

*Super Peta-scale Infrastructure for
Distributed e-health and
biomedical Research*

David MANSET

Director of Biomedical Applications
MAAT Gknowledge



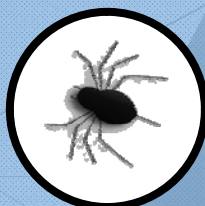
Content



SHARE Roadmap



EU-FP6 eHealth Portfolio Survey



A Leap Toward a Fully Operational European Healthgrid: SPIDeR



Facts

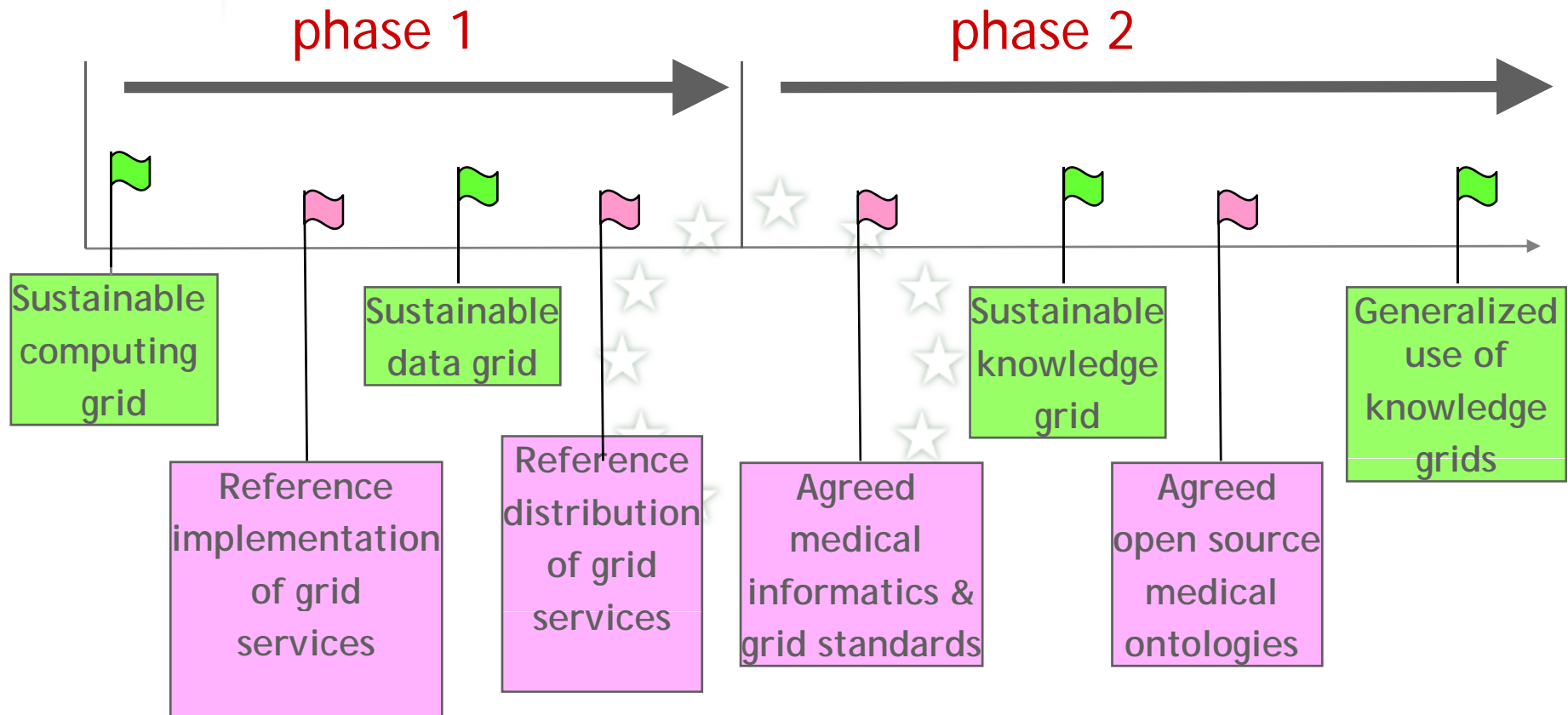


- is a roadmap for the wide adoption & deployment of grid technologies for healthcare and biomedical research in Europe
- Not focused only on technical challenges but has explored the wide range of ELSE issues (Ethical, Legal & Socio-Economic)
- Duration: January 2006 - March 2008
- www.eu-share.org for more information





Toward a Roadmap



The Share project (FP6-2005-IST-027694) is funded by the European Commission Information Society and Media DG under the 6th Framework Programme.





Starting from Users' Requirements



- The Computational example: Innovative Medicine
- The Data example: Epidemiology
- The Collaboration examples: Breast Cancer screening, Paediatrics, Neurology & VPH
- The Knowledge example: General Healthcare





Computational Grids






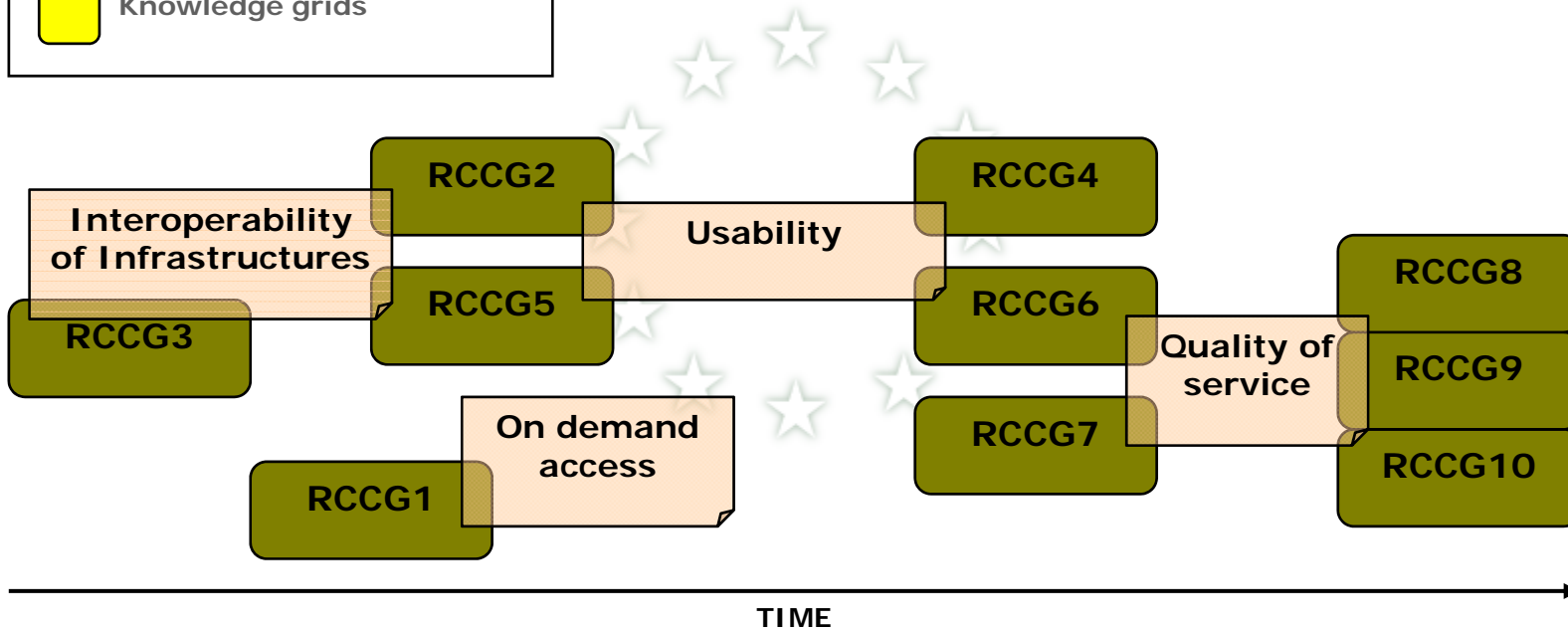
Challenge	Community	Description of the requirement
RCCG1	VPH	<ul style="list-style-type: none"> • Access to grid resources on demand.
RCCG2	VPH	<ul style="list-style-type: none"> • Transparent job submission to cluster and supercomputer grids. • Easy transfer of tasks between grid infrastructures
RCCG3	VPH	<ul style="list-style-type: none"> • Automatic migration of simulations between different scales.
RCCG4	VPH	<ul style="list-style-type: none"> • User friendly access. Lower barrier to adoption.
RCCG5	VPH	<ul style="list-style-type: none"> • Transparent access to different grids.
RCCG6	EPI	<ul style="list-style-type: none"> • Need for real fault-tolerant scheduling systems.
RCCG7	EPI	<ul style="list-style-type: none"> • Easily installed grid middleware for health environments. • Low maintenance and administration.
RCCG8	EPI	<ul style="list-style-type: none"> • Exploitation models and guaranteed QoS for services. • Advance resource reservation with pre-negotiated QoS.
RCCG9	EPI	<ul style="list-style-type: none"> • Need for scalable job scheduling system.
RCCG10	EPI	<ul style="list-style-type: none"> • Low latency/high performance services integrated.

The Share project (FP6-2005-IST-027694) is funded by the European Commission Information Society and Media DG under the 6th Framework Programme.



Research challenges for:

-  Computing grids
-  Data grids
-  Knowledge grids





Data Grids



Challenge	Community	Description of the requirement
RCDG1	EPI	<ul style="list-style-type: none"> • Easily installed grid middleware for health environments. • Low maintenance and administration.
RCDG2	EPI - VPH	<ul style="list-style-type: none"> • Data architectures/tools for private data dissociation, pseudo/anonymisation and encryption. • Automatic compliance with legal requirements.
RCDG3	EPI	<ul style="list-style-type: none"> • Exploitation models and guarantees QoS for services. • Advance resource reservation with pre-negotiated QoS.
RCDG4	EPI	<ul style="list-style-type: none"> • Scalable data cataloguing and data transfer.
RCDG5	VPH	<ul style="list-style-type: none"> • Storage services for easy upload/download of large binary objects.
RCDG6	VPH / EuroPhysiome	<ul style="list-style-type: none"> • Distributed data models and repositories multiscale data.
RCDG7	IMI	<ul style="list-style-type: none"> • Enhanced standards for data protection in web services environments.






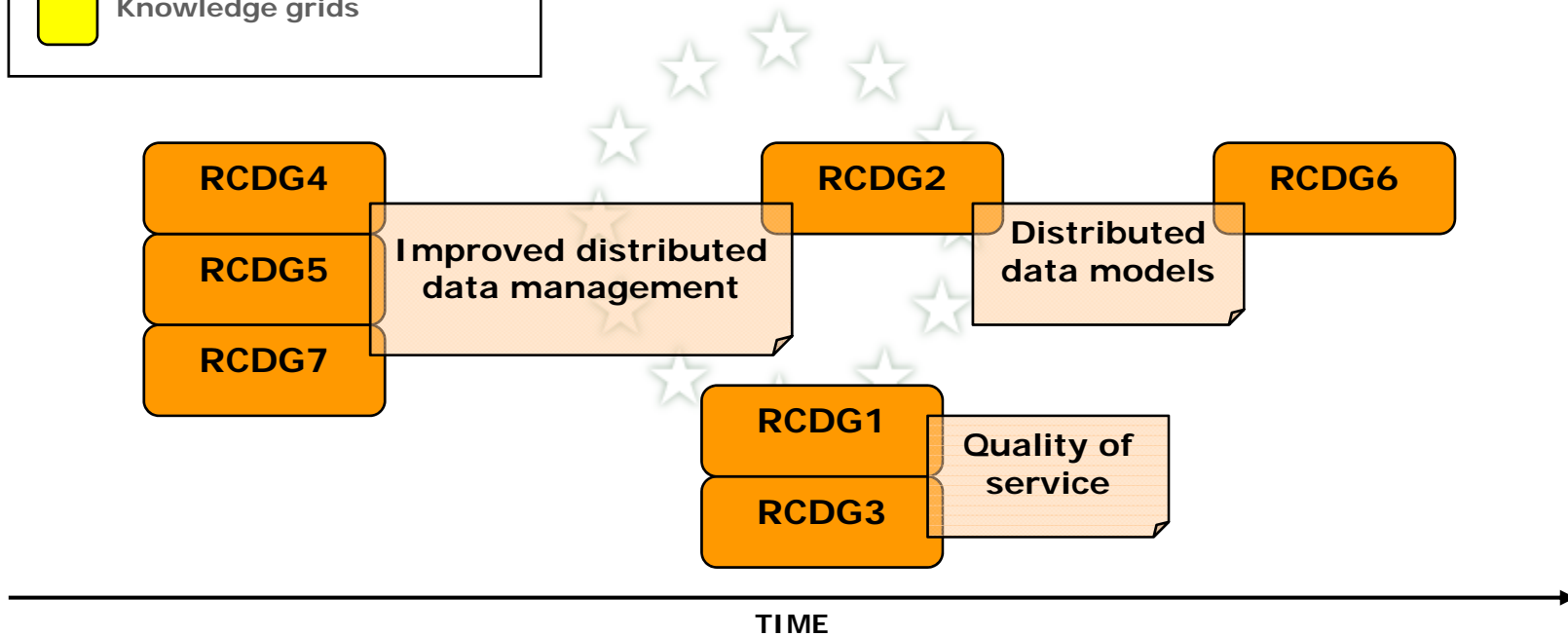


Data Grids



Research challenges for:

-  Computing grids
-  Data grids
-  Knowledge grids



The Share project (FP6-2005-IST-027694) is funded by the European Commission Information Society and Media DG under the 6th Framework Programme.

Courtesy the SHARE project – www.eu-share.org





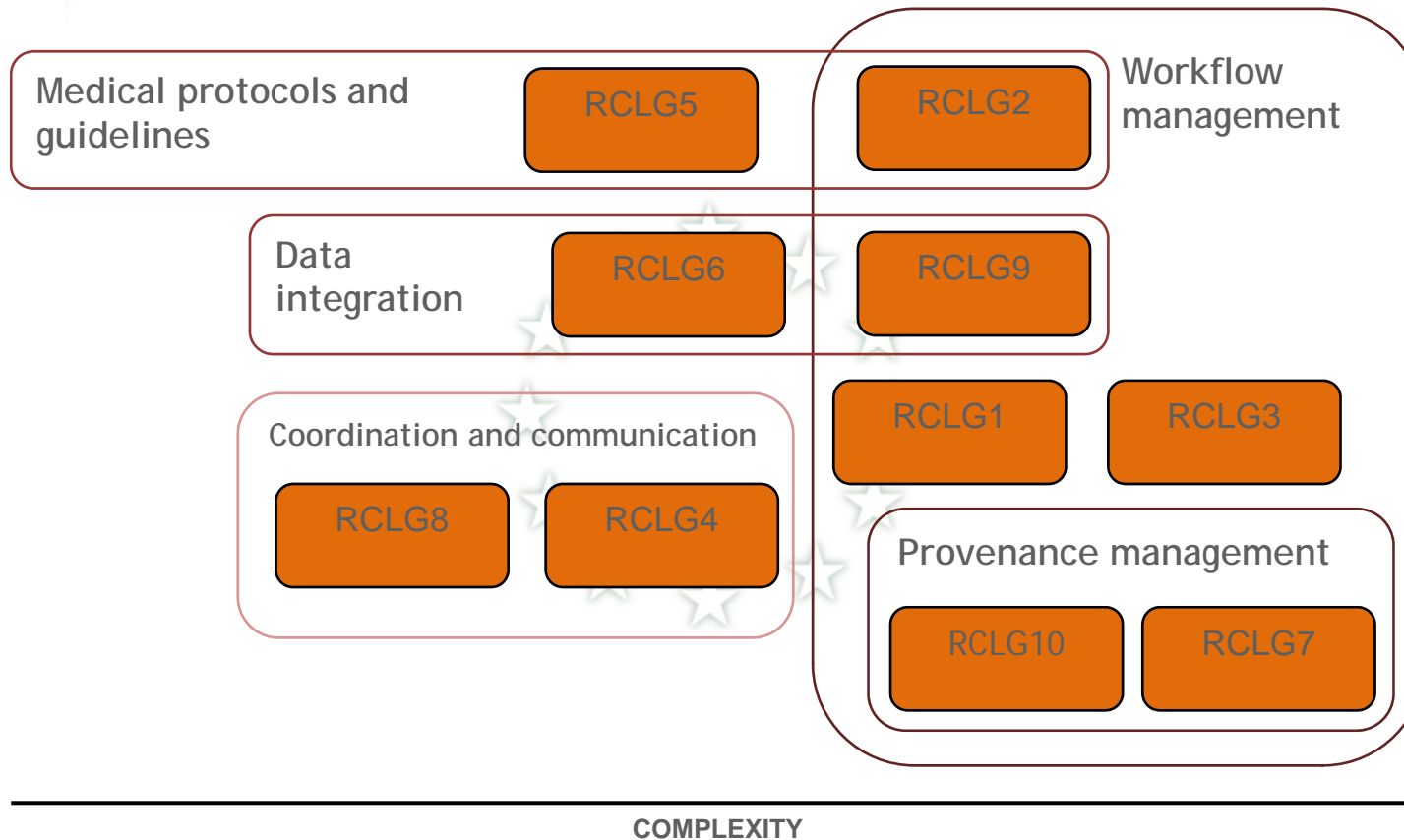
Collaboration Grids



Research challenge	Description of the health research challenges
RCLG1	Migration of e-science workflow engines to biomedical research to encompass end-to-end processes, e.g. stages on the road from drug discovery to clinical trial.
RCLG2	Natural mapping of healthcare/medical protocols to workflows for remote collaboration, education or quality control.
RCLG3	Certification of medical workflows, complying with relevant legal and ethical obligations, to ensure they are reliable, validated and updated when required.
RCLG4	Natural mapping of public health distributed decision support to facilitate coordinated action.
RCLG5	Natural mapping of guidelines, protocols and integrated care pathways to validate practice against constantly updated evidence base.
RCLG6	<i>Ad hoc</i> integration of heterogeneous sources of information where no prior coordination has been provided. Integration of different levels or modalities of medical data towards multidisciplinary diagnosis and treatment planning.
RCLG7	Workflow repositories to retain and maintain defined workflows and to enhance reuse, repurposing and recycling. Retain workflow histories and outcomes.
RCLG8	Support for persistent collaborations, esp. in relation to rights management and participant privileges.
RCLG9	Integration and management of workflows with implications in different domains, e.g. conflict between medical and ethical calls.
RCLG10	A forum for the discussion of health/medical workflows, including provenance data, and a broader means of discussion and communication between collaborators.

The Share project (FP6-2005-IST-027694) is funded by the European Commission Information Society and Media DG under the 6th Framework Programme.







Knowledge Grids



Challenge	Community	Description of the requirement
RCKG1	EPI	<ul style="list-style-type: none"> Knowledge-driven grid catalogues and integration based on the metadata.
RCKG2	IMI	<ul style="list-style-type: none"> Standards and models to expose web services (semantics), scientific services, properties of data sources, data sets, scientific objects, and data elements
RCKG3	IMI	<ul style="list-style-type: none"> Enhanced knowledge representation models and data exchange standards for complex systems
RCKG4	IMI	<ul style="list-style-type: none"> Develop new, domain-specific ontologies based on standard data representation models and reference ontologies
RCKG5	IMI	<ul style="list-style-type: none"> Advanced text mining tools to capture implicit information about complex objects, relationships and processes, as described in patents and literature
RCKG6	IMI	<ul style="list-style-type: none"> Standards and an expert tool (ontology/schema/rules negotiator) to expose properties of local sources in a federated environment
RCKG7	IMI-VPH	<ul style="list-style-type: none"> Standards and an expert tool (services/data negotiator) to guide users through the complexities of the data, data models, simulation and modelling tools.

The Share project (FP6-2005-IST-027694) is funded by the European Commission Information Society and Media DG under the 6th Framework Programme.





FP6 eHealth Survey

FP6 eHealth Portfolio

A Survey

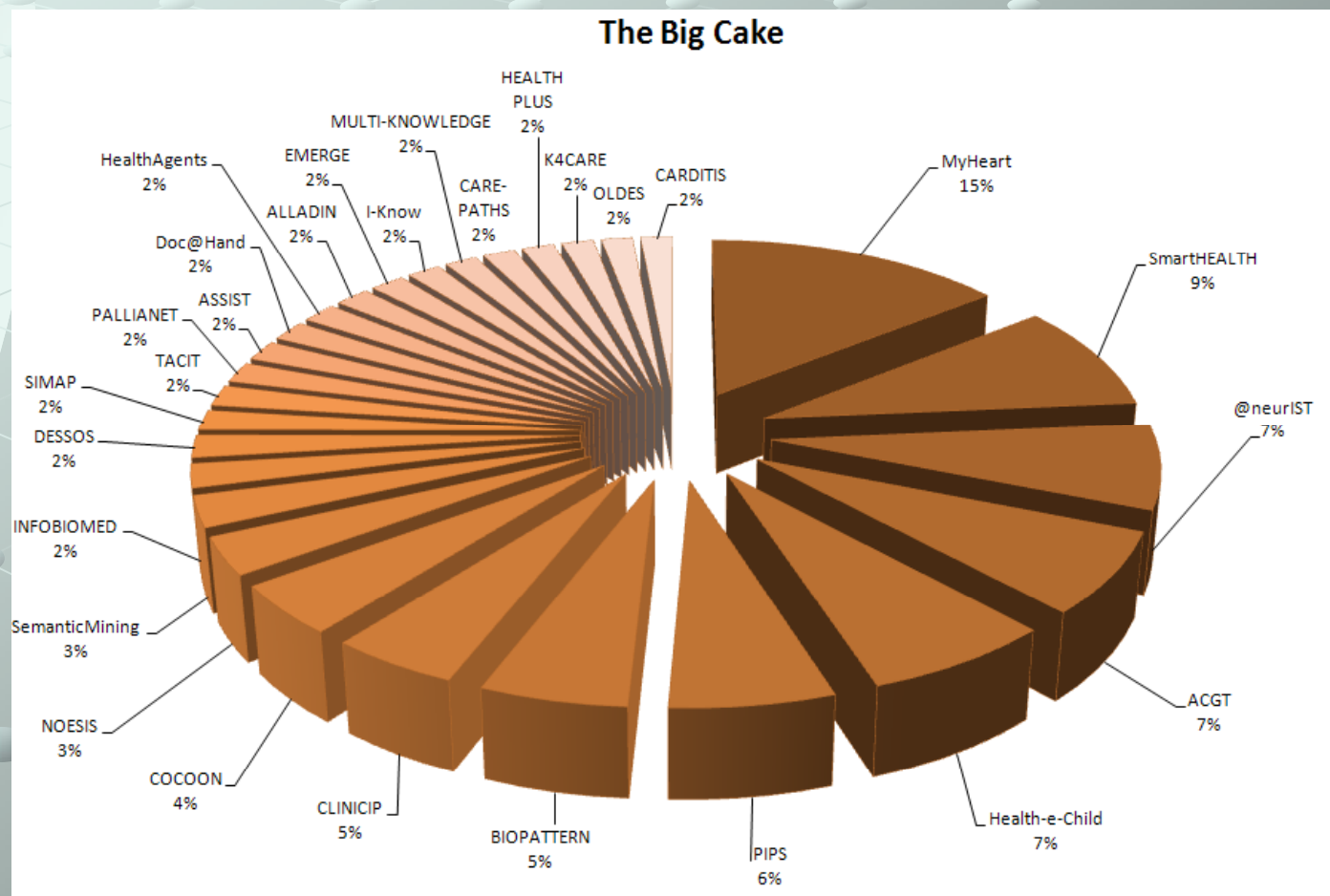




FP6 eHealth Survey

eHealth Projects Overview

The Big Cake

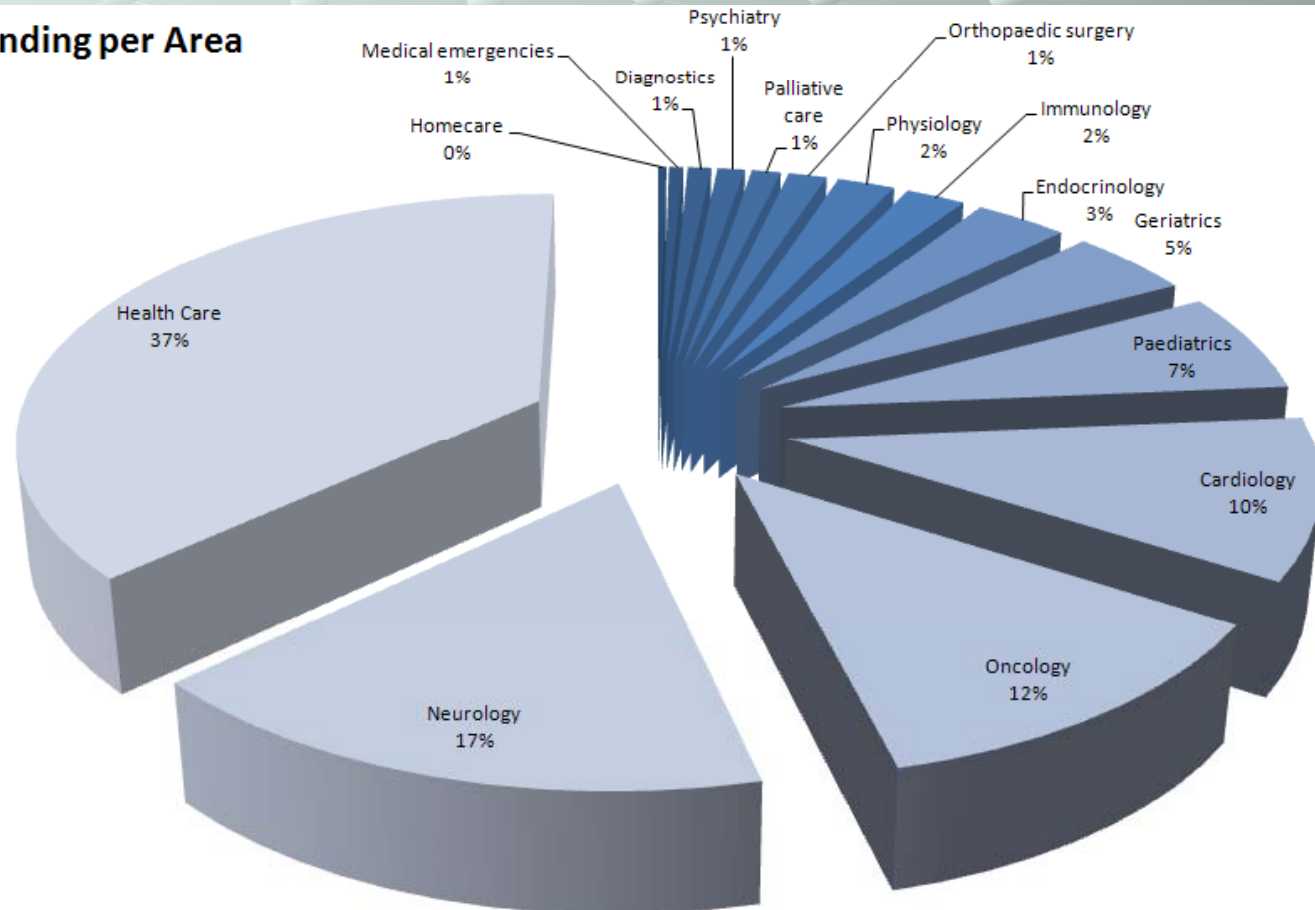




FP6 eHealth Survey

Medical Areas Overview

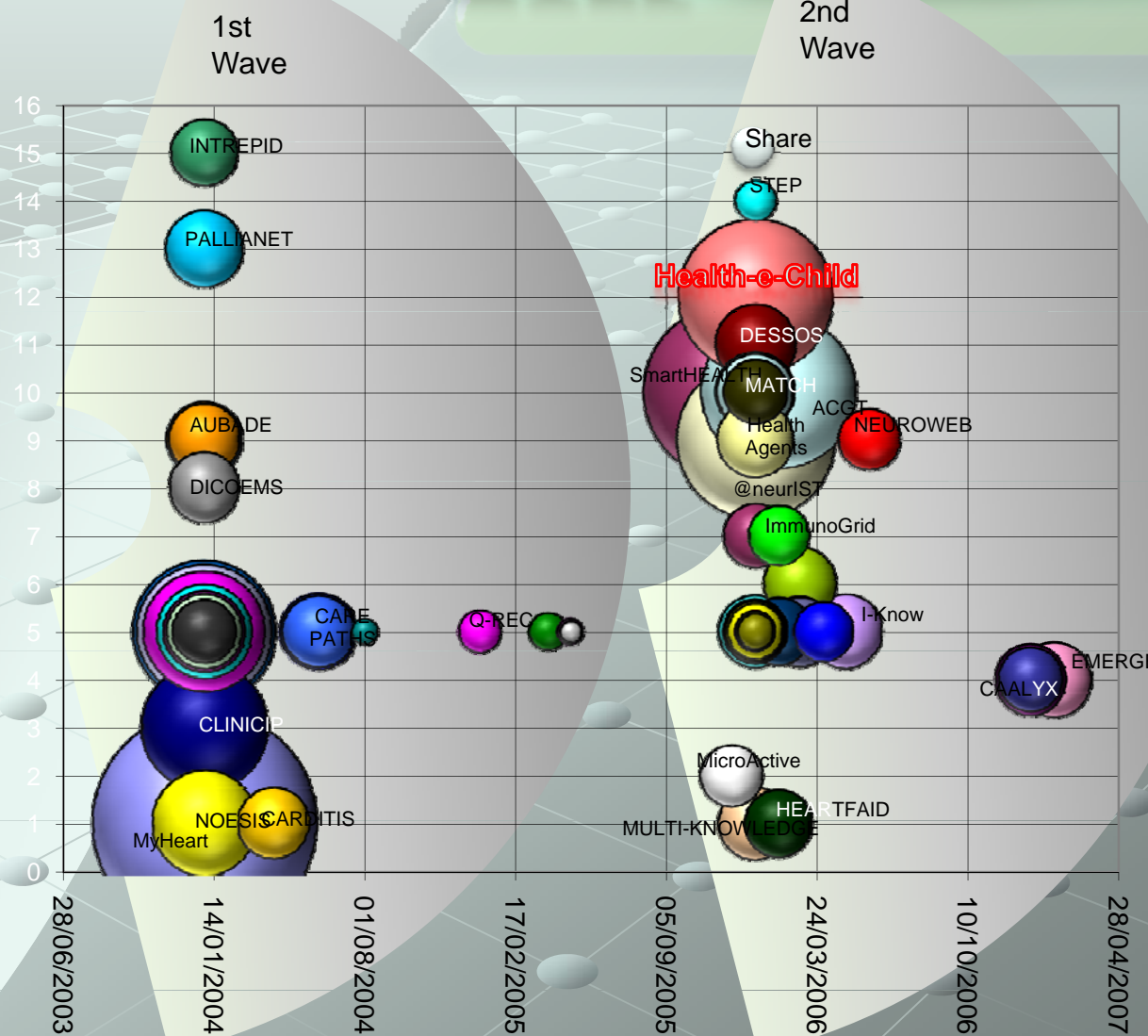
Funding per Area





FP6 eHealth Survey

Projects Waves



- MyHeart
- SmartHEALTH
- @neurlST
- ACGT
- Health-e-Child
- PIPS
- BIOPATTERN
- CLINICIP
- COCOON
- NOESIS
- SemanticMining
- INFOBIOMED
- DESSOS
- SIMAP
- TACIT
- PALLANET
- ASSIST
- Doc@Hand
- HealthAgents
- ALLADIN
- EMERGE
- I-Know
- MULTI-KNOWLEDGE
- CARE-PATHS
- HEALTH PLUS
- K4CARE
- OLDES
- CARDITIS
- AUBADE
- OFSETH
- VIROLAB
- DICOEMS
- LHDL
- INTREPID
- HEARTFAID
- MATCH
- RIGHT
- EuResist
- CAALYX
- ARTEMIS
- SAPHIRE
- MicroActive
- NEUROWEB
- ImmunoGrid

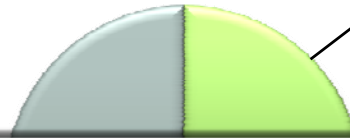


FP6 eHealth Survey

Projects Technologies

Regular Technologies

Centralised Databasing
25%

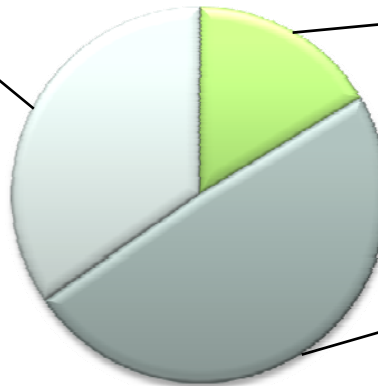


All Technologies

Grid Computing
35%

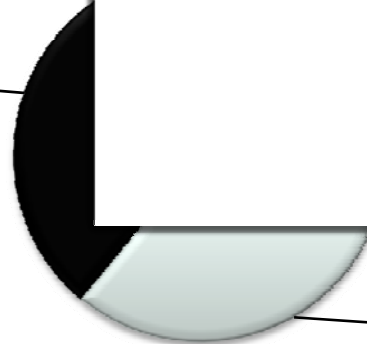
Centralised Databasing
16%

Decentralised Databasing
49%

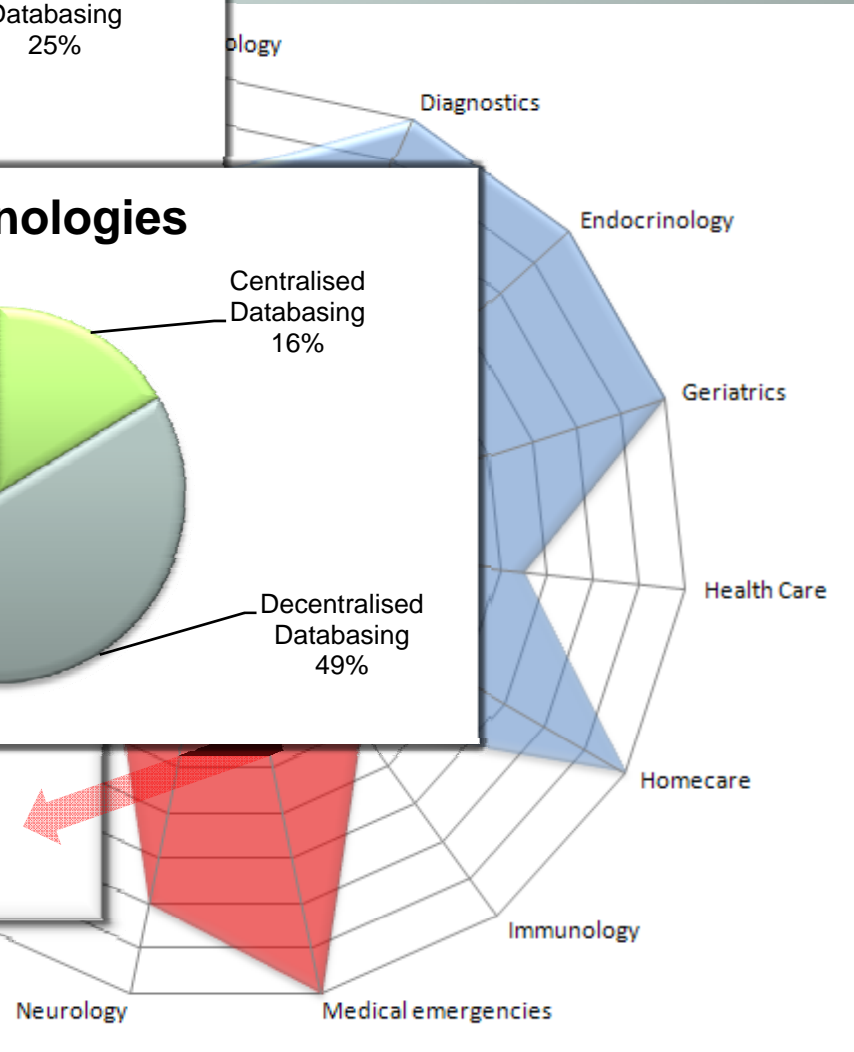


Grid

NA
39%



Grid / Globus
39%





- **Biomedical research in Europe is heterogeneous, fragmented and geographically dispersed**

→ **BUT rich in results... a goldmine!**

- **Production quality Infrastructures are very few, non interoperable nor sustainable**

→ **often such Infrastructures stop existing once projects are over...**

- **At least 80% of the community is developing service oriented tools**

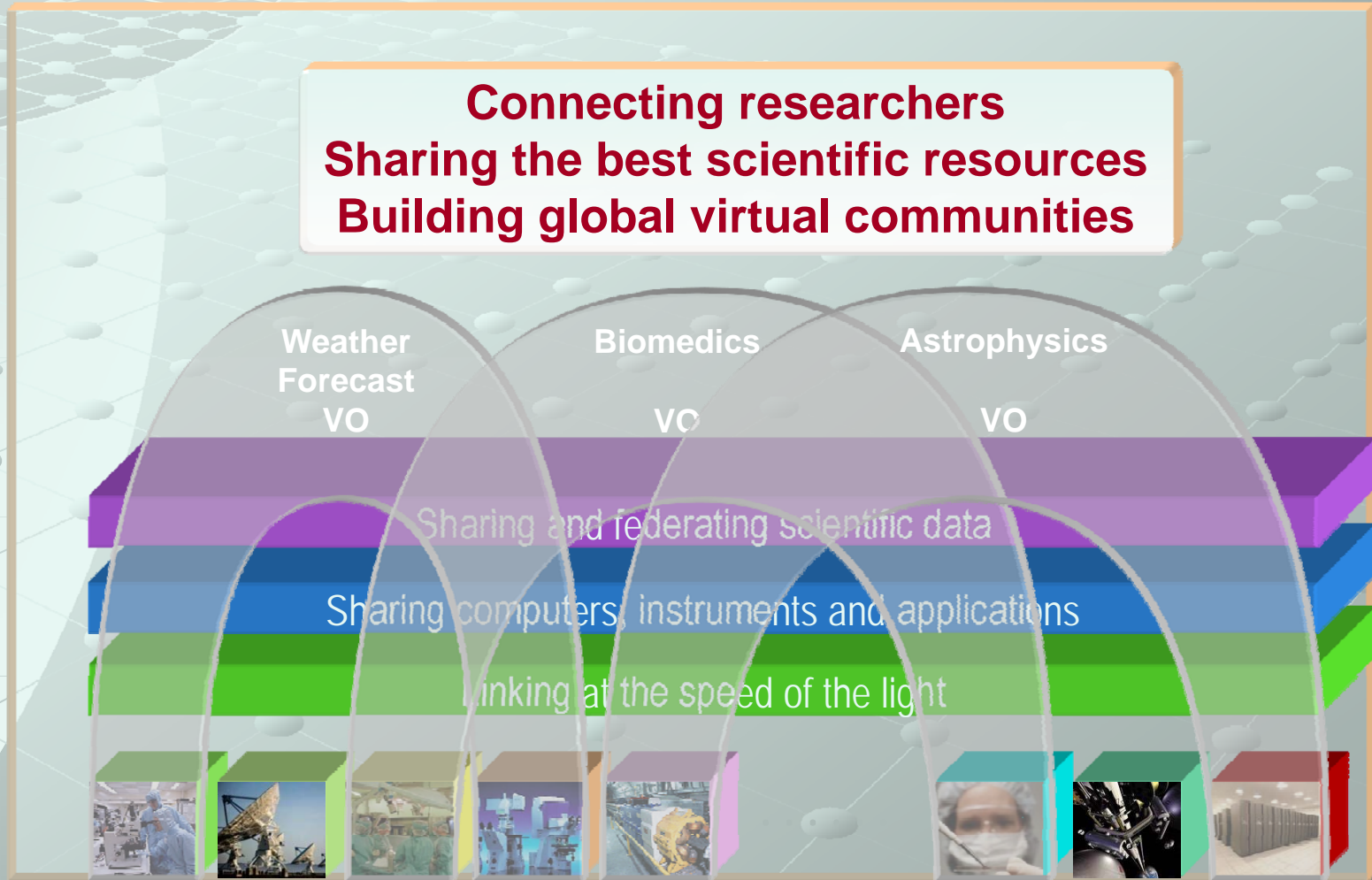
- **No European Infrastructure deployed in hospitals nor in (bio)medical centres**

→ **Technologies are often too complex or advanced to be adopted...**



e-Infrastructure - Vision

Connecting researchers
Sharing the best scientific resources
Building global virtual communities



**Courtesy of Maria Ramalho Natario*



e-Infrastructure - Reality



Commodity Data Services



Community Data Services



GRID
INFRASTRUCTURE

CONNECTIVITY
INFRASTRUCTURE



SPIDeR Objectives

(1) Strengthen Existing & Support New Scientific Data Infrastructures

- Data transfer, storage, curation, archiving
- Data access, interpret, preservation, certification

(2) Do not reinvent the wheel...

- Crystallise research result
- Promote technology reuse
- Capitalize knowledge

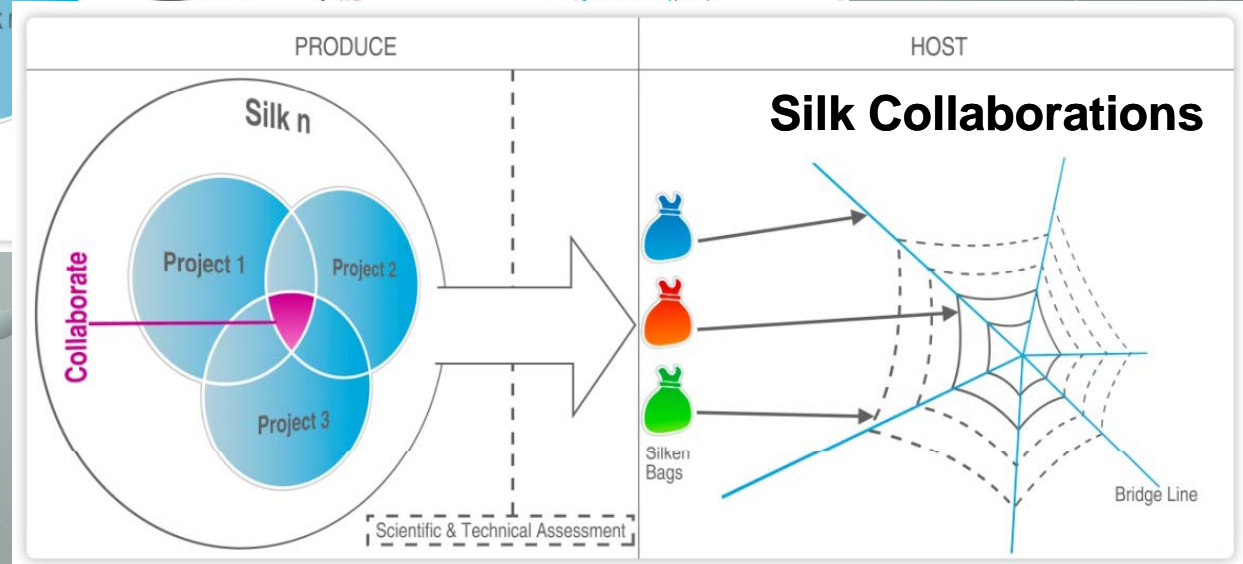
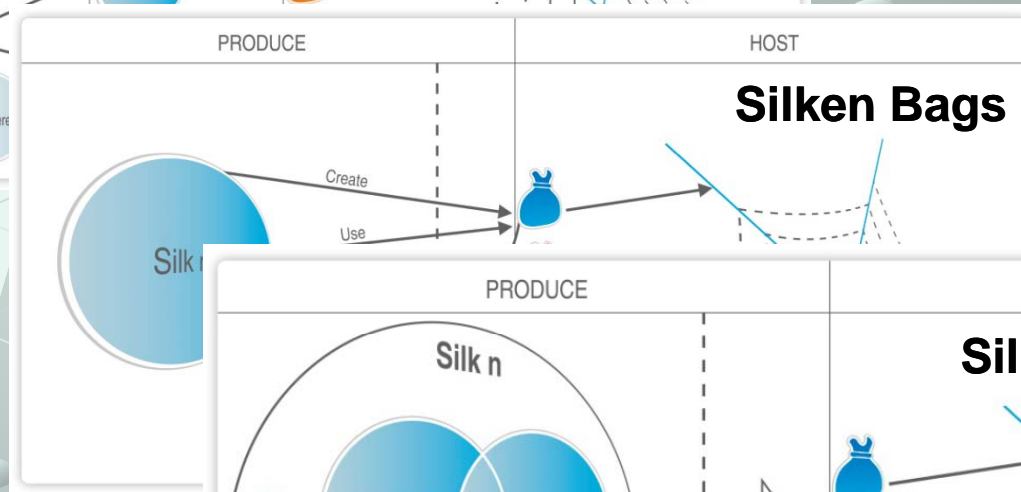
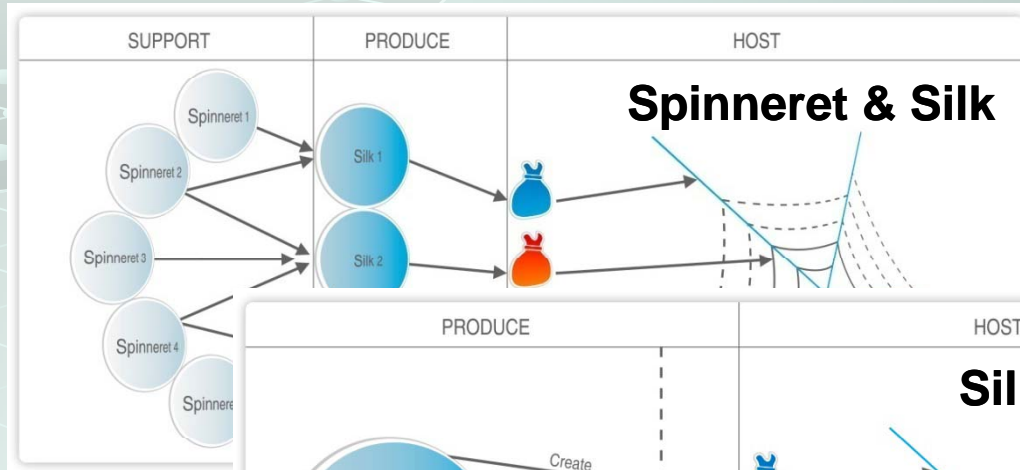


(3) Assess Infrastructure Offer Together with Concrete Projects

- Pairs/Tuples of existing/new projects to validate SPIDeR offer
- Connection of projects to SPIDeR anticipated in budget
- SPIDeR Workplan to follow and support projects waves

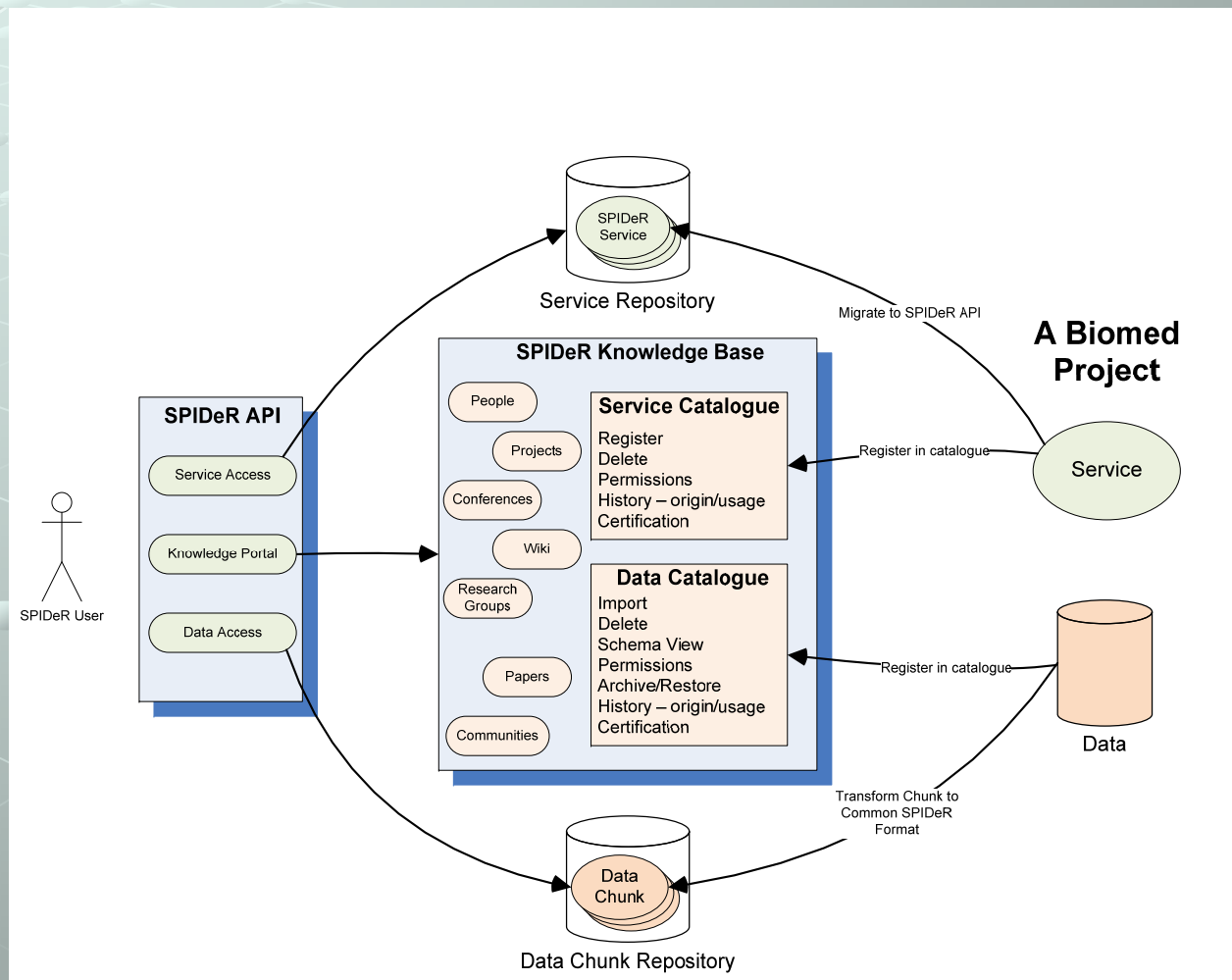


SPIDeR Anatomy (1)





SPIDeR Anatomy (2)



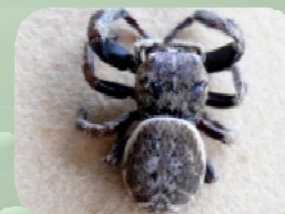


SPIDeR Offer (1/4)



Commodity Services

SPIDeR (DB) – Data Bank & Catalog



- Data cataloguing for existing and new projects
- Long term data storage
 - Guaranteed high Quality of Services
 - RAIDed distributed storage
 - Encrypted space
 - Advanced Management Services
 - Backup and recovery storage
 - Data provenance traceability
 - Data integration / migration

RCDG2

RCDG3

RCDG4

RCDG5

RCDG6

Data Grids

RCLG6

RCLG8

Collab Grids



SPIDeR Offer (2/4)



Commodity Services

SPIDeR (SC) – Service Shelf & Catalog



- Service cataloguing for existing and new projects
- Service-level agreement
 - ELSE concerns modelling
 - QoS-enabled discovery and composition
- Services advertisement to other projects
 - International service repository
 - Service documentation, through SPIDeR KB

RCCG1

RCCG3

RCCG4

RCCG5

RCCG7

RCCG8

RCLG2

RCLG3

RCLG5

RCLG7

RCLG8

RCLG9

Computing Grids

Collab Grids



SPIDeR Offer (3/4)



Community Services

SPIDeR (KB) – Knowledge Base



- Complete, maintained and public knowledge database
 - About projects, technologies, applications and
 - Other resources useful to community
- Translational knowledge database
 - Exploitable by both human and machine
 - Based on pre-existing SHARE Knowledge Database
 - Glues SPIDeR Services

RCKG1

RCKG2

RCKG3

RCKG4

RCKG6

RCKG7

RCLG10

Knowledge Grids

Collab Grids





SPIDeR Offer (4/4)



Community Services

SPIDeR (eT) – Education, Training and Collaboration



- Consultancy from SPIDeR partners/projects' expertise
 - On-demand consultancy for existing / new projects
 - On-demand scientific and technical audit
- On-line public collaboration portal
 - To submit new collaboration proposals
 - To avoid organisational problems
- On-line portfolio and e-Learning facilities
 - Based on SPIDeR KB, with public Web access
 - To discover SPIDeR offer and corresponding info
 - To subscribe to on-line training / events
 - To capitalise and disseminate knowledge

RCLG3

RCLG4

RCLG5

RCLG10

Collab Grids





SPIDeR Web (1)





SPIDeR Web (2)



- A Super Infrastructure of about
- 6000 CPUs,
 - 100TB,
 - 15 new sites dedicated to Biomedical Research

SPIDeR Conclusion (1/2)

(1) Strengthen Existing & Support New Scientific Data Infrastructures

- SPIDeR DB & SC

(2) Do not reinvent the wheel...

- SPIDeR KB & eT

(3) Assess Infrastructure Offer Together with Concrete Projects

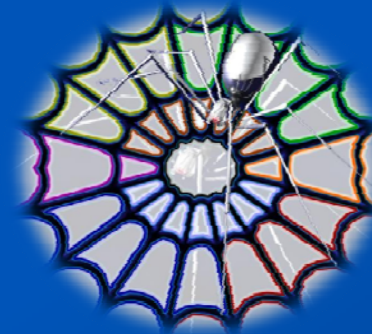
- SPIDeR DB:

(1) *Health-e-Child & @neurIST*

(2) *WISDOM-EU & WISDOM-LA & WISDOM-AS*

- SPIDeR SC:

(1) *neuGRID & CBRAIN & NeuroLOG*



SPIDeR Conclusion (2/2)

SPIDeR is hybrid approach bw US and European projects

- 1st of its kind
- Will turn fragments into continents for communities
- Will accompany successive waves of FP6 & FP7 projects

SPIDeR expected to run for 3 years

- M01-06: Requirements Analysis & Logistics Preparation
- M07-12: Infrastructure Deployment & Projects Selection
- M13-24: Stable Commodity Services & Beta Community Services
- M25-36: Final Release of both Commodity & Community Services

Consortium being defined

Thank you for your attention