1. PREFACE

Dear Sir/Madam,

In the course of their work, researchers routinely create, use and re-use digital information. The results of such research are usually published (e.g. journal articles) and preserved for future generations by publishers and/or libraries. But what happens to the underlying research data which is vital for verifying the research outcome, or which might prove of great value to other researchers in other contexts? Digital data are much more fragile than printed resources, as hardware and software platforms evolve and required background knowledge becomes unavailable. Which present-day computer can deal with an 8-inch floppy disk with a document in WordStar on it, or a tape with a raw satellite image in VICAR format?

The European Alliance for Permanent Access aims to work together to preserve research data for future re-use. Science and research publishers actively support and partake in the Alliance through the participation of the International Association of Scientific Technical and Medical publishers (Int STM Assocation). A number of Alliance partners have initiated a project, partly funded by the EU, entitled PARSE.insight which aims to gain insight into the requirements of researchers, data managers and funding agencies with regard to preserving research data.

PARSE.insight will help to shape the EU's preservation infrastructure, as part of the broader e-Research infrastructure, so that European citizens can continue to gain benefits from its digital resources, in all aspects of life.

In order to complete our picture of what the research community needs, we kindly request that you fill out this questionnaire, which may reach you via other members of the Alliance. It will take you about **20 minutes**. The research community as a whole will benefit from your cooperation!
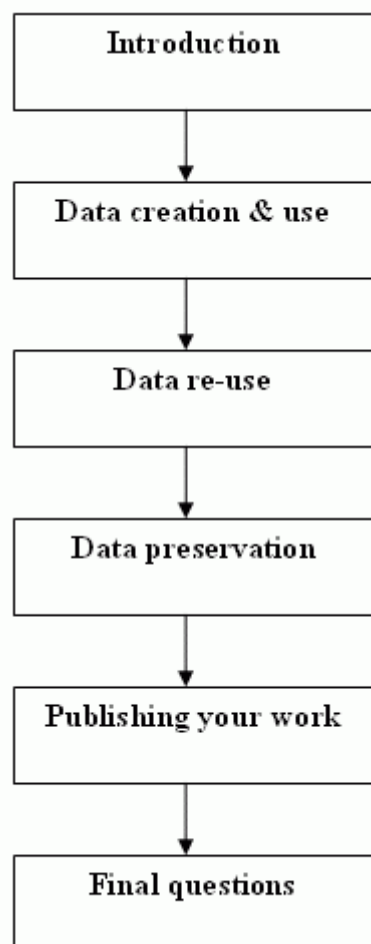
A word about terminology: in this questionnaire we make a distinction between **storing information**, routinely, in your day-to-day practice, on your computer or a faculty server, and **preserving information**, meaning data is specifically curated to be re-usable in the long term. In the latter case not only the data itself must be archived, but also data about the data: where did the data come from? How have they been stored? Which file formats have been used? What special terminology or other information is needed to interpret and use the data? etc.

Reponses are aggregated for analysis and made anonymous. However, if you're willing to take part in a more detailed interview which will further help shape Europe's preservation infrastructure, or if you wish to be informed about the results, you should enter your e-mail address at the end of the survey.

Thank you,

PARSE.insight team
Alliance for Permanent Access
Int STM Association

Flow of survey

```
┌─────────────────────────┐
│      Introduction       │
└─────────────────────────┘
             │
             ▼
┌─────────────────────────┐
│   Data creation & use   │
└─────────────────────────┘
             │
             ▼
┌─────────────────────────┐
│      Data re-use        │
└─────────────────────────┘
             │
             ▼
┌─────────────────────────┐
│   Data preservation     │
└─────────────────────────┘
             │
             ▼
┌─────────────────────────┐
│  Publishing your work   │
└─────────────────────────┘
             │
             ▼
┌─────────────────────────┐
│     Final questions     │
└─────────────────────────┘
```

2. Introduction

1. Name of organisation (e.g. name of research institute, company, or university):

2. Country (of your organisation):

3. Which is your primary research discipline?

Other (please specify)

4. How many years have you been involved in research?

- < 5 years
- 5 - 10 years
- 10 - 20 years
- > 20 years
- I have never been involved in research

5. What is your current role in research? (multiple answers possible)

- ⊖ I am a researcher

- ⊖ I am a lecturer with some research tasks

- ⊖ I am a research group leader or manager

- ⊖ I am a research director

- ⊖ Other (please specify)

4. Creation and use of digital research data

While doing research you create new research data either by performing experiments/making observations or by using analytical models. In addition, you may use existing digital research data. The following questions address the creation and use of digital research data.

6. Please indicate which of the following digital research data you use: (multiple answers possible)

- Standard office documents (text documents, spreadsheets, presentations)

- Network-based data (web sites, e-mail, chat history, etc.)

- Databases (DBASE, MS Access, Oracle, MySQL, etc.)

- Images (JPEG, JPEG2000, GIF, TIF, PNG, SVG, etc.)

- Structured graphics (CAD, CAM, 3D, VRML, etc.)

- Audiovisual (multimedia) data (WAVE, MP3, MP4, Flash, etc.)

- Scientific and statistical data formats (SPSS, FITS, GIS, etc.)

- Raw data (device specific output)

- Plain text (TXT in various encodings)

- Structured text (XML, SGML, etc.)

- Archived data (ZIP, RAR, JAR, etc.)

- Software applications (modelling tools, editors, IDE, compilers, etc.)

- Source code (scripting, Java, C, C++, Fortran, etc.)

- Configuration data (parameter settings, logs, library files)

- Other, please specify below.

5. Creation and use of digital research data (cont.)

7. Do you have non-digital research-related material (e.g. survey forms, laboratory notebooks, design models), which you intend to digitise in the near future?

　j-n　Yes

　j-n　No

8. Please provide us with an estimate of the volume of stored digital research data of your current research project (excluding back ups) as well an estimate of its volume in 2 and 5 years.

	0MB	1-100MB	100MB-1GB	1GB-1TB	1TB-1PB	1PB-10PB	>10PB	Don't know
Current	jᴄ	jᴄ	jᴄ	jᴄ	jᴄ	jᴄ	jᴄ	jᴄ
In 2 Years	jᴄ	jᴄ	jᴄ	jᴄ	jᴄ	jᴄ	jᴄ	jᴄ
In 5 Years	jᴄ	jᴄ	jᴄ	jᴄ	jᴄ	jᴄ	jᴄ	jᴄ

9. Do you as a rule assign any additional information to your digital research data? (multiple answers possible)

- Administrative information (e.g. creator, date of creation, filename, provenance)

- Technical information (e.g. encoding type, description, file format, settings, software utilities)

- No

- Other (please specify)

7. Data re-use

The following questions address the re-use of digital research data.

10. Regarding your current research, do you collaborate with researchers from other projects?

- ◯ Yes, with researchers from other projects in my discipline
- ◯ Yes, with researchers from other projects outside my discipline
- ◯ Yes, both within my discipline and cross-disciplinary
- ◯ No

11. Which of the following applies to the digital research data of your current research: (multiple answers possible)

- ◌ My data is openly available for everyone.
- ◌ My data is openly available for my research discipline.
- ◌ My data is openly available for my research group / colleagues in research collaboration.
- ◌ Access to my data is temporarily restricted.
- ◌ My data is available for a fee.
- ◌ My data could be made available with appropriate changes (e.g. anonymous clinical data)
- ◌ I do not share my data, but I would like to do so in the future.
- ◌ I do not share my data and I do not want to share it in the future.

12. Increasingly, awareness is growing that data should be shared as well as publications. Do you experience or foresee any of the following problems in sharing your data? (multiple answers possible)

- ◌ Fear to lose scientific edge
- ◌ Incompatible data types
- ◌ Restricted access to data archive
- ◌ Legal issues
- ◌ Lack of technical infrastructure
- ◌ Misuse of data
- ◌ Lack of financial resources
- ◌ No problems foreseen
- ◌ Other (please specify)

[   ]

13. Do you presently make use of research data gathered by other researchers in your discipline?

    ○ Yes

    ○ No

8. Data re-use (cont.)

14. Would you like to make use of research data gathered by other researchers in your discipline?

- ⊙ Yes

- ⊙ No

9. Data re-use (cont.)

15. Do you presently make use of research data gathered by other researchers in OTHER disciplines?

jn   Yes

jn   No

16. Would you like to make use of research data gathered by researchers from OTHER disciplines?

jn Yes

jn No

17. How do you locate and access digital research data? (multiple answers possible)

- ⊝ Via colleagues

- ⊝ Via institutional database and search facilities

- ⊝ Via general search engines (Google, Yahoo, etc.)

- ⊝ Via data centers or archives (World Data Center, DANS, National Archives, etc.)

- ⊝ Via formal literature (articles and book chapters)

- ⊝ N/A (I do not use data from others)

- ⊝ Other (please specify)

```
5


6
```

18. Did you ever need digital research data gathered by other researchers that was not available?

- ◌ Yes

- ◌ No

- ◌ Don't know

19. If your digital research data is unusable, how easily can you recreate it?

- ◌ Very easily

- ◌ With same effort as initially created

- ◌ Hardly

- ◌ Not possible

- ◌ Don't know

20. Do you know of digital research data that has become unusable within your organisation?

- ◌ No

- ◌ Yes, please specify below

Unusable digital research data:

```
5




6
```

12. Data re-use (cont.)

21. Why did the data become unusable? (multiple answers possible)

- Hardware to interpret the data is no longer available.

- Software to interpret the data is not longer available.

- Lack of contextual information (e.g. manuals, notes, experience).

- Don't know

- Other (please specify)

5
6

13. Data preservation

After you have been using your digital data during your research, you probably want to preserve the data for a certain period of time. As mentioned in the introduction, preserving information means data is specifically curated and maintained for re-use in the future. Not only the data itself must be archived, but also data about the data: where did the data come from? How has it been stored? How to use it? etc.

The following questions address the preservation of digital research data.

22. Please indicate how important you think the following reasons for preserving digital data are:

	Very important	Important	Slightly important	Not important
If research is publicly funded, the results should become public property and therefore properly preserved.	j୬	j୬	j୬	j୬
It will stimulate the advancement of science (new research can build on existing knowledge).	j୬	j୬	j୬	j୬
It may serve validation purposes in the future.	j୬	j୬	j୬	j୬
It allows for re-analysis of existing data.	j୬	j୬	j୬	j୬
It may stimulate inter-disciplinary collaborations.	j୬	j୬	j୬	j୬
It potentially has economic value.	j୬	j୬	j୬	j୬
It is unique.	j୬	j୬	j୬	j୬

23. How important do you regard the following threats over the next 10 years?

	Very Important	Important	Slightly Important	Not Important	Don't Know
Users may be unable to understand or use the data e.g. the semantics, format or algorithms involved	jn	jn	jn	jn	jn
Lack of sustainable hardware, software or support of computer environment may make the information inaccessible	jn	jn	jn	jn	jn
Evidence may be lost because the origin and authenticity of the data may be uncertain	jn	jn	jn	jn	jn
Access and use restrictions (e.g. Digital Rights Management) may not be respected in the future	jn	jn	jn	jn	jn
Loss of ability to identify the location of data	jn	jn	jn	jn	jn
The current custodian of the data, whether an organisation or project, may cease to exist at some point in the future	jn	jn	jn	jn	jn
The ones we trust to look after the digital holdings may let us down	jn	jn	jn	jn	jn

24. Do you think that an international infrastructure for data preservation and access should be built to help guard against some of these threats?

jn Yes

jn No

jn Don't Know

25. What do you think this infrastructure should look like?

```
                              5

                              6
```

15. Data preservation (cont.)

26. Apart from an infrastructure, what do you think is needed to guarantee that valuable digital research data is preserved for access and use in the future? (multiple answers possible)

- ⊖ Training
- ⊖ More knowledge/expertise
- ⊖ More resources (financial and operational)
- ⊖ More digital repositories/archives
- ⊖ Nothing else
- ⊖ Other (please specify)

[   ]

27. Which of the following standards or guidelines that are used in digital preservation are you familiar with? (multiple answers possible)

- ⊖ OAIS
- ⊖ Dublin Core
- ⊖ PREMIS
- ⊖ OAI-PMH
- ⊖ OAI-ORE
- ⊖ METS
- ⊖ MPEG21-DIDL
- ⊖ NISO
- ⊖ None of the above
- ⊖ Other (please specify)

[   ]

28. How do you presently store your digital research data for future access and use, if at all? (multiple answers possible)

- On my computer at work (local directory on computer)

- Organisational server (departmental or organisational directory)

- On my computer at home (local directory on computer)

- Portable storage carrier (e.g. hard disk, CD/DVD, Floppy, USB)

- I submit my data to the digital archive (repository) of my organisation

- I submit my data to the digital archive (data centre) of my discipline, namely (please specify in the text box below):

- I submit my data to the journal when I submit the manuscript

- I submit my data to an external web service (e.g. Google Base, Mozy), namely (please specify in the text box below):

- I do not store digital research data.

- Other (please specify in the text box below)

.

29. To which of the following facilities would you be willing to submit digital research data in the near future? (multiple answers possible)

- Digital archive of organisation

- Digital archive (data centre) of your discipline

- Publisher (data combined enclosed with scientific paper).

- External web service (e.g. Google Base, Mozy)

- I do not want to submit digital research data to external facilities

- Other (please specify)

16. Data preservation (cont.)

30. When submitting data to an external facility, are you required to:

	Yes	No	Don't Know
comply to a standard data format?	j	j	j
transfer copyrights?	j	j	j
supply additional information such as manuals, software?	j	j	j

If applicable, please specify standard data format:

31. If you do not want to submit data to an external facility, why not? (multiple answers possible)

- ∈ I do not know of any digital archives (repositories or data centres) to which I can submit data.

- ∈ I do not believe my digital research data is secure at a data centre, journal site or repository.

- ∈ I am not aware of any submission procedures.

- ∈ I do not want to run the risk of anyone else being able to access and use my digital research data.

- ∈ The submission procedures are too complicated and therefore it takes too much time to submit data.

- ∈ Submitting digital research data costs money and therefore is not attractive to me.

- ∈ I am afraid my data will be misused (wrong interpretation, biased statements, etc.)

- ∈ I am afraid I will lose my copyrights.

- ∈ Don't know.

- ∈ Other (please specify)

18. Data preservation (cont.)

32. Is there a preservation facility for preserving digital research data which can be used by all projects within your discipline?

- ⃝ No

- ⃝ Don't know

- ⃝ Yes (please specify)

  ```
                                                      5
                                                      6
  ```

33. Do you know whether there are plans within your research community to build a digital archive (repository or data centre) for long-term storage and access in the nearby future?

- ⃝ Yes, in less than a year

- ⃝ Yes, in 1 to 3 years

- ⃝ Yes, in 3 to 5 years

- ⃝ Yes, but it will take more than 5 years

- ⃝ Don't know

34. Does your organisation have

	Yes	No	Don't Know
a policy or procedures regarding the preservation of digital research data?	◌	◌	◌
guidelines or recommendations regarding the preservation of digital research data?	◌	◌	◌

35. Which of the following organisations provide guidelines or recommendations with regard to the preservation of digital research data in your discipline? (multiple answers possible)

- Learned societies
- Research associations
- Research councils
- Publishers (journals)
- Funding organisation
- Government
- None of the above
- Don't know

36. Are you legally required to preserve digital research data?

- Yes
- No
- Don't Know

37. Does your funding organisation require you to preserve digital research data?

- Yes
- No
- Don't Know

20. Data preservation (cont.)

38. Does your funding organisation provide any mandatory procedures for managing and preserving digital research data?

jn Yes

jn No

jn Don't Know

21. Data preservation (cont.)

39. What do you see as the main threats to your currently preserved digital storage data? (multiple answers possible)

- ⊝ Lack of structural funding
- ⊝ Lack of technical support
- ⊝ Natural disasters
- ⊝ Political instability
- ⊝ Continuity of organisation
- ⊝ Human errors
- ⊝ Don't know
- ⊝ Other (please specify)

40. Who, in your opinion, should pay for preservation of digital research data? (multiple answers possible)

- ⊝ My organisation
- ⊝ Research community (researchers collectively)
- ⊝ Government (national funding)
- ⊝ European Union (European funding)
- ⊝ Commercial organisation (profit data centre)
- ⊝ Don't know
- ⊝ Other (please specify)

41. Who, in your opinion, should pay for the preservation of publications? (multiple answers possible)

- ⊝ My organisation
- ⊝ Research community (researchers collectively)
- ⊝ Government (national funding)
- ⊝ European Union (European funding)
- ⊝ Commercial organisation (publishers, etc.)
- ⊝ Don't know
- ⊝ Other (please specify)

22. Publishing your work

When you finish your research project, you would like to present the results to your community and beyond. The following questions address the way you publish your results.

42. How do you publish your research results usually? (multiple answers possible)

- Books
- Journal of publisher (available by subscription)
- Open access journal
- Grey literature (e.g. technical reports, white papers, preprints)
- Conference proceedings
- Websites (e.g. Research Blogging)
- Other (please specify)

43. Do you think it is useful to link underlying digital research data with formal literature?

- Yes
- No

44. Do you want to be credited when your underlying digital research data is used by others?

- Yes
- No

45. Do journals to which you typically submit your work require you to include relevant digital research data (i.e. data used to create tables, figures, etc.)?

- Yes
- No

46. Online research collaboratories are virtual collaborations between researchers for sharing research data and new insights, or for debating about hot topics in science. Do you make use of an online research collaboratory?

- Yes
- No

23. Publishing your work (cont.)

47. Do you think that the information generated by these online collaboratories should be preserved (e.g. chat logs, wiki's)?

j⊓ Yes

j⊓ No

24. Final questions

48. Do you think the following initiatives would be useful for raising the level of knowledge about preservation of digital research data?

	Very useful	Useful	Slightly useful	Not useful	Don't know
Workshops on preservation of digital material	j⌒	j⌒	j⌒	j⌒	j⌒
User-oriented training sessions on digital preservation	j⌒	j⌒	j⌒	j⌒	j⌒
International knowledge platform/forum on digital preservation	j⌒	j⌒	j⌒	j⌒	j⌒
Development of guidelines/manuals on how to preserve digital data	j⌒	j⌒	j⌒	j⌒	j⌒

49. Would you be prepared to be interviewed by telephone after completion of this survey in order to help to shape the European preservation infrastructure?

j⌒　No

j⌒　Yes. Please provide your email address.

50. May we share the data collected in this survey with closely related projects in the field of digital preservation?

j⌒　No

j⌒　Yes

51. Are there others to whom we should send this questionnaire?

j⌒　No

j⌒　Yes. Please provide the email address of a person to whom we may send this questionnaire.

52. Do you want to be informed about the results of this survey?

j⌒　No

j⌒　Yes

Please provide your email address, unless provided above:

25. Thank you!

Thank you very much for answering our questions. With your answers we are able to outline the current state of preservation and access of digital data in science.

If you entered your email address we will inform you about the final results of this project. In any case, you can visit our website www.parse-insight.eu.

Please click 'done' to finish this survey.