
Harmonização das Bases de Dados sobre Investigação em Curso Problemas e Vantagens

MARIA DE FÁTIMA BISCAIA

Junta Nacional de Investigação Científica e Tecnológica

Membro do Grupo de Trabalho Europeu sobre Bases de Dados de Investigação

A COORDENAÇÃO dos esforços em Investigação e Desenvolvimento (I&D), a promoção de uma maior mobilidade dos investigadores e a cooperação entre o mundo da investigação e o mundo industrial são objectivos de uma importância vital para a Comunidade Europeia.

A facilitação do acesso à informação sobre os projectos de I&D em curso pode contribuir determinante para estes objectivos. Efectivamente a informação sobre quem está a investigar o quê, onde e com que meios, é de grande utilidade para todos os países e torna-se cada vez mais uma questão de interesse supranacional. Por um lado, as necessidades de investigação científica sentidas num país correspondem, frequentemente, a necessidades sentidas em vários países, ao mesmo

tempo, o que desde logo sugere a vantagem de unir esforços, isto é, trabalhar cooperativamente. Por outro lado os recursos actualmente exigidos pela investigação científica, sobretudo em alguns domínios estratégicos, são tão elevados que os seus custos têm de ser partilhados por vários países.

A estas questões, que apelam tão claramente à necessidade de racionalização e eficácia de meios, vem juntar-se a necessidade óbvia de *comunicação* dos novos conhecimentos produzidos, sem a qual não será possível evitar duplicações e fazer progredir, verdadeiramente, o Conhecimento.

Foi esta constatação que, em 1987, levou a CCE/DGXII e o *Liaison Committee* do Conselho de Reitores dos Estados Membros da CE a proporem a criação de um Grupo de Tra-

balho *ad hoc* com o mandato de estudar e propor medidas com vista à harmonização e melhoria do acesso às bases de dados sobre projectos de I&D em curso. Este Grupo de Trabalho foi efectivamente criado com representantes de todos os Países Membros e exerceu o seu mandato durante o ano de 1988. Em Novembro deste mesmo ano foi apresentado o Relatório Final — *Towards Harmonization of Databases on Research in Progress* — em que, partindo de uma análise das diversas situações nacionais, se tiram conclusões e se apresentam as recomendações consideradas necessárias para atingir os objectivos visados. Estas recomendações podem sintetizar-se em duas propostas essenciais:

- Adopção do Formato Comum Europeu (Common European Research Information Format: CERIF) definido no estudo.
- Utilização de um sistema de classificação comum, multilingue e multidisciplinar.

Em Maio de 1991 as propostas formuladas neste Relatório foram aprovadas pela Comissão e em 13 de Julho do mesmo ano foram publicadas no *Jornal Oficial* como *Recomendação da CCE aos Estados Membros*, aconselhando a sua aplicação progressiva até Janeiro de 1993 (Recomendação 91/337/CEE).

Pela importância desta questão para o acesso aos novos conhecimen-

tos produzidos em ciência e tecnologia, parece importante referir algumas das acções desenvolvidas pelo Grupo de Trabalho, assim como o seguimento que foi dado às recomendações por ele formuladas.

A coordenação dos esforços em Investigação e Desenvolvimento (I&D), a promoção de uma maior mobilidade dos investigadores e a cooperação entre o mundo da investigação e o mundo industrial são objectivos de uma importância vital para a Comunidade Europeia.

A facilitação do acesso à informação sobre os projectos de I&D em curso pode contribuir determinantemente para estes objectivos. Efectivamente a informação sobre quem está a investigar o quê, onde e com que meios, é de grande utilidade para todos os países e torna-se cada vez mais uma questão de interesse supranacional.

O Grupo de Trabalho

O «Workshop on European Research Databases» (*Workshop on European Research Databases, Brussels, 30/09/1987-02/10/1987, organized by the Directorate General for Science, Research and Development — Commission of the European Communities, in cooperation with the Liaison Committee*

of Rectors' Conferences of Member States of the European Communities) demonstrou, com muita evidência, a necessidade de harmonizar progressivamente, a nível europeu, as bases de dados sobre I&D existentes a nível institucional ou nacional, por forma a tornar possível uma maior cooperação e coordenação.

Por isso foi decidido pela Comissão, em estreita cooperação com o *Liaison Committee* dos Conselhos de Reitores dos Países Membros, criar um Grupo de Trabalho com peritos de todos os países da CE, com o seguinte mandato:

1. Elaborar os critérios a que as bases de dados sobre I&D devem obedecer para satisfazerem as necessidades dos vários utilizadores envolvidos na formulação e na execução das políticas científicas nacionais.
2. Estudar, com vista a um contexto de harmonização, as questões de:
 - normalização e quadro de trabalho comum;
 - adopção de uma linguagem e de um sistema de classificação comum, multilingue e multidisciplinar;
 - acesso orientado e harmonizado.
3. Formular propostas e recomendações com vista à harmo-

nização pretendida e à melhoria da permuta de informações sobre investigação científica e desenvolvimento tecnológico entre os Países Membros da CE.

A estas questões, que apelam tão claramente à necessidade de racionalização e eficácia de meios, vem juntar-se a necessidade óbvia de comunicação dos novos conhecimentos produzidos, sem a qual não será possível evitar duplicações e fazer progredir, verdadeiramente, o Conhecimento.

Para a realização deste mandato o Grupo de Trabalho começou por fazer uma prospecção dos objectivos e das necessidades reais e potenciais dos utilizadores interessados por este tipo de informação, seguida de uma análise aprofundada da situação existente em cada país no domínio das bases de dados sobre projectos de I&D em curso.

Objectivos e Necessidades

A informação fiável e actualizada sobre os projectos de I&D em curso é considerada indispensável como instrumento de política científica e tecnológica, tanto a nível nacional como a nível europeu.

Algumas das vantagens reconhecidas a esta informação são consensualmente declaradas:

- identificação clara de quem investiga o quê, onde e com que meios;
- possibilidade de evitar duplicações e repetição de erros;
- identificação de lacunas nos esforços de I&D desenvolvidos numa instituição, num país ou a nível internacional;
- possibilidade de fazer comparações e de estabelecer relações de cooperação;
- identificação de parceiros e de fontes de financiamento;
- alargamento da comunicação científica ao sector industrial, contribuindo para a melhor utilização industrial dos resultados de I&D;
- identificação de interesses comuns a nível transnacional e detecção de tendências novas a explorar cooperativamente;
- facilitação do fluxo transnacional do *know-how* científico e tecnológico;
- identificação de novos conceitos emergentes e de novas oportunidades para a inovação tecnológica;
- captação de dados para a previsão e a avaliação de ciência e tecnologia;
- possibilidade de estimular a cooperação a nível europeu e de progressivamente coordena-

nar a política comunitária de investigação científica e desenvolvimento tecnológico.

Este conjunto de possibilidades e vantagens evidencia, só por si, a necessidade de mobilizar esforços e meios para compatibilizar e desenvolver coordenadamente os serviços de informação sobre projectos de I&D em curso a nível comunitário. Também as questões de acesso carecem de ser melhoradas por forma a facilitar a comunicação aos utilizadores finais deste tipo de informação. Os estudos feitos sobre estes utilizadores, actuais e potenciais, permitem tipificá-los nas seguintes categorias:

- os decisores políticos (*policy makers*) com intervenção na área da investigação científica e do desenvolvimento tecnológico, a nível institucional, nacional ou internacional;
- os gestores de ciência e tecnologia, a nível global ou sectorial;
- os investigadores, os tecnólogos e os agentes da inovação industrial;
- os *information brokers* e todas as empresas de consultoria em I&D e em áreas de ligação do desenvolvimento económico e social com as actividades de I&D;
- todos os serviços de documentação e informação em domínios científicos e tecnológicos.

O Ponto da Situação

Pareceu muito importante ao Grupo de Trabalho fazer uma análise detalhada da situação existente em cada país no âmbito dos sistemas de informação sobre projectos de I&D em curso.

A informação fiável e actualizada sobre os projectos de I&D em curso é considerada indispensável como instrumento de política científica e tecnológica, tanto a nível nacional como a nível europeu.

Esta análise permitiu realizar um estudo comparativo em que foram ventilados os aspectos fundamentais das bases de dados já operacionais, planeadas ou em protótipo: conteúdo e âmbito, tipo de dados registados, estrutura dos formatos, níveis de desagregação, linguagens utilizadas, frequência das actualizações, equipamento e formas de acesso.

Os resultados deste estudo evidenciaram que, apesar de algumas características comuns, existia um certo grau de incoerência e falta de homogeneidade nos sistemas de informação dos diferentes países, o que tornava difícil, ou mesmo impossível, a sua comunicabilidade a nível europeu e uma pesquisa eficaz da informação necessária aos utilizadores finais.

Impunha-se portanto definir uma estratégia de harmonização que aproveitasse ao máximo os recursos já existentes e tornasse possível a intercomunicabilidade de um conjunto de dados essenciais a nível europeu. As opções a fazer deveriam constituir um modelo a seguir nas bases de dados que viessem a ser criadas posteriormente, favorecendo desde logo a compatibilidade e a facilidade de acesso a todos os países da CE.

Estratégias de Harmonização

Entre os vários cenários que foram analisados para a configuração dos objectivos pretendidos a nível comunitário, o Grupo de Trabalho decidiu optar por uma estratégia descentralizada de ligação em rede dos sistemas existentes ou a ser criados ulteriormente.

Esta abordagem foi consensualmente considerada a mais flexível e a que melhor correspondia aos interesses nacionais, sem contudo prejudicar a intercomunicabilidade dos dados a nível europeu.

Impunha-se, naturalmente, a necessidade de introduzir alterações nos sistemas de informação vigentes, orientando-os para uma norma comum europeia, o que implicaria alguns custos e inevitáveis resistências. No entanto, esta estratégia foi considerada preferível a um modelo centralizador que optasse pela consti-

tuição de uma base de dados europeia, tratada e formatada centralmente, com os dados de todos os países.

A interconexão dos sistemas nacionais já existentes ou em preparação teria, entre outras, a vantagem de aproveitar os esforços já feitos, deixando aos Países Membros a responsabilidade pela gestão e actualização da informação, mediante o respeito de alguns princípios normativos básicos que pudessem garantir a sua comunicabilidade europeia. As mudanças a introduzir traduzir-se-iam também em benefícios para os próprios países, pelo impacto maior dos seus esforços em I&D e pela possibilidade de acesso aos resultados de outros países, com todas as vantagens daí resultantes.

Os resultados deste estudo evidenciaram que, apesar de algumas características comuns, existia um certo grau de incoerência e falta de homogeneidade nos sistemas de informação dos diferentes países, o que tornava difícil, ou mesmo impossível, a sua comunicabilidade a nível europeu e uma pesquisa eficaz da informação necessária aos utilizadores finais.

Tornava-se necessário, portanto, estudar e definir os critérios de normalização que deveriam ser seguidos para garantir a compatibilidade pretendida.

Após um trabalho longo de estudo e discussão, o Grupo de Trabalho veio a propor a utilização de um Formato Comum Europeu que foi detalhadamente descrito em todos os aspectos de conteúdo e estrutura.

Foi também recomendada a utilização de um sistema de classificação comum, multilingue e multidisciplinar.

Formato Comum Europeu

Com base na definição de «projecto de investigação» consensualmente adoptada, o Grupo de Trabalho desenvolveu um formato para o registo de um conjunto de dados considerados de interesse comum a nível comunitário. Este «Common European Research Information Format (CERIF) foi apresentado com todo o possível detalhe e normas práticas de utilização.

Ao definir este formato foram tidos em conta todos os princípios normativos já existentes sobre formatos de comunicação de dados, designadamente da UNESCO e da ISO.

O CERIF é um formato flexível e aberto, isto é, define um conjunto mínimo de dados nucleares, com carácter obrigatório (de interesse comum essencial), mas apresenta também campos opcionais que poderão ou não ser preenchidos. Por outro lado, cada base de dados local pode acrescentar outros registos para corresponder a necessidades específi-

cas, conquanto exista esse conjunto mínimo de dados nucleares para ser utilizável num intercâmbio europeu.

O CERIF foi portanto concebido com o objectivo de garantir uma grande flexibilidade na manipulação dos diversos tipos de dados de projectos de I&D, nas diferentes bases de dados existentes localmente. Poderá igualmente optar-se por qualquer outro formato que seja convertível, através de um programa adequado, para o CERIF.

O CERIF é um formato flexível e aberto, isto é, define um conjunto mínimo de dados nucleares, com carácter obrigatório (de interesse comum essencial), mas apresenta também campos opcionais que poderão ou não ser preenchidos. Por outro lado, cada base de dados local pode acrescentar outros registos para corresponder a necessidades específicas, conquanto exista esse conjunto mínimo de dados nucleares para ser utilizável num intercâmbio europeu.

Os dados nucleares considerados de interesse essencial na descrição de um projecto de investigação são os seguintes:

- Título (recomenda-se que seja bilingue)
- Resumo (recomenda-se que seja bilingue)

- Endereço (incluindo o nome da unidade de investigação, o departamento e o instituto)
- Director do projecto
- Nome(s) do(s) principal(ais) investigador(es)
- Data de início e data prevista para a conclusão
- Pontos de identificação
 - palavras-chave livres
 - termos controlados (facultativo)
 - códigos do plano de classificação adoptado (recomendado)
- Financiamento directamente relacionado com o projecto
 - fonte de financiamento
 - orçamento anual; em opção: repartição por pessoal, funcionamento, equipamentos
- Pessoal: número de investigadores ETI (*Equivalente a Tempo Integral*).

Dentro do espírito de flexibilidade que presidiu à concepção do CERIF, optou-se por campos de dados de comprimento variável, com a possibilidade de campos repetíveis.

Questões Terminológicas e de Classificação

Num contexto multilingue como a CEE, foi extremamente difícil encontrar consensos sobre instrumentos

comuns, para tratamento e pesquisa da informação contida nas bases de dados locais sobre projectos de I&D em curso. Mas se por um lado esta matéria se revelava particularmente sensível, por outro lado reconhecia-se a imprescindibilidade de encontrar soluções de compromisso para não pôr em causa o objectivo fundamental visado: a comunicação da informação com um grau máximo de fiabilidade, apesar das diferentes línguas utilizadas.

Seria excessivamente longo descrever todas as soluções examinadas e os diferentes vocabulários e planos de classificação analisados e testados. Importa sobretudo referir que, finalmente, acabou por ser adoptada a *Classificação dos Domínios Científicos* utilizada no Serviço de Programação Científica da Bélgica (*Research Classification Scheme*) que foi traduzida em todas as línguas da Comunidade, com o compromisso de ser permanentemente actualizada.

Ainda com o mesmo objectivo de facilitar a permuta transnacional da informação, foi recomendado que o título e o resumo de cada projecto fosse apresentado na língua nacional e também numa outra língua da Comunidade, de larga utilização no domínio científico em causa.

Resultados e Avaliação

Conforme foi já mencionado, as recomendações formuladas pelo

Grupo de Trabalho foram aprovadas pela Comissão em 6 de Maio de 1991 e, em 13 de Julho do mesmo ano, foram publicadas no *Jornal Oficial* como *Recomendação aos Países Membros (Recomendação da Comissão relativa à harmonização na Comunidade das bases de dados no domínio da investigação e do desenvolvimento tecnológico — 91/337/CEE —, Jornal Oficial das Comunidades Europeias, n.º L189, 13 de Julho de 1991)*. O texto desta Recomendação inclui a descrição integral do CERIF e a Classificação dos Domínios Científicos adoptada.

Num contexto multilingue como a CEE, foi extremamente difícil encontrar consensos sobre instrumentos comuns, para tratamento e pesquisa da informação contida nas bases de dados locais sobre projectos de I&D em curso. Mas se por um lado esta matéria se revelava particularmente sensível, por outro lado reconhecia-se a imprescindibilidade de encontrar soluções de compromisso para não pôr em causa o objectivo fundamental visado: a comunicação da informação com um grau máximo de fiabilidade, apesar das diferentes línguas utilizadas.

Foi sem dúvida importante a apresentação desta Recomendação no *Jornal Oficial* da CE e também a força persuasiva com que foi formulada.

No entanto, como é do conhecimento de todos os que trabalham nesta área, a normalização de sistemas de informação não pode fazer-se de um dia para o outro, sobretudo quando incide em sistemas pré-existentes, respeitando outras regras e outras normas. Mudar de um formato para o outro é um trabalho que exige tempo para avaliar e tempo para mudar. Quando se trata de classificações e vocabulários controlados a dificuldade torna-se ainda mais complexa.

Parece portanto importante e oportuno avaliar o impacto que teve a referida Recomendação.

Em Setembro de 1992, a CCE/DGXII organizou uma reunião do Grupo de Trabalho anteriormente constituído, com o objectivo de fazer um balanço dos progressos realizados após a publicação da Recomendação. Nesta reunião participaram também representantes dos países da EFTA. Aliás um dos pontos agendados consistia mesmo em conhecer a situação destes países no domínio dos sistemas de informação sobre projectos de I&D em curso, tendo em vista o alargamento da acessibilidade desta informação a um mais largo espaço europeu.

Durante esta reunião foi possível constatar que o CERIF tinha provocado uma reflexão aprofundada e movimentos de aproximação em todos os países. Os níveis de aproximação eram variáveis, indo da adopção efectiva do CERIF, ao desenvolvi-

mento de esforços para a sua utilização ou à aplicação de formatos convertíveis.

Foram manifestadas algumas dificuldades na recolha dos dados financeiros constantes na zona dos campos nucleares, afigurando-se ser esta questão particularmente sensível em alguns países. Como solução possível apontou-se para a transferência destes dados para a zona de campos opcionais.

No caso da França a utilização do CERIF parece levantar maiores dificuldades, pelo facto de o financiamento para I&D ser feito por domínios científicos e não por projectos.

Os representantes dos países da EFTA descreveram sistemas já existentes ou em preparação que seguem de perto o CERIF ou que são convertíveis, tendo expressado a vontade de cooperação nesta matéria de reconhecimento interesse para a investigação científica e para o desenvolvimento tecnológico a nível europeu.

As questões de classificação e terminologia levantaram, porém, problemas maiores e mais complexos. A *Classificação dos Domínios Científicos* adoptada tem-se revelado de difícil aplicação em alguns países que consideram que não se adapta às suas necessidades e que preferem, por isso, manter as suas próprias classificações ou vocabulários controlados. Concluiu-se portanto que esta questão deveria ser novamente estudada, com vista à procura de soluções mais adequadas e consensuais.

Finalmente, e como balanço global, concluiu-se que a Comissão deverá valorizar os esforços de harmonização já feitos pelos Países Membros e estimular novos esforços mediante a resolução das questões técnicas de ligação em rede, comunicação e facilitação do acesso aos utilizadores finais.

A Situação Portuguesa

A Recomendação 91/337/CEE foi largamente divulgada em Portugal. A Junta Nacional de Investigação Científica e Tecnológica (JNICT), por reconhecer a importância desta Recomendação, enviou-a a todos os seus departamentos, chamando a atenção para os objectivos visados e para a necessidade de aplicar progressivamente as regras comuns recomendadas.

Esta Recomendação foi também publicada no Boletim Informativo mensal da JNICT, *ID: Investigação & Desenvolvimento* (n.º 23/24, Setembro/Outubro de 1991) que é dirigido a toda a comunidade científica e que, nessa altura, atingia já mais de 7000 destinatários.

Além disso, com o apoio da JNICT, o Secretariado Europa 92 fez um *spot* legislativo sobre esta Recomendação para a televisão portuguesa, com o objectivo de sensibilizar o público, de uma maneira muito acessível, para a importância da normalização das bases de dados sobre investigação em

curso. Este *spot* legislativo foi transmitido, durante o mês de Setembro de 1991, nos dois canais da RTP.

Estas acções de divulgação tiveram, portanto, o objectivo de explicar a importância da Recomendação 91/337/CEE e as vantagens da sua utilização para uma melhor comunicação dos resultados da investigação científica e do desenvolvimento tecnológico. Esta Recomendação passou assim a constituir um novo quadro de referência para o qual se torna necessário caminhar.

As bases de dados sobre projectos de I&D financiados pela JNICT seguem já, muito de perto, o CERIF, pelo menos no que respeita a zona de dados nucleares.

Por outro lado, no quadro da cooperação comunitária, Portugal participa em dois sistemas de informação sobre projectos de I&D: AGREP (Permanent Inventory of Agricultural Research Projects) e BIOREP (Permanent Inventory of Biotechnology Research Projects).

O *focal point* para o sistema AGREP é o Instituto Nacional de investigação Agrária (INIA). Este sistema de informação abrange as actividades de I&D directamente relacionadas com a Agricultura no seu sentido lato (Agricultura, Pecuária, Sivicultura, Piscicultura em águas interiores e Indústrias agrícolas). Foram já introduzidos na base de dados os projectos de I&D financiados pelo INIA e está a ser preparada uma participação a nível nacional.

Para o sistema BIOREP o *focal point* português foi inicialmente a Escola Superior de Biotecnologia, no entanto essa responsabilidade foi transferida para a JNICT em Maio de 1992. Foram já introduzidos na base de dados todos os projectos de I&D financiados pela JNICT e, muito brevemente, serão introduzidos os projectos de outras entidades financiadoras.

O sistema BIOREP adoptou já o CERIF; o sistema AGREP contém já os dados essenciais e deverá adoptá-lo brevemente. Tratando-se de programas cooperativos europeus, estes sistemas poderão influenciar fortemente a harmonização, provocando um efeito de arrastamento que tenderá a alargar-se a projectos de I&D noutros domínios científicos e também em domínios transdisciplinares.

Esta mesma observação é extensiva a outros sistemas europeus como, por exemplo, a base de dados PROJETOS do sistema comunitário CORDIS (*Community Research and Development Information Service*).

Em Portugal, como nos outros países da CE, está em curso uma progressiva harmonização das bases de dados sobre investigação em curso. No entanto, essa harmonização só poderá tornar-se efectiva se forem criados estímulos suficientemente fortes e propiciados os meios financeiros necessários. Um outro factor que parece determinante para o sucesso desta harmonização é a facilitação do acesso aos utilizadores finais. Tudo o que venha a ser feito nesse sentido poderá ter uma importância decisiva para os objectivos visados.