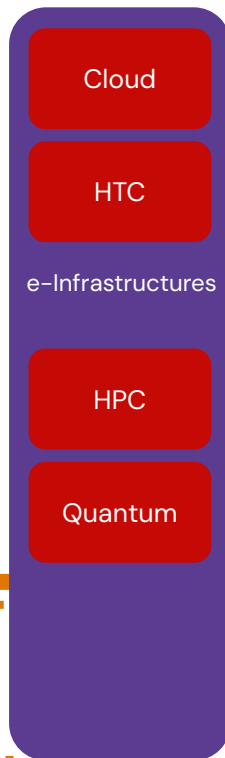
Partners: 9 partners + 1 affiliated

Budget: 2,449,542.50€

Funding Source: Horizon Europe – Call [HORIZON-INFRA-2023-DEV-01-05](#)

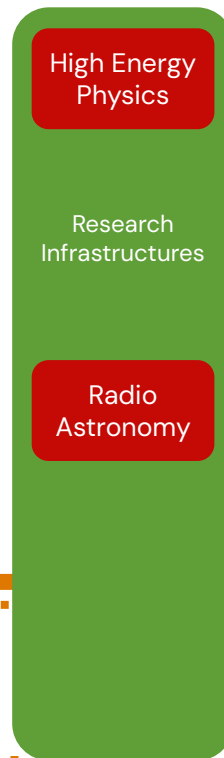
- *Preparation of common strategies for future development of RI technologies and services within broad RI communities*

Consortium Overview



**SPECTRUM
project**

**Community
of
Practice**



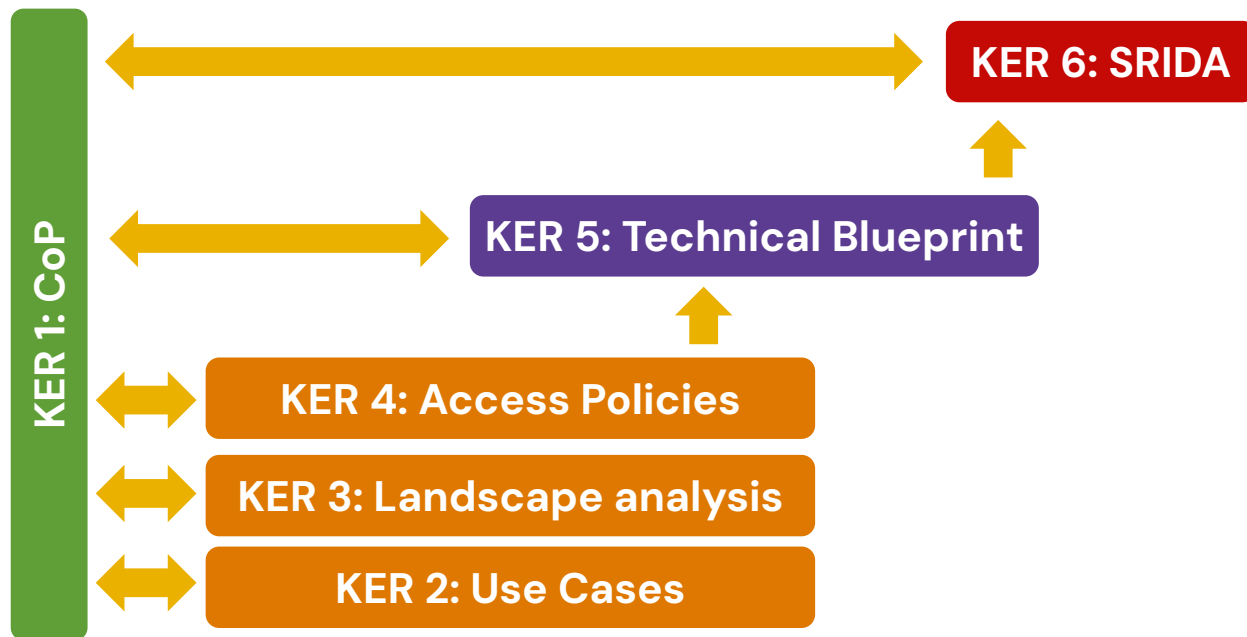


Main Expected Results

Community of Practice (CoP)	Compendium of Use Cases	Landscape analysis
<ul style="list-style-type: none">• Cross-disciplines WGs with experts from HEP, RA and digital infras• Knowledge Base• Collaboration platform	<ul style="list-style-type: none">• From science case to technical challenges, requirements, gaps• Both technical and policy aspects	<ul style="list-style-type: none">• Existing approaches, services, technical solutions and policies for the federation of data and compute infrastructures
Access Policies Recommendations	Technical Blueprint for Compute & Data Continuum	SRIDA
<ul style="list-style-type: none">• Existing access policies across the continuum• Development of recommendation for Interoperable access	<ul style="list-style-type: none">• Proposal for a compute-and-data trans-continuum infrastructure design and architecture	<ul style="list-style-type: none">• Vision and overall goals• Research, innovation and deployment roadmap• For data-intensive science and infras in HEP & RA



Relationship among Key Exploitable Results (KER)



Primary Target Groups

TG1 – Scientific Communities / Research Infrastructures

Scientific Communities in HEP and RA and other relevant domains.

Participate in CoP and WGs, Implement outputs

TG2 – Computing & Data Service Providers/e-Infrastructures

Existing Data, HPC, HTC, Cloud, Quantum

Participate in CoP and WGs. Implement outputs

TG3 – Policy Makers / Funding Bodies

European Commission, EuroHPC, EOSC,
National Authorities, ESFRI...

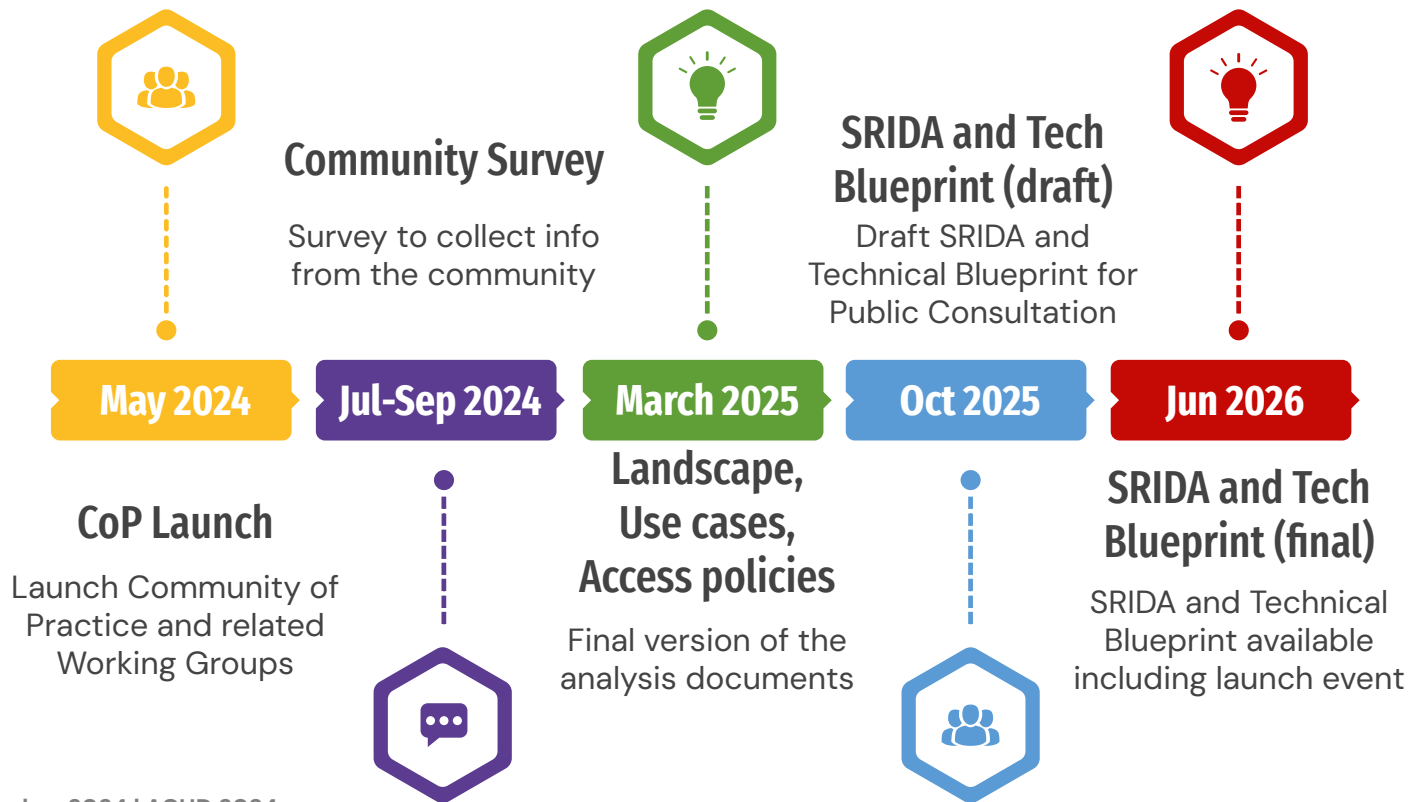
Provide inputs, align policies

TG4 – ‘Long Tail of Researchers’

Inputs to use cases through represented communities



Timeline





Collaboration with other initiatives

JENA initiatives

- Comes from the 3 main committees for HEP, Nuclear Physics and Astroparticle Physics (Joint ECFA-NuPECC-APPEC)
- Workshop in 2023 and launch of working groups after SPECTRUM was submitted
 - Timescale similar to ours
 - WG leaders in JENA are already members of SPECTRUM WGs.

JENA Working groups:

WG1: HPC

WG2: Software

WG3: Data

WG4: AI

WG5: Training, Dissemination, Education

The Upgrade o the European Strategy for Particle Physics

- Happens every 5-10 years
- It was scheduled in 2026, moved 1 y earlier
- Will receive contributions/inputs from the community by March 2025
 - Same time scale as SPECTRUM, and also a target for JENA



Current Outputs

SPECTRUMCoP: List of WGs



75 participants
6 Working Groups
~50 unique individuals

More information:

<https://www.spectrumproject.eu/spectrumcop>

WG1: Data Management and Access

Chair: BAGNASCO Stefano (INFN, ET, Virgo)

- Data Management
- Data Access Protocols
- Data Archiving
- Security

WG2: Workflow management and organization

Chair: DELL'AGNELLO Luca (INFN CNAF-T1)

- Resource Discovery and Workflow Submission
- Resource Allocation
- Complex Workflows

WG3: Compute Environment

Chair: BOZZI Concezio (INFN, LHCb, JENA-HPC)

- Expected Tools and Services
- Facility Expectations
- Edge Services
- Library Provisioning

WG4: SW tools

Chair: SWINBANK John (Astron)

- Machine Learning Frameworks
- Multithreading Frameworks
- Multi-Node Tools
- Compilers, toolchains, ...
- Quantum computing tools and frameworks
- Code Management Practices

WG5: Scientific Use cases

Chairs: FERRARI Chiara (CNRS, OCA), GIRONE Maria (CERN)

- Typical Use Cases
- Requirements and Needs
- Best Practices Collection
- Data Fluxes and Paths

WG6: Facilities

Chair: HOPPE Hans-Christian (Jülich)

- HPC Centers
- Access to Quantum Computing Hardware
- Access to Commercial and Public Clouds
- Sustainability
- Security

SPECTRUMCoP: Community Survey

Community consultation to understand current best practices and future needs in large-scale and data-intensive scientific computing

Authentication and Authorization

Processing Needs

Data Management Needs

Training and careers

Expected Computing Environment

Use cases

Software development, distribution, policies

LINK:

https://ec.europa.eu/eusurvey/runner/SPECTRUM-JENA_Survey1



First wave of data collection
Jul 2024– Oct 2024

The survey is still open, so you are still on time to contribute!



SPECTRUMCoP: Community Survey

Survey: No. of Replies from first wave by area

Which are the categories which better describe your role(s)?	Answers	Fraction (%)
Experimental High Energy Physics (HEP)	30	29.7%
Experimental Gravitational Waves (GW)	6	5.9%
Observational Astroparticle (not RA or GW)	9	8.9%
Observational Radio Astronomy (RA)	8	7.9%
Theoretical High Energy Physics (HEP)	15	14.9%
Other physics related domains (please specify below)	6	5.9%
Other non-physics related research domains (please specify below)	4	4.0%
Experimental Nuclear Physics (NP)	9	8.9%
Theoretical Nuclear Physics (NP)	12	11.9%
Theoretical Gravitational Waves (GW)	2	2.0%
Total	101	100%



Use Case Analysis

The Template

- **Goal:** Identifying current and future requirements of HEP and RA communities
- Template “co-design” with the survey and “internal” use-cases (from CERN/LHC and RA) as initial examples
- Start from the scientific challenge and drill down into the technical challenges

Structure for each identified use-case:

Scientific challenge

Storage

Data transport

Compute

Workflow management

Access and analysis

Gap analysis

Non-technical challenges



Use Cases: Current draft list (16)

High Energy Physics (7)

1. CMS
2. ATLAS
3. LHCb
4. ALICE
5. AI-based particle flow reconstruction
6. LLMs for the CERN accelerator complex
7. Detector Simulation Digital Twins (CNRS/IJCLab)

Radio Astronomy (5)

1. SKA Regional Centre Network – A Measurement of the Power Spectrum of 21cm HI Fluctuations during the Epoch of Reionization and Cosmic Dawn
2. SKA Regional Centre Network – Star Formation History of the Universe
3. Extragalactic Surveys with LOFAR
4. Fast radio bursts and transients
5. ALMA

Astrophysics (1)

1. JIVE

Gravitational Waves (1)

1. Virgo Digital Twin

Theoretical Physics (1)

1. Lattice-QCD

Neuroscience (1)

1. eBrains

Use Case Analysis: Initial findings

- Unprecedented Data Volumes
- Heterogeneous Computing Needs
- Complex Workflow Management
- Data Federation and Accessibility
- Scalability Challenges

Final analysis to be published in March 2025

DRAFT



What's Next → Coming in March 2025

Community of Practice (CoP)	Compendium of Use Cases	Landscape analysis
<ul style="list-style-type: none">• Cross-disciplines WGs with experts from HEP, RA and digital infras• Knowledge Base• Collaboration platform	<ul style="list-style-type: none">• From science case to technical challenges, requirements, gaps• Both technical and policy aspects	<ul style="list-style-type: none">• Existing approaches, services, technical solutions and policies for the federation of data and compute infrastructures
Access Policies Recommendations	Technical Blueprint for Compute & Data Continuum	SRIDA
<ul style="list-style-type: none">• Existing access policies across the continuum• Development of recommendation for Interoperable access	<ul style="list-style-type: none">• Proposal for a compute-and-data trans-continuum infrastructure design and architecture	<ul style="list-style-type: none">• Vision and overall goals• Research, innovation and deployment roadmap• For data-intensive science and infras in HEP & RA

Acknowledgement



- Sergio Andreozzi
- Patricia Ruiz
- Gwen Franck
- Xavier Salazar



- Raymond Oonks
- Kristen Lutz



- Hans-Christian Hoppe
- Luis Cifuentes



- Fabio Affinito



- Maria Girone
- David Southwick
- Eric Wulff



- Tommaso Boccali
- Luciano Gaido



- Jean-Pierre Vilotte
- Jeff Wagg



- John Swinbank
- Hanno Holties



- Chiara Ferrari



- Corentin Lefevre
- Valentin Jay Blanchard

How can you contribute?

Join the CoP

https://www.spectrumproject.eu/spectrum_cop

Follow us on LinkedIn

<https://www.linkedin.com/company/spectrum-project-eu>

Fill the Survey

https://ec.europa.eu/eusurvey/runner/SPECTRUM-JENA_Survey1

Subscribe to the Newsletter

<https://www.subscribepage.com/spectrumproject>



Thank you! Questions?



spectrumproject.eu



linkedin.com/company/spectrum-project-eu



Funded by
the European Union

SPECTRUM is funded by the European Union - Grant Agreement Number 101131550