**** The COAST logo is based on the painting "Birds Eye View of Sea Coast" by Leonardo Da Vinci***

Content Aware searching and streaming in Future Internet

Theodore Zahariadis

Project Technical Coordinator

zahariad@synelixis.com

Search Engines Cluster Meeting
Brussels, 4 February 2010

Project Summary
COntent Aware Searching and sTreaming

COAST aims to offer fast content-aware retrieval, delivery and streaming, while meeting network-wide Service Level Agreements (SLAs).

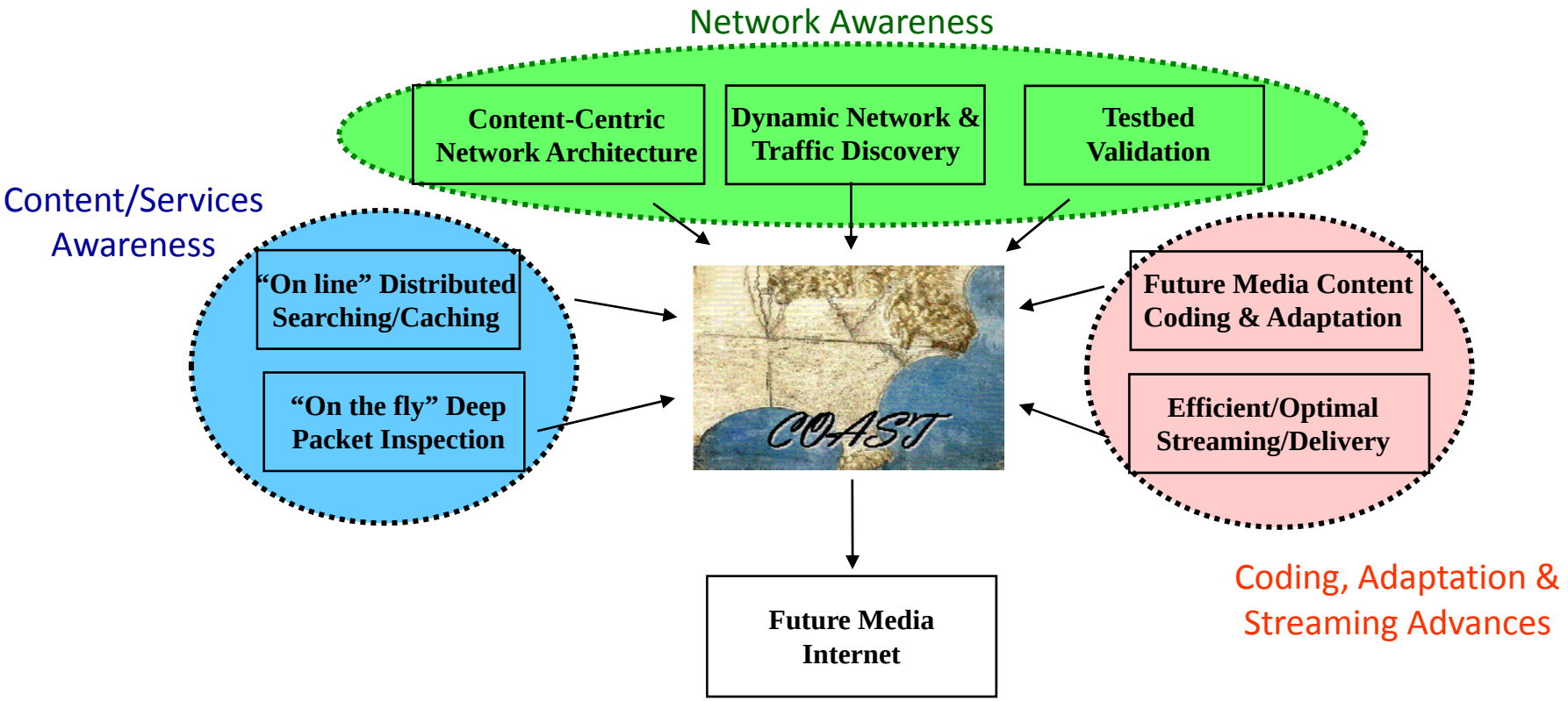
COAST will focus on 3 innovation pillars:

- **Content & Services Searching & Indexing.** Based on intelligent nodes, COAST will
  - *"on the fly" identify/classify content and identify Web services* via inspection of the traffic that flows through the nodes
  - *discover "on line", where services are located and content is located/cached*.

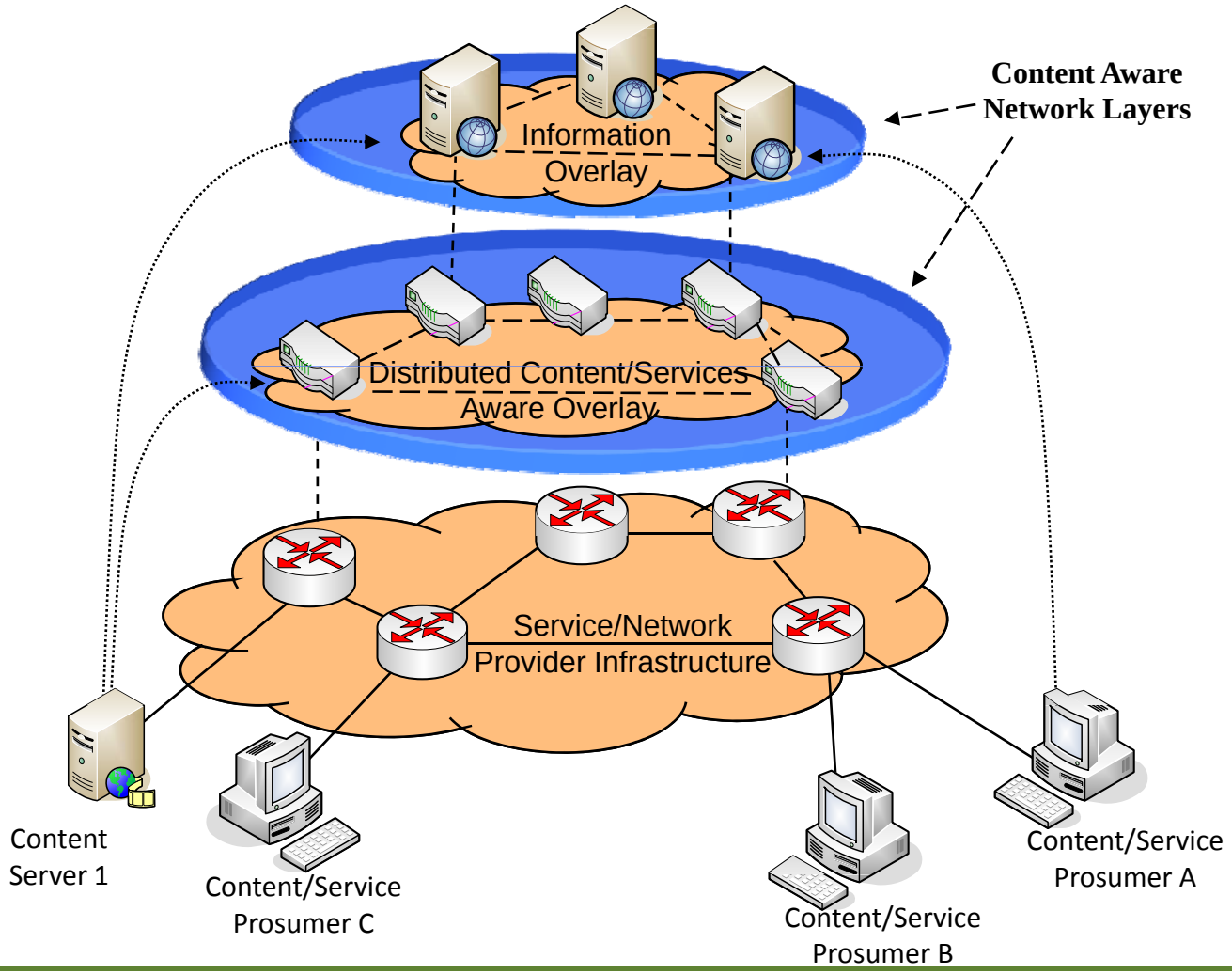
- **Content-Aware Delivery Network Architecture**. COAST will
  - *discover the underlying network infrastructure*, the *user terminal* and the *user needs*
  - *construct content-aware overlays* to offer distributed, robust and network-/service-provider friendly content delivery, leading to improved PQoS.

- **Future media content adaptation and enrichment.** COAST will provide for scalable, high-definition *3D/free-viewpoint video with interactive virtual panning/zooming*, which will be on-the-fly adapted, enriched and optimized.

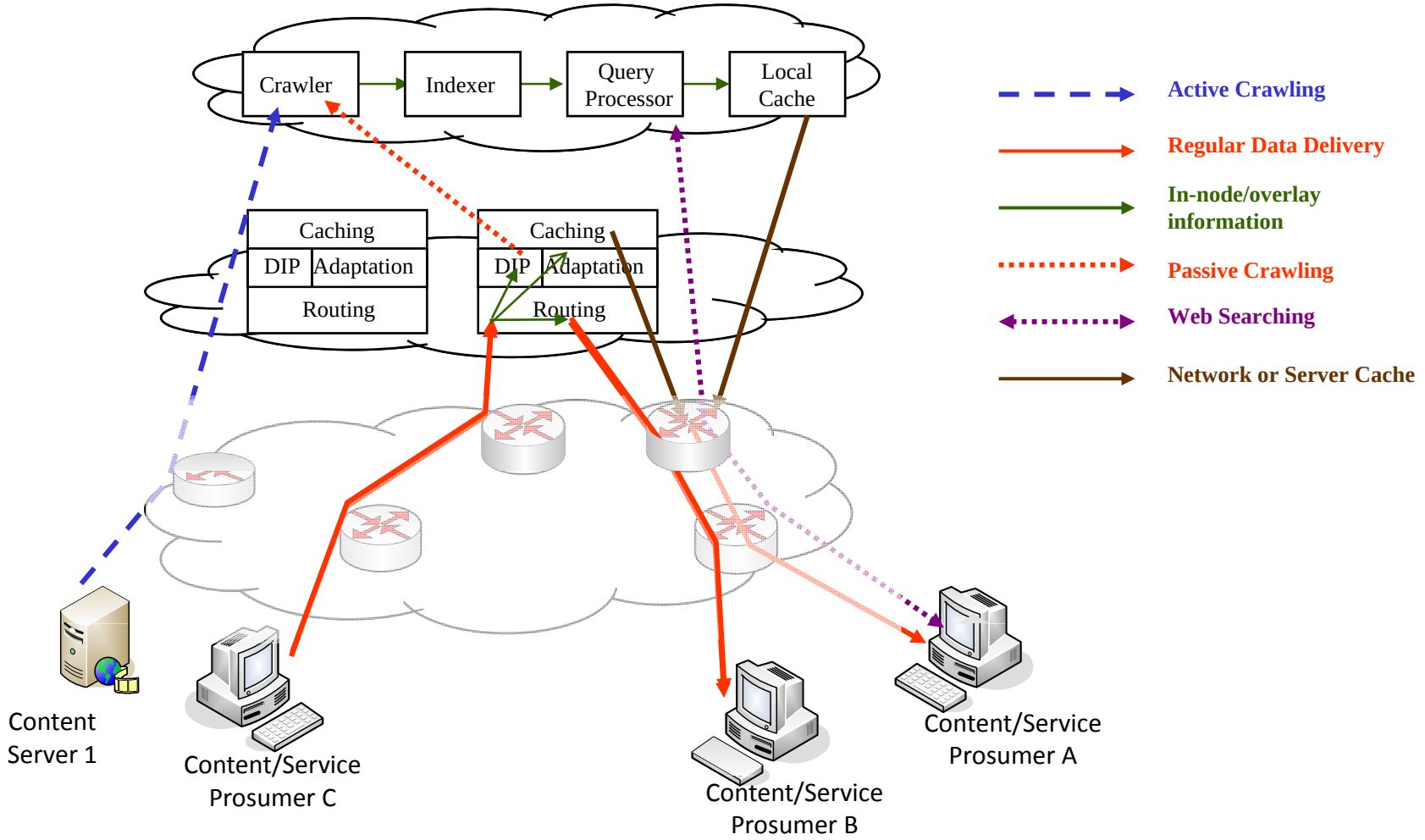
COAST Approach



Network Architecture



Searching & Caching



Search in COAST

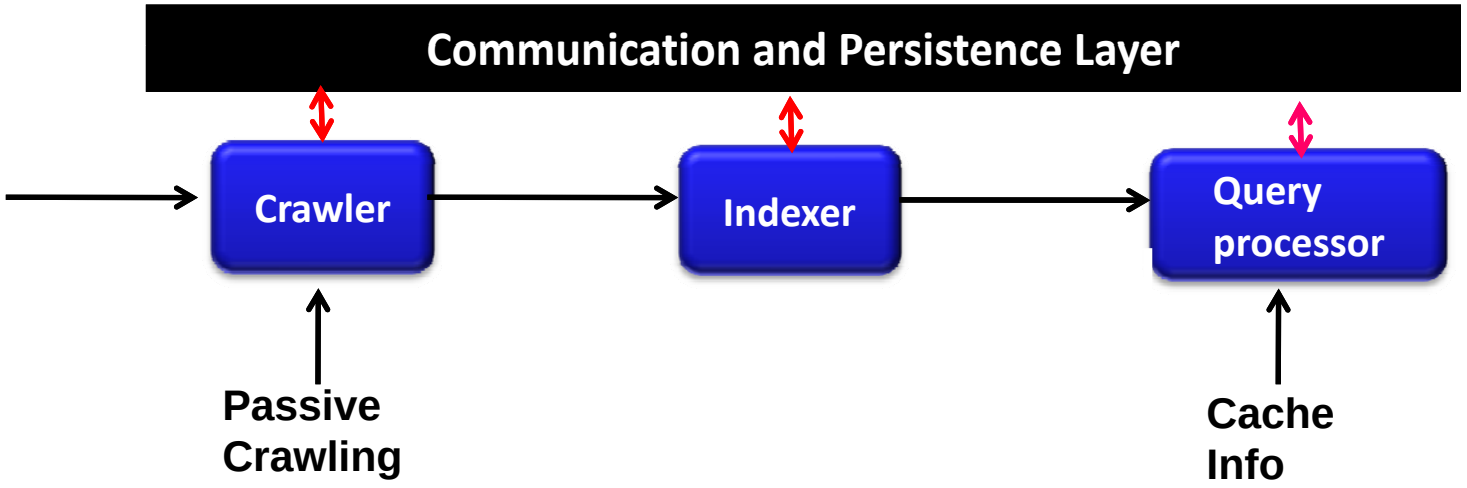
- **Main partners:** Yahoo! Research and Fundación Barcelona Media
- **Context:** Distributed computing and search at Yahoo!
    - Search engine caching [SIGIR'07, SPIRE'07, SIGIR'08, WWW'10]
    - Multi-site search engines [SIGIR'09, CIKM'09]
    - Distributed coordination and large scale processing [CIDR'09, PODC'09]
- **In COAST**
    - Multi-site search engines: Multiple data centers geographically spread
    - Focus on scalability, locality, and cost
- **Multi-site search engines**
    - **Goal 1**: Minimize the amount of resources in each site
    - **Goal 2**: Minimize the amount of communication between sites
    - **Goal 3**: Minimize latency when processing queries
    - Different from *federated search*: all sites are capable of processing queries without consulting others
    - Different from *P2P* search: there is no churn or adversarial behavior

Enhancements to search

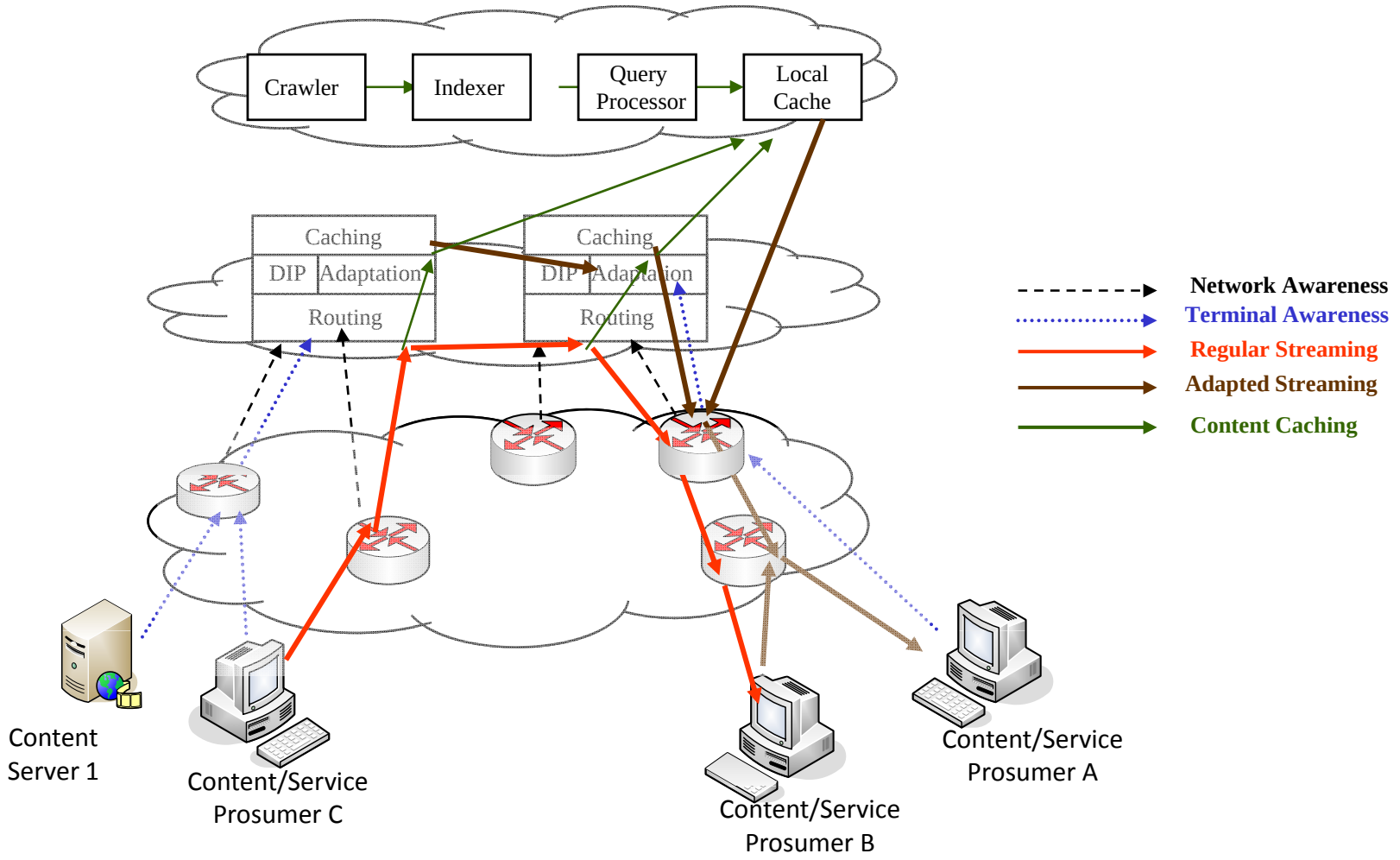
- **Passive crawling**
  - Information discovered through deep-packet inspection
  - User statistics
    - *E.g.*, popular objects
  - Unknown documents
    - Hidden Web: Pages generated dynamically
  - Web services
    - Discover new registries
- **Cache information**
  - Cached copy of objects in intelligent nodes
  - Incorporate cache information into query processing
  - Redirect users to closest copy

Search: Implementation

- **Build upon open-source software**
  - Lucene: Indexing and query processing
  - Nutch: Crawling
  - Hadoop: Data crunching (as part of the indexer)
- **High-level plan**
  - Step 1: Bare bones engine
  - Step 2: Incorporate enhancements
  - Step 3: Distributed version



Network/Terminal Awareness, Streaming & Adaptation



Who is “COAST”

- ST Microelectronics (Italy)
- Synelixis Solutions Ltd (Greece)
- Yahoo! Iberia (Spain)
- NEC Europe Ltd (UK)
- Telefonica I+D (Spain)
- Fraunhofer HHI (Germany)
- Politecnico di Torino (Italy)
- Technische Universität Berlin (Germany)
- Fundacion Barcelona Media (Spain)
- University of California, Los Angeles (USA)
- Seoul National University (S. Korea)

9 Participants from 5 EU countries + 1 from USA and 1 from S. Korea
3 Industries, 1 Operator, 1 SME, 4 Research Institutes, 2 non-EU participants