

Open Access Scientific Repositories

ERA-NET PathoGenoMics Workshop

Lisboa, 4th July 2011

Luis Magalhães
President **Knowledge Society Agency**
Portugal



UMIC
Knowledge Society
Agency

Knowledge Society Agency Mission

- To coordinate information society policies and its mobilization through research, qualification and awareness activities
- To promote emerging technologies such as ICT and Nanotechnology
- To develop and fund e-Science



UMIC
Knowledge Society
Agency

Knowledge Society Agency in EU

Represents Portugal mostly next to DG Information Society and Media:

- ❖ HLG on Digital Agenda for Europe (2010-2020)
- ❖ HLG on Internet Governance
- ❖ ICT National Research Directors Forum
- ❖ e-Infrastructures Policy Forum
- ❖ Future Internet Forum of Member States
- ❖ National ICT Directors WG on FET Flagships
- ❖ AAL Joint Program Committee and AAL Association General Assembly
- ❖ ICT CIP – ICT Committee, component of the Competitiveness and Innovation Program
- ❖ . . .



UMIC

Knowledge Society
Agency

Knowledge Society Agency at OECD

Represents Portugal as:

- ❖ Bureau of ICCP – Information Computers and Communication Policy Committee
- ❖ Chair of WPIIS – Working Party on Indicator for the Information Society
- ❖ Vice-Chair of WPIE – Working Party on Information Economy
- ❖ Steering Committee of HL Meeting “The Internet Economy: Generating Innovation and Growth”, 28-29 June 2011



UMIC
Knowledge Society
Agency

Knowledge Society Agency Main Operational Projects

Incubated eGovernment and developed transversal large scale projects

- ❖ Citizen's Portal (2004-2007)
- ❖ Enterprise Portal (2006-2007)
- ❖ Full Creation of Enterprises Online (2006)
- ❖ e-ID Citizen Card (2005-2007)
- ❖ Public Administration Interoperability Platform (2006-2007)

spinned off to AMA – Agency for Public Services Modernization, 1st May 2007

Incubated the National Public eProcurement Program

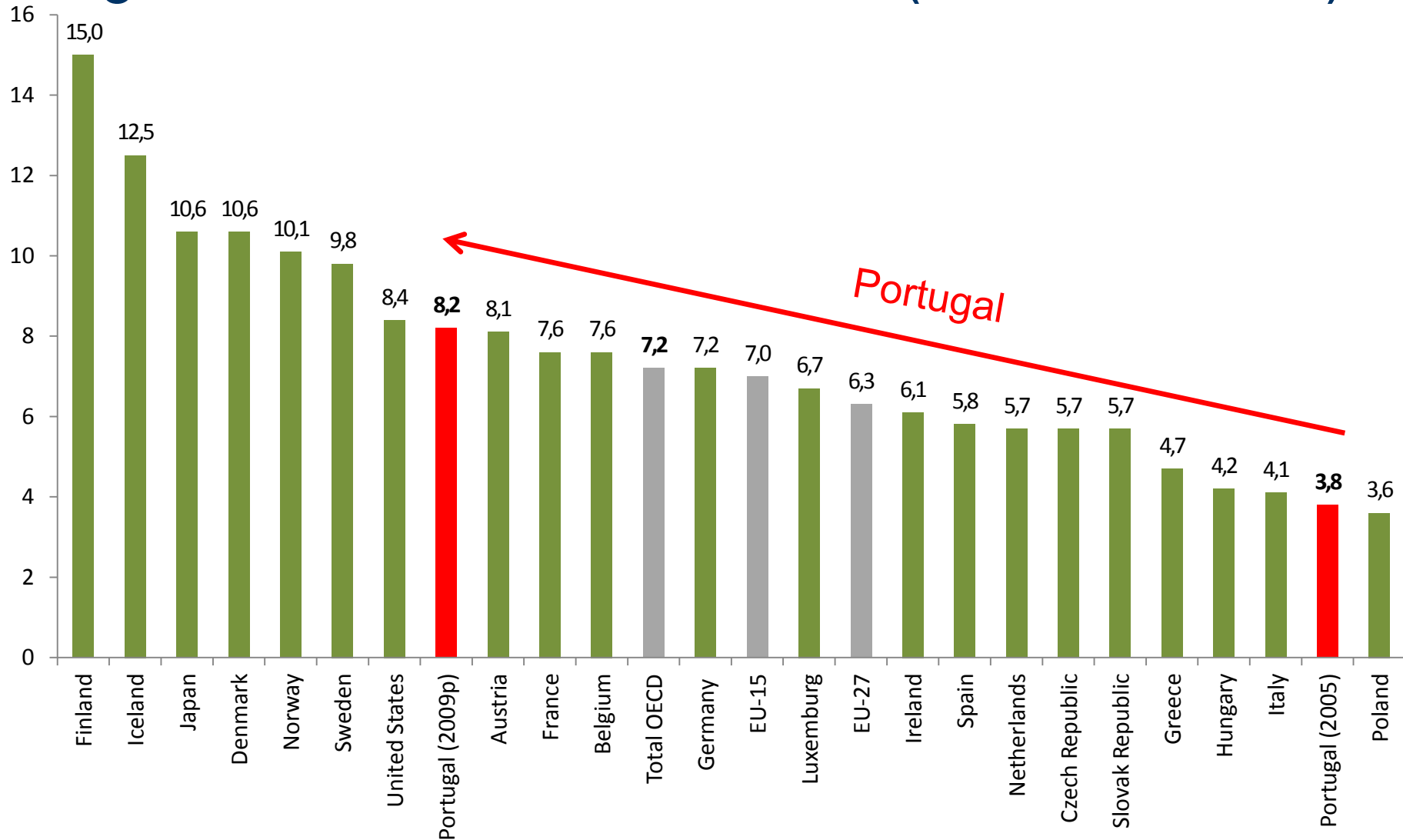
spinned off to National Agency of Public Procurement in Ministry of Finance, 9th May 2007

More recent challenges – towards **knowledge and innovation**:

- ➔ **e-Science**
- ➔ **International partnerships in S&T**
- ➔ **Health and biomedical sciences information** for citizens on the Web
- ➔ **Emerging Technologies**, such as **Future Internet** and **Nanotechnology**

Context of High S&T Growth in Portugal more than doubling in 2005-2009

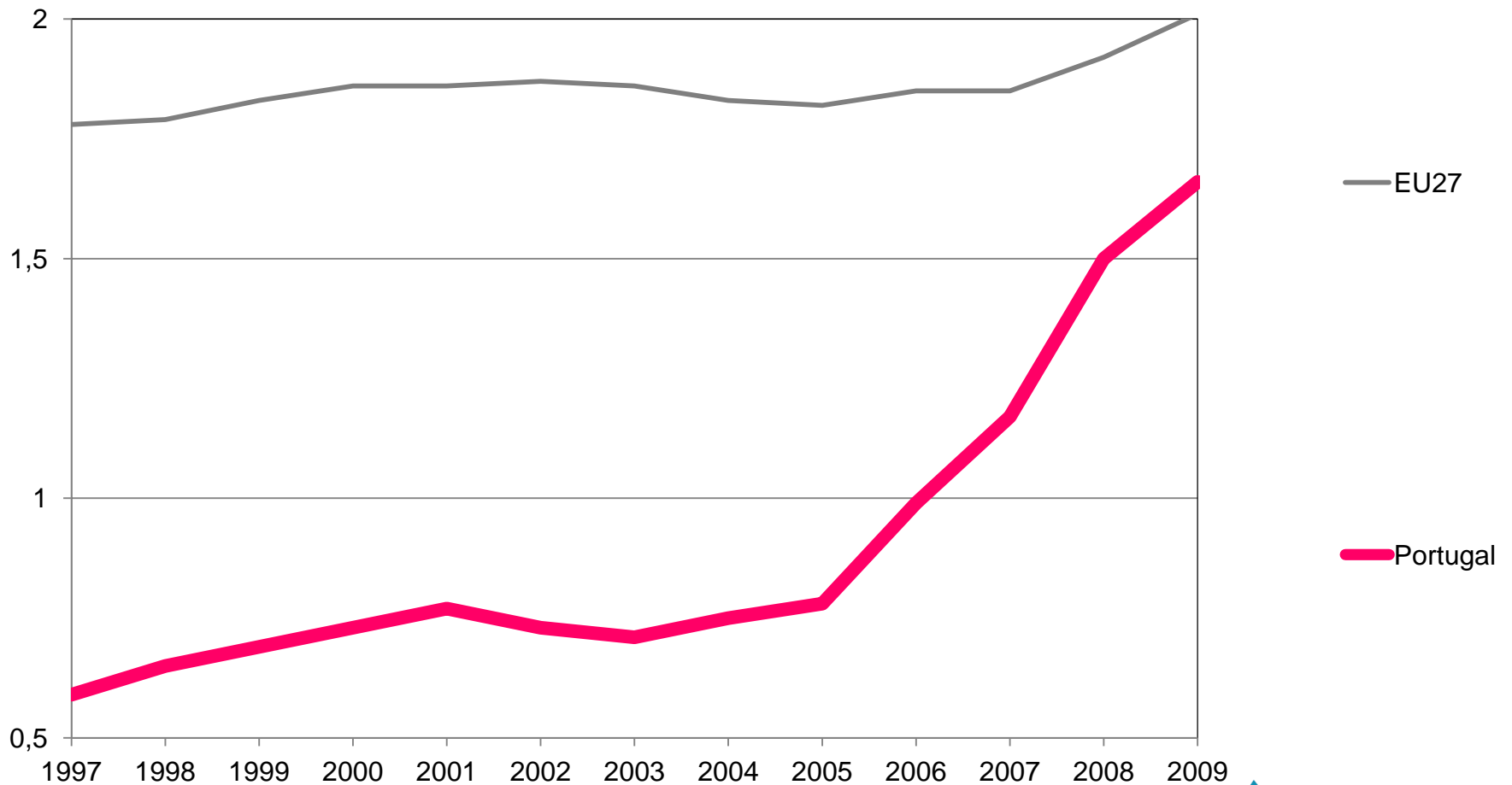
High Growth in Researchers (% labor force)



Note: Data for 2008, except for Portugal whose data are for 2005, 2007, 2008, 2009

Source: OECD.

R&D Intensity, in Portugal



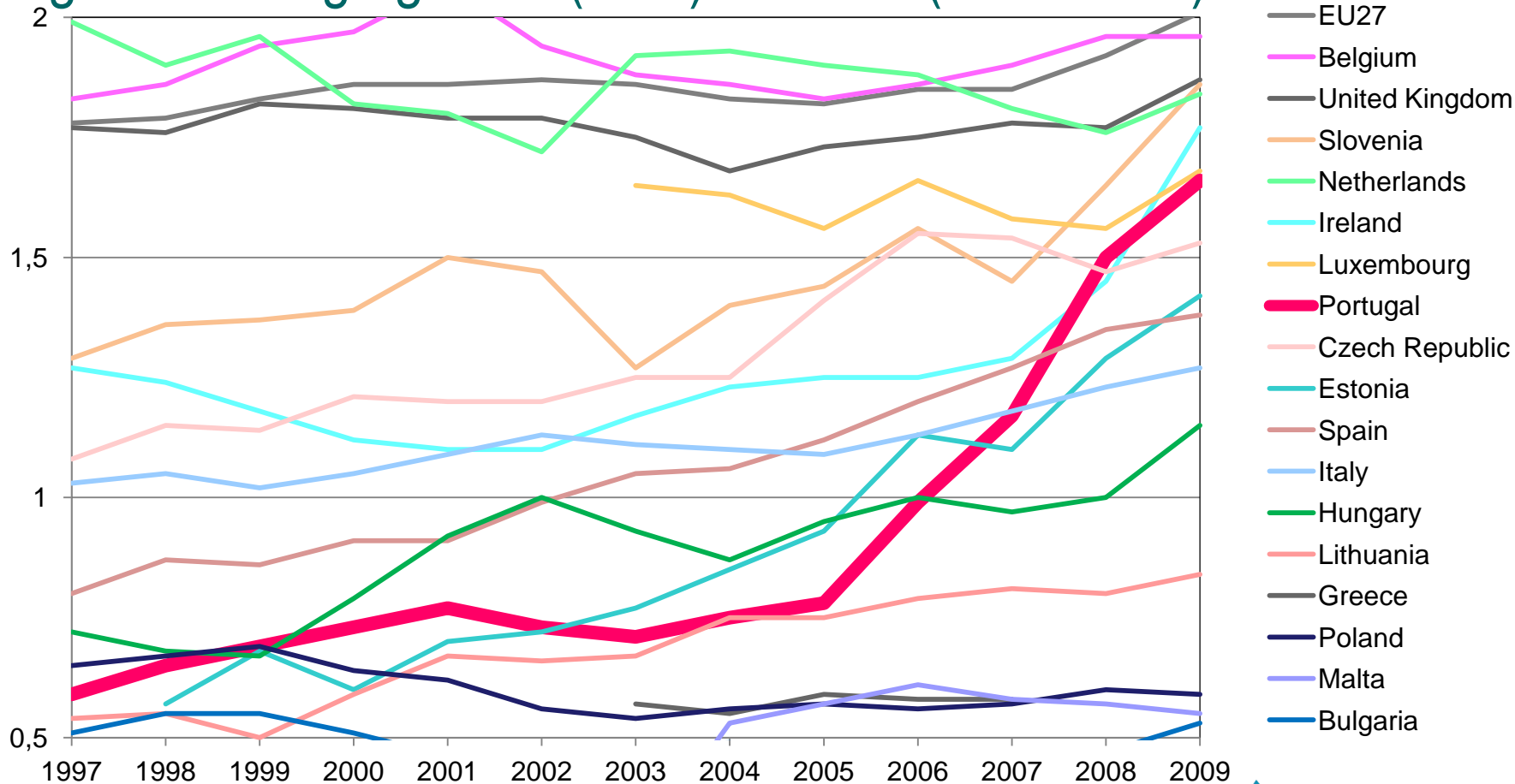
Source: EUROSTAT



UMIC
Knowledge Society
Agency

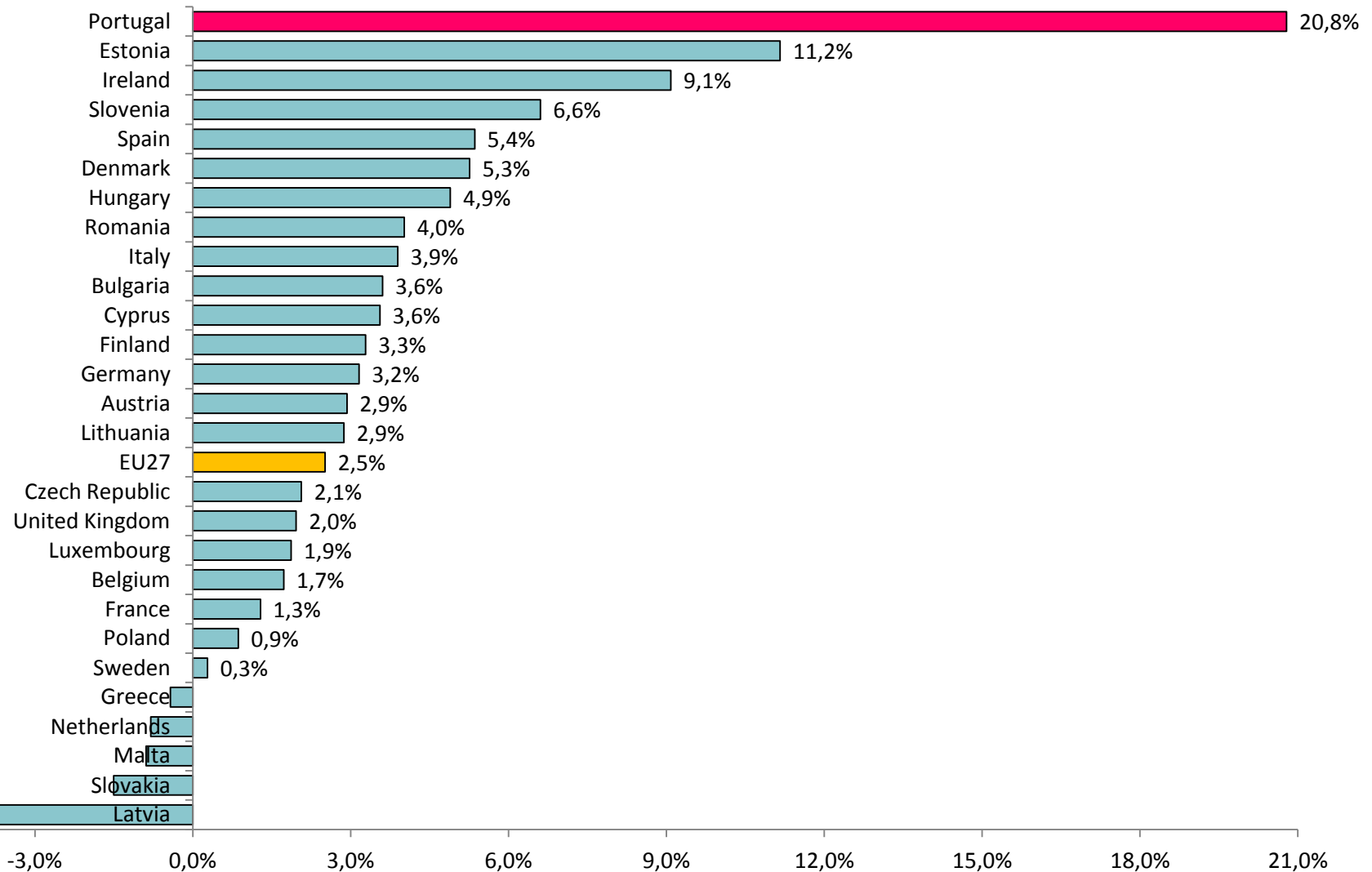
R&D Intensity, in Portugal

Highest average growth (21%) of UE27 (total=2.5%) in 2005-2009



Source: EUROSTAT

Average Annual Growth of R&D Intensity 2005-2009



Source: EUROSTAT

In this Context of High S&T Growth
Strong National e-Science Strategy

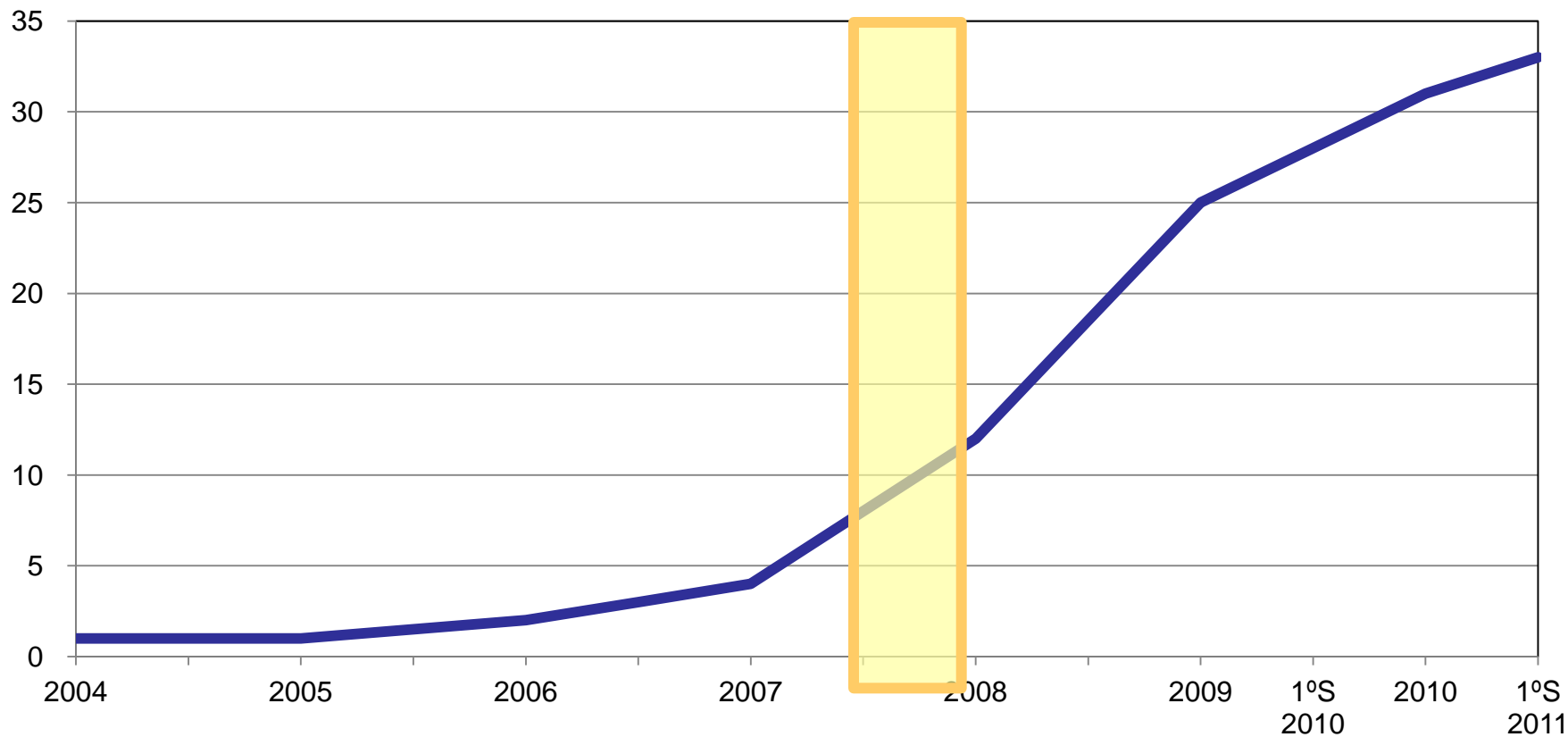
National platforms provided by NREN,
with distributed services for research and higher education,
with high economies of scale,
at zero cost for public user institutions.

Infrastructure • Content • Distributed Computing • Cooperative Work at a Distance

National e-Science Strategy

- **National Research and Education Network** as a public NGN, presently with dark fiber (48 fibers) of the NREN to 55% of Higher Education System, at 10 Gbps and scalable
- **e-U: Virtual Campus** wireless access and roaming integration of all Higher Education *campi*
- **CERT.PT**, internationally accredited CSIRT since 2002, coordinates national CSIRTs network
- **b-on: Knowledge Library Online** planned in 1999, with 17,100 scientific journals, 18,200 e-books, 12,400 proceedings and transactions titles, 10 referential data bases, free access in all public Higher Education and Scientific Institutions, “*big deal*” at national scale, 6.8 million downloads of full text in 2010
- **RCAAP: Open Access Scientific Repository of Portugal**, presently with 33 institutions, including all 14 public universities, and >59,000 documents, integration w/ Brazil OASIS.BR
- **Zappiens.pt: HD video repository of scientific educational and cultural interest**, with Creative Commons licensing and DRM since 2008. In 2010 Brazil started Zappiens.br
- **Archive of the Portuguese Web**, periodic snapshots of web under .pt. Research on new search tools for web archives and on measuring the Web
- **INGRID: National GRID Initiative** (2,100 CPUs, 1 PetaByte of disc memory), integrated w/ Spanish GRID (IBERGRID), and part of EGI – European Grid Initiative
- **IBERCIVIS: Voluntary Computing at the service of science** jointly w/ Spain
- **Tools for collaborative work at a distance**
 - HD Videoconferencing and immersive rooms
 - VoIP for all Higher Education and Scientific System, allowing simple collaborative video- and tele- conferencing from the desktop and without intermediaries
 - **National platform for scientific and educational digital content for Medicine** and **Future Internet**, to be further extended.

Institutional Repositories in the Open Access Scientific Repository of Portugal

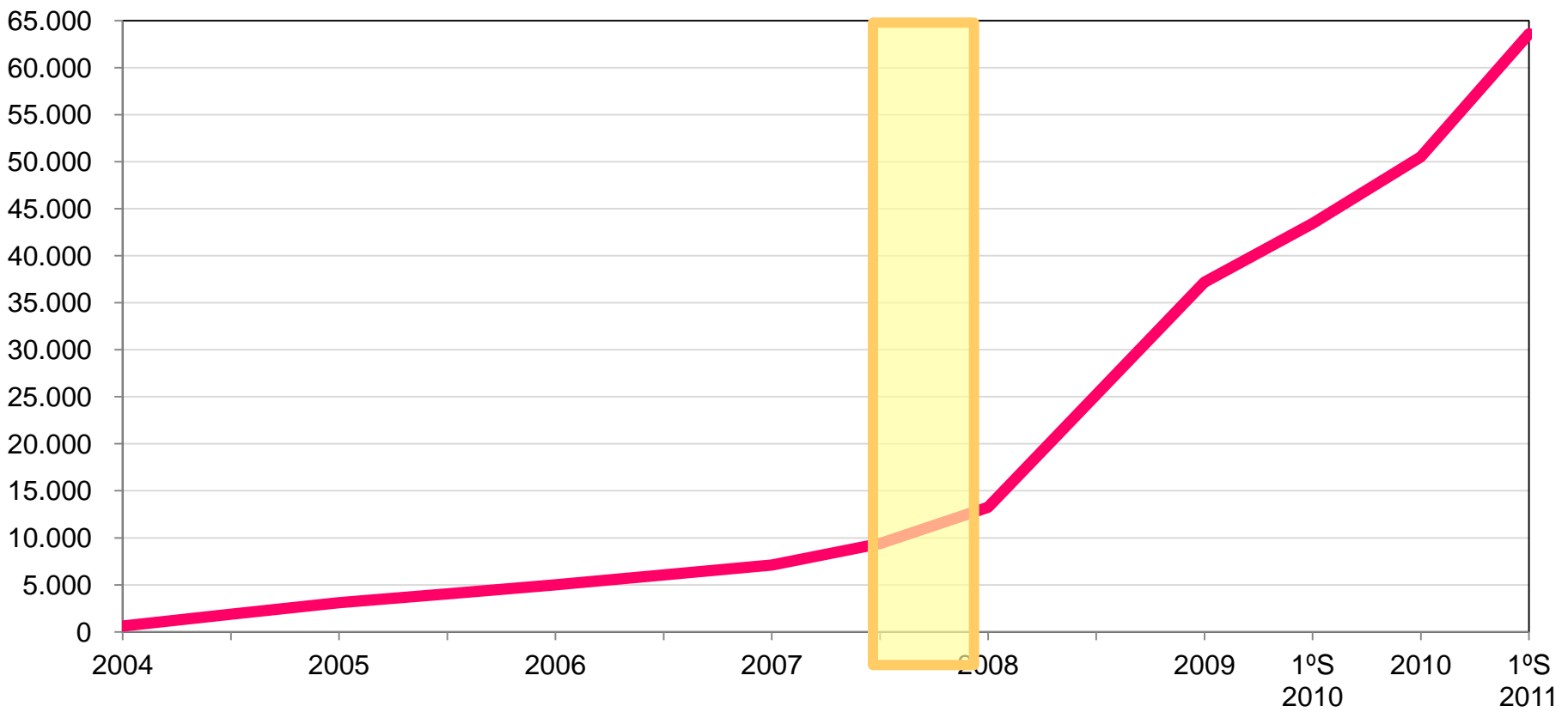


Source: FCCN



UMIC
Knowledge Society
Agency

Documents in the Open Access Scientific Repository of Portugal



Source: FCCN



UMIC
Knowledge Society
Agency

Open Access Scientific Data Repositories

- CERN LHC experiment distributed data and processing through Grid Computing
- RCAAP Pilot Project on Scientific Data Repositories: 2010
- Open call for open access scientific data repositories: 18-31 May 2011, aiming at 2 projects



UMIC
Knowledge Society
Agency

Open Access Scientific Data Repositories

OECD Principles and Guidelines for Access to Research Data from Public Funding, 2007

Principles:

→ *Flexibility:*

Take into account the **rapid and often unpredictable changes in IT**, the **characteristics of each research field** and the **diversity of research systems**, **legal systems** and **cultures**.

→ *Transparency:*

- ❖ Research organizations and government research agencies should **actively disseminate information on research data policies to individual researchers, academic associations, universities and other stakeholders**.
- ❖ Whenever relevant, **all members of the various research communities should assist in establishing agreements on standards for cataloguing data**. The application of existing standards should be considered, whenever appropriate.
- ❖ Information on data management and access should be communicated among data archives and data producing institutions, **so that best practices can be shared**.

Open Access Scientific Data Repositories

OECD Principles and Guidelines for Access to Research Data from Public Funding

Principles (continued):

→ *Legal conformity:*

Subscribe to **professional codes of conduct**.

→ *IP protection:*

- ❖ **The fact that there is private sector involvement in the data collection should not, in itself, be used as a reason to restrict access to the data.**

Consideration should be given to **measures that promote noncommercial access and use** while protecting commercial interests, such as delayed or partial release of such data, or the voluntary adoption of licensing mechanisms.

- ❖ In those jurisdictions in which government research data and information are protected by IP rights, **the holders of these rights should nevertheless facilitate access to data** particularly for public research or other public-interest purposes.

Open Access Scientific Data Repositories

OECD Principles and Guidelines for Access to Research Data from Public Funding

Principles (continued):

→ *Formal responsibility:*

Access arrangements should promote **explicit, formal institutional practices**, such as rules and regulations, regarding the **responsibilities of the various parties**: authorship, producer credits, ownership, dissemination, usage, financial arrangements, ethical rules, licensing terms, liability, sustainable archiving.

Access arrangements, should be developed in consultation with representatives of all directly affected parties. In **collaborative research programs or projects**, and especially in **international scientific co-operation** or in **research projects based on public/private partnerships** where there are differences in regulatory frameworks, **the parties involved should negotiate research data sharing arrangements as early as possible** in the life of the research project, **ideally at the initial proposal stage.**

Long-term sustainability of the infrastructure required for data access is particularly important. Research institutions and government organizations should take formal responsibility for ensuring that **research data are effectively preserved, managed and made accessible**, in order that they can be put to efficient use **over the long term.**

Open Access Scientific Data Repositories

OECD Principles and Guidelines for Access to Research Data from Public Funding

Principles (continued):

→ *Professionalism:*

Institutional arrangements should be **based on the relevant professional standards and values embodied in the codes of conduct of the scientific communities involved.**

In current research practice, the initial data-producing researcher or institution is sometimes rewarded with temporary exclusive use of the data. **The rules for such incentive arrangements should be developed and explicitly stated by the funding sources in co-operation with the affected research communities.**

Project and program planning activities, at all levels, should expressly acknowledge data issues at the earliest stages to take into consideration funding and technical assistance for the essential organization and curation of data sets. Attention should be paid to **incentives** and the **development of professional expertise** in all areas of **research data management.**

Open Access Scientific Data Repositories

OECD Principles and Guidelines for Access to Research Data from Public Funding

Principles (continued):

→ *Interoperability:*

Technological and semantic interoperability is a key consideration in enabling and promoting international and interdisciplinary access to and use of research data.

Access arrangements, should pay due attention to the relevant **international data documentation standards**. Governments and research institutions should **co-operate with international organizations charged with developing new standards**.

The standards employed should be explicitly mentioned as this is the first requirement for interoperability.

→ *Quality:*

Data managers, and data collection organizations, should pay particular attention to ensuring **compliance with explicit quality standards**. **Institutions and research associations should engage with their research community on their development**. **Universal data quality standards are not practical. Standards should be developed in consultation with researchers.**

Open Access Scientific Data Repositories

OECD Principles and Guidelines for Access to Research Data from Public Funding

Principles (continued):

Data access arrangements should **describe good practices for methods, techniques and instruments** employed in the **collection, dissemination and accessible archiving** of data **to enable quality control by peer review** and **other means of safeguarding quality and authenticity**.

The origin of sources should be documented and specified in a verifiable way. Such documentation should be **incorporated into the metadata of the data sets**. Developing such metadata is important for **enabling scientists to understand the exact implications of the data sets**.

Whenever possible, **access to data sets should be linked with access to the original research materials**, and **copied data sets should be linked with originals**, as this facilitates validation of the data and identification of errors within data sets.

Research institutions and professional associations should develop appropriate practices with respect to the citations of data and the recording of citations in indexes, as these are important indicators of data quality.

Open Access Scientific Data Repositories

OECD Principles and Guidelines for Access to Research Data from Public Funding

Principles (continued):

→ *Security:*

The **data, metadata and descriptions**, should be **protected against intentional or unintentional loss, destruction, modification and unauthorized access in conformity with explicit security protocols.**

Data sets and the **equipment** on which they are stored should be **protected from environmental hazards** such as heat, dust, electrical surges, magnetism, and electrostatic discharges.

→ *Efficiency:*

The data archiving community should **carry out cost-benefit assessments periodically** and constantly **develop and refine retention protocols** to ensure that those data sets with the greatest potential utility are preserved and made accessible.

Open Access Scientific Data Repositories

OECD Principles and Guidelines for Access to Research Data from Public Funding

Principles (continued):

Specialized support services, for example through **collaboration with non-academic specialists** on specific research projects or the **engagement of data management specialist organizations**, **should be considered** as a means to ensure the cost-effective production, use, management and archiving of research data.

Insufficient incentives for researchers or database producers may lessen their efforts on data-related activities. The **development of new reward structures** and the **adaptation of existing ones**, including **recognition of data management activities in tenure and promotion review**, should be considered as a way to address this problem.

→ **Accountability:**

The **performance of data access arrangements** should be subject to **periodic evaluation by user groups, responsible institutions** and **research funding agencies**. Although **each party is likely to use somewhat different evaluation criteria**, the sum total of the results should provide a comprehensive picture of the value of data and of data access regimes.

Open Access Scientific Data Repositories

OECD Principles and Guidelines for Access to Research Data from Public Funding

Principles (continued):

→ *Sustainability:*

Due consideration should be given to the sustainability of access to data as a **key element of the research infrastructure**. This means taking **administrative responsibility for measures to guarantee permanent access to data** that require long-term retention. This can be a **difficult task, given that most research projects and their funding have a limited duration**. Research funding agencies and research institutions, therefore, should consider the **long-term preservation of data at the outset of each new project** and **determine the most appropriate archival facilities**.

Open Access Scientific Data Repositories

RCAAP Scientific Data Repositories: State of the Art, July 2010

→ *Recommendations to funding agencies:*

- ❖ Include data curation planning in **research projects evaluation criteria**.
- ❖ Define policy and procedures **mandating the deposit** of research data in sustainable open access repositories.
- ❖ Consider as **eligible in project funding** expenses with data curation and sharing.
- ❖ **Open a funding line for R&D projects in curation and management of research data.**

→ *Recommendations to research institutions:*

- ❖ **Diagnose** the research data practices in the institution.
- ❖ **Provide infrastructures and services** for the curation of research data, **preferably in collaboration with other institutions** and **using existing data repositories**.
- ❖ **Assign responsibility of research data curation** to an organizational unit in the institution and **provide it with the necessary resources**.



UMIC

Knowledge Society
Agency