RADIATION RESEARCH 175, 526–531 (2011)
0033-7587/11 $15.00

DOI: 10.1667/RR2471.1

DATA ARCHIVE RESOURCE

The European Radiobiological Archives: Online Access to Data from Radiobiological Experiments

M. Birschwilks,[a,1] M. Gruenberger,[b] C. Adelmann,[a] S. Tapio,[a,c] G. Gerber,[d,2] P. N. Schofield[b] and B. Grosche[a]

[a] *Federal Office for Radiation Protection, Department of Radiation Protection and Health, 85764 Neuherberg, Germany;* [b] *Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3DY, United Kingdom;* [c] *Helmholtz Zentrum München, German Research Centre for Environmental Health, 85764 Neuherberg, Germany; and* [d] *Brussels, Belgium*

Birschwilks, M., Gruenberger, M., Adelmann, C., Tapio, S., Gerber, G., Schofield, P. N. and Grosche, B. The European Radiobiological Archives: Online Access to Data from Radiobiological Experiments. *Radiat. Res.* 175, 526–531 (2011).

For financial and ethical reasons, the large-scale radiobiological animal studies conducted over the past 50 years are, to a large extent, unrepeatable experiments. It is therefore important to retain the primary data from these experiments to allow reanalysis, reinterpretation and re-evaluation of results from, for example, carcinogenicity studies, in the light of new knowledge in radiation biology. Consequently, there is an imperative need to keep these data available for the research community. The European Radiobiological Archives (ERA) were developed to fulfill this task. ERA has become a unique archive, including information from almost all European long-term studies carried out between the 1960s and the 1990s. The legacy database was originally developed in a manner that precluded online use. Therefore, strong efforts were made to transform it into a version that is available online through the web. This went together with quality assurance measures, including first the estimation of the rate of non-systematic errors in data entry, which at 2% proved to be very low. Second, every data set was compared against two external sources of information. Standardization of terminology and histopathology is a prerequisite for meaningful comparison of data across studies and analysis of potential carcinogenic effects. Standardization is particularly critical for the construction of a database that includes data from different studies evaluated by pathologists in different laboratories. A harmonized pathology nomenclature with modern standard pathology terms was introduced. As far as possible, references for the various studies were directly linked to the studies themselves. Further, a direct link to the JANUS database was established. ERA is now in a position where it has the potential to become a worldwide radiobiological research tool. ERA can be accessed at no cost at https://era.bfs.de. An ID and password can be obtained from the curators at era@bfs.de.

[1] Address for correspondence: Federal Office for Radiation Protection, Department of Radiation Protection and Health, Ingolstaedter Landstrasse 1, 85764 Oberschleissheim, Germany; e-mail: mbirschwilks@bfs.de.

[2] Retired.

INTRODUCTION

The quantitative assessment of radiation risk is currently based on knowledge gained from epidemiological studies of radiation-exposed populations, complemented by data from animal experiments, as well as fundamental *in vitro* cellular and biophysical studies. Recent developments in molecular and genetic research are, to an increasing extent, providing new opportunities to quantify the risks of radiation exposure even at an individual level (*1–3*). The retrospective analysis of earlier epidemiological and animal studies has therefore become an important approach for the modeling and evaluation of newly discovered risk factors [e.g. ref. (*4*)].

In the mid-1980s, the Radiation Protection Programme of the European Commission and the European Late Effects Project Group (EULEP) embarked on an initiative to collect and collate data covering all available information on European long-term radiobiological animal studies (*5, 6*). The Office of Biological and Environmental Research of the U.S. Department of Energy and, more recently, the Japanese Late Effects Group have started similar efforts to archive the American and Japanese data in the U.S. National Radiobiology Archives (NRA) and Japanese Radiobiological Archives (JRA), respectively. Combined into the International Radiobiology Archives (IRA), this database includes, in addition to two human cohort studies, nearly all radiation biology studies using animals carried out between 1960 and 1998 in Europe, the U.S., and Japan, involving a total of more than 460,000 animals (*7, 8*). The database contains individual information on each animal studied and includes data from the following species: dog, monkey, mouse, rat, other rodents, and livestock. It also includes information from human cohort studies, in particular the Spiess series and the cohort of German ankylosing spondylitis patients treated with ^{224}Ra. The possibility of returning to mine the existing data in new ways provides enormous added value to the original findings and funding of these studies.

The data from ERA documents were stored in a file-based relational database. It was the aim of our efforts to transform this legacy database into an online version and to make the information easily accessible to the scientific community. At the same time, quality assurance measures were taken and the database was updated. This paper describes the content of the online version of ERA and the various steps that were taken to ensure its quality.

MATERIAL AND METHODS

The Source

ERA documents were originally stored in a file-based relational database implemented in MS-ACCESS 2000 that contained information about the purpose of the study, the executing laboratory with the participating scientists, the type of treatment such as radiation, chemical or biological agents, dosimetry, end points such as the survival rate or pathological and molecular biological characteristics, species, strain, sex and age of the animals used, a brief summary of the results, and selected publications. This legacy database was described in detail by Gerber, who initiated it and undertook, with the support of others, the archiving process (*8*). It was not, however, possible to query the legacy database without local access to the data files. The online version of ERA, which is presented here, is based on the previous version for local access.

In the course of converting the legacy database into an online version, a number of tasks were conducted: the data were validated and it was *a priori* decided that the transformation would not be done if the percentage of non-systematic data entry errors was above 5%; pathology nomenclature was standardized; as far as possible, data and references were updated.

Data Quality and Literature

A first validation and quality check of the existing database was conducted by re-entering randomly selected data from four different studies, comprising more than 4,000 individual records. In this validation exercise the records of the database were compared both to the original paper records and to the stored electronic data [described in detail by Tapio *et al.* (*9*)]. A second quality check was done by comparing the contents of the printed version of the legacy database (*5*) with those from the legacy database as of December 2003 and those of the beta version of ERA. This was done by checking the entire dataset manually, defining inconsistencies, and solving problems by constant feedback between Prof. Gerber and one of the authors of this paper (CA).

Many of the primary data in the archive exist in the published literature and/or internal institutional reports. It is important that these sources be readily available together with the original publications and pointers to the studies using the data. An important aim of the project, therefore, was to complement the two existing reference libraries by including links to original articles, reviews, internal reports, etc. and supplementary information whenever feasible. Whenever possible, links to PubMed® references are given and, if there were no legal constraints, PDFs of original papers or reports are included in the database.

Standardization of the Pathology Nomenclature

Because accurate descriptions of the end points are crucial for any future use of the database, meticulous standardization of terminology has been essential. Human pathology was originally recorded using either ICD-8, the International Statistical Classification of Diseases and Related Health Problems, published by the World Health Organization (*10*) or the Klinischer Diagnosenschlüssel (KDS) (*11*). Both classifications contain anatomically and etiologically predicated terms. Rodent pathology is variously represented in the European, American and Japanese archives by the equivalent human nomenclature or by local formalisms, which are effectively unstructured controlled vocabularies. The NRA beagle data were captured using the Systematized Nomenclature of Dog (SNODOG), an unpublished local modification of SNOVET, the Systematized Nomenclature of Veterinary Medicine (*12*) adapted to the dog.

In the 1990s, Prof. Gerber with the assistance of Prof. Goessner translated the local rodent nomenclature terms into the "DIS-ROD" terms (Disease-Rodent) on the basis of the work carried out by the EULEP pathology panel (*13*); ICD-8 and KDS terms were mapped onto ICD-9. While ICD-9 remains a well-used and understood disease classification for humans, the DIS-ROD codes used for rodents are not systematic, and so the decision was taken to map the DIS-ROD codes onto the now standard nomenclature for rodent pathology, the MPATH ontology. The MPATH ontology was developed for the Pathbase database by a group of veterinary and human pathologists and anatomists working extensively with laboratory mice as a description framework for histological images of tissue lesions generated in response to underlying genetic or extrinsic damage (*14*). For rodents, the original pathology nomenclature, DIS-ROD, was replaced by the formalism of Mouse Pathology (MPATH) and Mouse Anatomy (MA). A pathology panel checked a cross section of histopathological slide material and compared the original diagnoses with currently accepted diagnostic criteria.

RESULTS

To make the legacy system accessible for a broad community as well as to enable interoperability with other databases, it has been converted to a relational database with an intuitive user interface. ERA can be accessed at no cost at https://era.bfs.de. An ID and password can be requested from the curators at era@bfs.de.

Data Quality and Literature

Validation of the existing legacy database was performed by re-entering randomly selected data sets from four different studies, comprising more than 4,000 individual records. There was a 2% rate of non-systematic errors. This rate was low compared to the *a priori* threshold level of 5%. If the rate of non-systematic errors had been higher, the project would have been stopped. Systematic errors and systematically questionable information were corrected.

The very detailed comparison of the printed version (*5*) with the electronic version, as of December 1996, and with the beta version of ERA showed few errors; most errors were systematic phase shifts or keystroke errors that were corrected. In a few instances, inconsistent or questionable coding was detected. These problems could be fixed.

Whenever possible, references to each study have been linked to original articles, reviews, etc. by the PubMed ID or to the available PDF file. Overall, ERA includes 734 citations: 368 are linked to PubMed, 134 are linked

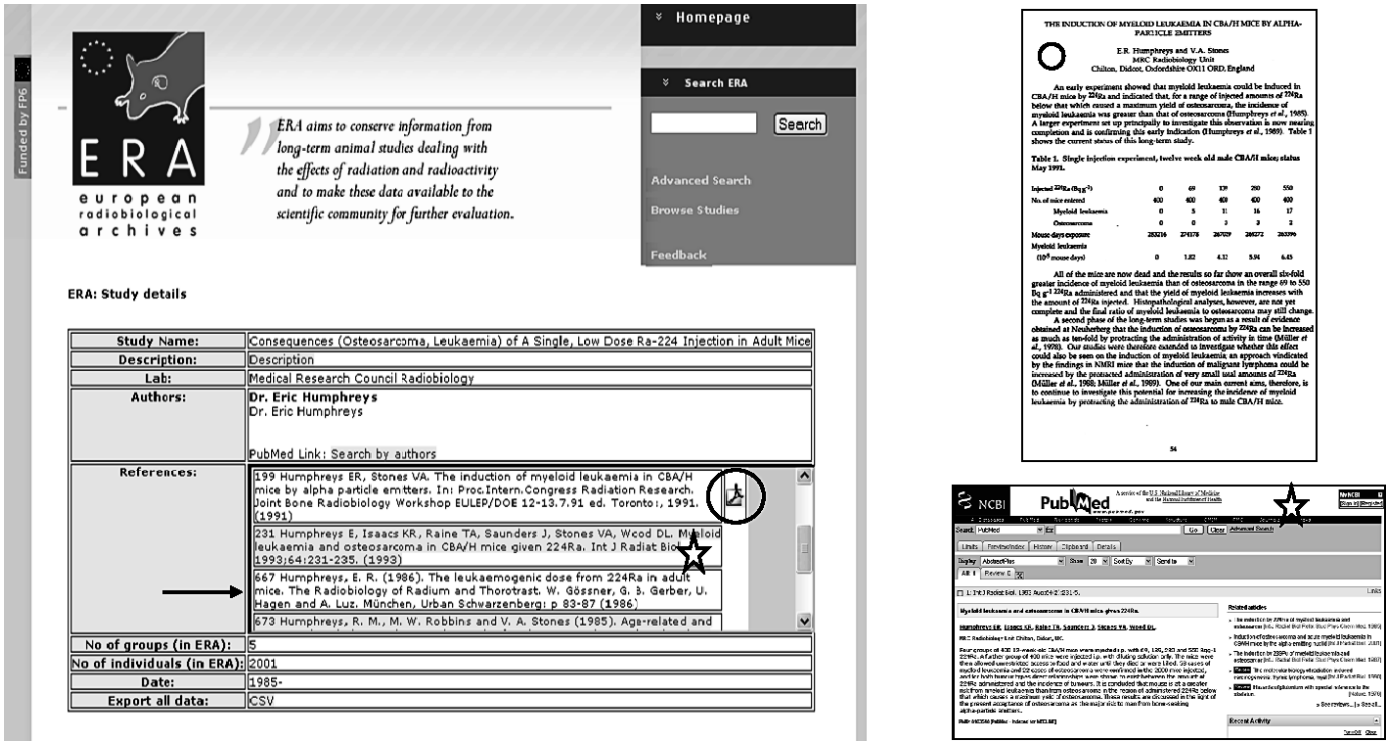

FIG. 1. References linked to a PDF file (circle), to PubMed (the whole citation is highlighted, marked with a star), or without any linkage (marked with an arrow).

to an available PDF file, and 8 are linked both to PubMed and to a PDF file. A further 14 references are available at BfS on request. The respective information is given in ERA.

Figure 1 shows an example of references linked to a PDF file (circle), to PubMed (the whole citation is highlighted, marked with a star), or without any linkage (marked with an arrow).

Standardization of the Pathology Nomenclature

ERA now uses the relevant ontology terms – Mouse Anatomy (MA) (*15*) and Mouse Pathology (MPATH) (*16, 17*). Now that the legacy database schema and the data have been successfully transformed into the new versions, disaggregated data may be searched using MA and MPATH through look-up tables (part of the database back-end), but the legacy coding is still available and can also be used. Queries are performed through a set of hypertext pre-processor (PHP) scripts. The accuracy of the primary data input was assessed and improved. The original rodent pathology nomenclature was recoded to replace the local DIS-ROD formalism with Mouse Pathology (MPATH) and Mouse Anatomy (MA) ontology terms. The pathology nomenclature was standardized. Human pathology was originally coded using either the ICD-8 or KDS. Both terms were mapped onto ICD-9. The NRA beagle data were captured using the Systematized Nomenclature of Dog (SNODOG).

The Content of ERA

ERA includes 151 studies from 21 laboratories in Europe, 143 studies from 11 laboratories covered by the National Radiobiology Archives (NRA), and 40 studies from 14 laboratories covered by the Japanese Radiobiology Archives (JRA). For each of the studies the following information is given: (a) purpose of the study; (b) laboratory and investigators; (c) type of treatment such as radiation, chemical or biological agents; (d) dosimetry; (e) end points such as the survival rate or pathological and molecular biological characteristics; (f) species, strain, sex and age of the animals used; (g) summary of the results; and (h) selected publications. Overall, ERA includes information on about 460,000 subjects on different levels. Table 1 describes the level of information and its source.

Further, a collection of zoomable histopathological slides from some of the experiments is included. Last but not least, links have been made to other relevant databases such as Pathbase (*17*) and JANUS (*18*). In the latter case it is important to note that the JANUS studies are dynamically linked to the studies in ERA using web services, integrating the two databases.

The User Interface

ERA has become a standardized, readily searchable database that provides a quick and easy way to find information successfully. The search is simplified by

TABLE 1
Content of ERA

		Europe	U.S.	Japan	Total
Laboratories		21	11	14	46
Studies		151	143	40	334
Groups		4,623	1,861	367	6,851
Individuals	Total	241,077	190,471	31.976	463,524
Animals and humans[a]	With data	97,719	115,801	5,835	219,355
Humans[a] **only**	Total	8,490	0	2,439	10,929
	With data	4,274	0	2,439	6,713
Organisms		mice, rats, dogs (beagles), cats, monkeys, guinea pigs, Chinese hamsters, pigs, rabbits			
Exposure	Radiations	*photons* (X rays <1 MeV; γ rays) *neutrons* [fission neutrons, other neutrons (accelerator): low-medium energy, high energy, neutrons from ^{252}Cf] *α particles, β particles*			
	Radionuclides	e.g.: ^{90}Sr injected, ^{238}Pu inhaled or injected, ^{239}Pu inhaled, injected or oral			
	Other substances	protective or biological agents; food; chemical substances (oral inhaled or injected)			

[a] Human data are included in the database but due to ethical and data protection issues not yet publicly available.

semantically prescribed vocabularies and ontologies [organism, treatment, different IDs (MA, MPath, DisRod, DisFam etc.)]. The database reports the results visually as a miniature document.

The user has three possible ways (quick search, browse studies, advanced search) to perform a query and receive the desired results. A quick search is performed by typing a term into the search box of the homepage. This box searches only study names and descriptions and not the individual or group data. It is possible to run either an AND or an OR query when combining different terms. When browsing the studies, the user passes successively through the entire data sets of all laboratories with the respective studies. Browsing can be useful for becoming familiar with the content of ERA.

With an advanced search, various parameters can be defined, e.g. organism, pathology, exposure (see Fig. 2). A series of look-up tables and links will allow the user to identify the experiments that include data on the search terms as well as to extract individual animal data that meet the search criteria. If more than one search parameter is set, the query will produce results that are the intersection of the selected search parameters. Results are given in a tabular format (Fig. 3a). Studies are commonly arranged in rows. Selecting a study will give the study details (Fig. 3b), including the study groups. Selecting a group will show the individual study subjects (Fig. 3c).

Data can be exported to CSV files and downloaded for further analyses by clicking on the CSV button.

The digital high-resolution pathological slides (whole-slide images; WSI) can be browsed either by using the ''high resolution images'' link under the heading ''Resources'' (Fig. 4a) or by using the links given in the study description (Fig. 4b).

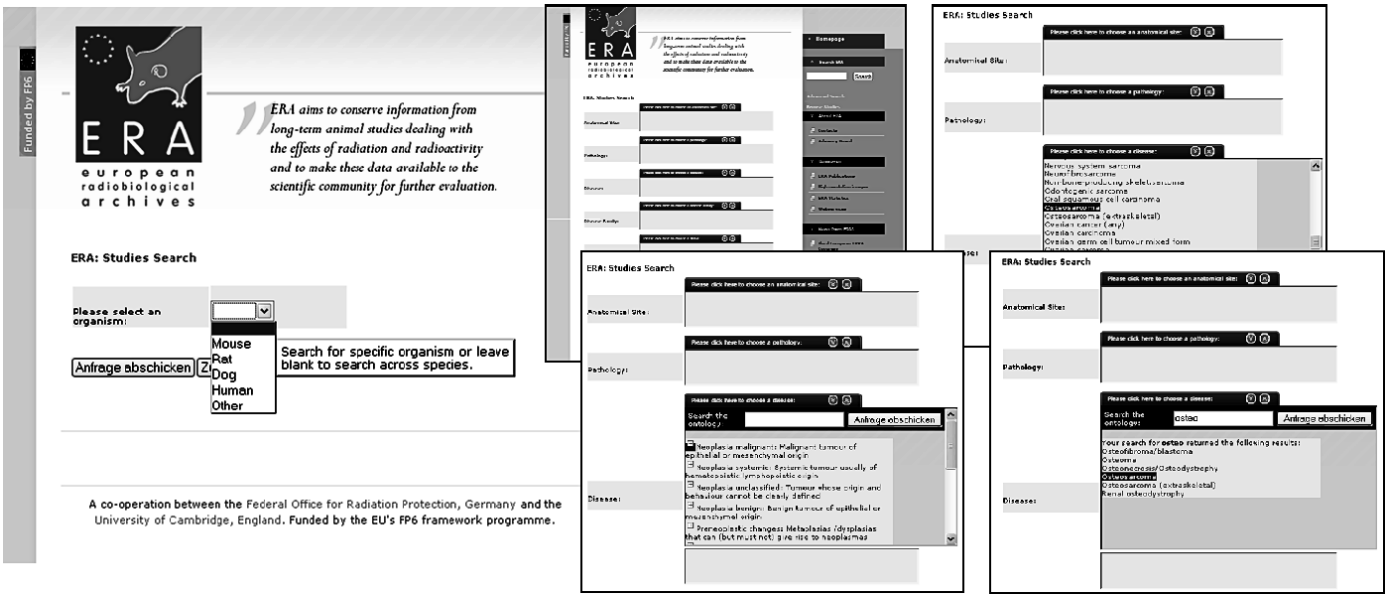

FIG. 2. Screenshots: how to perform a search.

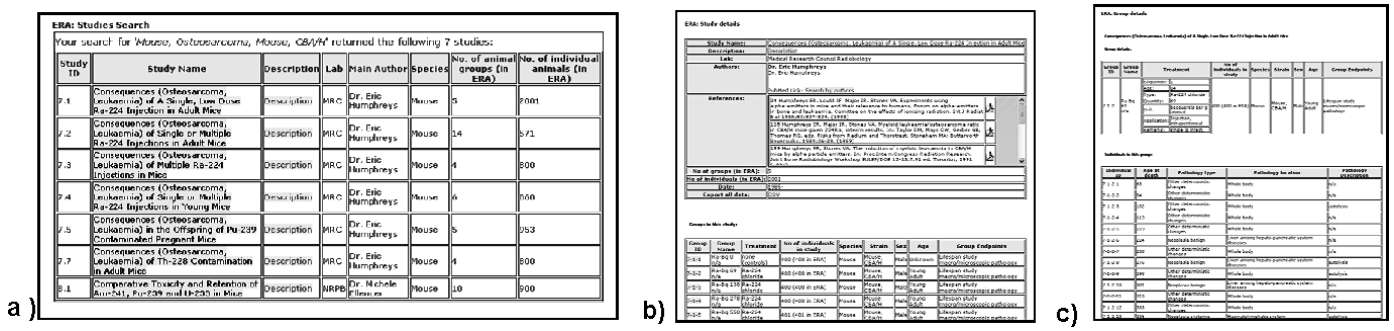

FIG. 3. Screenshots of (panel a) results given in a tabular format after performing an advanced search (panel b) of the page that give study details of a selected study (panel c) of the group details also given in a tabular format.

DISCUSSION

Molecular and genetic research is now providing us with the opportunity to quantify the risks of radiation exposure at the individual level. The retrospective analysis of earlier epidemiological and animal studies will be an important resource for modeling and evaluating such new risk parameters. With great foresight the EU and EULEP had created a database collecting and collating data from almost all of the available animal radiation biology studies carried out in Europe, the U.S. and Japan between 1960 and 1998 and data from two human cohort studies. This database is called the European Radiobiology Archives (ERA). Based on this legacy database we conducted quality assessment measures, standardized the pathology nomenclature, updated ERA's content, and made nearly all data available online. All steps were successful, and we are certain that ERA will become an important tool for the scientific community. ERA acts as a repository for historical primary data from unique and unrepeatable experiments. It allows the access to original data that are not available in the publications that arose from the studies.

All possible efforts were taken to eliminate errors and any inconsistencies that might have been present in the legacy database, even though this was found to be a problem in only a very low percentage of the data. It is important, however, to note that as with all community databases, ERA cannot accept any responsibility or liability for the correctness of the data and can only vouch for the accuracy of transfer from the original sources to the database. As with publications, the correctness of the original data is the responsibility of the various investigators who conducted the studies. Mechanisms are available for user feedback from the ERA homepage if problems are found in the data or their presentation.

The data contained in ERA include a few human cohort studies. The access to these is currently restricted and will only be made accessible once ethical and data

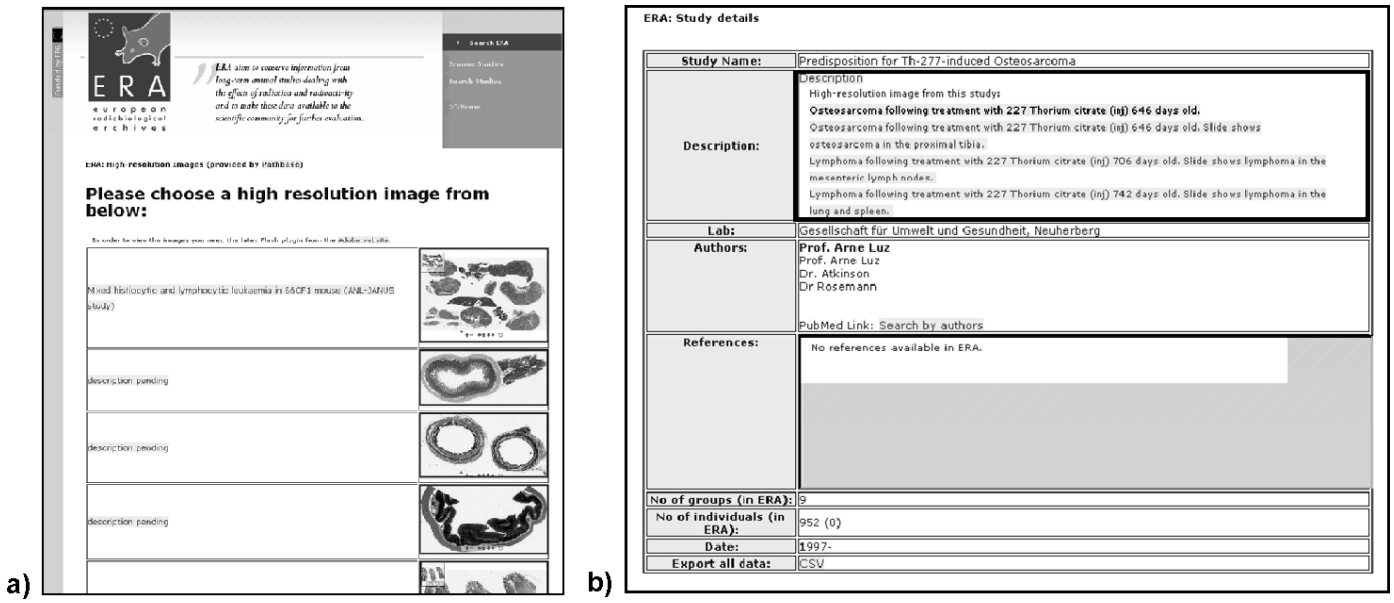

FIG. 4. Panel a: Table of digitized printed pathological slides (browsable high-resolution images). Panel b: High-resolution images that are available are referred under the description in the study details table.

protection issues have been settled. Further, for a few studies we have received original data sets in ASCII code. These data sets will be added to ERA to allow the user to work with both data sets.

In conclusion, ERA has the potential to become an important tool for the worldwide radiation research community, and the authors are committed to its constant improvement. Close collaboration with the relevant Japanese groups as well as the American National Archives will be continued.

ACKNOWLEDGMENTS

This work was supported by the European Commission under Contract FI6P 28275. We would like to thank all persons and experts who have contributed to the compilation, validation and review of the data included in the database. The authors are grateful for the support they received from members of the Advisory Board and the Pathology Panel. Advisory Board: Mike Atkinson (chair), Harmen Bijwaard, Leticia Quintanilla-Martinez, Georg Gerber, Shin Saigusa, Soile Tapio, Chuck Watson. Pathology Panel: Jonathan Bard, Pierre Dubus, Irene Esposito, Leticia Quintanilla-Martinez, Laurence Fiette, Björn Rozell, Ji-Ying Song, John Sundberg, Martin van der Valk, Madhuri Warren. Further we thank Gayle Woloschak and Ben Haley from Northwestern University in Chicago for the linkage of the JANUS data to ERA.

Received: October 14, 2010; accepted: November 15, 2010; published online: January 25, 2011

REFERENCES

1. A. L. Brooks, Developing a scientific basis for radiation risk estimates: goal of the DOE Low Dose Research Program. *Health Phys*. **85**, 85–93 (2003).
2. K. Sankaranarayanan and J. S. Wassom, Ionizing radiation and genetic risks XIV. Potential research directions in the post-genome era based on knowledge of repair of radiation-induced DNA double-strand breaks in mammalian somatic cells and the origin of deletions associated with human genomic disorders. *Mutat. Res*. **578**, 333–370 (2005).
3. S. D. Bouffler, B. A. Bridges, D. N. Cooper, Y. Dubrova, T. J. McMillan, J. Thacker, E. G. Wright and R. Waters, Assessing radiation-associated mutational risk to the germline: repetitive DNA sequences as mutational targets and biomarkers. *Radiat. Res*. **165**, 249–268 (2006).
4. H. Bijwaard, Mechanistic models of bone cancer induction by radium and plutonium in animals compared to humans. *Radiat. Prot. Dosimetry* **122**, 340–344 (2006).
5. G. B. Gerber, C. R. Watson, T. Sugahara and S. Okada, *International Radiobiology Archives of Long-term Animal Studies I. Description of Participating Institutions and Studies*. Report DOE/RL-96-72, U.S. Department of Energy, Richland, WA, 1996.
6. G. B. Gerber, R. R. Wick, C. R Watson, W. Gössner and A. M. Kellerer, International radiobiology archives of long-term animal studies: structure, possible uses and potential extension. *Radiat. Environ. Biophys*. **38**, 75–79 (1999).
7. G. B. Gerber and R. R. Wick, The European Radiobiology Archives ERA, present state and future developments. *Radiat. Prot. Dosimetry* **112**, 529–530 (2004).
8. G. B. Gerber, R. R. Wick, A. M. Kellerer, J. W. Hopewell, V. Di Majo, N. Dudoignon, W. Gössner and J. Stather, The European Radiobiology Archives ERA: content, structure and use illustrated by an example. *Radiat. Prot. Dosimetry* **118**, 70–77 (2006).
9. S. Tapio, P. N. Schofield, C. Adelmann, M. J. Atkinson, J. L Bard, H. Bijwaard, M. Birschwilks, P. Dubus, L. Fiette, G. Gerber, M. Gruenberger, L. Quintanilla-Martinez, B. Rozell, S. Saigusa, M. Warren, C. R. Watson and B. Grosche, Progress in updating the European Radiobiology Archives. *Int. J. Radiat. Biol*. **84**, 930–936 (2008).
10. WHO, *Manual for International Classification of Diseases and Health Related Problems*, 9th Revision. WHO, Geneva, 1992.
11. H. Immich, *Klinischer Diagnosenschlüssel*. Schattauer, Stuttgart, 1968.
12. J. L. Palotay, SNOMED-SNOVET: an information system for comparative medicine. *Med. Inform. Lond*. **8**, 17–21 (1983).
13. P. Bannasch, *Pathology of Neoplasia and Preneoplasia in Rodents*. Schattauer, Stuttgart, New York, 1994.
14. P. N. Schofield, J. B. Bard, C. Booth, J. Boniver, V. Covelli, P. Delvenne, M. Ellender, W. Engstrom, W. Goessner and A. Ward, Pathbase: a database of mutant mouse pathology. *Nucleic Acids Res*. **32**, D512-5 (2004).
15. T. F. Hayamizu, M. Mangan, J. P. Corradi, J. A. Kadin and M. Ringwald, The Adult Mouse Anatomical Dictionary: a tool for annotating and integrating data. *Genome Biol*. **6**, R29 (2005).
16. P. N. Schofield, M. Gruenberger and J. Sundberg, Pathbase and the MPATH ontology: community resources for mouse histopathology. *Vet. Pathol*. **47**, 1016–1020 (2010).
17. P. N. Schofield, J. B. Bard, J. Boniver, V. Covelli, P. Delvenne, M. Ellender, W. Engstrom, W. Goessner, M. Gruenberger and A. Ward, Pathbase: a new reference resource and database for laboratory mouse pathology. *Radiat. Prot. Dosimetry* **112**, 525–528 (2004).
18. Q. Wang, T. Paunesku and G. Woloschak, Tissue and data archives from irradiation experiments conducted at Argonne National Laboratories over a period of four decades. *Radiat. Environ. Biophys*. **49**, 317–324 (2010).