www.eurocris.org

Institutional Repositories and CRIS infrastructure, concepts and organisation

Keith G Jeffery

President, euroCRIS

keith.jeffery@.stfc.ac.uk

©euroCRIS/Keith G Jeffery

Agenda

- Introduction – speaker
- Requirements
- Institutional Repositories
- CRIS Purpose & Stakeholders
- CERIF-CRIS at the centre of the Organisation
- e-Infrastructure
- Synthesis
- Role of euroCRIS

Who ?

- Director, IT & International Strategy
  - Strategy, advice
    - International
    - UK Government
    - UK Research Councils
    - STFC
    - STFC Departments
  - SSC Project Design Authority
- President ERCIM
- President euroCRIS
- Chair,Alliance for the Permanent Access to the Records of Science

euroCRIS
Current Research Information Systems

2006-

Director, IT & International Strategy

1999-2006

Director IT & Head BITD

(IT, library, photorepro; > 1000 servers, 360000 users)

Agenda

- Introduction – speaker
- Requirements
- Institutional Repositories
- CRIS Purpose & Stakeholders
- CERIF-CRIS at the centre of the Organisation
- e-Infrastructure
- Synthesis
- Role of euroCRIS

The R&D Process

Workprogramme

Proposal

Project

Results

Exploitation

WealthCreation

Nirvana

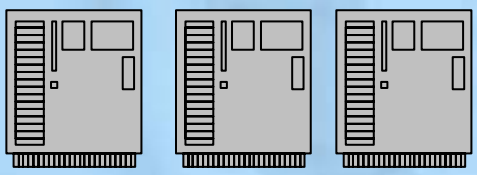
Requirement

Actors	Roles
• Researcher	• Review existing material • ideas, techniques • Research products
• Research Manager • Funding Agency • Policymaker	• Evaluation/decision-making • researcher, organisation
• Innovator	• Search for innovative ideas
• Educator • Student	• Discover teaching material
• Media	• Input to ‘stories’ for public • interest, ethics

Requirement

application

middleware



- User:
- Fast
- Easy
- Homogeneous
- Sharing
- Legal
- Cost-effective

Technical:

GRIDs/SOA

Formalised metadata

Canonical syntax/semantics

Benefits

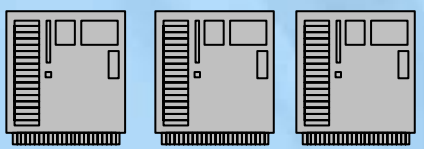
- Faster research turnround – more progress
- Originator improved quality – access & review
- Community improved quality – access & review
- Improved innovation
- Improved education
- Improved public engagement
- Improved PR for institution
- => wealth creation / quality of life improvement

Scenario

Not only work with the e-literature repository but also…..

application

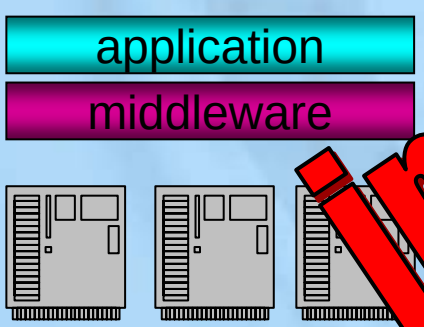
middleware



- CRIS
- project, person, organisational unit, research output (products, patents, publications), funding, facilities, equipment, events……
- e-Research repository
- research datasets, software
- e-Research
- control experiments, take data, visualistaion, in-silico experiments (simulation)
- e-Process
- Workflows, research applications, travel requests, claims

Scenario

Not only work with the e-literature repository but also…..



- CRIS
- project, person, organisatio[illegible] output (products, paten[illegible] funding, facilities, equipment[illegible]
- e-Research [illegible]
- research d[illegible]
- e-R[illegible]
- [illegible]ents, take data, visualistaion, in-[illegible]o [illegible]riments (simulation)
- [illegible]ocess
- workflows, research applications, travel requests, claims

in one environment

Agenda

- Introduction – speaker
- Requirements
- Institutional Repositories
- CRIS Purpose & Stakeholders
- CERIF-CRIS at the centre of the Organisation
- e-Infrastructure
- Synthesis
- Role of euroCRIS

Repositories

- Document / article repositories
- - simple metadata (discovery, description)
- - ePrints, DSpace, Fedora, ePubs….
- e-Research repositories
- - more complex metadata (discovery, description, usage control, software parameters…)
- - ‘homebrew’ systems – portals to research datasets and software

IRs and CRs

- Institutional Repository
  - IP of organisation in one place
  - PR and brand recognition
  - Mandating
  - Workflow and management /quality control
  - Automated evaluation
- Central Repository
  - Convenient for subject domain (one stop shop)

Policy
OA Benefits

- Ethics: public access to publicly funded research
- Research Impact: greater access and use
- Costs and economic benefit: reduced costs and clear benefits to economy of open access
- Metrics: easier to get real metrics of usage
- Added value: link OA repositories to CRIS etc
- Just reward: overcomes publishers profiting from scholarly work provided free

Policy
Barriers to OA

- Loss of publisher income: publishers fear catastrophic cancellations of subscriptions
- Copyright: transfer author ➔ publisher so cannot re-use (in fact mostly can)
- Access Difficulties: DC metadata insufficient
- Completeness: 8-15% fill: need mandates and better workflowed input/update systems

Policy

- Many OA declarations (Budapest onwards)
- Increasing use of green institutional OA repositories – publisher permissions (embargoes)
- Publishers offering OA – but author / institution pays (gold)

- Note: for highly productive institutions gold costs more than subscription models

Mandates

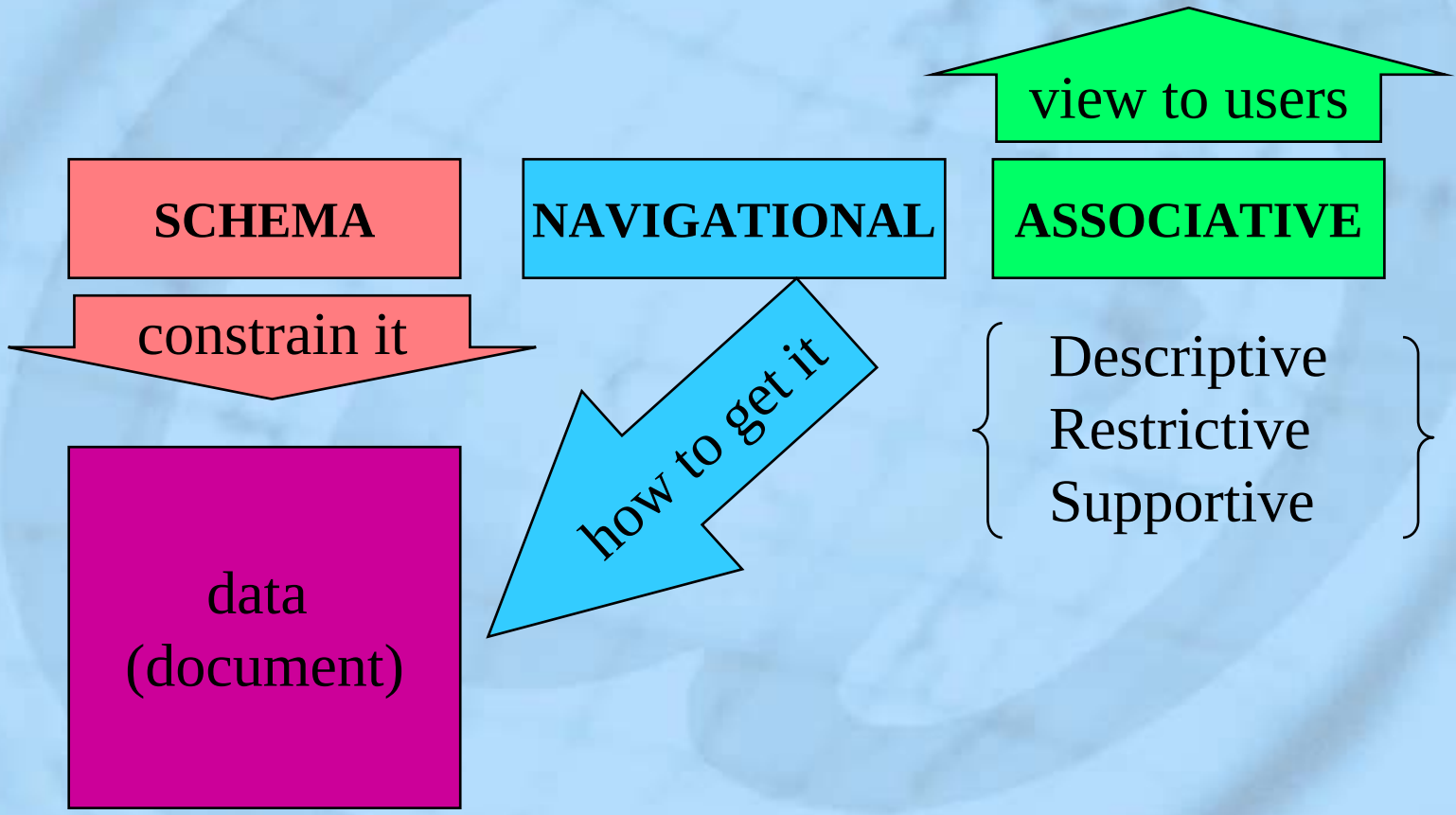
- Progressively more mandates – institutional and funding organisations

- The preferred, optimal and recommended procedure is :
- immediately upon acceptance for publication the metadata and full article are deposited in an institutional repository.
- if the publisher does not demand an embargo period both are set to open access;
- if an embargo period is demanded then only the metadata is made visible until the end of the embargo period.
- Of course, associated with the metadata record there can be (and ePrints provides) a ‘request button’ so that the material can be sent automatically to any researcher who requests it under the usual ‘fair use’ conditions.

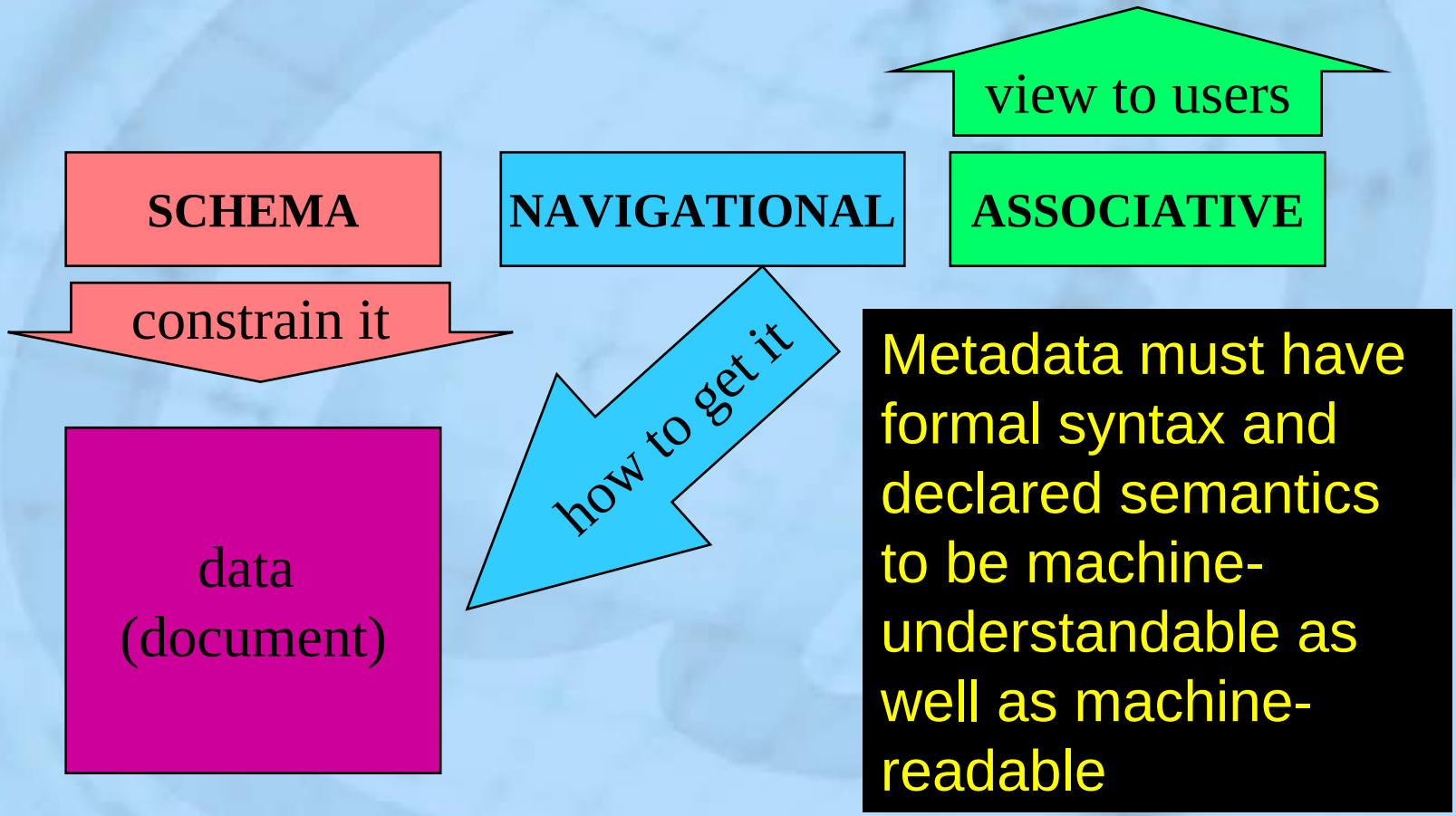
Integration

- Need all funding organisations to mandate OA in institutional repositories
- Resistance from publishers (including learned societies as publishers)
- Engage with them to find new business models

Classification of Metadata



Classification of Metadata



Dublin Core

- Simple DC
- elements, some syntax, no semantics
- Qualified DC
- better syntax, namespaces
- More Recent proposals (2007)
- abstract data model, RDF

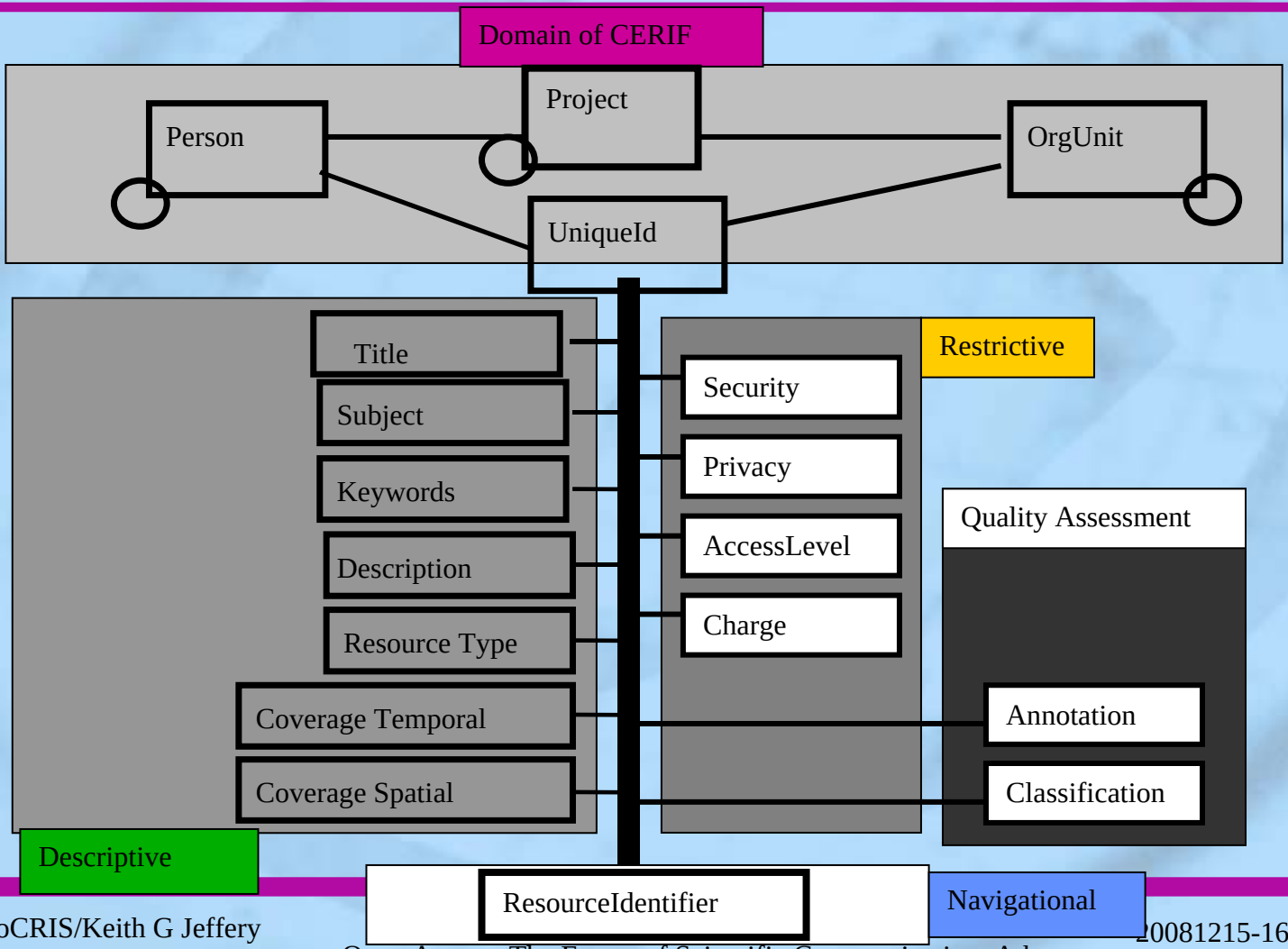
DC Problems

- In parallel (1999-present) criticism of DC:

1. Syntax and semantics not sufficiently formal
2. <creator>,<contributor>,<publisher> are ROLES of person or organisational unit not base entities
3. <relation> : extremely general
4. <source> : is a variant of a role-based relationship object<>object
5. <coverage> recently separated into geographic and temporal but needs formalisation
6. Formalised version of DC proposed 1999, considered, now in CERIF

7. Note: recent (2007) work on DC and SWAP going in this direction

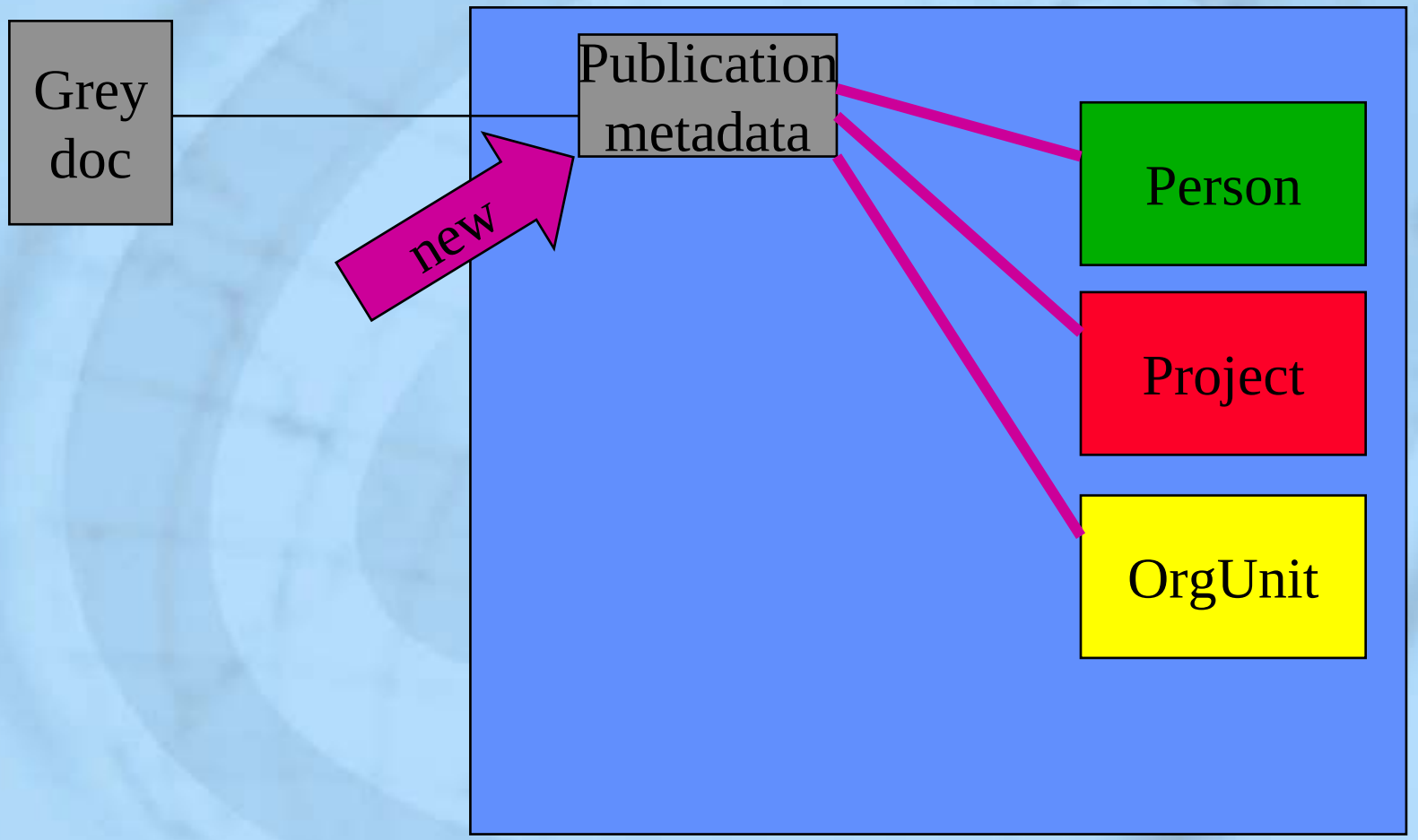
Publications



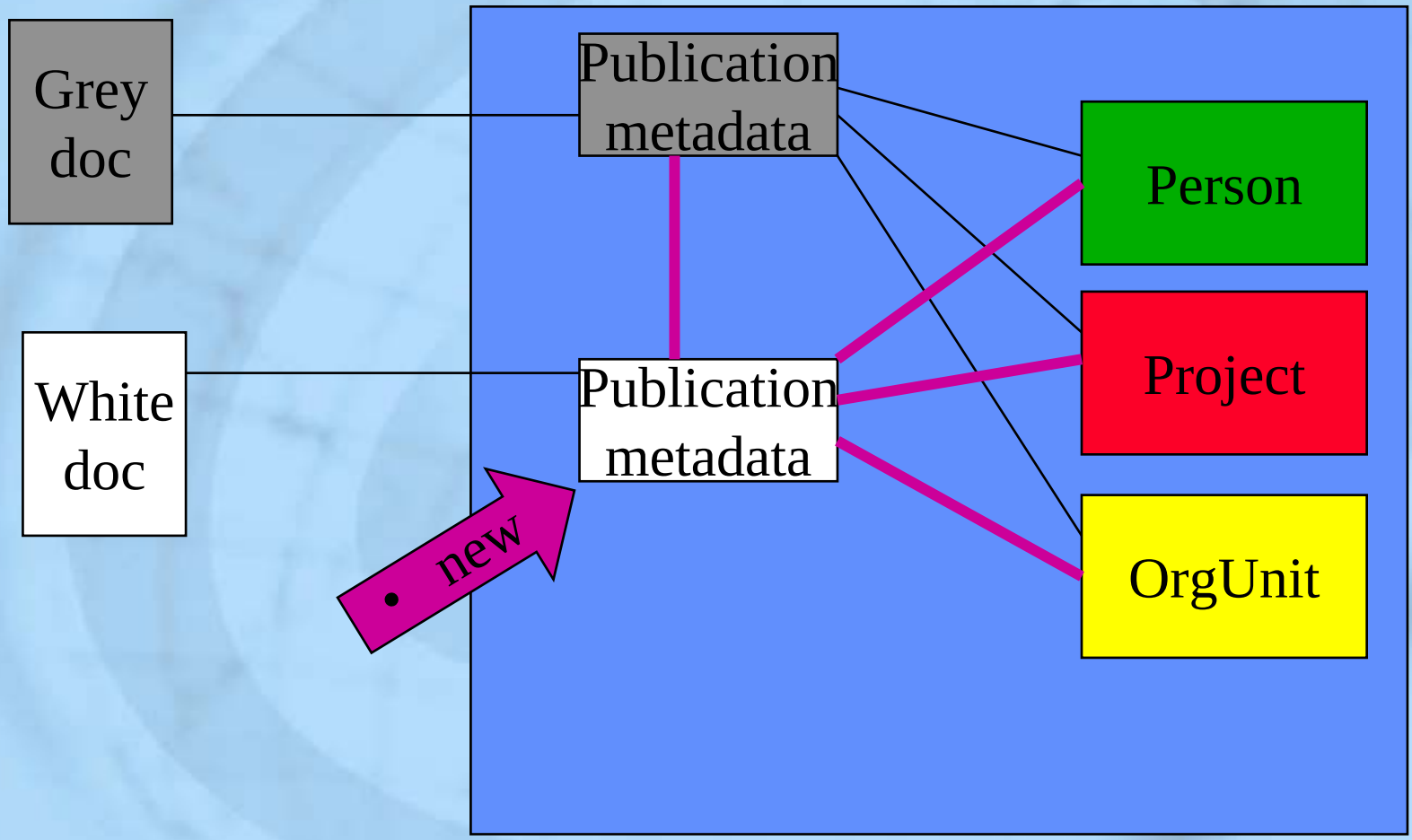
But the problem with metadata is

- It takes too much effort for the researcher to put it in (many web-form-screens)
- So have to input incrementally, no repetition, using the workflow..
- And not re-keying data stored already elsewhere in other (linked-up) systems

Progressive Recording
Grey Document



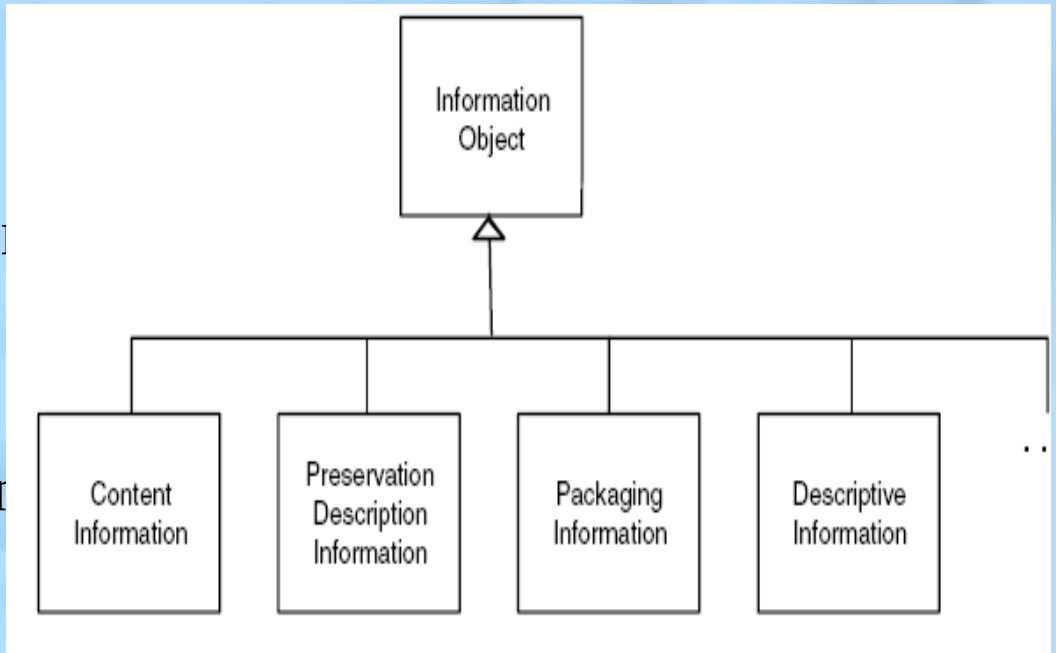
Progressive Recording
White document



Curation

- Problem
- fast changing media: need media conversion
- digital fading: need for refresh
- metadata to understand late

- Answer
- OAIS : but provides only ar architecture: no interoperation metadata



Speculation

- Author deposits in green OA IR
- Push technology informs learned society
- Referees access and record reviews
- Learned society places ‘kitemark’

- Or anyone can referee and record review?

- Note: JISC OJIMS (Overlay Journal Infrastructure for Meteorological Sciences) exploring this space from March 2007

Agenda

- Introduction – speaker
- Requirements
- Institutional Repositories
- CRIS Purpose & Stakeholders
- CERIF-CRIS at the centre of the Organisation
- e-Infrastructure
- Synthesis
- Role of euroCRIS

CRIS

"a Current Research Information System, commonly known as "CRIS", is any information tool dedicated to provide access to and disseminate research information" (www.eurocris.org)

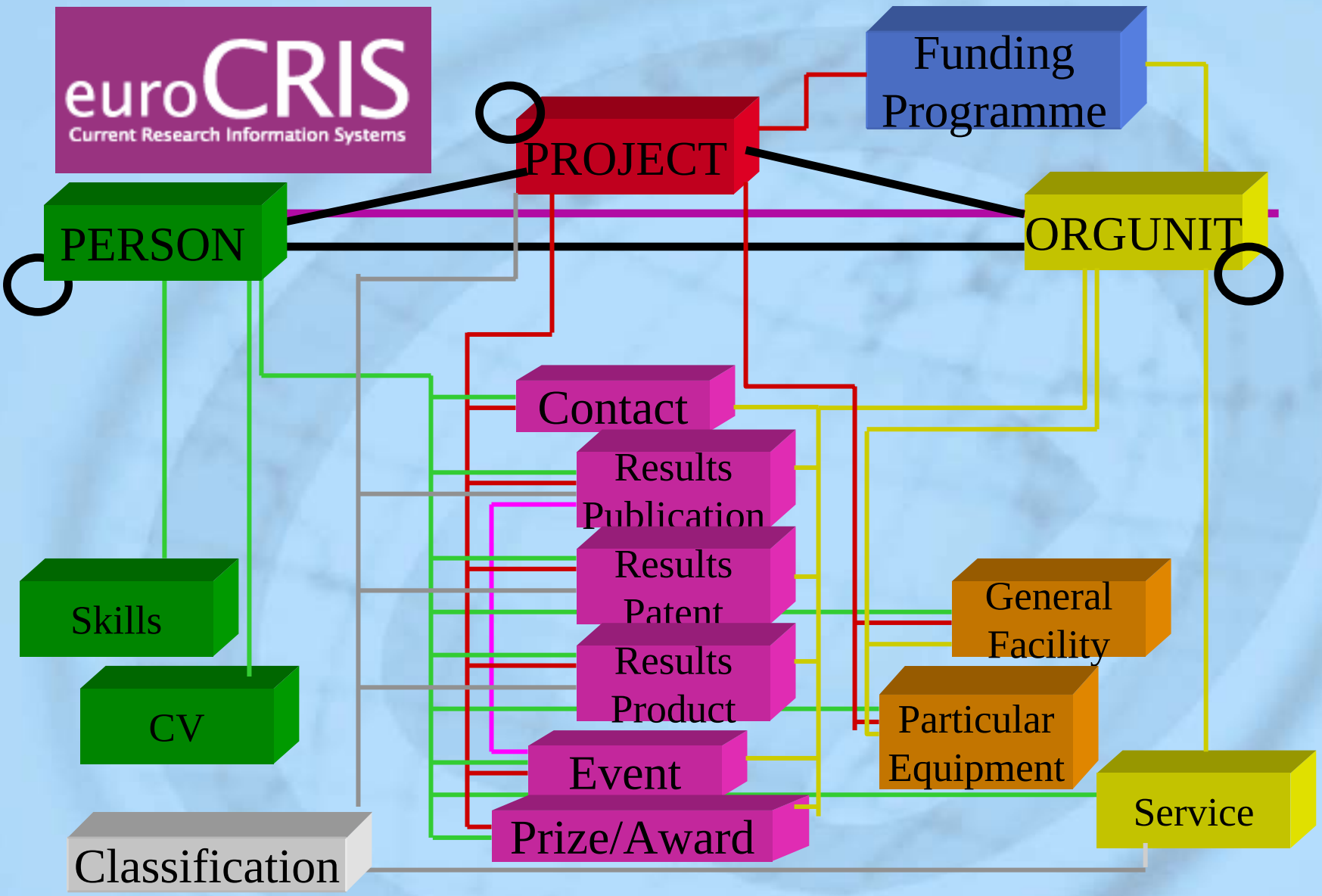
- A CRIS consists of
  - a datamodel describing objects of interest to R&D
  - a tool or set of tools to manage the data

Purpose CRIS

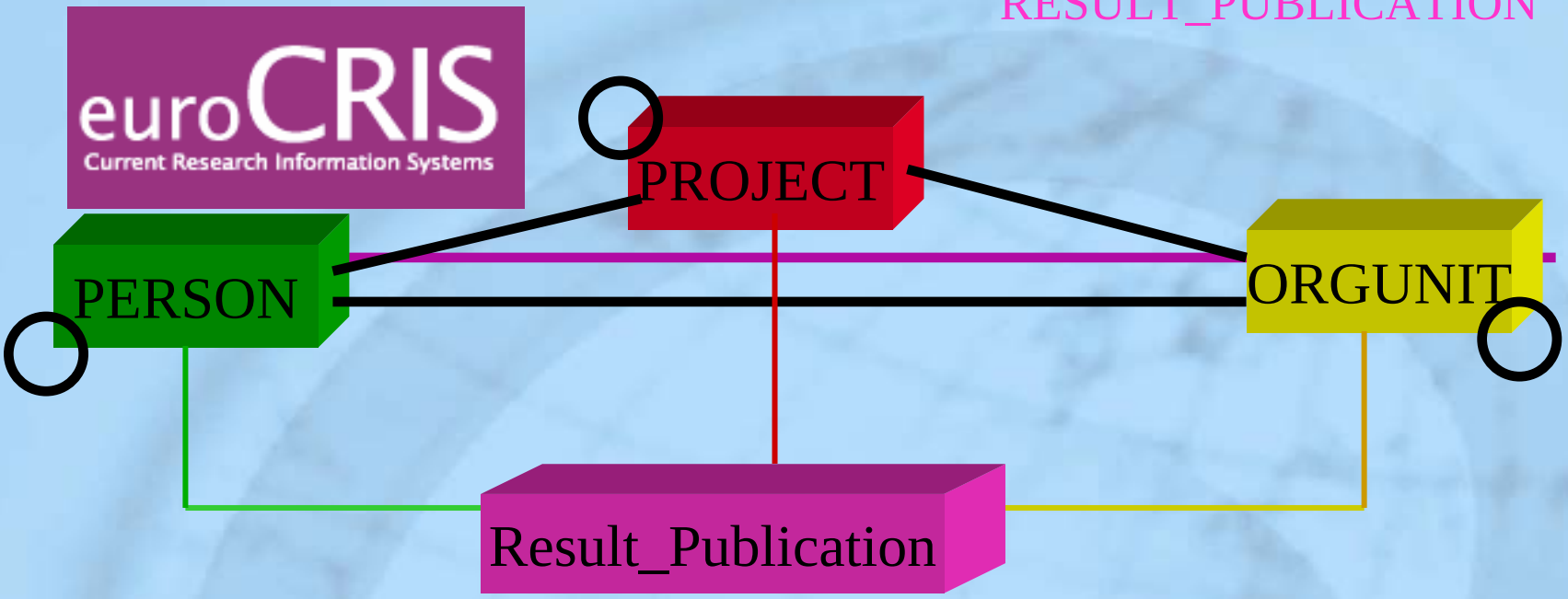
- To assist users in their recording, reporting and decision-making concerning the research process
- whether developing programmes, allocating funding, assessing projects, executing projects, generating results, assessing results or transferring technology

The Users

- Research and Development Information
  - For the political decision-makers
  - For the funding organisations
  - For the entrepreneurs
  - For the researchers
  - For the innovators
  - For the media
  - For the general public

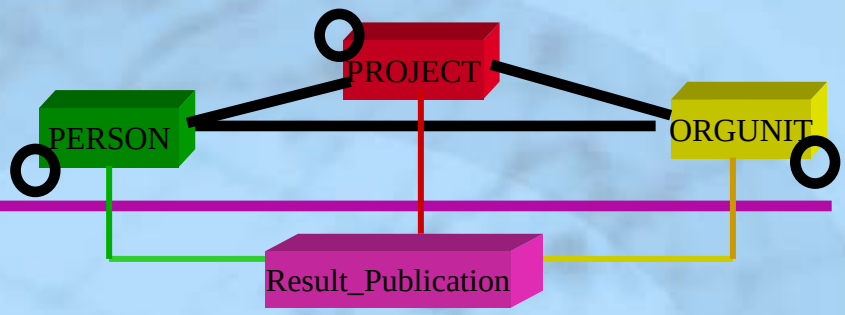


CERIF: EU Recommendation to Member States



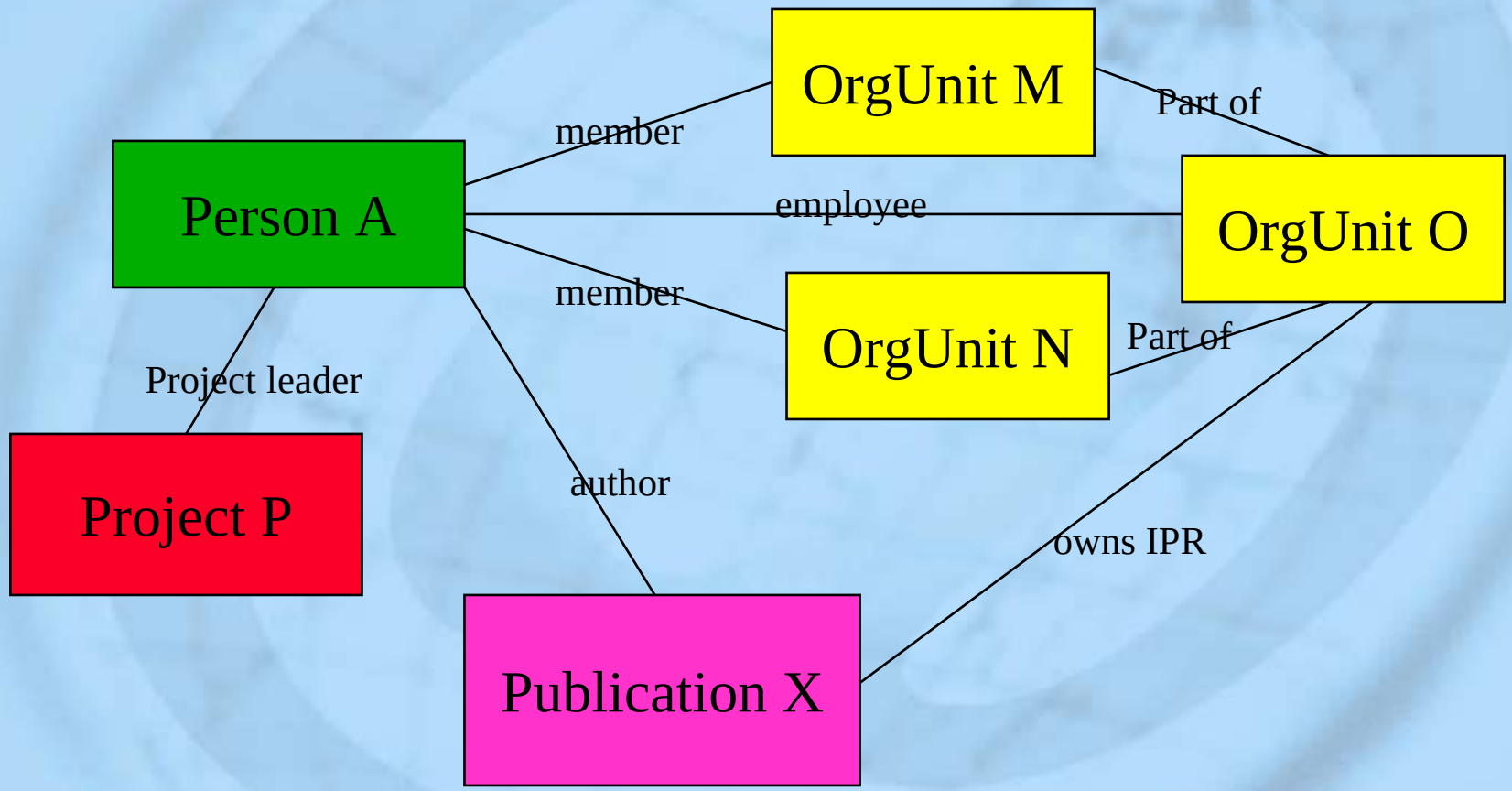
Concepts:
(1) temporally-bound role linking relations
(2) >1 linking relation : Result_Publication and other entities
(3) PERSON role may be author, co-author, editor, reviewer….
(4) ORGUNIT role may be publisher, IPR or copyright owner..
(5) PROJECT role may be the source of the idea

RESULT_PUBLICATION



Can Express:
Person A (DT1 - DT2) (is author of) Publication X
Orgunit O (DT1 - DT2) (is owner of IPR in) Publication X
Person A (DT1 - DT2) (is employee of) Orgunit O
Person A (DT1 - DT2) (is project leader of) Project P
Person A (DT1-DT2) (is member of) Orgunit M
Person A (DT1-DT2) (is member of) Orgunit N
Orgunit M (DT1-DT2) (is part of) Orgunit O
Orgunit N (DT1-DT2) (is part of) Orgunit O

Result_Publication Instance Diagram



Agenda

- Introduction – speaker
- Requirements
- Institutional Repositories
- CRIS Purpose & Stakeholders
- CERIF-CRIS at the centre of the Organisation
- e-Infrastructure
- Synthesis
- Role of euroCRIS

Requirement

- Researcher
  - should provide a view of everything of interest to the researcher in a structured manner which appears logical to the researcher in order to optimize the productive time of the researcher.
- Organisation
  - should provide the information required for decision-making to the benefit of the organisation.
- World-at-large
  - Selected views of the systems described above for researchers or organisations may be made available as information to others for purposes such as publicity, education (of scholars and of the general public) or offerings for technology transfer and commercialisation.

CERIF Characteristics

- extensible while preserving backward continuity to allow guaranteed interoperation between CERIF-CRIS
  - by adding new base entities and then link entities to integrate with the structure.
- link to any other system
  - using the link entities.
- normalized to avoid replication of data and to improve performance.
  - and consequent update integrity problems

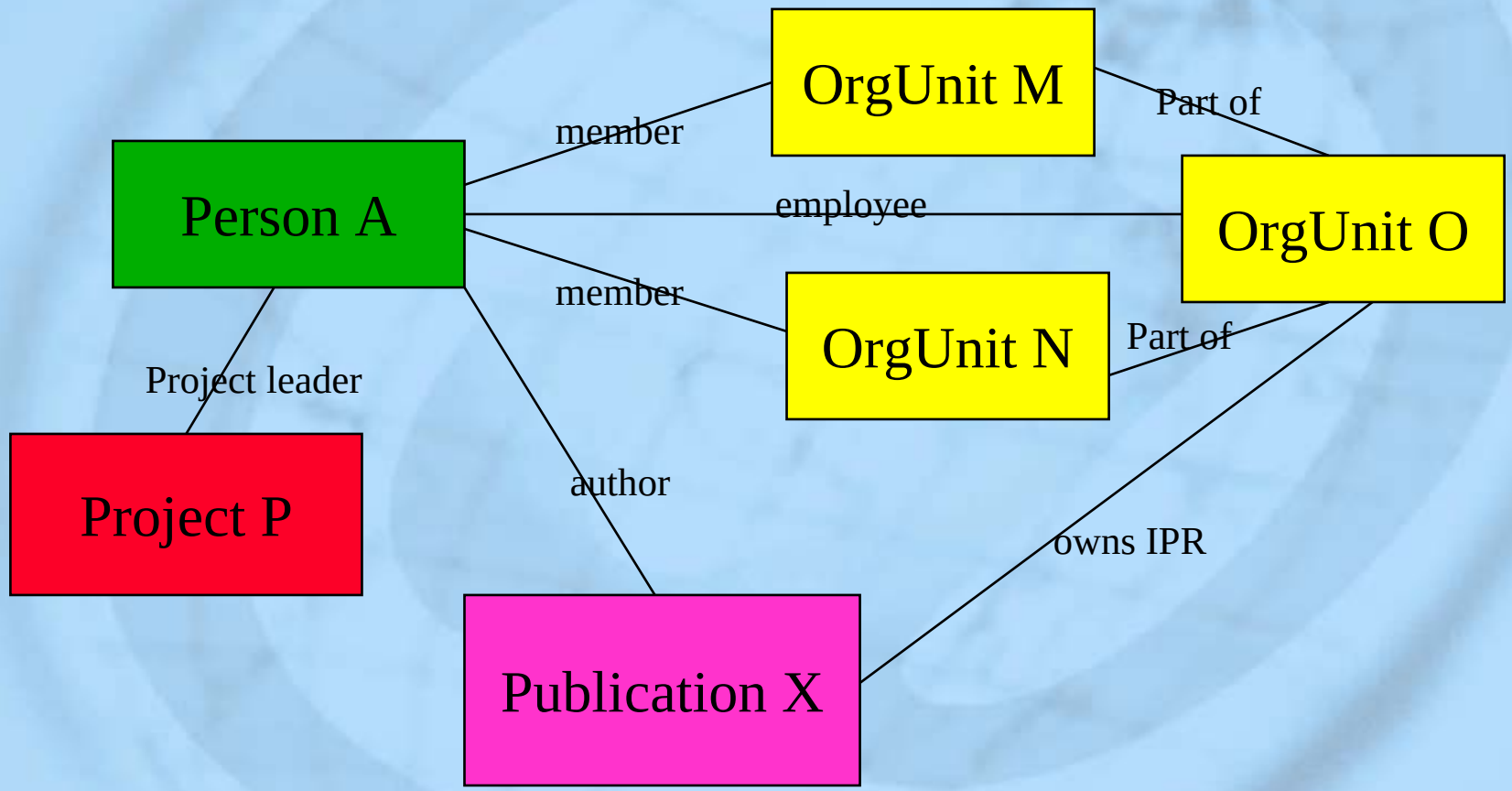
CERIF Characteristics

- implementable using any technology from hypermedia to information retrieval (semi-structured) and on to knowledge-based systems.
- follows formally first order logic
  - and so is available for deduction and induction leading to greater potential utilization of the data
  - Is scalable because machine-understandable as well as machine-readable.
- includes lookup tables (used also as classification tables)
  - improved data integrity by validation at input/update time
  - permits intelligent user interfaces to utilise the information to provide user assistance.

CERIF The Key

- The key to the design is the separation of base entities from link entities.
- The base entities, once populated, are rarely amended but may be appended with new information.
- The link entities are where the main update activity takes place since they record new relationships between records in the base entities.
- These new relationships may be input or they may be generated by deduction or induction.

Result_Publication Instance Diagram



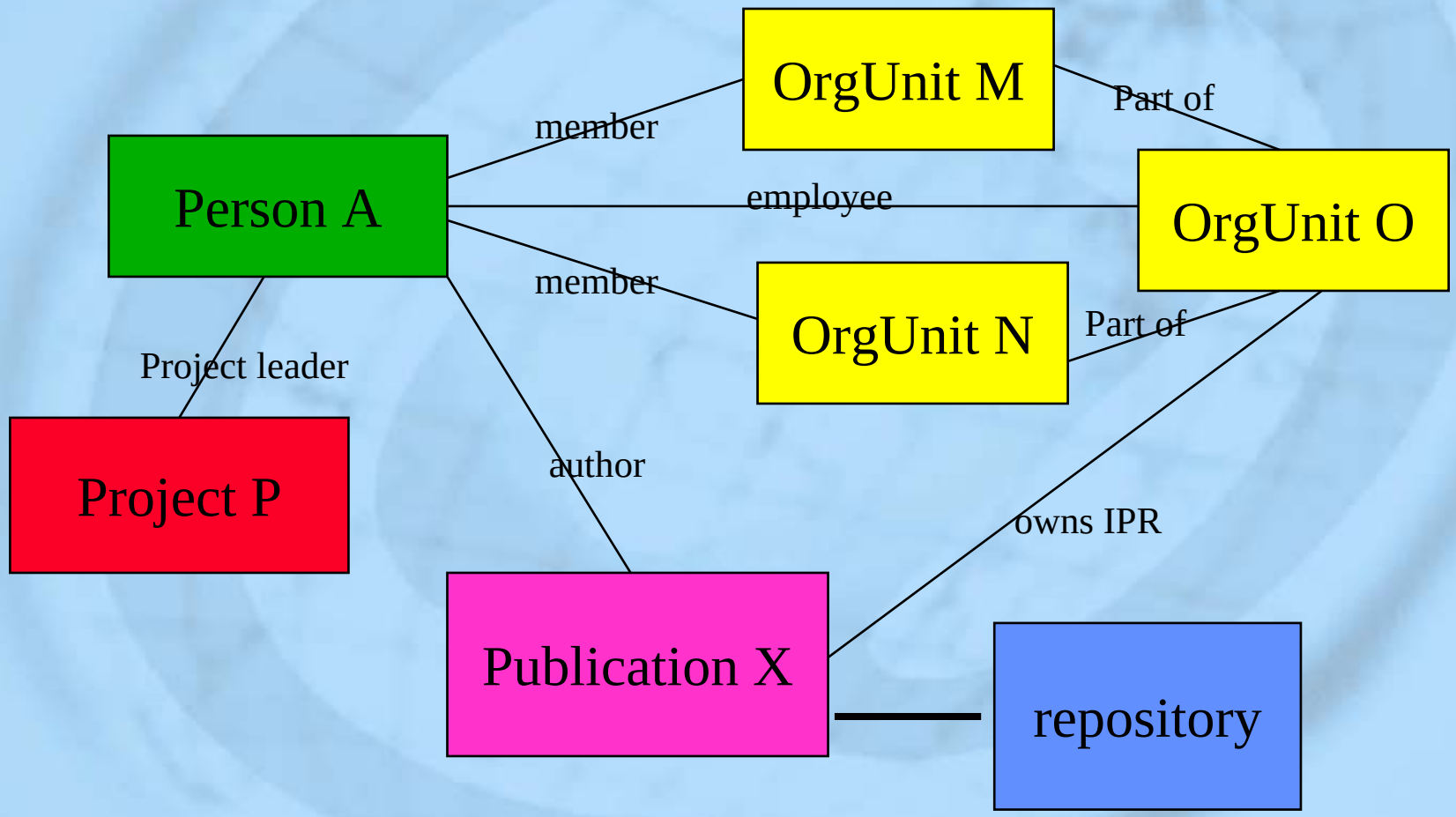
Linkages From CERIF
Staying with this example:

- CERIF does not only provide strong, role-typed, timestamped within-links
- But also provides the facility for strong, role-typed, timestamped outward-links

Linkages From CERIF Staying with this example:

- publication X full-text (or multimedia) is not stored within the CERIF data model but in an institutional repository or publisher's online database. CERIF provides the direct linkage to the full text.

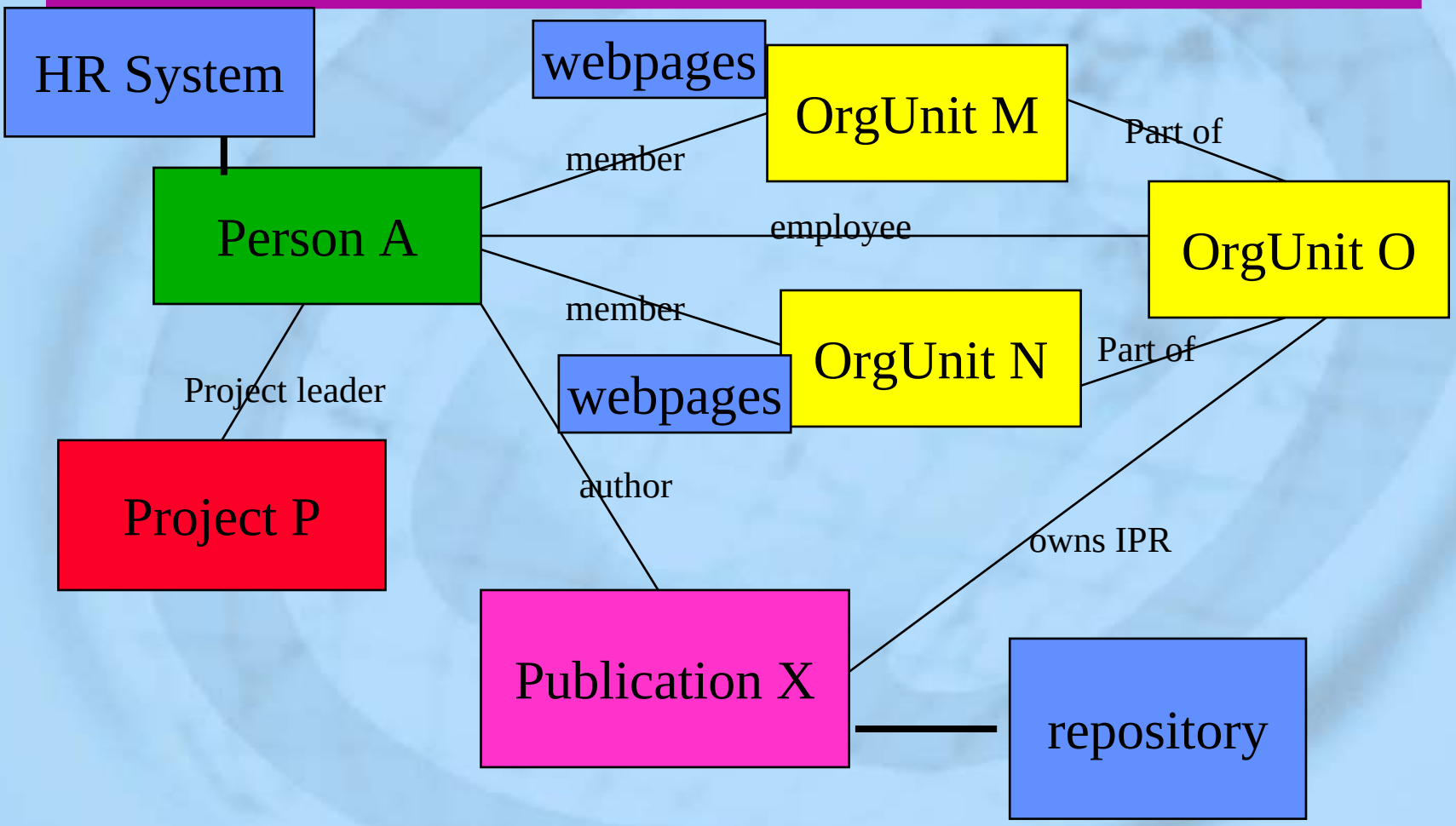
Result_Publication Instance Diagram



Linkages From CERIF Staying with this example:

- more information about Person A may be found in the HR (human resources) system of OrgUnit O, or on web-pages associated with either OrgUnit M or N.

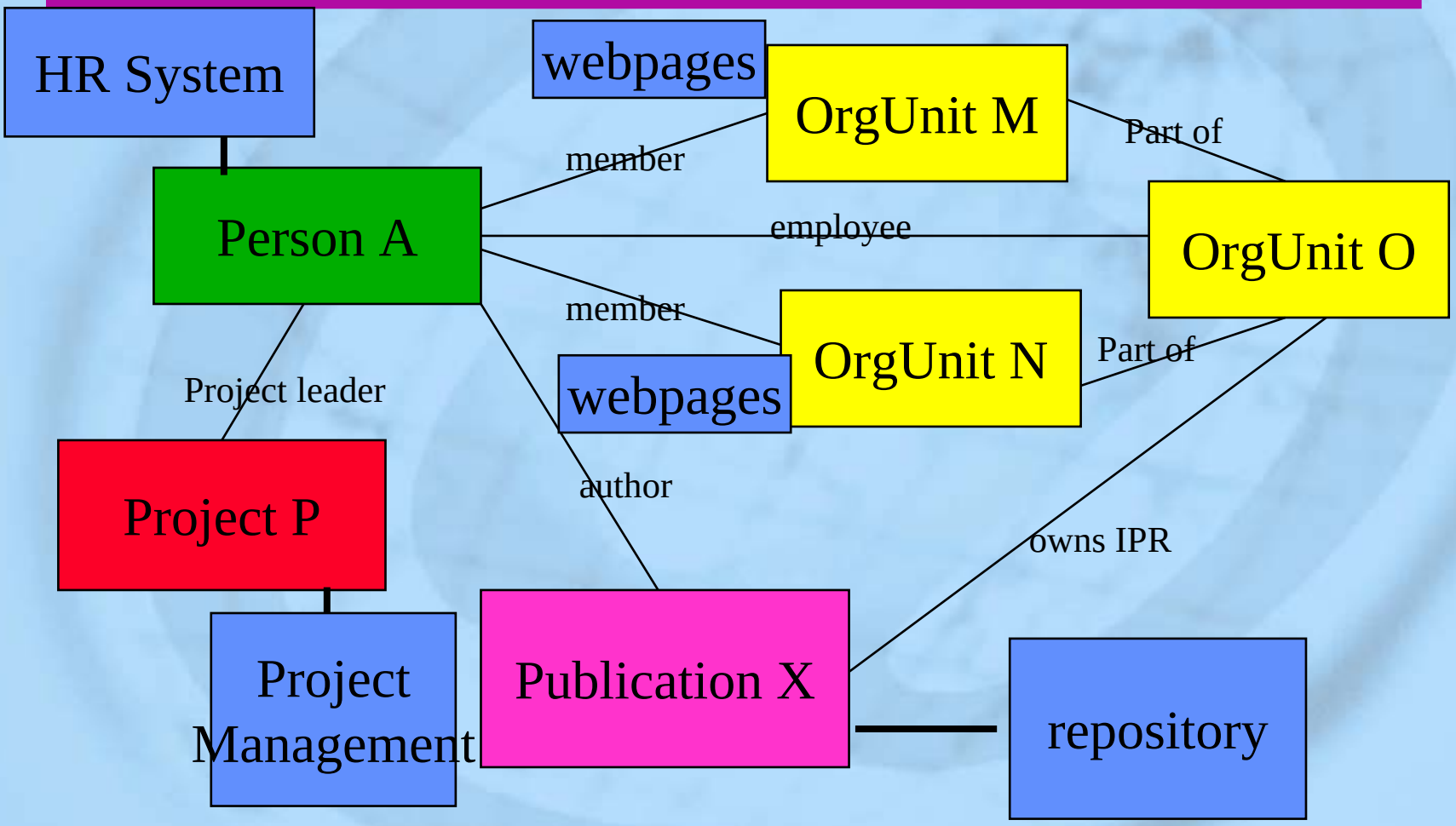
Result_Publication Instance Diagram



Linkages From CERIF
Staying with this example:

- the full project management information associated with Project P may be accessed in the project management system of Organisation O,

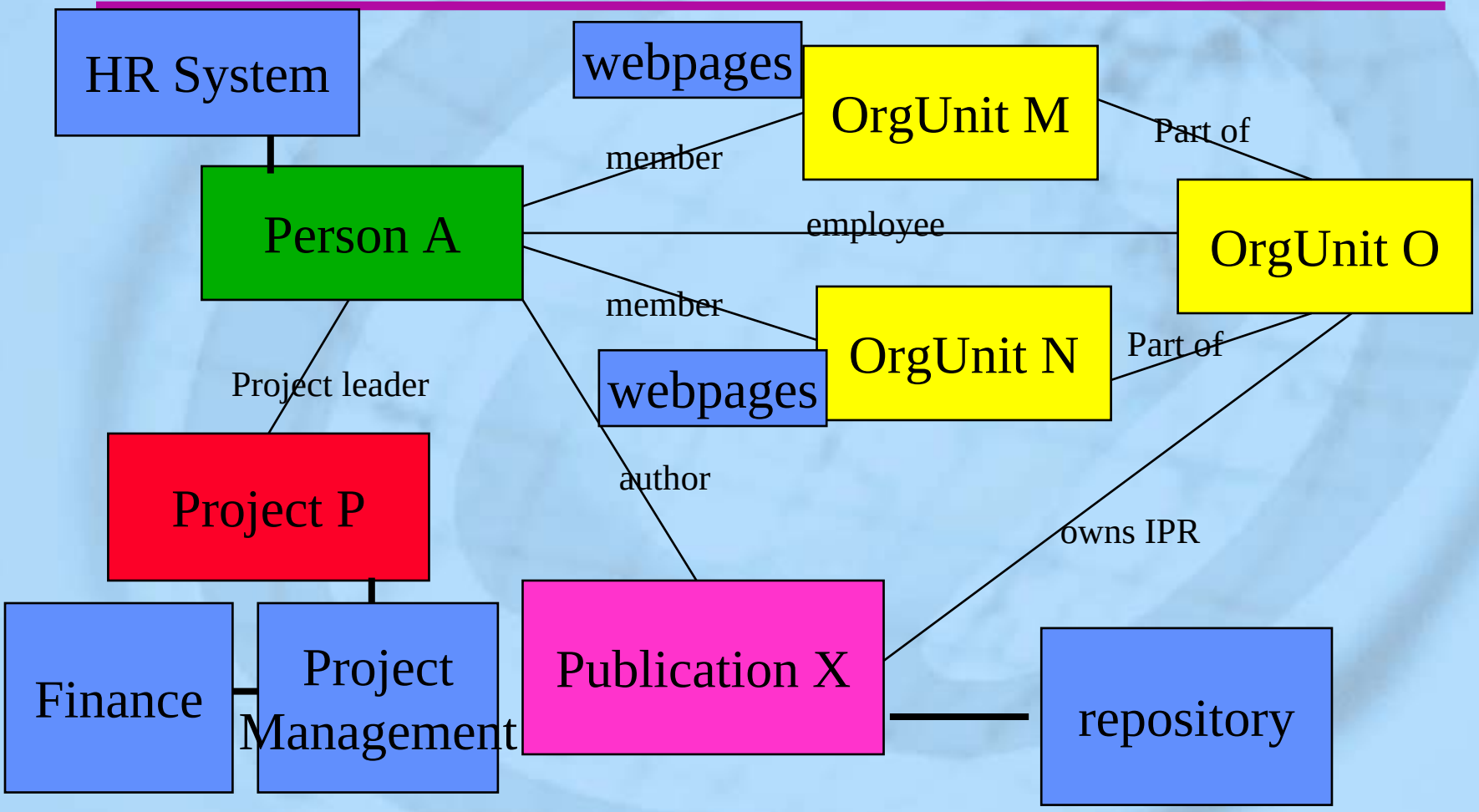
Result_Publication Instance Diagram



Linkages From CERIF Staying with this example:

- and from thence financial information may be found in the financial systems of Organisation O.

Result_Publication Instance Diagram



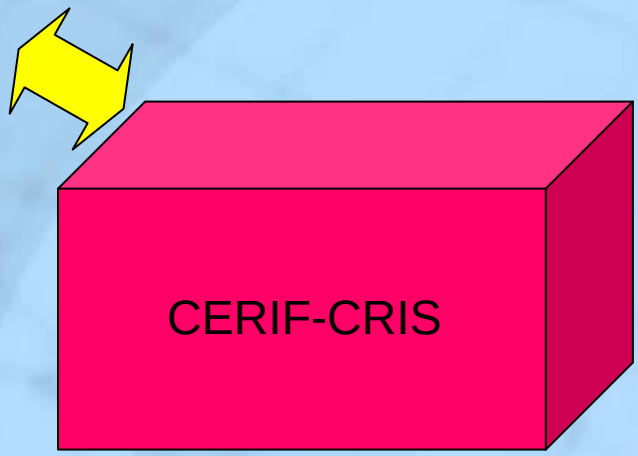
The Problem

- the traditional divide between
  - the individual researcher or research group view of the world
    - peer recognition
  - the organisation management view of the world
    - governance and value for money
- the traditional fierce independence of researchers and unwillingness to provide information on their activity
  - a quest for curiosity-led academic research freedom
  - despite possible advantages in cooperating with the management of an organisation
  - the view that the IT system provided is inadequate and they could have designed it better!

The Solution CERIF-CRIS plus Links

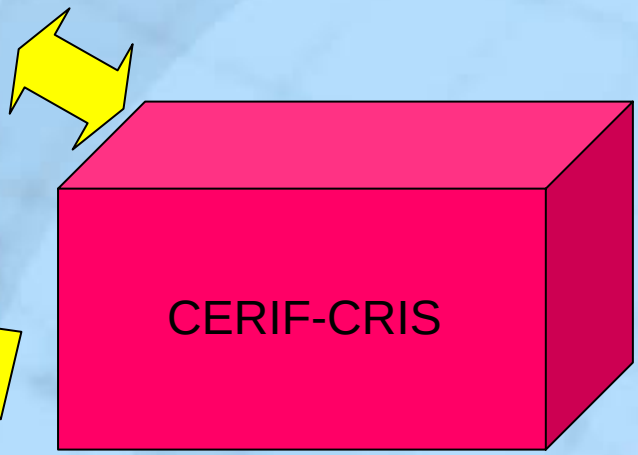
- CERIF: Person
  - Link to organisation HR system
- CERIF: OrgUnit
  - Link to organisational webpages
  - Link to catalogue of organisations (eg D&B)
- CERIF: Project
  - Link to organisational project management system
  - Link to funding organisation(s) records on the project
- CERIF: Funding
  - Link to funding organisation programme
- CERIF: Event
  - Link to e.g. conference webpage
- CERIF: Contact
  - Link to customer relationship management system
- CERIF: Result_Publication
  - Link to repository or publisher online database
- CERIF: Result_Patent
  - Link to patent database(s)
- CERIF: Result_Product
  - Link to e-research portal to datasets, software
- CERIF: Facility
  - Link to webpages of facility
- CERIF: Equipment
  - Link to webpages of equipment
- etc

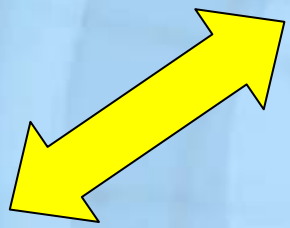
CERIF-CRIS at One Organisation

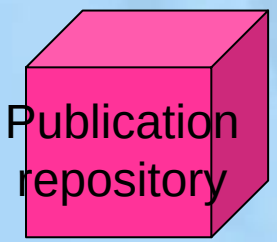


Managing Research Information at a researching or research funding organisation: decision support

CERIF-CRIS at One Organisation

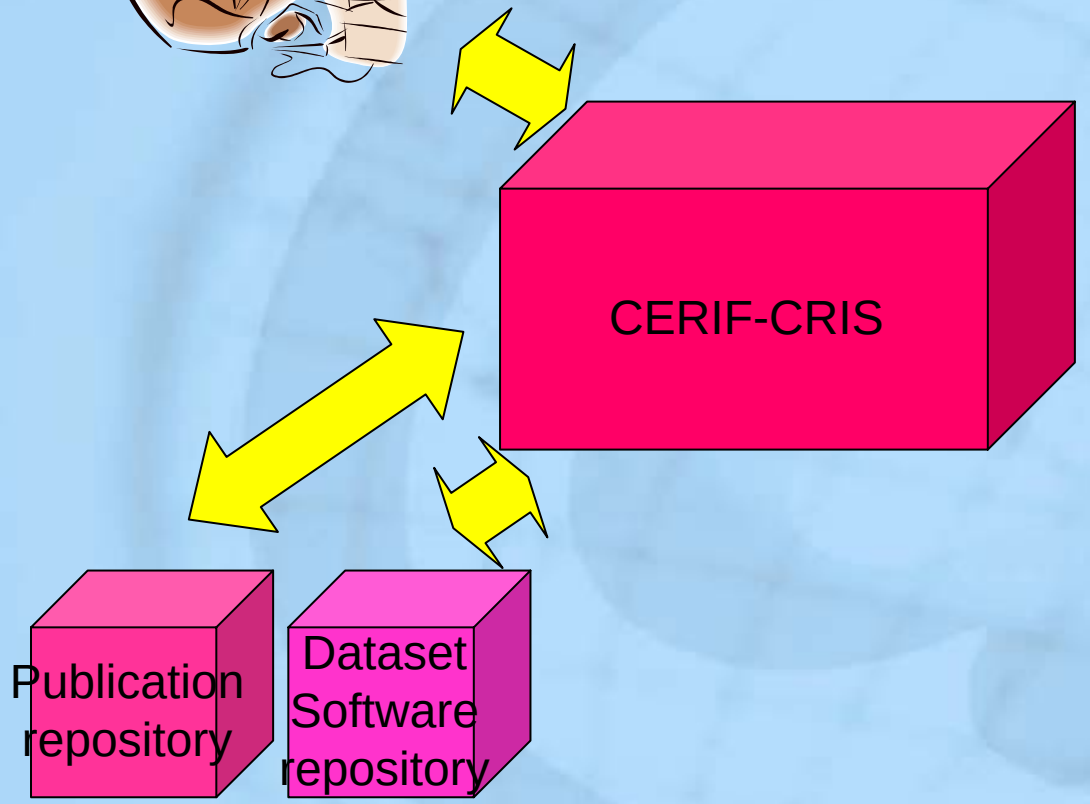






With associated scholarly publications providing deeper information on the research; metadata in the CERIF-CRIS

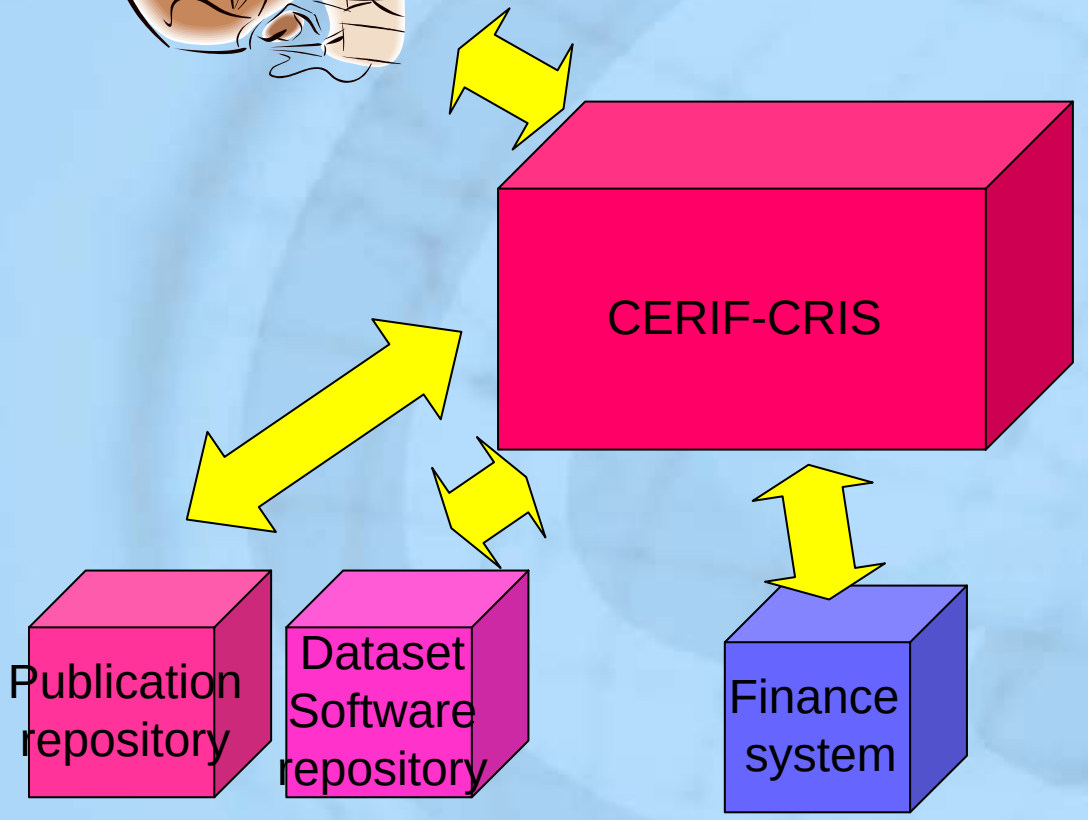
CERIF-CRIS at One Organisation



And research datasets and software to allow detailed examination of the research method; metadata in the CERIF-CRIS

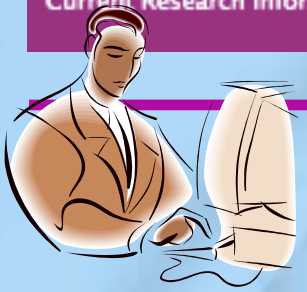
Note: metadata for products and patents stored in CERIF-CRIS; detail elsewhere (e.g. national or international system)

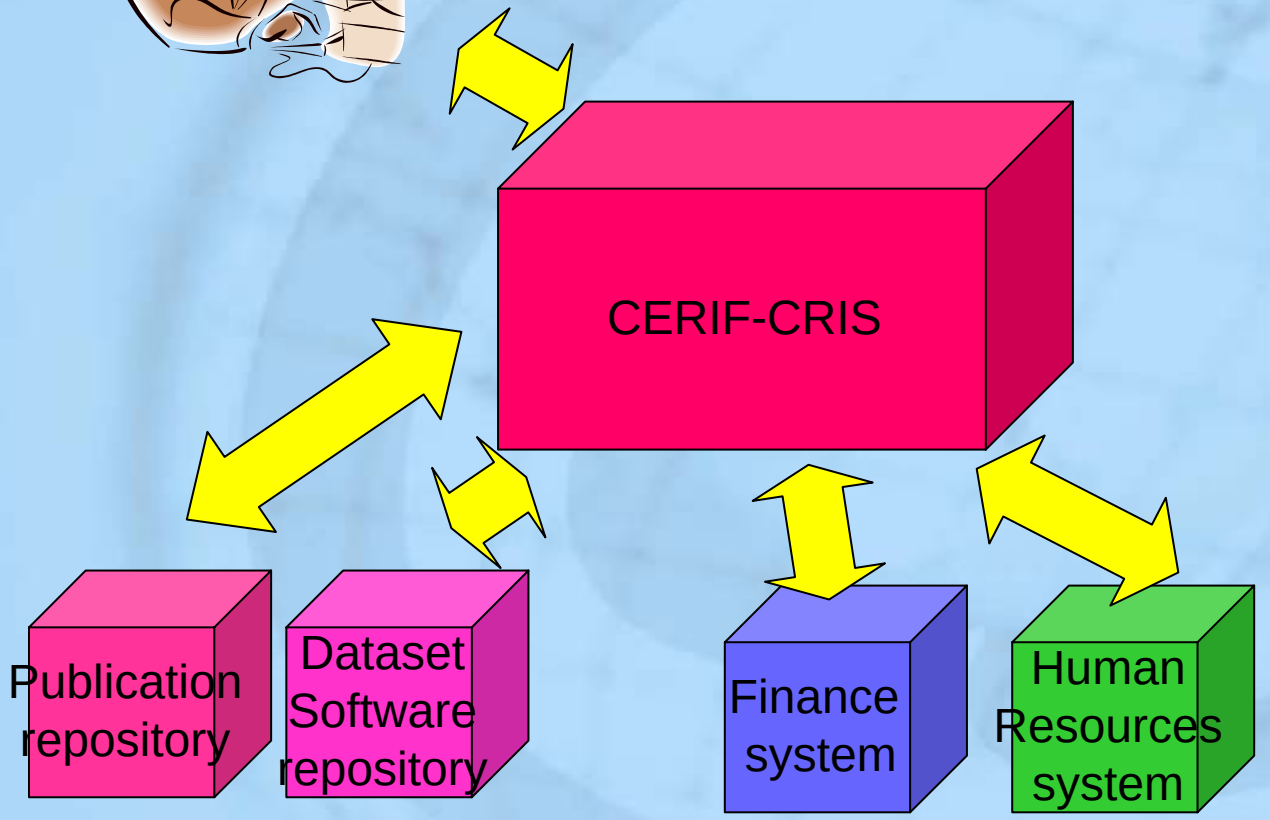
CERIF-CRIS at One Organisation



With financial information related to research activity to assess value for money

CERIF-CRIS at One Organisation

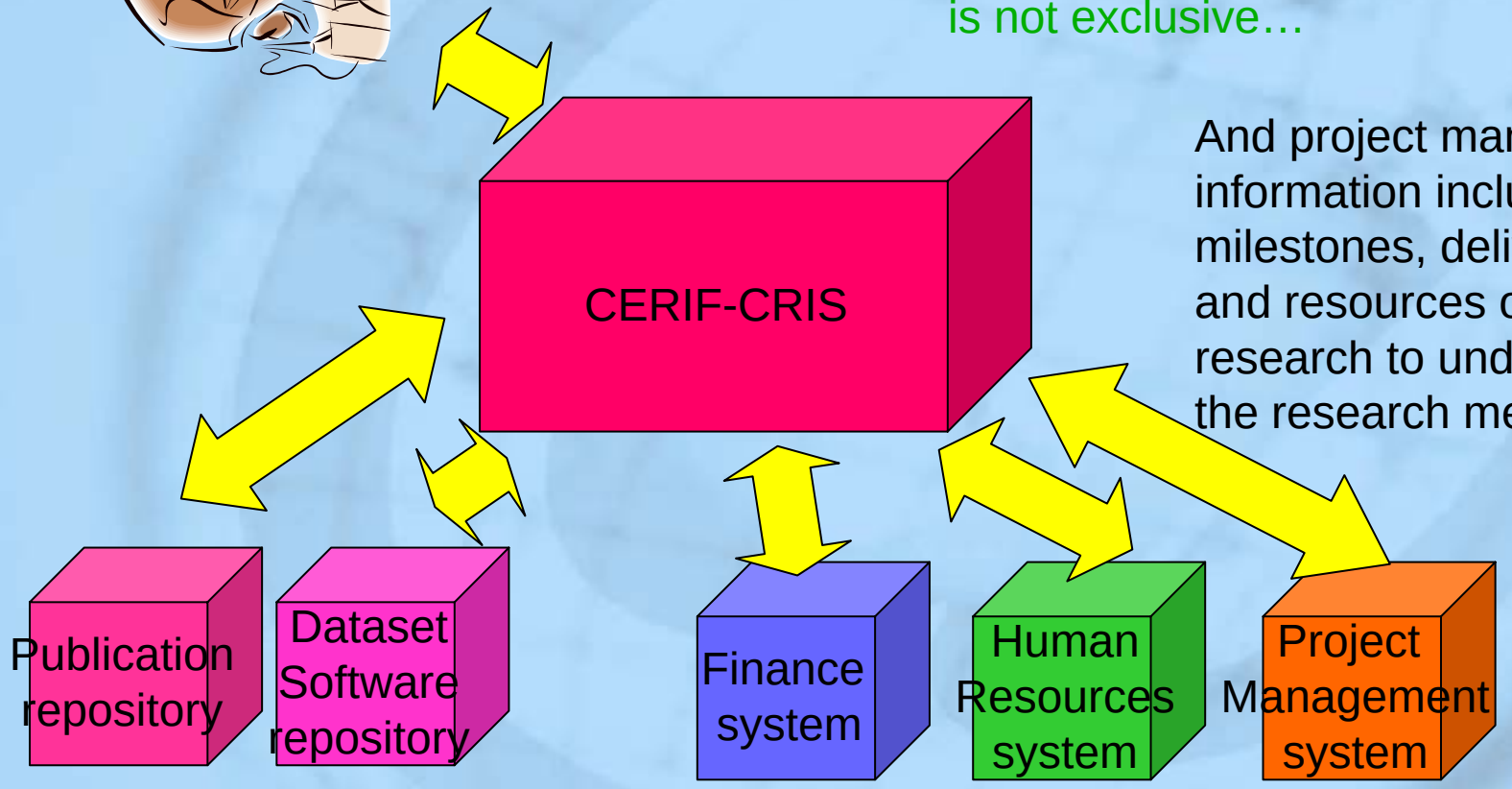




And human resource information related to the research activity to ensure appropriate skills and resource availability

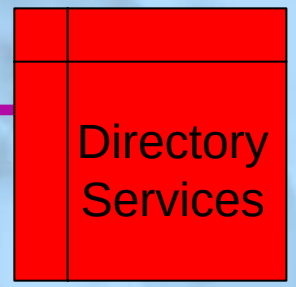
CERIF-CRIS at One Organisation

This list of organisational ICT systems is not exclusive…

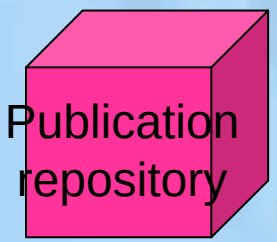


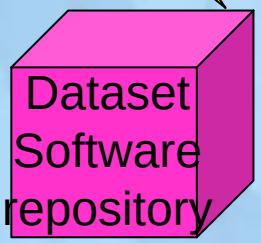
And project management information including milestones, deliverables and resources of the research to understand the research method

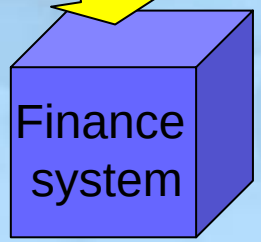
CERIF-CRIS at One Organisation

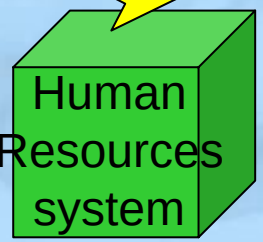


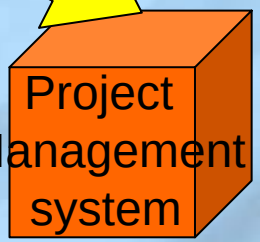
And directory services to control research workflow, messaging, authentication, authorisation, access











CERIF-CRIS at One Organisation

Web pages

Directory Services

CERIF-CRIS

And generation of intranet (organisation), DMZ (trusted business partners) and extranet (public) web-pages

Publication repository

Dataset Software repository

Finance system

Human Resources system

Project Management system

CERIF-CRIS at One Organisation

Web pages

Directory Services

CERIF-CRIS

This is fine for one organisation but research is international, so…

Publication repository

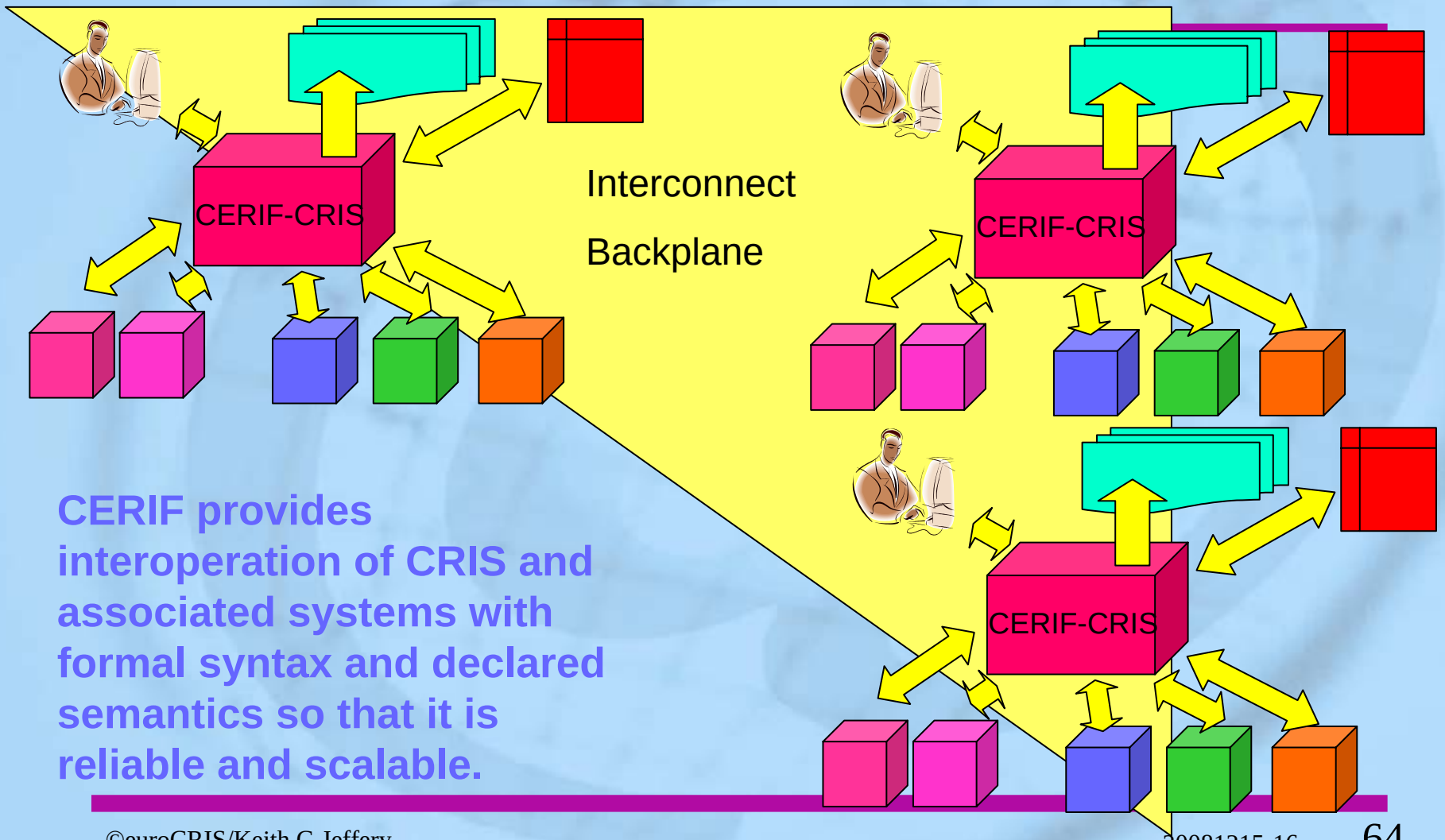
Dataset Software repository

Finance system

Human Resources system

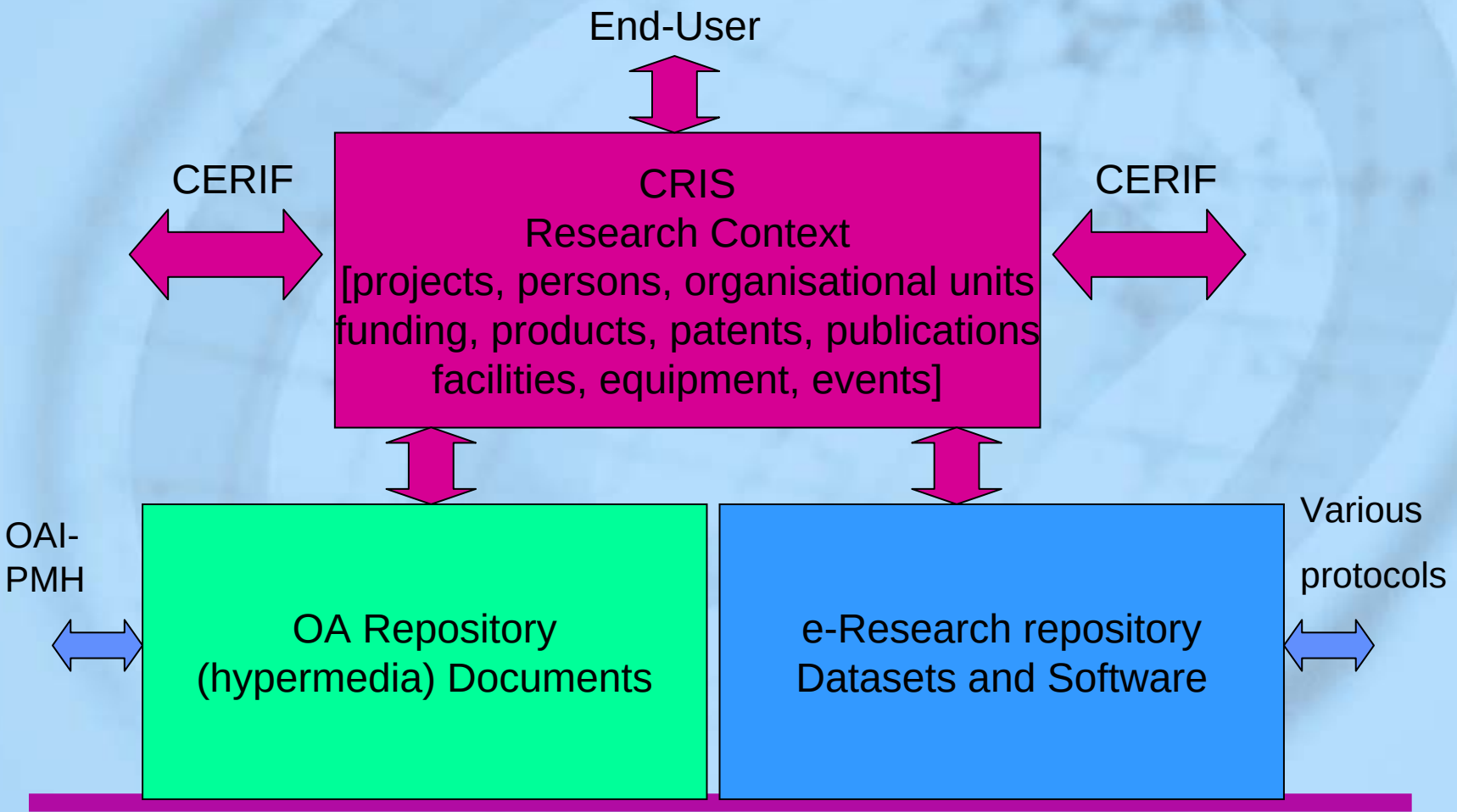
Project Management system

CERIF Interoperation

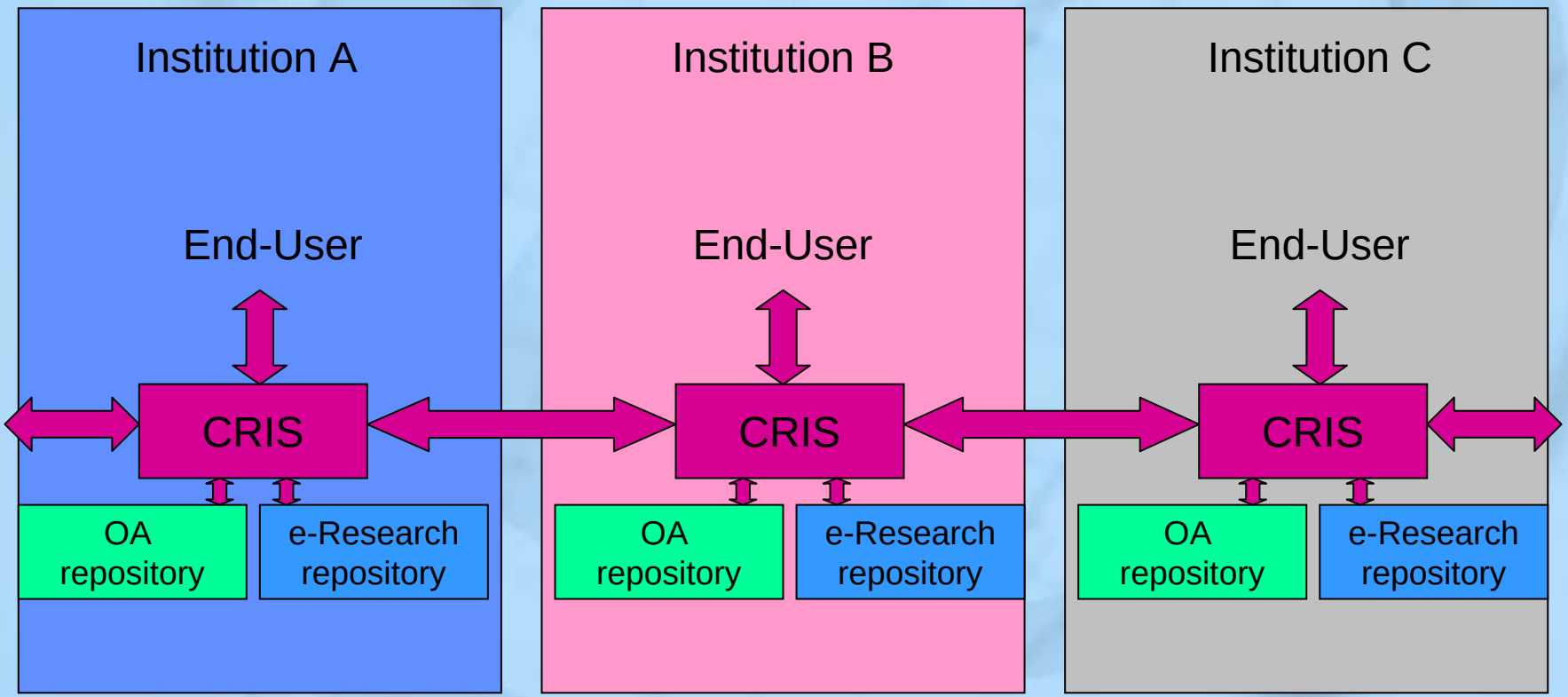


CERIF provides interoperation of CRIS and associated systems with formal syntax and declared semantics so that it is reliable and scalable.

CRIS + Repositories at 1 institution



….and multiple institutions



Roles of CERIF-CRIS: Re-iteration

- Research information system for decision-support

Roles of CERIF-CRIS: Re-iteration

- Research information system for decision-support
- Metadata (index) to scholarly publications (white and grey) in a repository

Roles of CERIF-CRIS: Re-iteration

- Research information system for decision-support
- Metadata (index) to scholarly publications (white and grey) in a repository
- Metadata (index) to research datasets and software in a repository

Roles of CERIF-CRIS: Re-iteration

- Research information system for decision-support
- Metadata (index) to scholarly publications (white and grey) in a repository
- Metadata (index) to research datasets and software in a repository
- Access view to financial information of an organisation

Roles of CERIF-CRIS: Re-iteration

- Research information system for decision-support
- Metadata (index) to scholarly publications (white and grey) in a repository
- Metadata (index) to research datasets and software in a repository
- Access view to financial information of an organisation
- Access view to human resource information of an organisation

Roles of CERIF-CRIS: Re-iteration

- Research information system for decision-support
- Metadata (index) to scholarly publications (white and grey) in a repository
- Metadata (index) to research datasets and software in a repository
- Access view to financial information of an organisation
- Access view to human resource information of an organisation
- Access view to project management information of an organisation

Roles of CERIF-CRIS: Re-iteration

- Research information system for decision-support
- Metadata (index) to scholarly publications (white and grey) in a repository
- Metadata (index) to research datasets and software in a repository
- Access view to financial information of an organisation
- Access view to human resource information of an organisation
- Access view to project management information of an organisation
- (and to other relevant organisation systems)

Roles of CERIF-CRIS: Re-iteration

- Research information system for decision-support
- Metadata (index) to scholarly publications (white and grey) in a repository
- Metadata (index) to research datasets and software in a repository
- Access view to financial information of an organisation
- Access view to human resource information of an organisation
- Access view to project management information of an organisation
- (and to other relevant organisation systems)
- Provision of directory service information for authentication, authorisation, workflow, cooperative working…

Roles of CERIF-CRIS: Re-iteration

- Research information system for decision-support
- Metadata (index) to scholarly publications (white and grey) in a repository
- Metadata (index) to research datasets and software in a repository
- Access view to financial information of an organisation
- Access view to human resource information of an organisation
- Access view to project management information of an organisation
- (and to other relevant organisation systems)
- Provision of directory service information for authentication, authorisation, workflow, cooperative working…
- Generation of web pages presenting the organisation on intranet, DMZ and extranet directly or from other organisational systems through the CERIF-CRIS

Roles of CERIF-CRIS: Re-iteration

- Research information system for decision-support
- Metadata (index) to scholarly publications (white and grey) in a repository
- Metadata (index) to research datasets and software in a repository
- Access view to financial information of an organisation
- Access view to human resource information of an organisation
- Access view to project management information of an organisation
- (and to other relevant organisation systems)
- Provision of directory service information for authentication, authorisation, workflow, cooperative working…
- Generation of web pages presenting the organisation on intranet, DMZ and extranet directly or from other organisational systems through the CERIF-CRIS
- Interoperation with other CERIF-CRIS (and their associated systems) to give a global view of research information

Take-Home Message

Make the CERIF-CRIS the centre of the research organisation to

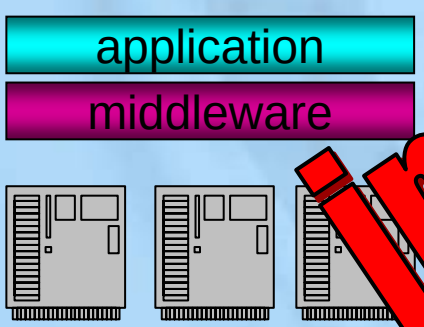
a) Integrate all other systems

b) Interoperate with external systems

Recall the requirement scenario

Not only work with the e-literature repository but also…..

- CRIS
- project, person, organisatio[illegible] output (products, paten[illegible] funding, facilities, equipment[illegible]
- e-Research [illegible]
- research d[illegible]
- e-R[illegible]
- [illegible]ents, take data, visualistaion, in-[illegible]o [illegible]riments (simulation)
- e-[illegible]ocess
- workflows, research applications, travel requests, claims



in one environment

Where we are

- The foregoing integrates
  - CRIS
  - Repositories of publications
  - Repositories of research datasets / software
  - Legacy systems (e.g. finance, HR…) of an organisation
- And allows organisation to organisation interoperation

- But what about the ‘researcher workbench’ for integrated scientific and administrative activities…

Agenda

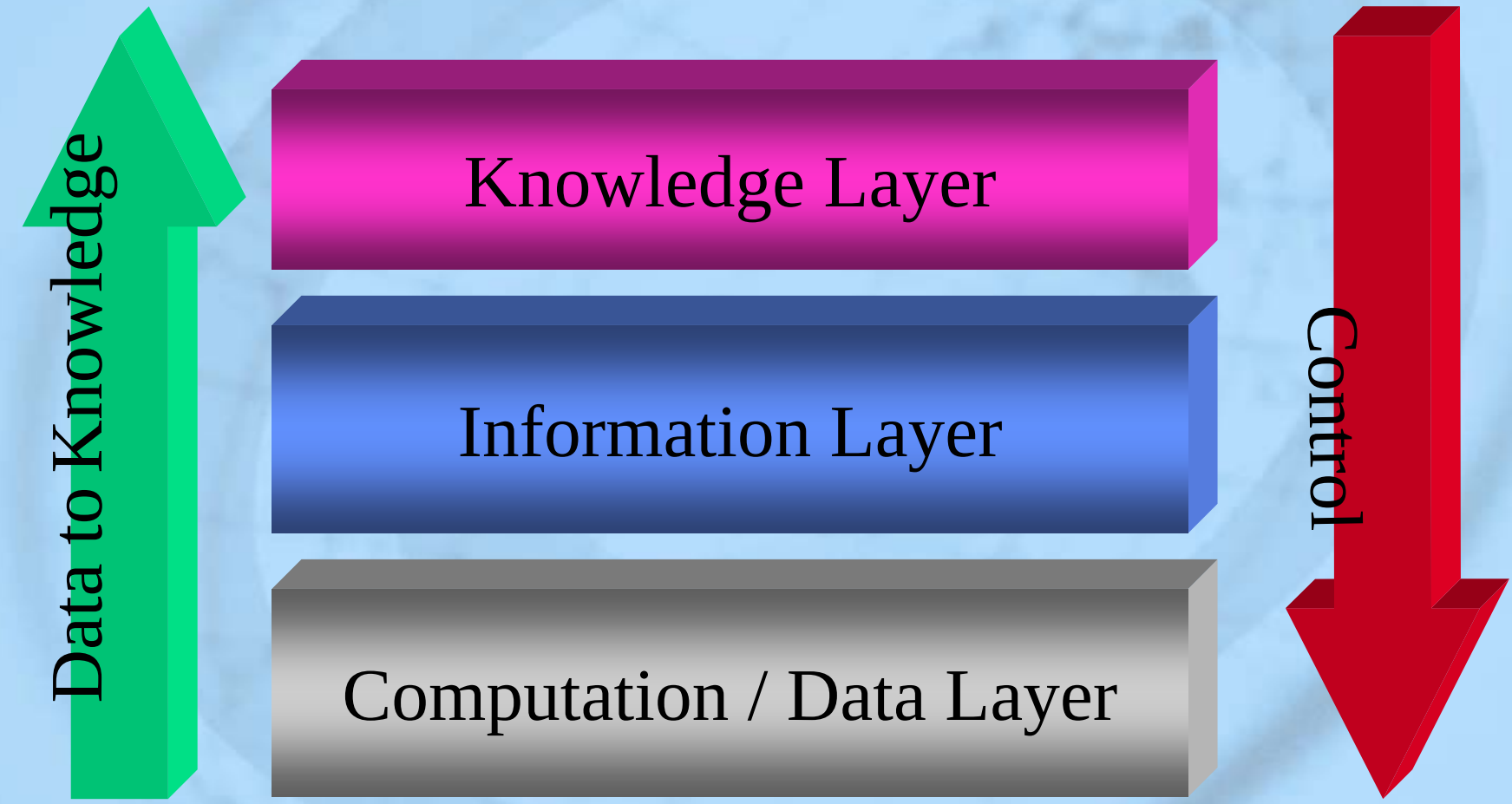
- Introduction – speaker
- Requirements
- Institutional Repositories
- CRIS Purpose & Stakeholders
- CERIF-CRIS at the centre of the Organisation
- e-Infrastructure
- Synthesis
- Role of euroCRIS

Workflow on the GRIDs surface

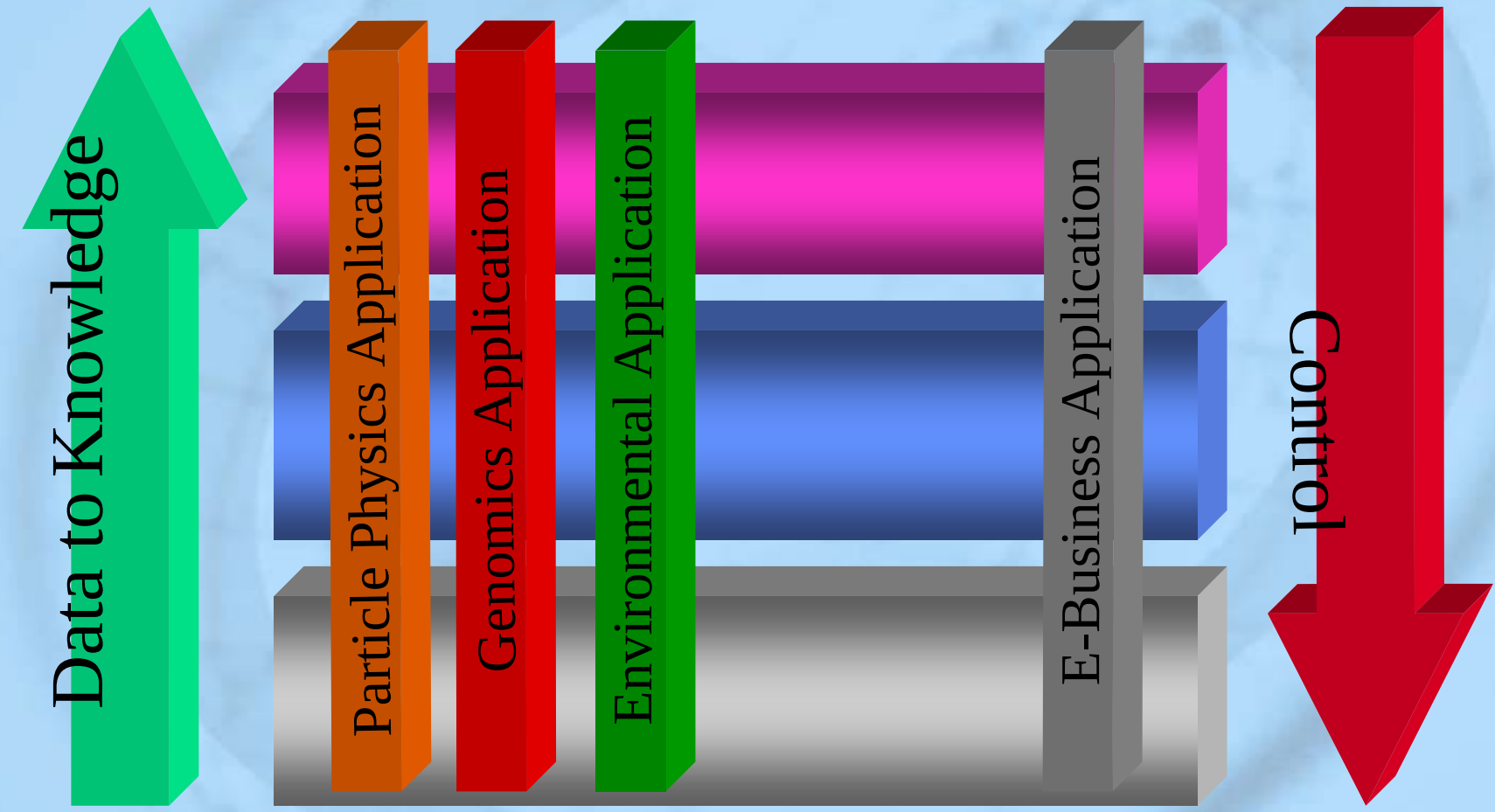
- GRIDs ‘surface’ provides
  - Computational capabilities of GRID
  - Information presentation capabilities of WWW
  - Information management capabilities
- and environment for workflow

- It can support the research process

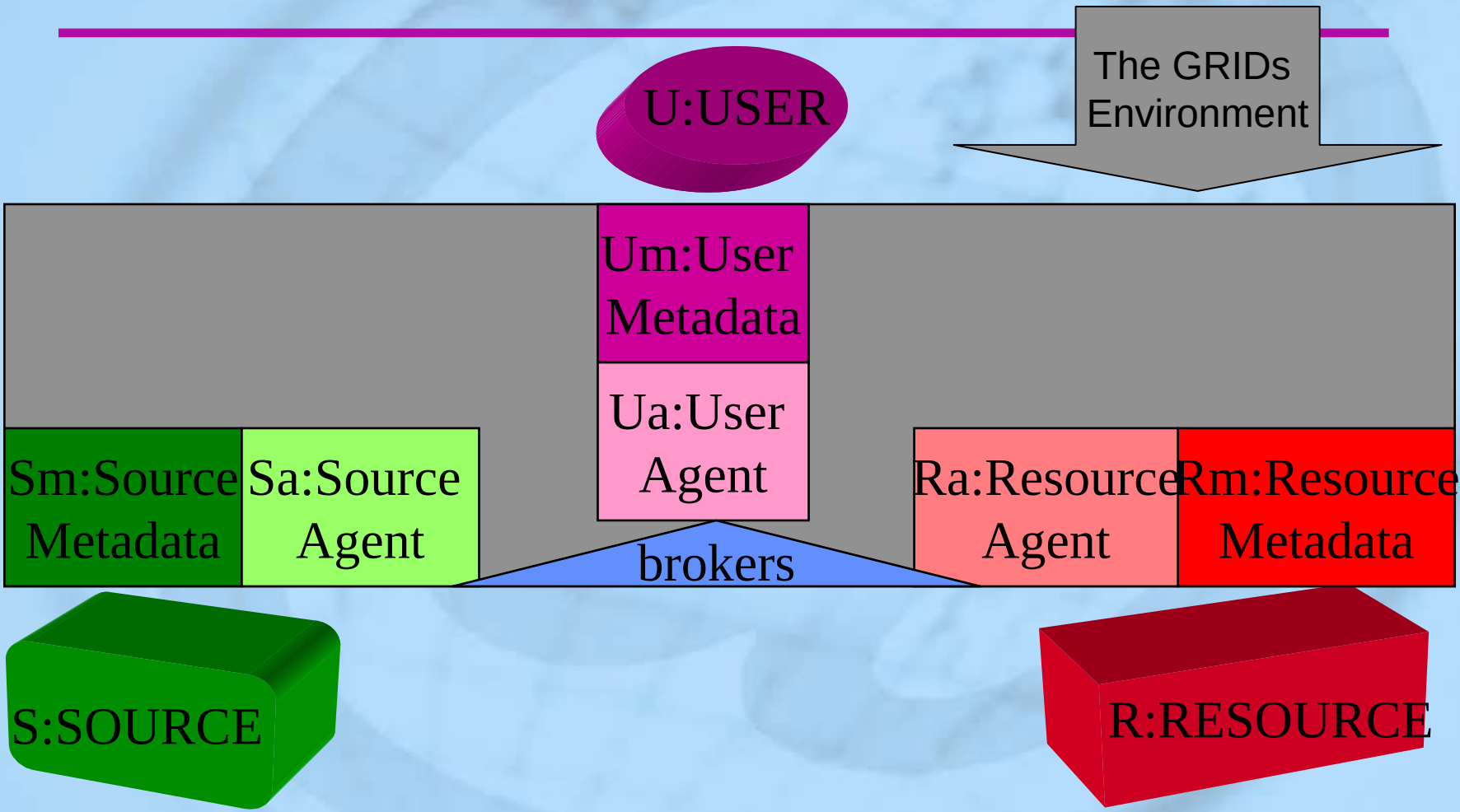
The GRIDs Architecture



The GRIDs Architecture



A POSSIBLE ARCHITECTURE



A Brief History of GRIDs

- 1G: custom-made architecture machines to user
  - Pioneering metacomputing
- 2G: proprietary standards and interfaces
  - I-WAY ➔GLOBUS, UNICORE, CONDOR, LEGION➔ AVA
- 2.5G: added in FTP, SRB, LDAP, AccessGRID
- 3G: adopted W3C concepts for open interfaces – OGSA / OGSI: note especially OGSA/DAI
  - But built on 2.G foundations

1999

e-Science Apps

2002

e-Science R&D

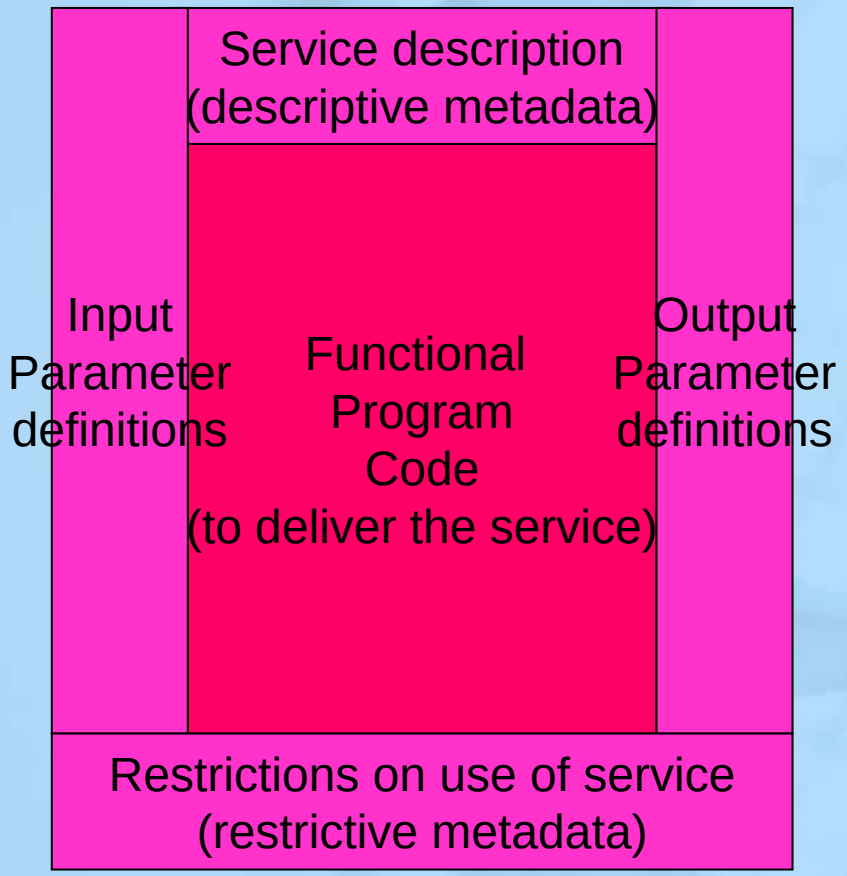
2007

G-Lite towards EGI

But…..

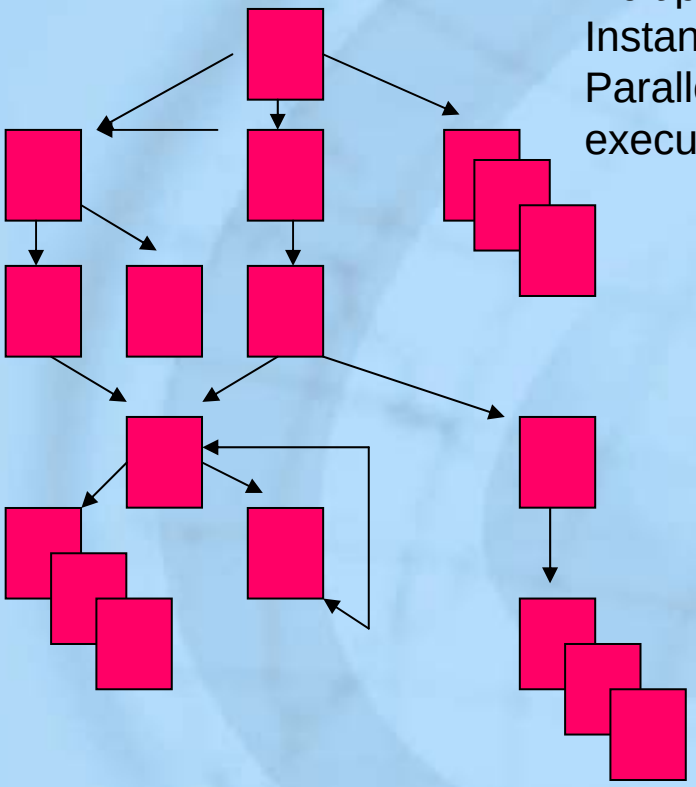
- This comes nowhere near the requirements as originally defined for GRIDs
- Too low-level (programmer not end-user level)
  - Insufficient representativity
  - Insufficient expressivity
  - Insufficient resilience
  - Insufficient dynamic flexibility

Services: Challenges 1



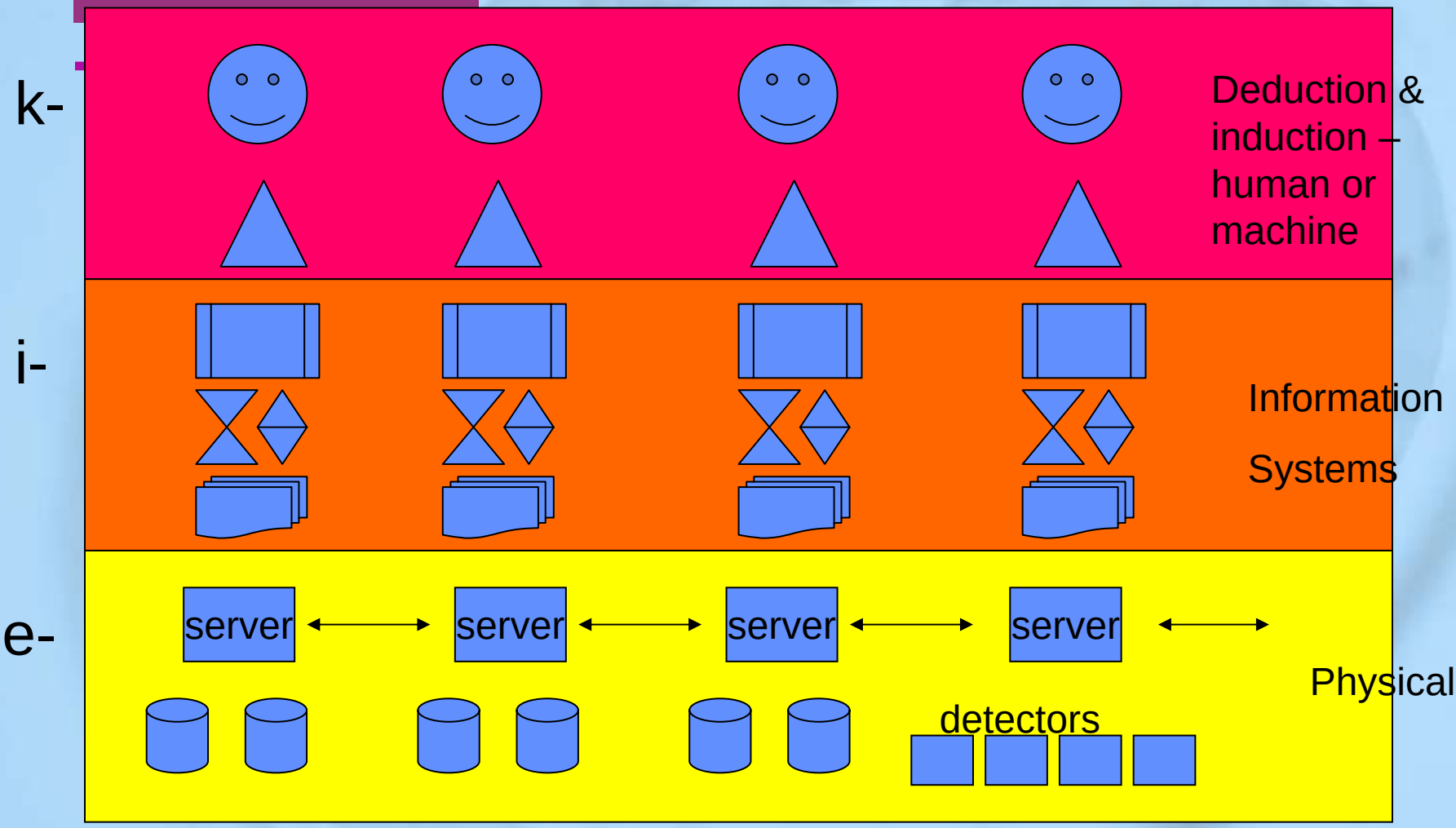
- Description
  - Location
  - Requirements matching
  - Composing
  - Utilising
- ➔ metadata

Services: Challenges 2

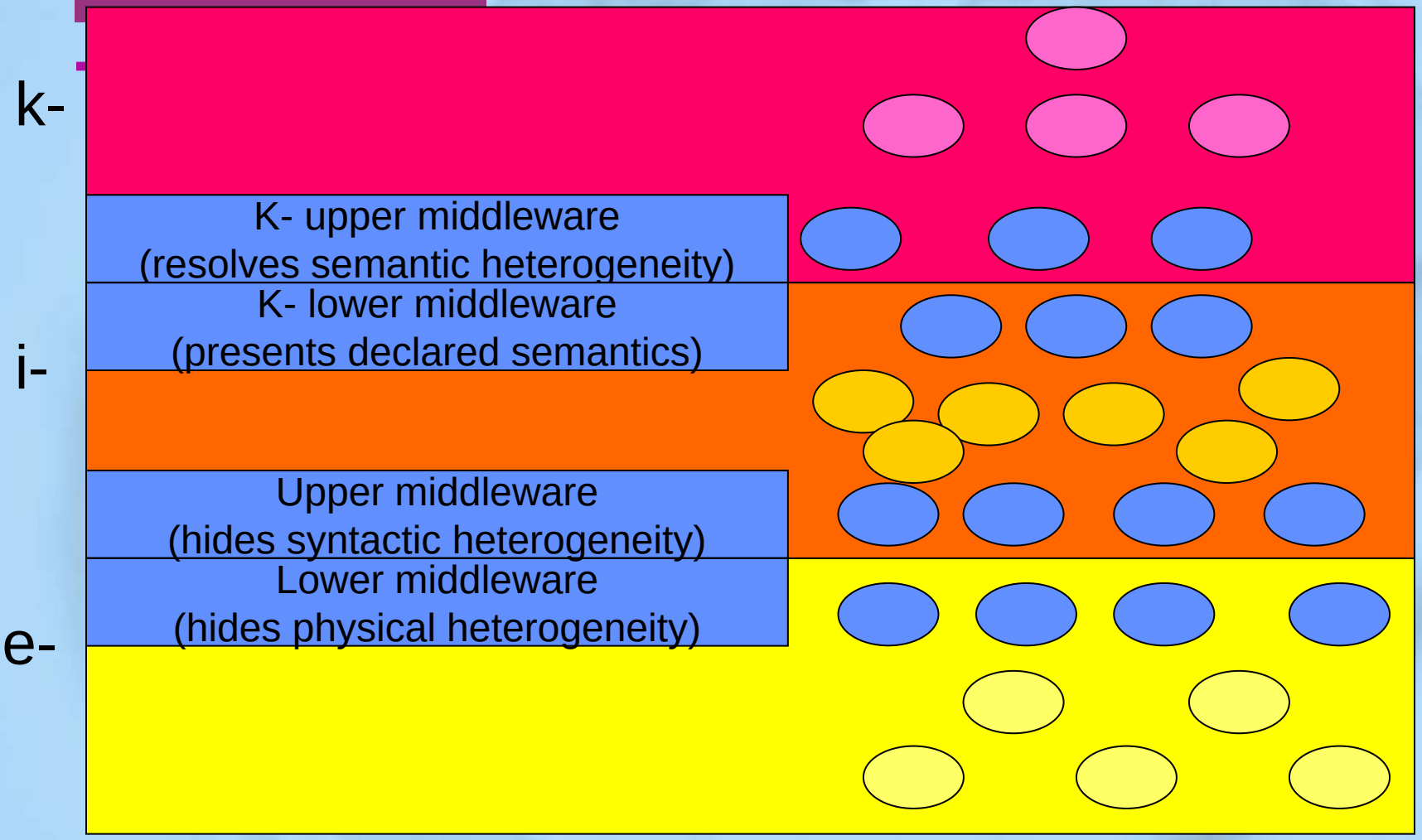


- Composition
  - End-to-end FR satisfaction
  - End-to-end NFRs satisfaction
  - Avoiding emergent properties
  - Conditions of use of services
    - Processes
    - wrapped with data
    - wrapped with processing, storage etc
    - wrapped with real estate
    - wrapped with staff

e-,i-,k-infrastructure



Middleware – and as SOKUs



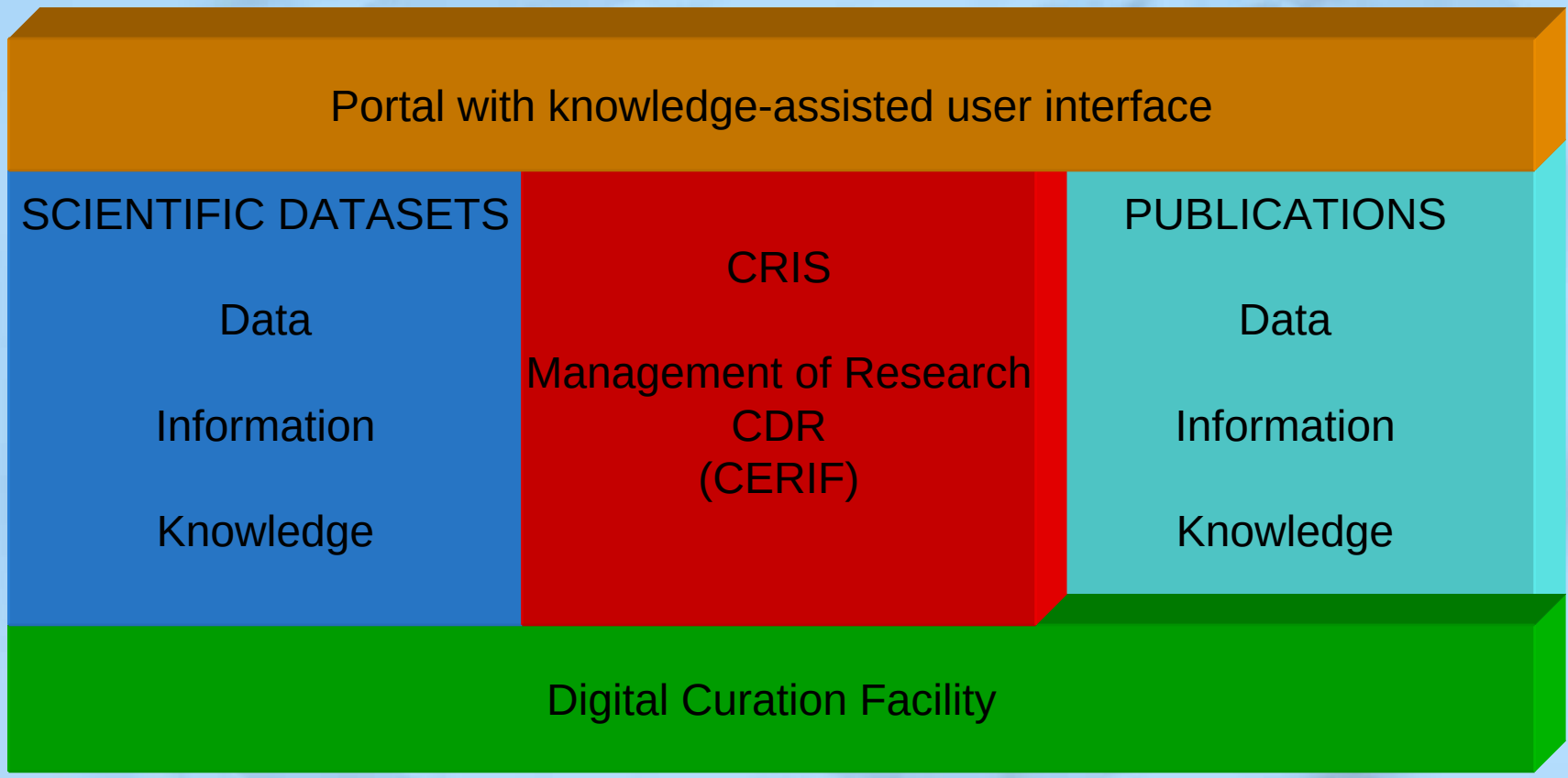
Agenda

- Introduction – speaker
- Requirements
- Institutional Repositories
- CRIS Purpose & Stakeholders
- CERIF-CRIS at the centre of the Organisation
- e-Infrastructure
- Synthesis
- Role of euroCRIS

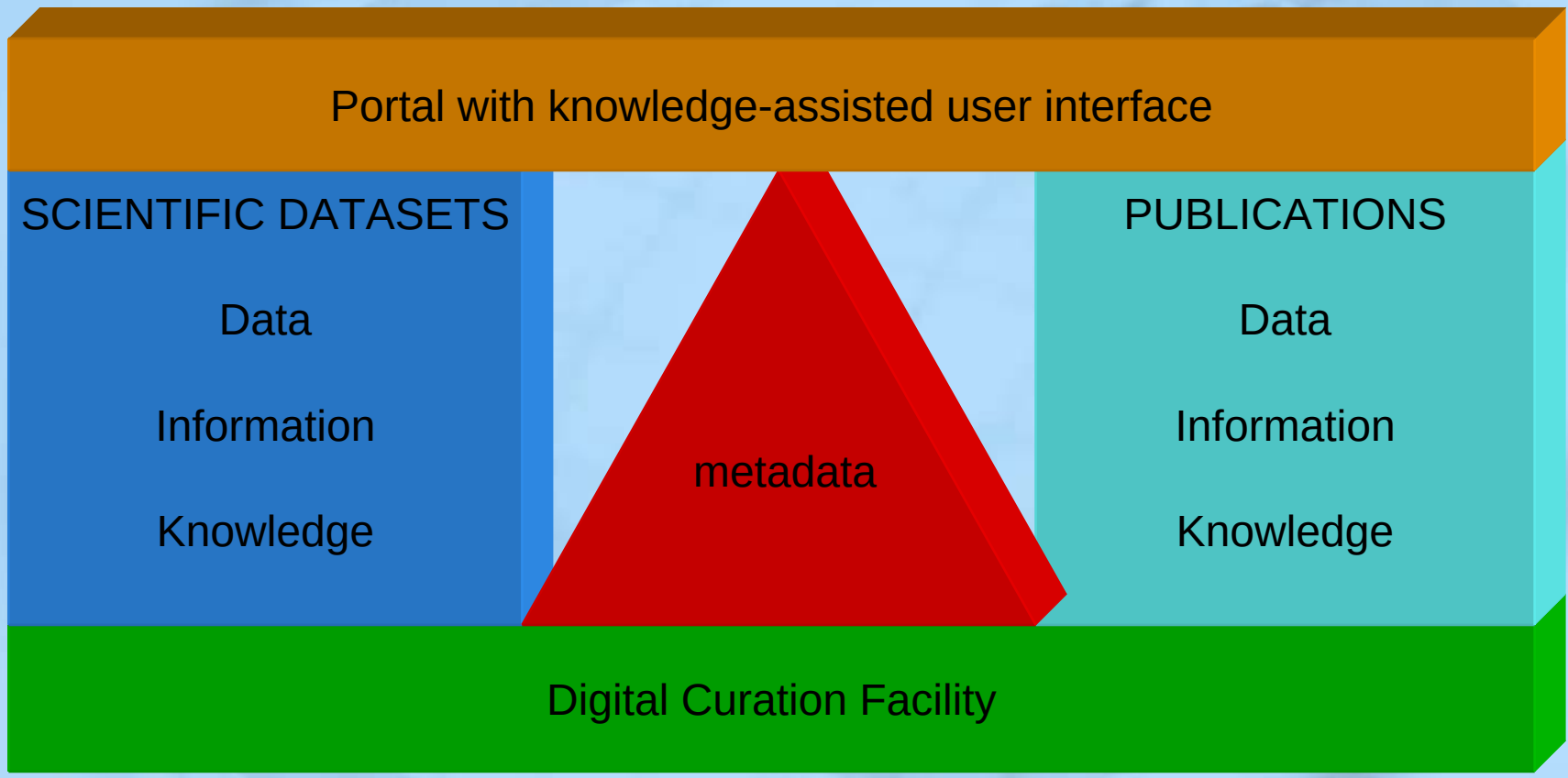
Overall : The Way Forward

SCIENTIFIC DATASETS	CRIS	PUBLICATIONS
Data	Management of Research	Data
Information		Information
Knowledge		Knowledge

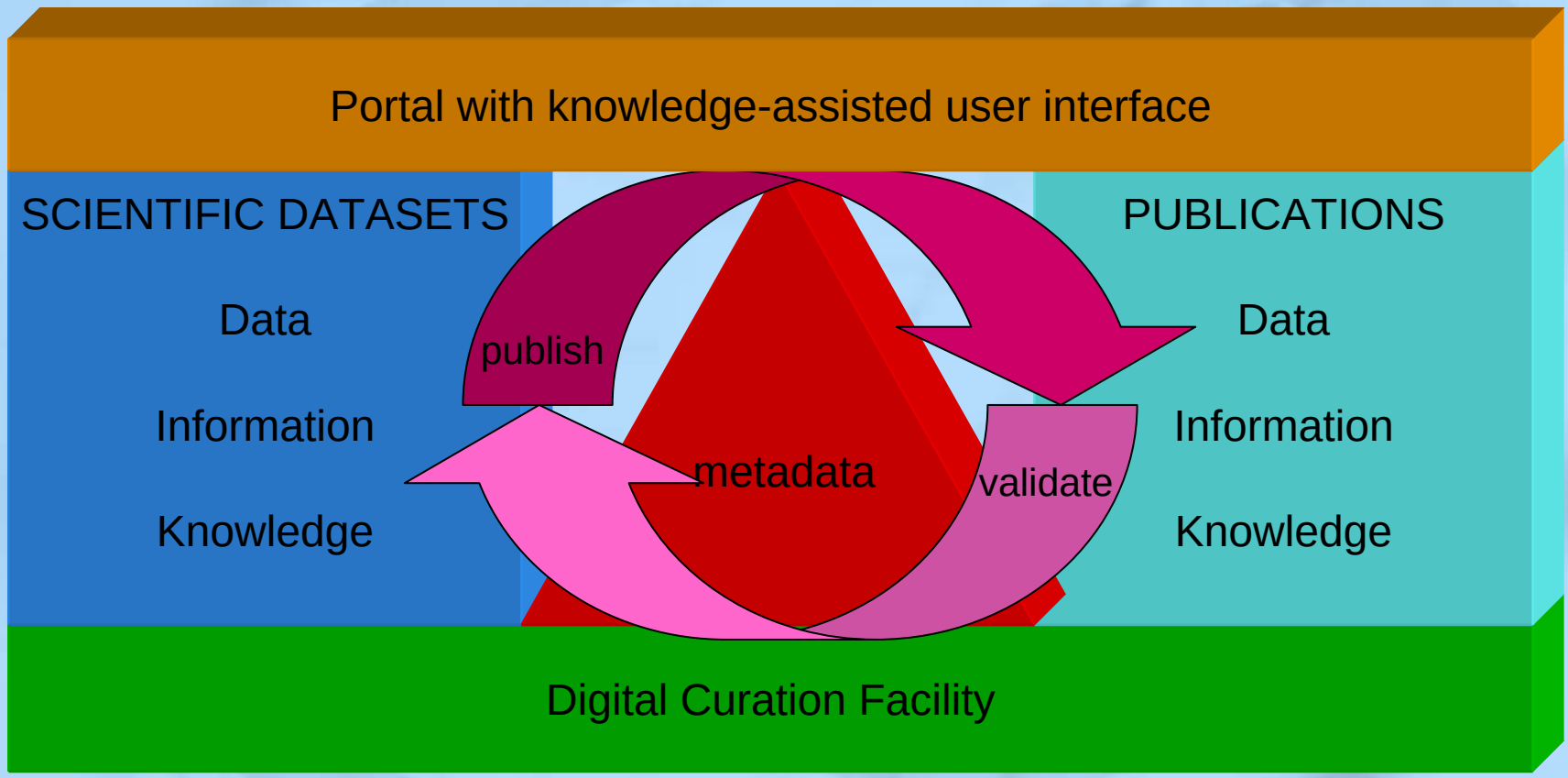
Overall : The Way Forward



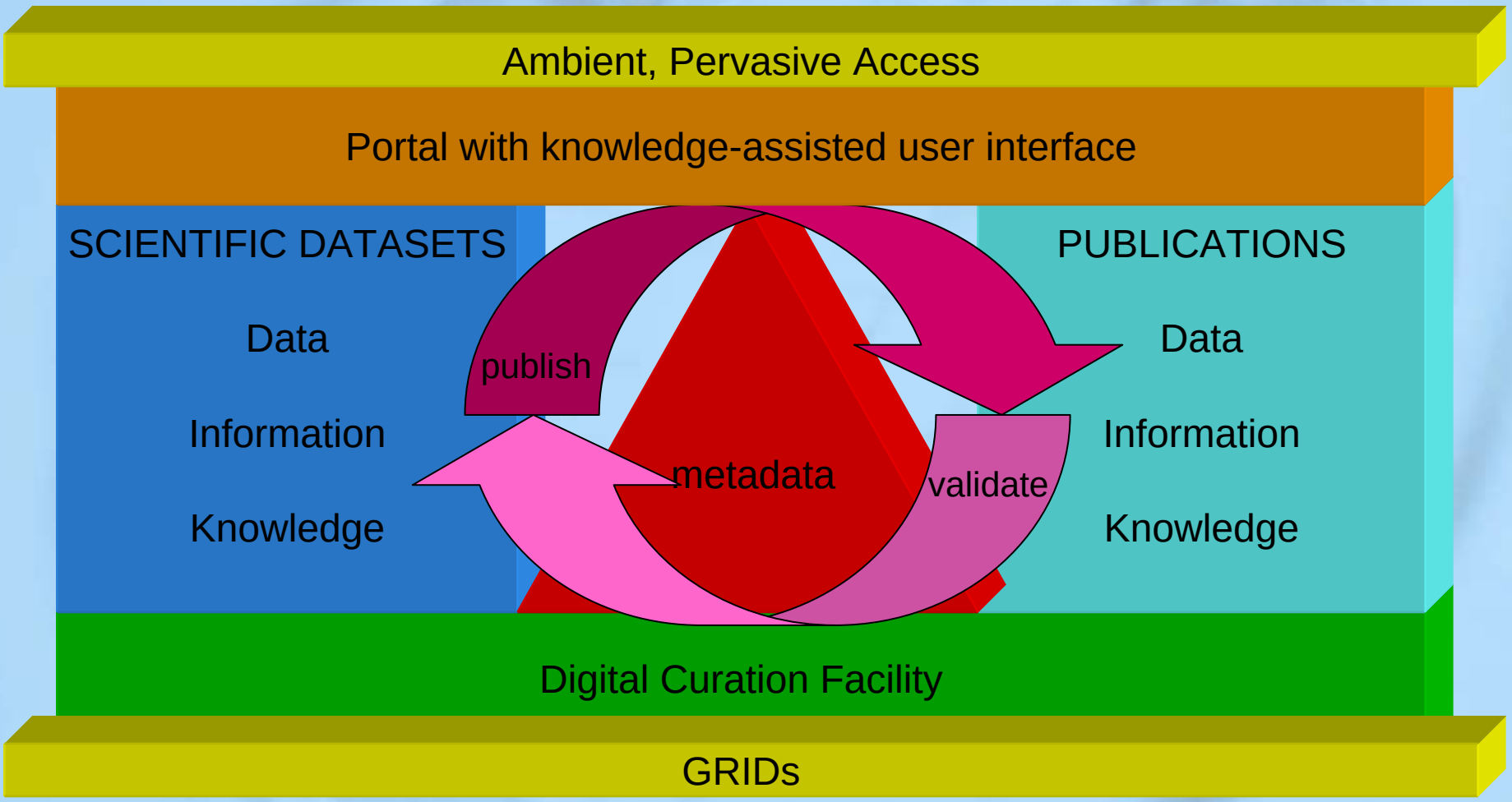
Overall : The Way Forward



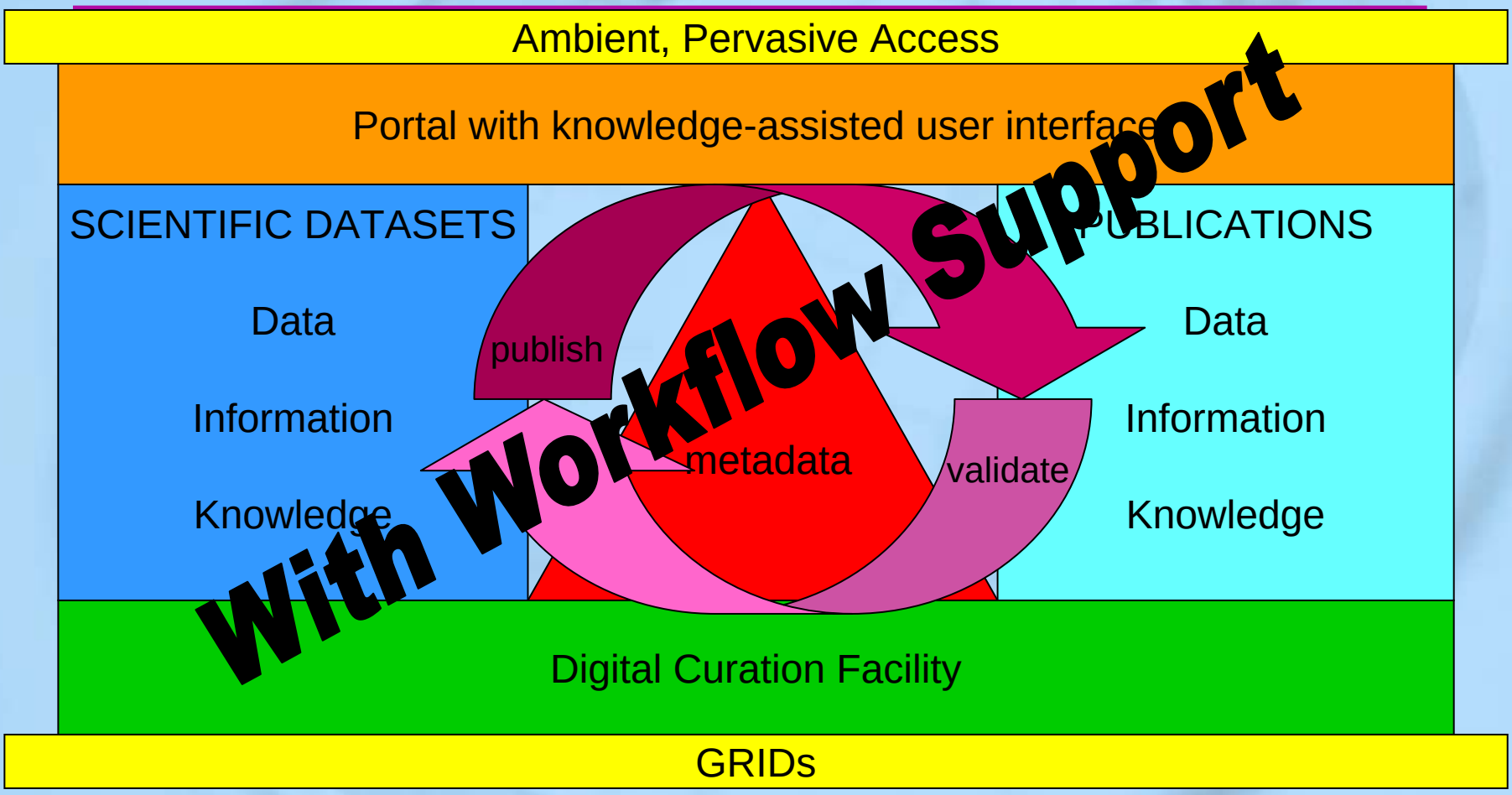
Overall : The Way Forward



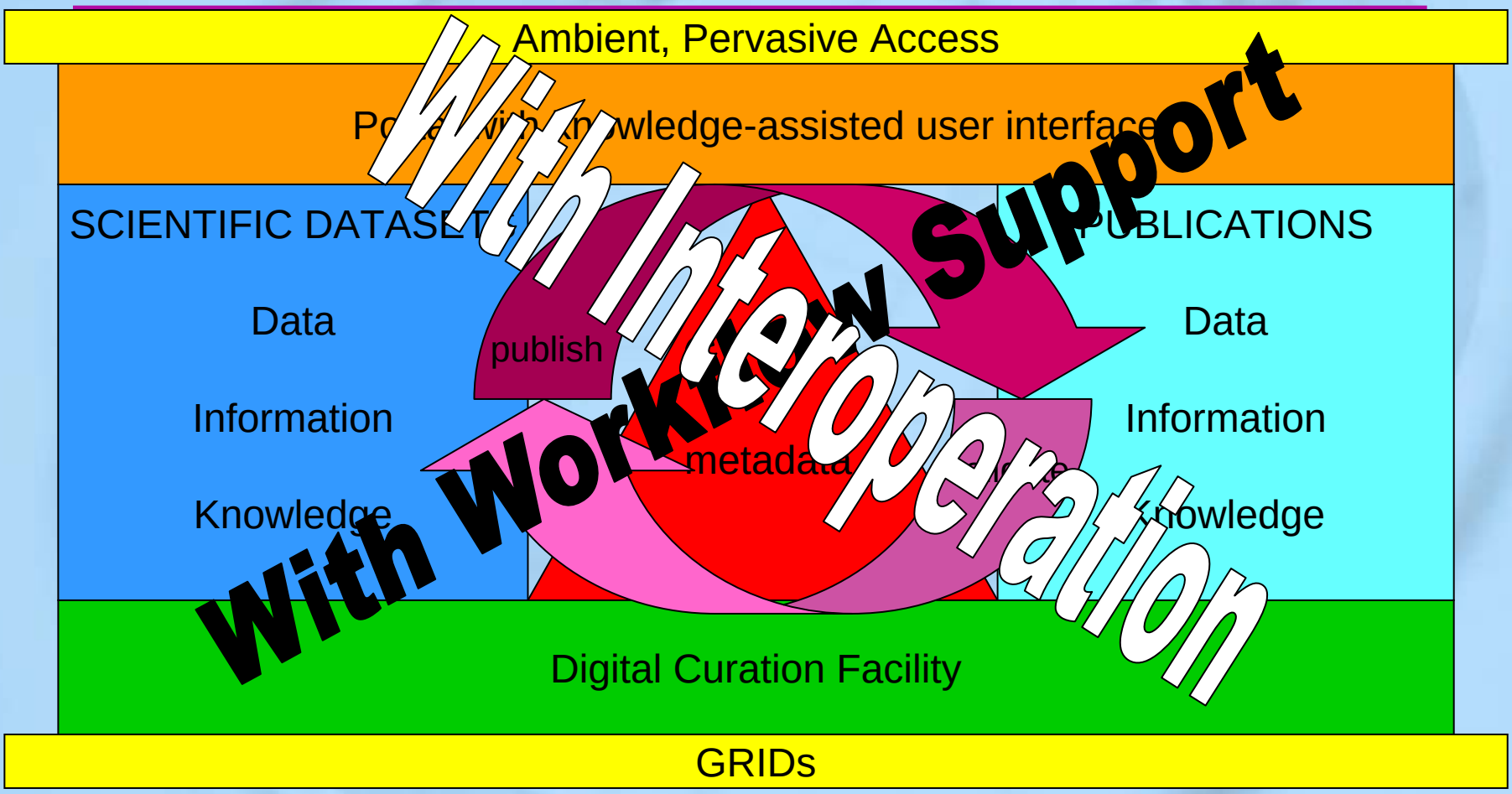
Overall : The Way Forward



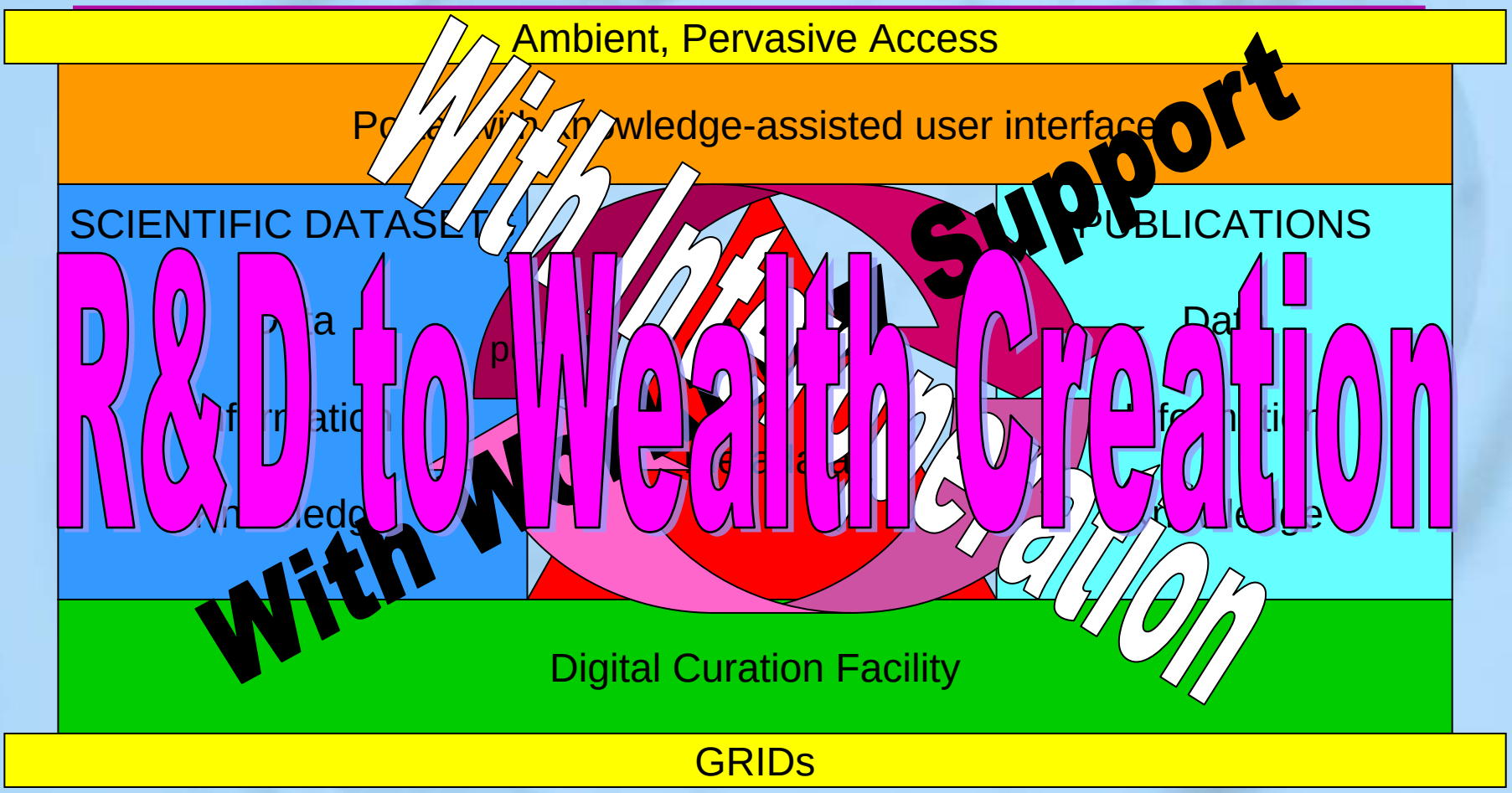
Overall : The Way Forward



Overall : The Way Forward



Overall : The Way Forward



Three Steps to Nirvana

The Perfect CRIS

Workflow on the GRIDs Surface

Metadata and Data Exchange Standards

Complete Process ICT Support

Agenda

- Introduction – speaker
- Requirements
- Institutional Repositories
- CRIS Purpose & Stakeholders
- CERIF-CRIS at the centre of the Organisation
- e-Infrastructure
- Synthesis
- Role of euroCRIS

euroCRIS
The Role

- It is the role of euroCRIS to:
  - Promote and improve communication and interaction between global CRIS;
  - Maintain and publish the CERIF (Common European Research Information Format) recommendation and any standards endorsed by euroCRIS;
  - Organize and run the CRIS series of conferences with associated workshops and other events;

euroCRIS The Role

- Provide a source of expertise in CRIS to members and to others under business arrangements made at the time;
- Develop euroCRIS guidelines;
- Nurture the CRIS community by events, a newsletter, an online discussion forum and other appropriate mechanisms;
- Provide a forum for exploring and exploiting new and emerging concepts and technologies (including data quality, standards, etc.);
- Establish a one-stop portal / gateway to international CRIS resources. (eurocris charter)

euroCRIS
Current Research Information Systems

Prof. Keith G Jeffery

President, euroCRIS

www.eurocris.org