
E-Science – The Vision and its Effects on Data-, Knowledge-, and Information Management within Grid-based Virtual Digital Libraries



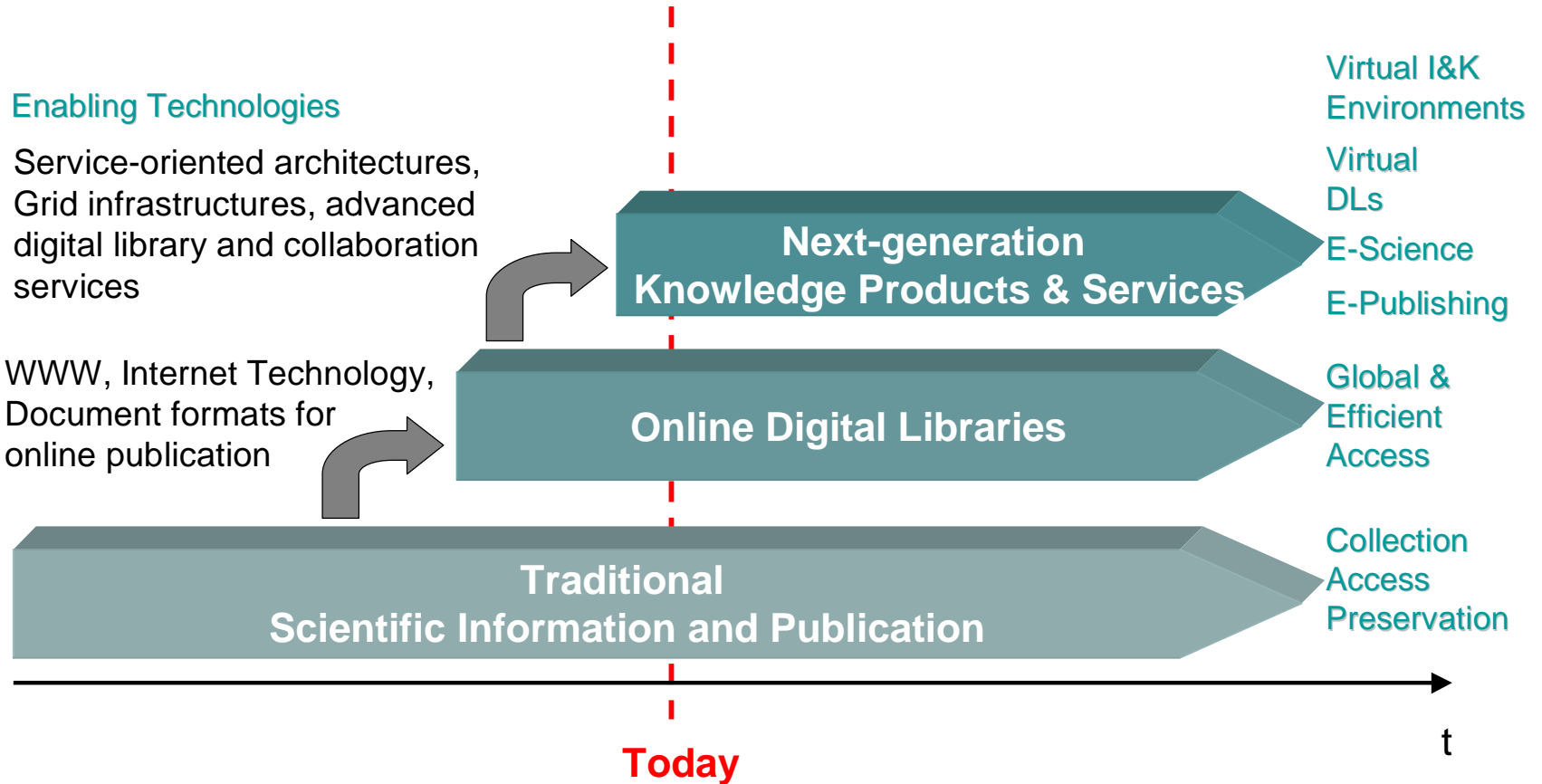
Fraunhofer **Integrated Publication
and Information
Systems Institute**

Matthias Hemmje
Claudia Niederée
Andreas Wombacher

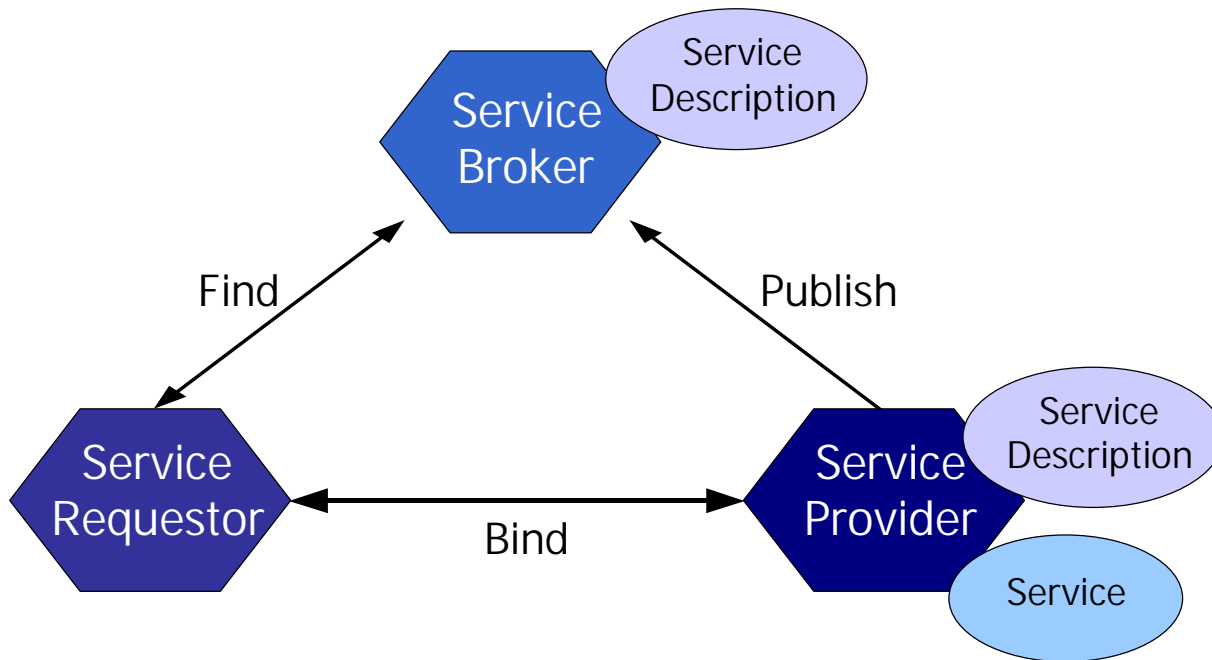
www.ipsi.fraunhofer.de/

- Motivation: Next-Generation knowledge products and services
- Trends in next generation E-Science architectures
 - Service oriented architectures
 - Grid-based infrastructures
- Future Information and Publication Scenarios and Use Cases
- Technology challenges
- Grid-based infrastructures for digital libraries and e-Science support
- Activity Fields

Motivation



An abstract description of how pieces of systems interact to achieve the desired results.



Technology Trend: The Grid and Grid Services

Idea

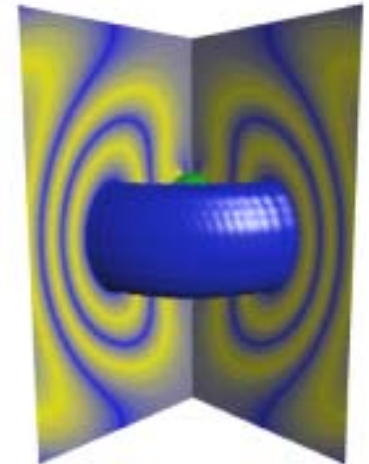
- Worldwide usage of idle resources
 - Computational power
 - Storage
 - Services
- Born from the scientific requirements on huge amount of storage and computational power

Vision

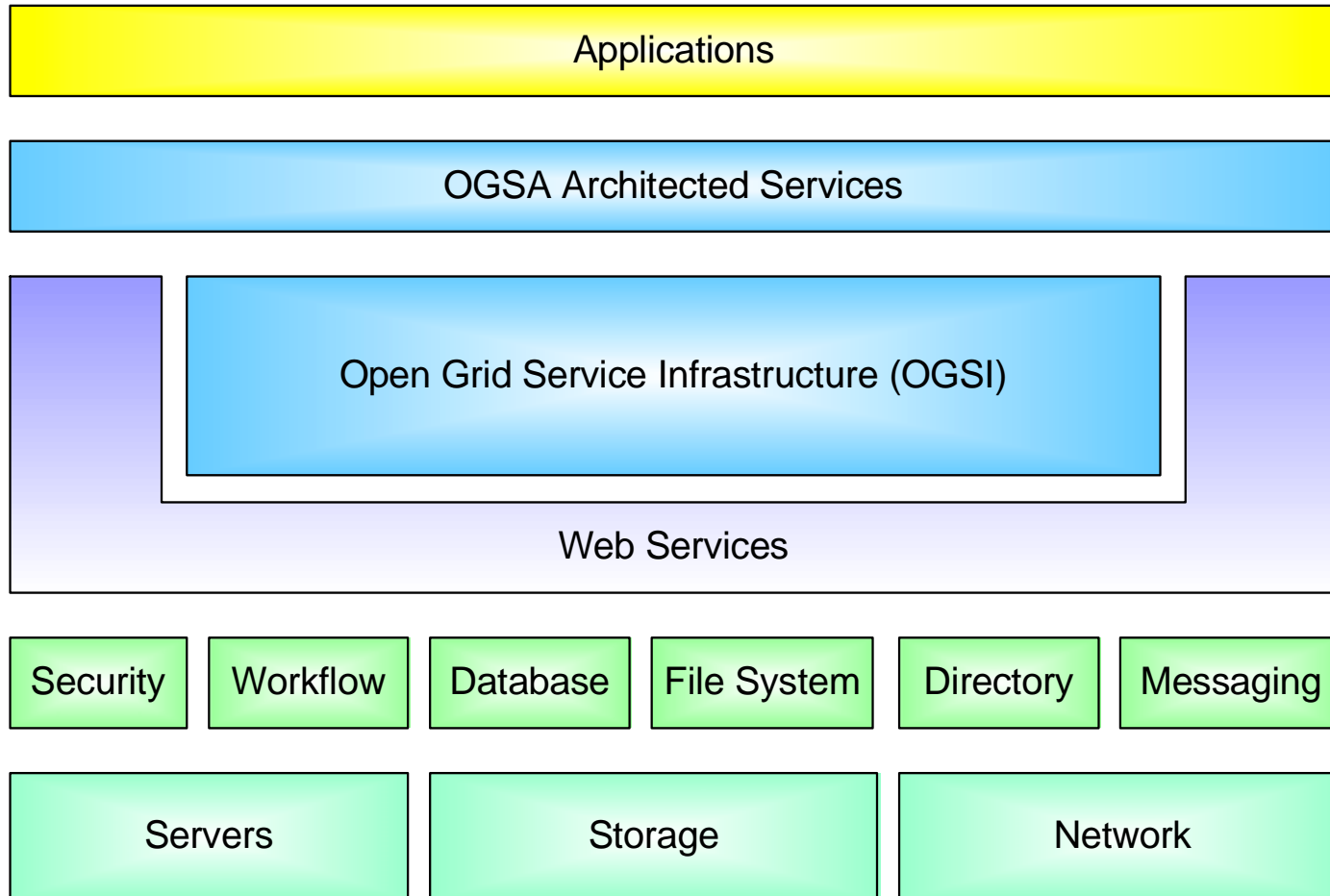
- Consume IT resources from the internet as easy as electrical power from the electricity grid

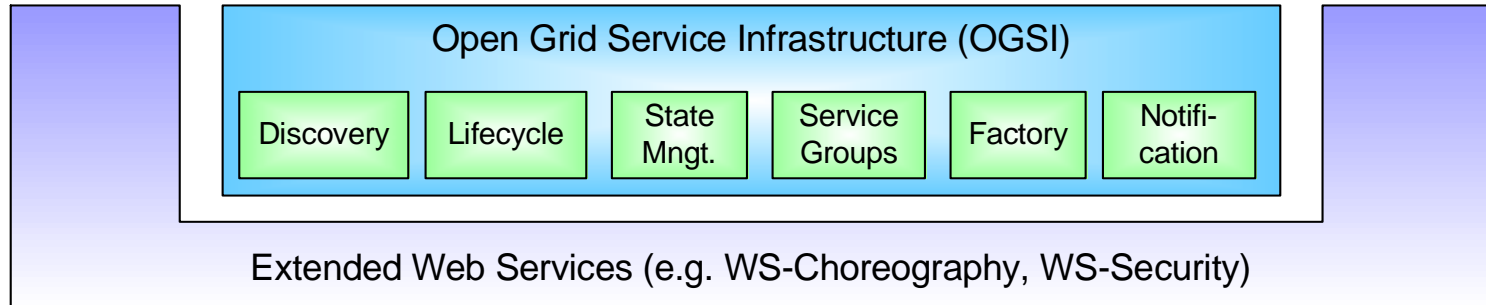
Advantages

- Dynamic allocation of resources
- Cross-organizational resource sharing
- Resource owner still have the full control
- Security infrastructure



Open Grid Service Architecture





Dynamic creation of services

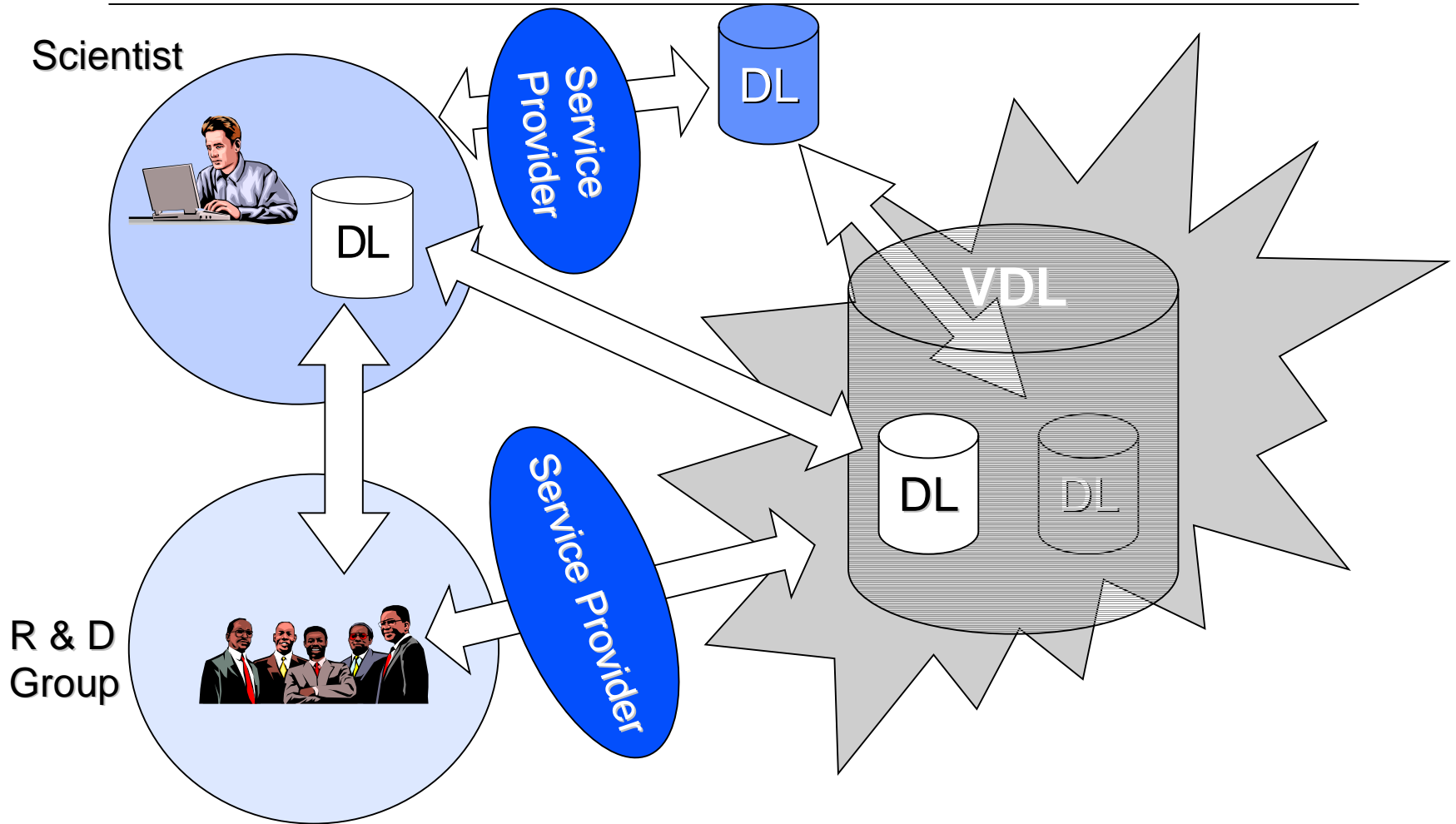
- Temporal Services have a lifetime
- Management of creation and destruction of services

Support of service states

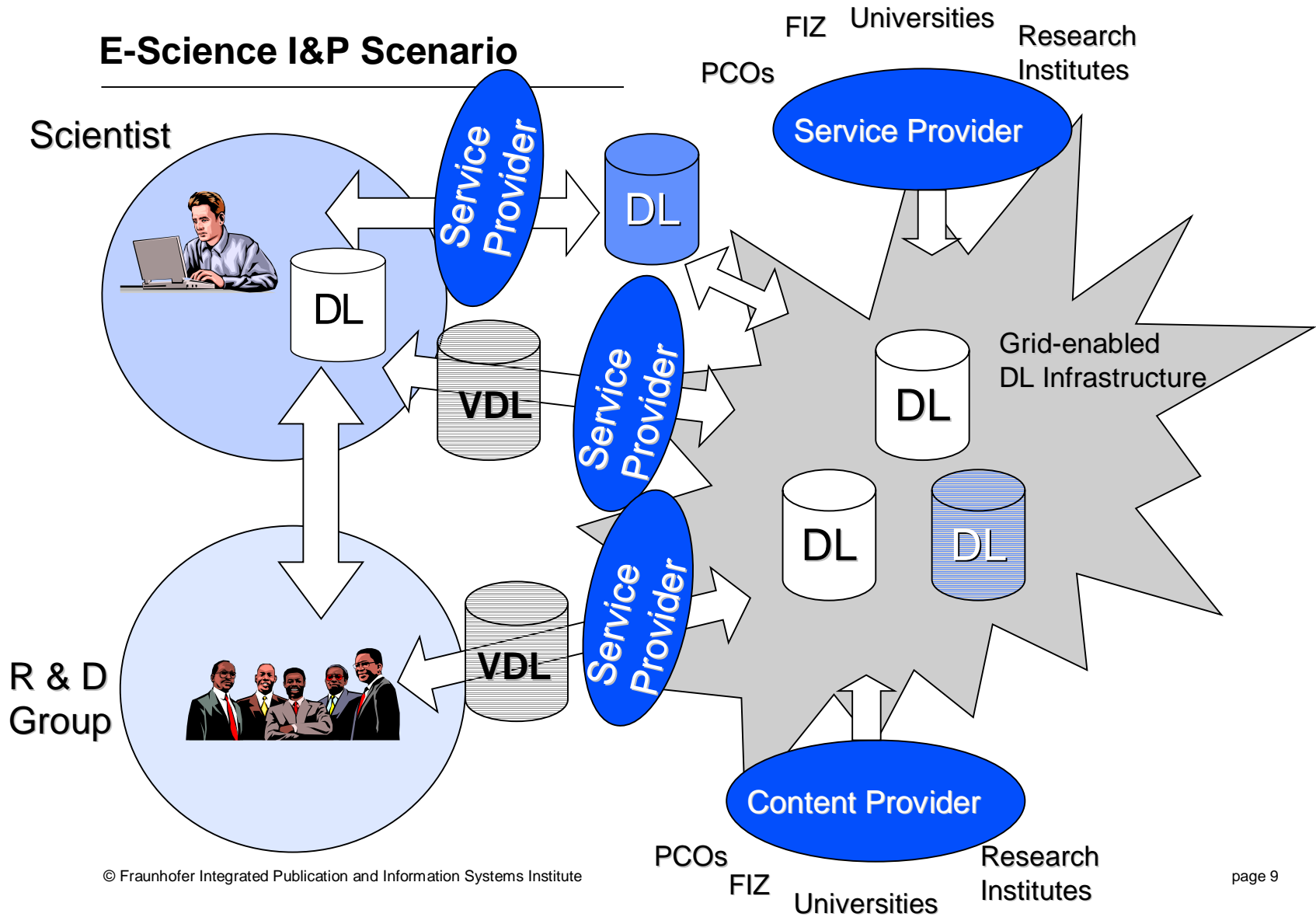
Collection of similar services (Service Groups)

Monitoring and Notification of service status

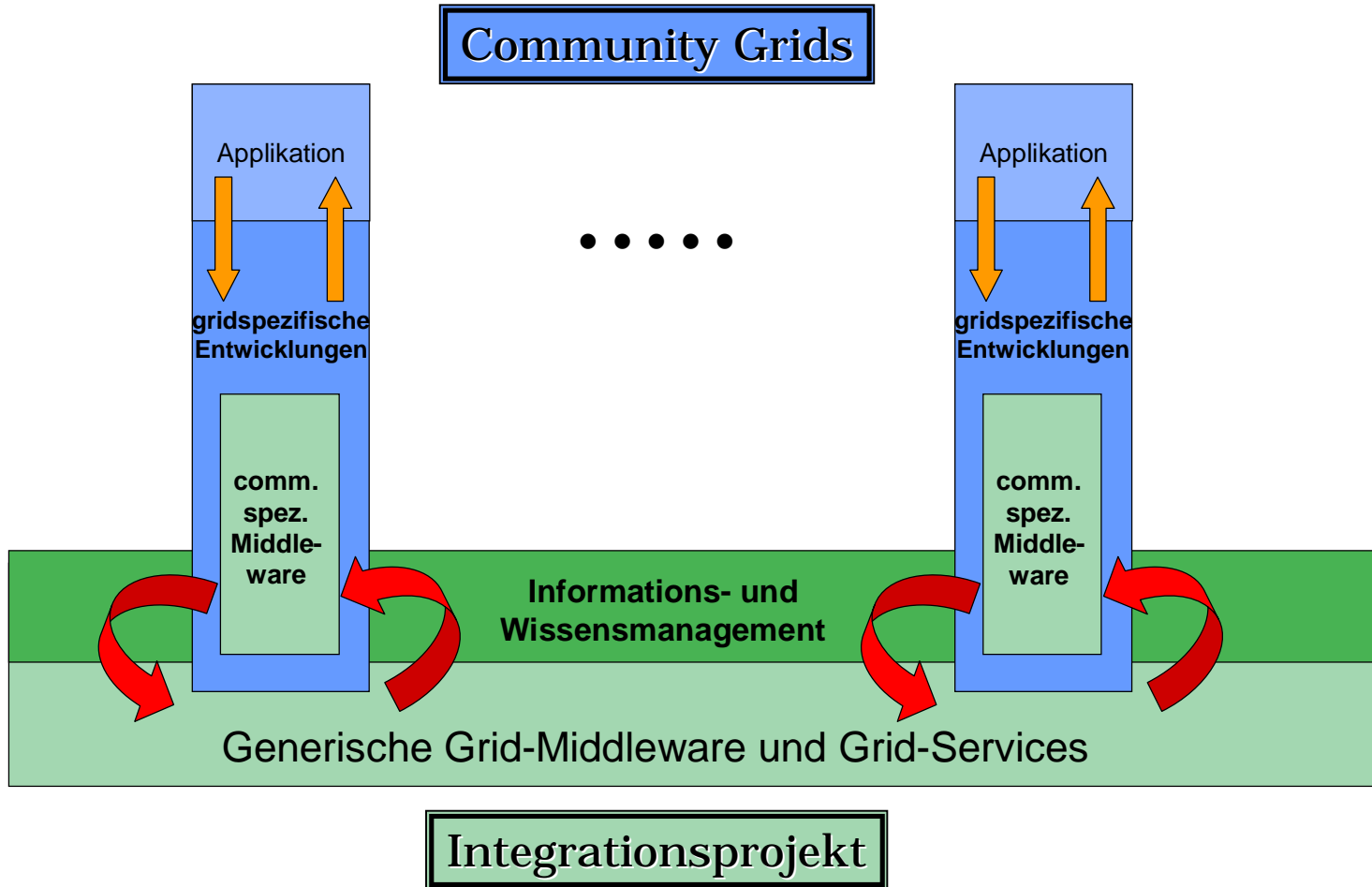
Traditional Information and Publication Scenario



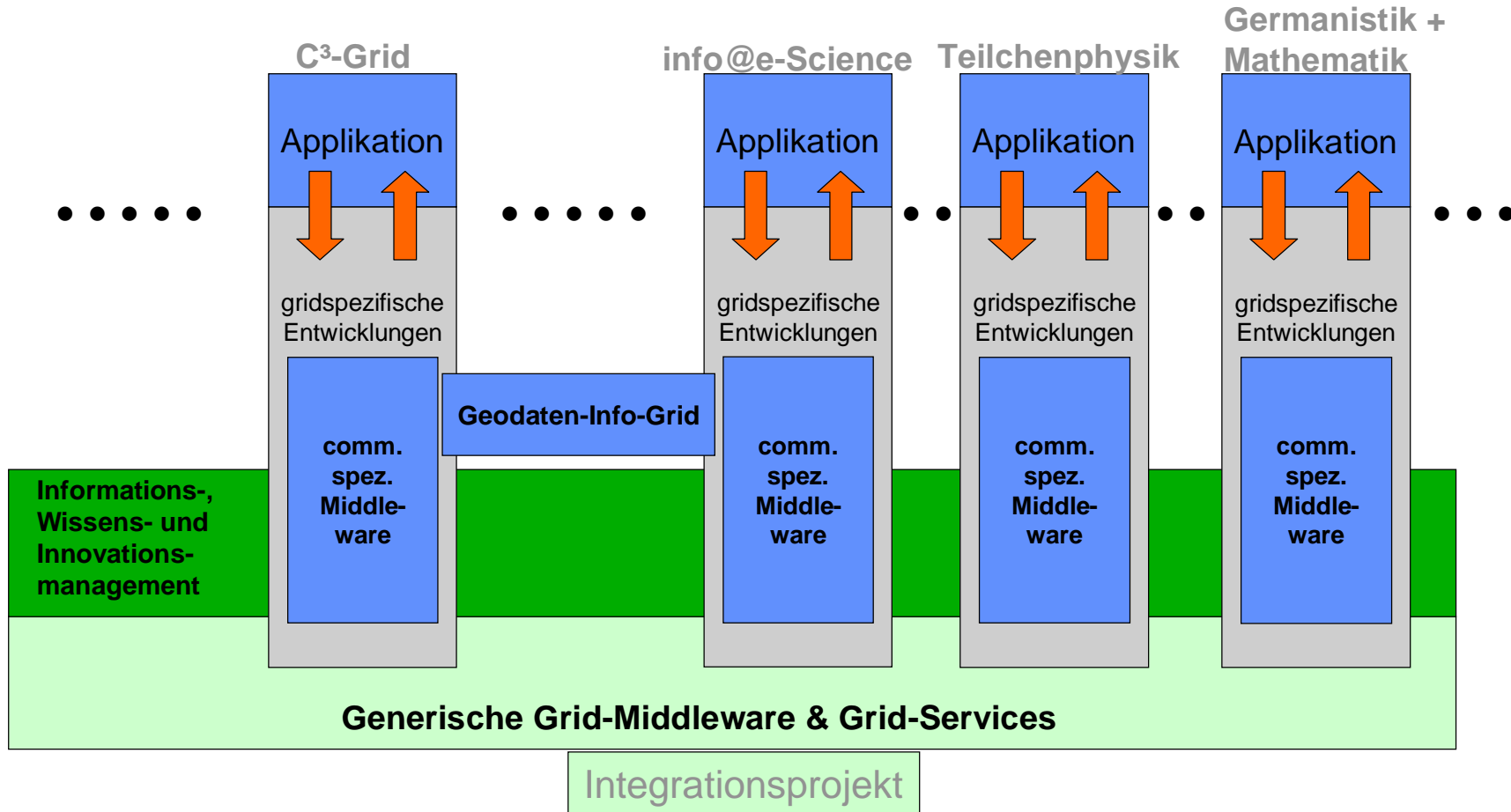
E-Science I&P Scenario



General R&D Integration Approach



Initial R&D Communities interested to get going ...



Andreas V. is biology **researcher** and

- is looking for raw data from experiments that support his new hypothesis assay for his gene research problem; **[discovery of scientific raw data]**
- is part of a distributed team of researchers from different disciplines that work on a research project **[scientific collaboration]**

While preparing her paper on experiences in a recent environmental project **scientific author** Barbara W.

- is looking for relevant contributions on environmental projects in last years proceedings to update her knowledge on state of the art related to her contribution **[state of the art]**
- is looking for sensor data and satellite images on the pollution of the Mediterranean sea **[multimedia data]**

While reviewing a paper on an ontology for learning content for the Semantic Web Conference, **reviewer** Michael W.

- is looking for related work on RDF schemata for learning content to verify the papers claim that no such schemata exist; **[confirmation of claims]**

In preparing next years ECDL [conference organizer](#) Claudio M.

- is looking into the conference program of the US and the Asian Digital library conference to achieve a differentiating factor for the European counterpart; [\[differentiation from other conferences\]](#)
- is looking for appropriate recipients for sending out the call for paper and the call for participation and for new persons for the program committee; [\[addressing the right people\]](#)

[Librarian](#) Martha B. from the NIT Library

- is identifying trends in the library's domain to re-structure the library catalogue and to plan future collection policy [\[adapting library focus\]](#)

[Division Manager](#) Karin M. of the Max-Fraunhofer Research Institute

- is looking for open research issues to make up the research roadmap of her institute [\[open research issues\]](#)

Resulting Knowledge Products & Services within info@e-Science for E-Science

For example:

Next-Generation Publishing Services for authors:

- service for preliminary rated classification of the paper or abstract into the conference structure/domain ontology for conference and track selection
- browsable domain model views with contributions connected
- state of the art surveys based on semantic analysis of contributions on a topic, reviews on the state of the art presented in documents, and on document structure analysis
- tailored environments for scientific information, publication and related collaboration;

Next Generation Services for Reviewers

- semantically annotated overview over publications in a domain;
- annotation of the conference program based on domain ontology;
- state of the art surveys based on semantic analysis of contributions on a topic and reviews on the state of the art presented in documents;

Goal: Grid-based Infrastructures within info@e-Science for Virtual Digital Libraries and E-Science support

- Grid-based digital library infrastructure
 - enables cost-effective access to DL technologies → extended clientele for DL technology (see below: creation of Virtual Digital Libraries)
 - supports metadata management and metadata brokering
 - enables the integration of DL services and information collections as GRID services
 - flexible definition of DL and community workflows
 - advanced data and information services

- Services for the creation and management of virtual digital libraries in support of research teams (virtual organizations)
 - selection and integration of relevant resource collections
 - selection and adaptation of DL services as required
 - definition of project-specific processes for information selection, information processing, information publishing etc.

- Innovative value-added services in support of the different stakeholders of the e-Science and the e-Publishing processes, e.g.
 - semantic annotation of content (based on RDF, RDF Schema, OWL)
 - collaboration services (based on CSCW Technologies)
 - advanced services for context-based personalized access
 - workflow support (based on BPEL, BPML)
- Development and exploitation of adequate standards (based on XML, RDF, OWL, SOAP, WSDL, BPEL, BPML, XML Signature, WSS SAML)
- Flexible and dynamic architectures
 - on-demand access to resources (services, computational power, etc.)
 - dynamic definition, execution and control of tailored information handling processes (e.g. publishing, information enrichment, reviewing)
 - tailored configuration of DL services

E-Science R&D Approach and D-Grid Activity Fields

Track I

Exploitation of existing services and technologies;
adoption of solutions from other domains;
establishment of basic D-GRID infrastructure;

Track II

Adoption and Adaptation of D-GRID and DL technology that is currently under development

Track III

R&D for Advanced E-Science Information, Collaboration and Publication Support, i.e., approaches that require further integrated research w.r.t. VDLs

Fine

Thank you very much for your attention.

Prof. Dr.-Ing. Matthias Hemmje

Fraunhofer-IPSI
Virtual Information
and Knowledge Environments
Dolivostraße 15
D-64293 Darmstadt, Germany

University of Duisburg-Essen
Faculty of Engineering
Institute for Computer Science
Chair for Media Informatics
47048 Duisburg

email: Matthias.Hemmje@ipsi.fhg.de
phone: +49-6151-869-844
fax: +49-6151-869-6844

Gartner Group (technical definition)

- Web Services are loosely coupled software components that interact with one another dynamically via standard Internet technologies.

Forrester Research (business definition)

- Web Services are automated connections between people, systems and applications that expose elements of business functionality as a software service and create new business value.

