EARTO Background Note:
Overview of US Federal Agencies Data Sharing Policies

5 December 2016

Introduction: US data sharing policy legislation at federal level

At a time when the policy debate in the EU has been very much focused on the open access to research data and as a follow-up to the EARTO paper on Open X, this background note aims at giving an overview of the US policy in this field, analysing the data sharing policies of the US Federal Agencies. Indeed, this could give an interesting perspective to the EU debate, also given the fact that Research and Innovation is now very globalised and that competition happens at international level.

One of the key conclusions of this analysis is that the wording "open data" is rarely used by US federal agencies. Instead they generally talk about "data sharing" policy and "access rights".

Two specific pieces of legislation at US federal level are worth mentioning before reviewing the different US Federal Agencies policies in terms of data sharing: the Small Business Act and the White House Order of February 2013.

- **Small Business Act**

In the US Small Business Innovation Research (SBIR) and Small Technology Transfer Research (STTR), US R&I programmes dedicated to SMEs for which $2.2 billion is set aside annually by US Federal agencies, data sharing is never an issue. In compliance with the Small Business Act of 1953 which created the Small Business Administration: it is forbidden to open the data of those projects before four years (4) after the end of the programmes and Data Management Plan is never mandatory. Indeed, in order to develop and to compete, SMEs need the exclusive rights associated to Intellectual Property.

- **White House Order**

In a White House Order of February 2013 on "Increasing Access to the Results of Federally Funded Scientific Research"[1], the objective is set on "*access to scientific data in digital formats*".

The objective is to maximize access to data, but while:

i. *protecting confidentiality and personal privacy,*
ii. *recognizing proprietary interests, business confidential information, and intellectual property rights and avoiding significant negative impact on intellectual property rights, innovation, and U.S. competitiveness, and*
iii. *preserving the balance between the relative value of long-term preservation and access and the associated cost and administrative burden.*

The White House Order also aims to ensure that data management plans are developed for publicly funded research "*as appropriate*", or that researchers "*explain why long term preservation and access cannot be justified".*

Finally, the general provisions of this report stipulate that "*consistent with the America COMPETES Reauthorization Act of 2010, nothing in this memorandum, or the agency plans developed pursuant to it, shall be construed to authorize or require agencies to undermine any right under the provisions of title 17 or 35 United States Code.".* This means that open access policies must be compliant with the US patent law (USC Title 35) and with the US copyright law (USC Title 17). In other words, fundamental legal rights, including intellectual property rights, cannot be altered through open access policies.

[1] https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

US Federal agencies have to comply with this order and, indeed, federal agencies data sharing policies refer to it.

This note reviews the main data sharing and data management plans requirements of the main US Departments and Agencies funding RDI, as follows:

1. US Department of Defence (DoD)
2. Department of Energy (DoE)
3. National Institutes of Health (NIH)
4. National Institute of Standards and Technology (NIST)
5. National Aeronautics and Space Administration (NASA)
6. US Food and Drug Administration (FDA)
7. US Department of Agriculture (USDA)
8. U.S. Department of Transportation (DOT)

1. US Department of Defence (DoD)

The US department of Defense[2] policy on the dissemination and sharing of research results to which proposals need to conform include:

- *Conditions for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;*
- *If, for legitimate reasons, the data cannot be preserved and made available for public access, the plan will include a justification citing such reasons.*

The Defense Advanced Research Projects Agency (DARPA), which is part of the US Department of Defense, has no identified data sharing policy. Several sources indicate that in DARPA funded projects, data sharing and data management plans are not required[3]. The US therefore did not put in place any data sharing policies in the field of advanced technologies in the defense and security sector, even though many of those technologies could also have civil applications if the knowledge created in such programs could be background knowledge of other Research, Development, and Innovation (RDI) projects funded by other agencies.

It is important to note that:

- US funding for RDI is concentrated in a few departments and agencies. 50% of the total US annual federal budget for RDI, that is $69 billion, is allocated to the Department of Defense, including DARPA[4].
- The share of basic research in the global RDI Department of Defense's budget is less than 4%.
- Federal agencies fund 62% of the whole RDI of Public Research.

Comments: This could mean that a great part of RDI federally funded projects in the US are not subject to data sharing and Data Management Plans.

2. Department of Energy (DoE)

The US Department of Energy (DoE) affirm the following principle related to data management: "*Not all data need to be shared or preserved. The costs and benefits of doing so should be considered in data management planning."* [5].

- **About data confidentiality and sensitivity**, the DoE states that "*if the plan is not to share and/or preserve certain data, then the plan must explain the basis of the decision".* The DoE gives the examples of cost/benefit considerations, feasibility parameters, scientific appropriateness, or other limitations linked to confidentiality and personal privacy: "*DMPs must protect confidentiality, personal privacy, Personally Identifiable Information, and U.S. national, homeland, and economic security; recognize proprietary interests, business confidential information, and intellectual property rights; avoid significant negative impact on innovation, and U.S. competitiveness; and otherwise be consistent with all*

[2] http://www.dtic.mil/dtic/pdf/dod_public_access_plan_feb2015.pdf
[3] http://guides.nyu.edu/data_management/dmp_agencies
[4] https://www.fas.org/sgp/crs/misc/R43944.pdf
[5] http://science.energy.gov/funding-opportunities/digital-data-management

applicable laws, regulations, and DOE orders and policies. There is no requirement to share proprietary data."

- **About the cost of data sharing**, the DoE adds that: "*the DMP should provide cost/benefit considerations to explain why the scientific value in sharing and preserving data generated by the research does not justify the expense, and should describe how the research results can be validated if data are not shared or preserved".*

In the end, "*at a minimum, DMPs must describe how data sharing and preservation will enable validation of results, or how results could be validated if data are not shared or preserved.*". Indeed, the DoE gives the following definition of data sharing: *"Data sharing means making data available to people other than those who have generated them. Examples of data sharing range from bilateral communications with colleagues, to providing free, unrestricted access to the public through, for example, a web-based platform.*". In such context, DMPs could only state that the data will be shared only for bilateral communications with colleagues to validate the results.

The Advanced Research Projects Agency–Energy (ARPA-E)[6], which is part of the Department of Energy, has different options for data management, and states the principle that "*To promote rapid commercial application of the results of ARPA-E projects, it is anticipated that a majority of ARPA-E Project Teams will request that ARPA-E not publicly disclose any data generated under the project and will assert that generated data is "Protected Data" or "SBIR-STTR Data" for their DMP under Option 1 below."* [7]

- ***Option 1 (Non - SBIR-STTR Awards):*** *It is anticipated that all digital data generated will be protected as "Protected Data" and, therefore, will not be publicly shared during the applicable "Protected Data" five (5) year protection period. Because any digital data will be at least five (5) years old when it is no longer considered "Protected Data", the effort to release such data will exceed any potential impact or value of the actual release. If any data generated under this award is published, an effort will be made to also release any related digital data that is not "Protected Data" to the public at the time of publication.*

- ***Option 1 (SBIR-STTR Awards):*** *It is anticipated that all digital data generated will be protected as "SBIR-STTR Data" and, therefore, will not be publicly shared during the applicable or "SBIR-STTR Data" four (4) year protection period. Because any digital data will be at least or four (4) years old when it is no longer considered "SBIR-STTR Data", the effort to release such data will exceed any potential impact or value of the actual release. If any data generated under this award is published, an effort will be made to also release any related digital data that is not "SBIR-STTR Data" to the public at the time of publication.*

- ***Option 2 (All Awards):*** *Use this option if the Project Team plans to publicly disclose technical data or data during the data protection period and/or expects that some data generated in the course of the project will not be asserted as "Protected Data" or "SBIR-STTR Data" by any Team Member. Project Teams that select this option must submit below a DMP that meets the minimum requirements specified in Section 6.2(a)(ii) of the "Applicant's Guide to Award Negotiations with ARPA-E" available at http://arpa-e.energy.gov/?q=site-page/pre-award-guidance.*

Comments: ARPA-E seems to have a very flexible data sharing policy, allowing for instance not to share data, even in non SBIR-STTR projects. If the grantees choose option 1, they do not even need to provide a DMP or give any additional explanation why they do not want to share Data. Therefore, in the US, advanced technology projects like those funded by ARPA-E can easily choose Option 1 (no data sharing, no DMP), whereas for other more basic research projects a detailed DMP is mandatory in order to explain for example which data will be shared and which part will be protected. If confirmed, the possible transposal to H2020 would be to have DMPs as default regime for pillar I, and two options "data sharing & DMP" and "no data sharing & no DMP" at the same level for pillar II and III (or part of them), with a special case of "no data sharing & no DMP" for the SME instrument.

[6] https://arpa-e.energy.gov/?q=arpa-e-site-page/arpa-e-history

[7] http://arpa-e.energy.gov/sites/default/files/ARPA-E%20236,%20Award%20Negotiations%20Guide.pdf

3. National Institutes of Health (NIH)

In its final statement on sharing research data published in February 2003[8], the US National Institutes of Health (NIH) indicates that "*all investigator-initiated applications with direct costs greater than $500,000 in any single year will be expected to address data sharing in their application".* It also notes that the proposed data sharing plan will not be taken into account to determine the scientific merit or the priority of the proposal.

Besides, NIH recognizes that data sharing may be complicated or limited, in some cases, "*by institutional policies, local IRB rules, as well as local, state and Federal laws and regulations, including the Privacy Rule".*

About proprietary data, the NIH states that[9]:

- *Although Small Business Innovation Research (SBIR) applicants are also to address data sharing in their applications, under the Small Business Act, SBIR grantees may withhold their data for 4 years after the end of the award. The Small Business Act provides authority for NIH to protect from disclosure and nongovernmental use all SBIR data developed from work performed under an SBIR funding agreement for a period of 4 years after the closeout of either a phase I or phase II grant unless NIH obtains permission from the awardee to disclose these data. The data rights protection period lapses only upon expiration of the protection period applicable to the SBIR award, or by agreement between the small business concern and NIH.*
- *Issues related to proprietary data also can arise when cofunding is provided by the private sector (e.g., the pharmaceutical or biotechnology industries) with corresponding constraints on public disclosure. NIH recognizes the need to protect patentable and other proprietary data. Any restrictions on data sharing due to cofunding arrangements should be discussed in the data-sharing plan section of an application and will be considered by program staff. While NIH understands that an institution's desire to exercise its intellectual property rights may justify a need to delay disclosure of research findings, a delay of 30 to 60 days is generally viewed as a reasonable period for such activity.*

The NIH also gives its definition of restricted data as "*datasets that cannot be distributed to the general public, because of, for example, participant confidentiality concerns, third-party licensing or use agreements, or national security considerations*"[10].

4. National Institute of Standards and Technology (NIST)

The National Institute of Standards and Technology (NIST) was founded in 1901 and is now part of the U.S. Department of Commerce[11]. In its Plan for Providing Public Access to the Results of Federally Funded Research, the NIST establishes a plan to enable public access to the results of research funded wholly or in part by NIST "t*o the extent feasible and consistent with law, agency mission, resource constraints, U.S. national, homeland, and economic security"*.

Defined in Circular A-110 of the Office of Management and Budget, "research data" is defined here as the "*recorded factual material commonly accepted in the scientific community as necessary to validate research findings*." The following data is not considered research data in Circular A-110 and is therefore not covered by this plan:

- *Laboratory notebooks, results of preliminary analyses, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as laboratory specimens;*
- *Trade secrets, commercial information, or other materials necessary to be held confidential by a researcher until they are published, or similar information that is protected under law; and*
- *Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy.*

[8] https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html
[9] https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#rest
[10] Ibid.
[11] https://www.nist.gov/sites/default/files/documents/data/NIST-Plan-for-Public-Access.pdf

The plan also adds that "*NIST will protect confidentiality and personal privacy and will recognize proprietary interests, business confidential information, and intellectual property rights, avoiding significant negative impact on intellectual property rights, innovation, and U.S. competitiveness."*

5. National Aeronautics and Space Administration (NASA)

In the key principles of its Plan for Increasing Access to the Results of Scientific Research[12], the National Aeronautics and Space Administration (NASA) includes the facts that:

- *Proprietary interests, business confidential information, intellectual property rights, and other relevant rights will continue to be recognized and appropriately protected; and*
- *Protecting confidentiality and personal privacy are paramount, and no change will be made to existing policies that would reduce current protections.*

Research data is also defined as "*the recorded factual material commonly accepted in the scientific community as necessary to validate research findings*", and they do not include:

- *Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law; and*
- *Personal and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study."*

In general, about DMPs, "*All proposals or project plans submitted to NASA for scientific research funding will be required to include a DMP. The DMP should describe whether and how data generated through the course of the proposed research will be shared and preserved (including timeframe), or explain why data sharing and/or preservation are not possible or scientifically appropriate. At a minimum, DMPs must describe how data sharing and preservation will enable validation of published results or how such results could be validated if data are not shared or preserved."*

In terms of applicability, "*all researchers receiving federal funding would be required to submit DMPs; however, in some cases it is expected that the data will not be made public. Such data would include but are not limited to the following categories:*

- *Educational grants and grants to individual students;*
- *Work that is proprietary;*
- *Work that results in personally identifiable human subjects research;*
- *Export controlled data;*
- *Sensitive But Unclassified (SBU; CUI – Controlled Unclassified Information) data;*
- *National Security classified data; and*
- *SBIR/STTR contracts".*

6. US Food and Drug Administration (FDA)

In its plan to Increase Access to Results of FDA-Funded Scientific Research[13], the Food and Drug Administration (FDA) stipulates that "*In general, the agency is restricted from disclosing information by statute, regulation, and policy, including, but not limited to*

- *information that constitutes trade secret and confidential commercial information, or that otherwise must be protected to preserve intellectual property rights;*
- *privileged information, including information related to ongoing product reviews, regulatory decision-making, and ongoing criminal or administrative investigations;*
- *personal privacy information; and*
- *national security and other classified information".*

According to the OSTP Memo and OMB Circular A-110, the FDA defines "digital data" as "*the digitally recorded factual material that would be commonly accepted in the scientific community as necessary to validate published, peer-reviewed scientific articles".*
They also exclude from this definition:

- *preliminary materials underlying the data or factual information, including lab notebooks, preliminary analyses, drafts, plans for future research, peer-review reports, communications with colleagues, or physical objects such as lab specimens;*

[12] https://www.nasa.gov/sites/default/files/atoms/files/206985_2015_nasa_plan-for-web.pdf
[13] http://www.fda.gov/downloads/ScienceResearch/AboutScienceResearchatFDA/UCM435418.pdf

- *data shared with FDA but owned by other organizations (e.g., aggregate electronic healthcare data from other parties used by FDA in product safety monitoring pursuant to FDA's Sentinel program, WHO Medical Device Single Audit data)*
- *data received by FDA as part of an application for market authorization or application for exemption from marketing restrictions for investigational use;*
- *data obtained under licensing or data use agreements, or cooperative research and development agreements that include terms that restrict the release and/or sharing of the data;*
- *materials necessary to be held confidential by a researcher until published to ensure the acceptance of research for publication;*
- *data or information not available for disclosure pursuant to statute or regulation as described in Section I, above; and*
- *technical and administrative data.*

Regarding public access requirements, the FDA "*intends to increase public access to digital data supporting FDA-conducted or -funded research findings and to further the goals and requirements of the OSTP Memo while recognizing and protecting intellectual property rights and proprietary interests, including protections from disclosure of trade secret or confidential commercial information, and personal privacy information".*

FDA will aim to maximize access to digital data, but while

- *preserving the integrity of the data;*
- *adhering to applicable legal or regulatory restrictions on information disclosure, such as those identified in Section I, above; and*
- *balancing the value of public access to the data and the associated cost and administrative burden of modifying datasets to allow disclosure.*

7. US Department of Agriculture (USDA)

In its Public Access Implementation Plan[14], the USDA:

- *Recognizes proprietary interests, business confidential information, and intellectual property rights, and avoids significant negative impact on intellectual property rights, innovation, and U.S. competitiveness; and*
- *Protects confidentiality and personal privacy*

8. U.S. Department of Transportation (DOT)

Finally, the US department of Transportation (DOT) states in its Public Access Plan[15] that digitally formatted scientific data resulting from unclassified research supported wholly or in part by Federal funding needs to be stored and publicly accessible for search, retrieval, and analysis "*to the extent feasible and consistent with applicable law and policy; agency mission; resource constraints; U.S. national, homeland and economic security*".

It is also indicated that "*this plan requires that awardee(s) and/or the respective Operating Administration ensure Public Access to final research data, subject to the above restrictions and those imposed by data quality and the need to protect national/homeland security, individual privacy, and confidentiality".*

There also need to be *interactions among the awardee(s), the data repository, and the DOT grant manager to ensure that:*

- *Data meet minimum quality standards*
- *Data is appropriately evaluated for and secured to prevent disclosure of personally identifiable information, protect proprietary interests, confidentiality, and intellectual property rights*
- *Data is licensed in a manner that encourages both access and reuse*

Conclusion

This analysis shows that all US federal agencies have strong safeguards and limits to data sharing. For instance, (but not limited to), they:

[14] http://www.usda.gov/documents/USDA-Public-Access-Implementation-Plan.pdf

[15] https://www.transportation.gov/sites/dot.gov/files/docs/Official%20DOT%20Public%20Access%20Plan.pdf

- recognise proprietary interests, business confidential information, and intellectual property rights, and avoid significant negative impact on intellectual property rights, innovation, and U.S. competitiveness; and
- protect confidentiality and personal privacy

All US federal agencies, which fund 62% of the whole R&D&I in US Public Research, are aware that data management results in high costs and they would generally be bearing the costs. They therefore recommend to balance data sharing not only with economic, confidentiality, privacy and Intellectual Property Rights requirements, but also more generally with cost considerations. For instance:

- The Defense Advanced Research Project Agency foresees no data sharing nor DMP.
- The Advanced Research Projects Agency-Energy (ARPA-E) has the option "no data sharing, no DMP" at the same level than the option "DATA sharing, DMP" in its projects.
- The Department of Energy (DoE) has a quite flexible definition of data sharing and DMP: "*Data sharing means making data available to people other than those who have generated them. Examples of data sharing range from bilateral communications with colleagues, to providing free, unrestricted access to the public through, for example, a web-based platform".*

EARTO - European Association of Research and Technology Organisations *is a non-profit international association established in Brussels, where it maintains a permanent secretariat. The Association represents the interests of about 350 Research and Technology Organisations (RTOs) from across the European Union and "FP-associated" countries.*

EARTO Vision*: a European research and innovation system without borders in which RTOs occupy nodal positions and possess the necessary resources and independence to make a major contribution to a competitive European economy and high quality of life through beneficial cooperation with all stakeholders.*

EARTO Mission*: to promote and defend the interests of RTOs in Europe by reinforcing their profile and position as a key player in the minds of EU decision-makers and by seeking to ensure that European R&D and innovation programmes are best attuned to their interests; to provide added-value services to EARTO members to help them to improve their operational practices and business performance as well as to provide them with information and advice to help them make the best use of European R&D and innovation programme funding opportunities.*

EARTO Working Group Legal Experts*: is composed of 25 corporate legal advisers working within our membership. Established in autumn 2013, this Working Group has also worked on the revision of the State-Aid Rules & the GBER. Our experts also contributed to the setting-up of the DESCA Consortium Agreement model for Horizon 2020.*