Proposed AMENDMENTS for the

Proposal for the revision of the Directive 2003/98/EC
on the reuse of public sector information

9 October 2018

The following representatives of the library and research communities support these proposed amendments: EBLIDA, DCC, IFLA, and SPARC Europe and urge for their consideration. Please note that additions are in bold, and deletions are struck through.

Recital 23

Text proposed by the Commission	*Amendment*
"The volume of research data generated is growing exponentially and has potential for re-use beyond the scientific community. In order to be able to address mounting societal challenges efficiently and in a holistic manner, it has become crucial and urgent to be able to access, blend and re-use data from different sources, as well as across sectors and disciplines. Research data includes statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. It also includes meta-data, specifications and other digital objects. Research data is different from scientific articles reporting and commenting on findings resulting from their scientific research. For many years, the open availability and re-usability of scientific research results stemming from public funding has been subject to specific policy initiatives. Open access policies aim in particular to provide researchers and the public at large with access to research data as early as possible in the dissemination process and to enable its use and re-use. Open access helps enhance quality, reduce the need for unnecessary duplication of research, speed up scientific progress, combat scientific fraud, and it can overall favour economic growth and innovation. Beside open access, data management planning is swiftly becoming a standard scientific practice for ensuring data that is	"The volume of research data generated is growing exponentially and has potential for re-use beyond the scientific community. In order to be able to address mounting societal challenges efficiently and in a holistic manner, it has become crucial and urgent to be able to access, blend and re-use data from different sources, as well as across sectors and disciplines. Research data includes statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. It also includes meta-data, specifications and other digital objects. Research data is different from scientific articles reporting and commenting on findings resulting from their scientific research. For many years, the open availability and re-usability of scientific research results stemming from public funding has been subject to specific policy initiatives. Open access policies aim in particular to provide researchers and the public at large with access to research data as early as possible in the dissemination process and to enable its use and re-use. Open access helps enhance quality, reduce the need for unnecessary duplication of research, speed up scientific progress, combat scientific fraud, and it can overall favour economic growth and innovation. Beside open access, data management planning is swiftly becoming a standard scientific practice for ensuring data that is

findable, accessible, interoperable and re-usable (FAIR principles).

findable, accessible, interoperable and re-usable (FAIR principles), **as is the consistent provision of data management plans which must be further encouraged.**"

Justification

We strongly support the acknowledgement of the importance of the FAIR Data principles (findable, accessible, interoperable and reusable) for today's research management policy and practices in recital 23. FAIR is rightly specified in the context of open access to publicly funded research and relates to research data management needs and practices in recent years ensuring that outputs are not only made Open, where applicable, but more importantly that they are meaningful and reusable. To strengthen the implementation of this policy, and the FAIR principles when related to publicly-funded research, we suggest the inclusion of an obligation for a data management plan also to ensure longer term access to that data ahead of time.

Recital 24

Text proposed by the Commission	*Amendment*
"For the reasons explained above, it is appropriate to set an obligation on Member States to adopt open access policies with respect to publicly-funded research results and ensure that such policies are implemented by all research performing organisations and research funding organisations. Open access policies typically allow for a range of exceptions from making scientific research results openly available. On 17 July 2012, the Commission adopted a Recommendation on access to and preservation of scientific information, updated on 25 April 2018[1], and describing, among other things, relevant elements of open access policies. Additionally, the conditions, under which certain research results can be re-used, should be improved. For this reason, certain obligations stemming from this Directive should be extended to research data resulting from scientific research activities subsidised by public funding or co-funded by public and private-sector entities. However, in this context, concerns in relation to privacy, protection of personal data, trade secrets, national	"For the reasons explained above, it is appropriate to set an obligation on Member States to adopt open access policies with respect to publicly-funded research results and ensure that such policies are implemented by all research performing organisations and research funding organisations. Open access policies typically allow for a range of exceptions from making scientific research results openly available. On 17 July 2012, the Commission adopted a Recommendation on access to and preservation of scientific information, updated on 25 April 2018[2], and describing, among other things, relevant elements of open access policies. Additionally, the conditions, under which certain research results can be re-used, should be improved. For this reason, certain obligations stemming from this Directive should be extended to research data resulting from scientific research activities subsidised by public funding or co-funded by public and private-sector entities. **As a result, all wholly or majoritally publicly-funded research data should be made open as the default**

_____

[1]   C(2018)2375
[2]   C(2018)2375

security, legitimate commercial interests and to intellectual property rights of third parties should be duly taken into account. In order to avoid any administrative burden, such obligations should only apply to such research data that have already been made publicly available by researchers. Other types of documents held by research performing organisations and research funding organisations should continue to be exempt from the scope of application of this Directive."

option, and as closed as necessary as required by other bodies of law ~~However, in this context, concerns~~ subject ~~in relation~~ to privacy, protection of personal data, trade secrets, national security, legitimate commercial interests and to intellectual property rights of third parties ~~should be duly taken into account~~. **Data should also follow the FAIR data principles. It should be recalled that data - as 'mere facts' - are not subject to copyright under the Berne Convention. Where data is restricted, valid reasons for restrictions on access and/or re-use must be communicated to the public, where appropriate in the metadata, and that metadata itself should be openly available.** In order to avoid any administrative burden, such obligations should only apply to such research data that have already been made publicly available by researchers. Other types of documents held by research performing organisations and research funding organisations should continue to be exempt from the scope of application of this Directive."

Justification

Recital 24 specifies in what cases scientific research results may or may not made openly available. When research data results from publicly-funded research, it is responsible practice to make this data openly available by default. In addition, all data should adhere to the FAIR principles, irrespective of the mode of access, to enable reuse. We would therefore strongly encourage the introduction of a new principle here that specifies that all publicly-funded data is open by default, but restricted where the law or ethical considerations require it. This could help further ensure that more publicly-funded research data is made available as the standard way of practice, and more immediately.

This recital also expresses concerns related to restricting access to data with relation to sensitive and private data of various natures including privacy, personal data, trade secrets or IPR and specifies that these matters "should be duly taken into account" without specifying how, i.e. how these areas should be dealt with where publicly-funded research data is restricted. For example, clear explanations should be made public if data is restricted, and best efforts made to provide at least minimum data, i.e. metadata to make the discovery of data possible. This would also in effect also provide limited access to both digital and analogue data such as documents.
In summary, in cases where data cannot be made available for reasons of national security, related to privacy issues, etc. these reasons must be clearly expressed in machine-readable form.

Recital 28

Text proposed by the Commission	*Amendment*
"In order to get access to the data opened for re-use by this Directive, the use of suitable and well-designed Application Programming Interfaces (APIs) is needed. An API describes the kind of data can be retrieved, how to do this and the format in which the data will be received. It has different levels of complexity and can mean a simple link to a database to retrieve specific datasets, a web interface, or more complex set-ups. There is general value in re-using and sharing data via a suitable use of APIs as this will help developers and start-ups to create new services and products. It is also a crucial ingredient of creating valuable ecosystems around data assets that are often unused. The set-up and use of API needs to be based on several principles: stability, maintenance over lifecycle, uniformity of use and standards, user-friendliness as well as security. For dynamic data, meaning frequently updated data, often in real time, public sector bodies and public undertakings shall make this available for re-use immediately after collection by ways of suitable APIs."	"In order to get access to the data opened for re-use by this Directive, the use of suitable and well-designed Application Programming Interfaces (APIs) is needed. An API describes the kind of data **that** can be retrieved, how to do this and the format in which the data will be received. It has different levels of complexity and can mean a simple link to a database to retrieve specific datasets, a web interface, or more complex set-ups. There is general value in re-using and sharing data via a suitable use of APIs as this will help developers and start-ups to create new services and products. It is also a crucial ingredient of creating valuable ecosystems around data assets that are often unused. The set-up and use of APIs needs to be based on several principles: stability, maintenance over lifecycle, uniformity of use and standards, **openness,** user-friendliness as well as security. For dynamic data, meaning frequently updated data, often in real time, public sector bodies and public undertakings shall make this available for re-use immediately after collection by ways of suitable APIs. **APIs also need to be compatible with the FAIR principles, being self-descriptive and using open protocols**."

Justification

We strongly welcome the inclusion of reference to APIs to facilitate access to data in this Recital. However, we would urge that some guidelines be added to make the APIs compatible with FAIR data principles themselves, i.e. making APIs more easily interoperable for computers and humans. It is important to optimise API management to prevent today's plethora of different API approaches as seen in the ProgrammableWeb, an API registry that has over 20.000 APIs specifications. APIs should be self-descriptive and should use harmonised common elements such as protocols that are semantically enriched such as SmartAPI (https://smart-api.info).

Article 10 – paragraph 1
Availability and re-use of research data

Text proposed by the Commission	*Amendment*
"1. Member States shall support the availability of research data by adopting national policies and relevant actions aiming at making publicly funded research data openly available ('open access policies'). These open access policies shall be addressed to research performing organisations and research funding organisations."	1. Member States shall support the availability of research data by adopting national policies and relevant actions aiming at making publicly funded research data openly available ('open access policies'). **All wholly or majorily publicly-funded research data should be made open by default, with any restrictions subject to justification.** These open access policies shall be addressed to research performing organisations and research funding organisations.

Justification

Since publicly funded research is financed by the public purse, we believe that access to research data should be guaranteed to as many as soon as possible. We strongly believe that by making publicly-funded research openly available the default, this will stimulate more researchers to contribute and share their work for the benefit of research, industry and society. This will thereby help accelerate the pace of discovery, ensure that we do not miss scientific breakthroughs and improve the integrity of the scientific and scholarly record.

Article 10 – paragraph 2
Availability and re-use of research data

Text proposed by the Commission	*Amendment*
2. Research data shall be re-usable for commercial or non-commercial purposes under the conditions set out in Chapters III and IV, insofar as they are publicly funded and whenever access to such data is provided through an institutional or subject-based repository. In this context, legitimate commercial interests and pre-existing intellectual property rights shall be taken into account. This provision shall be without prejudice to point (c) of Article 1(2).	2. Research data shall be re-usable for commercial or non-commercial purposes under the conditions set out in Chapters III and IV, insofar as they are publicly funded and whenever access to such data is provided through institutional or subject-based repositories **or other local, national or international data infrastructure**. In this context, legitimate commercial interests, **including universities' knowledge transfer activities** and pre-existing intellectual property rights shall be taken into account. This provision shall be without prejudice to point (c) of Article 1(2).

Justification

Research data is stored in a range of manners. Various data formats such as XML, relational schemas, RDF, JSON and CSV are stored in file systems, database systems or triple stores. Non-digital data should also be made available and stored appropriately in existing data storage infrastructures. Normally, the decision on the choice of data storage infrastructure is based on community, institutional, national or international policies. In some cases researchers opt to store their data on an institutional repository, in other cases, national or domain-specific infrastructures are chosen. For instance, in the genetics field, it is common for genetics data to be stored at the European Genome-Phenome Archive (EGA). It is for this reason that the types of data storage media needs to be extended in Article 10.2.

Furthermore, Research Performing Organisations such as universities, supported by the European Commission and Member States, are increasingly engaging in knowledge transfer activities with the private sector. In such instances, jointly created or privately funded research, may require different rules to entirely publicly funded research data. For this reason we believe the final sentence of Article 10.2 is particularly important and should highlight specifically knowledge transfer activities.