|  |  |
| --- | --- |
| Question | CNECT comment |
| Does data holder X fall into the scope of the Open Data Directive and of the Implementing Regulation on high-value datasets? | We recommend asking two questions:   1. Is data holder X considered as public sector body, as defined in Article 2(1) of the Open Data Directive? 2. Are the datasets held by data holder X on the list of high-value datasets, as set out in the Annex to the Implementing Regulation on high-value datasets?   If the answer to both questions is affirmative, it is very likely that Article 3(1) of the Implementing Regulation on high-value datasets (“*Public sector bodies holding high-value datasets listed in the Annex shall ensure that the datasets described or referenced in the Annex are made available…*”) will apply to data holder X. |
| Which data holder should ensure conformity with the Implementing Regulation on high-value datasets if the same dataset is held by several bodies? | We suggest looking at this question through the eyes of reusers who need to be able to find a dataset from the list (including metadata description, etc.) and be able to reuse it under the conditions applicable to high-value datasets. The question of which body publishes the dataset then appears rather secondary. Practically speaking, if more than one institution hold the same high-value datasets, the one that does not publish them directly should indicate to reusers (publicly, online, e.g. via hyperlink) where the relevant datasets can be found. |
| Is the Member State free to choose any publicly documented, Union or internationally recognised open, machine-readable data format and any tool to create OpenAPI documentation? | Yes, MS are free to choose. In practice, MS will follow domain specific rules that are already in use. For data within the scope of the INSPIRE Directive, please follow the INSPIRE rules. Where the Implementing Regulation includes provisions on formats and metadata, it reflects what has already been achieved in concrete sectors, but it also allows for flexibility. |
| What if our high-value datasets contain data that we consider sensitive or confidential (due to statistical confidentiality, trade secrets or personal data). | It is indeed acceptable to not share sensitive/protected information or any other data that should not be made public. See the exceptions to the scope of covered data in Article 1(2) of the Open Data Directive. That said, the data holder should still try to publish the relevant dataset after removing the non-publishable information. |
| Can we continue to use our existing APIs? | There are many possible API solutions, so the MS have flexibility to choose what works best, also with respect to the needs of a given category of high-value datasets. The Commission has researched on APIs: (see reports under <https://publications.jrc.ec.europa.eu/repository/handle/JRC118082> and <https://publications.jrc.ec.europa.eu/repository/handle/JRC129940>). That said, this topic could be further discussed with other MS within a dedicated subgroup under INSPIRE MIG. |
| How about INSPIRE data? | Please follow the work of the relevant MIG subgroup, which aims to ensure that INSPIRE implementation is HVD compliant. |
| Is metadata description imposed by Article 3(5) only important for reporting by MS to the Commission? | Although there is a clear link with reporting indeed, the objectives of Article 3(5) go beyond reporting from Member States to the Commission: findability of HVDs is important for reusers too.  It appears that DCAT-AP-HVD as it is designed will work for both Article 3(5) and for the reporting under Article 5. We recommend (albeit just informally) using DCAT-AP-HVD.  Metadata does not need to be in English, data.europa.eu will ensure their machine translation. |
| Is Creative Commons BY 4.0 licence mandatory? | Please refer to Recital 12, to Article 4, and to particular mentions in the Annex. The Implementing Regulation is very flexible in letting you use Creative Commons BY 4.0 licence or any equivalent or less restrictive open licence. You can publish the data under both national and model licences, but then the question is what is the difference between the two, or why keep using both of them, if they are identical? We recommend choosing the approach that will be the clearest and simplest for reusers to follow. |
| Article 14(1) of the Directive and Article 3(1) of the Implementing Regulation state that HVDs must be available in a machine-readable format. If scanned documents are considered HVDs, do they all have to be converted to a machine-readable format? | In the company data category, there is one softening of the rule: “*machine-readability shall not be imposed to data that are maintained in formats that are not machine readable (e.g. scanned company documents and accounts) or to unstructured/non-machine readable data fields included as part of machine readable documents.*” This solution was found together with Member States within the Open Data Committee as an appropriate reaction to the fact that a lot of documents held by business registers have been filed by companies for years on paper, as scans or PDFs and that it could not be envisaged that business registers would be able to convert all those documents into machine readable formats. At the same time, the consensus was that it is important to include those documents on the list (in light of Article 14(2) of the Directive) and thus improve their reusability – which is the key objective of Article 14(1).  There is a similar softening in Article 2.2 a) of the Annex with respect to historical versions of environmental datasets.  That said, Member States can go beyond the provisions of the HVD Regulation and not take advantage of those softer requirements (see Recital 10).  The Implementing Regulation is a minimum list and Member states are encouraged to go beyond it, both in terms of substance and in terms of bodies that are in its scope. |
| Some countries publish statistical datasets according to a different breakdown that the one set out in the Annex to the Implementing Regulation. Do they need to generate additional datasets that according to the breakdowns specified in the annex?  Furthermore, for some datasets the breakdowns are not specified in a dedicated table in the Annex of the IR, but there are references to legal acts. | No. The Implementing Act applies to documents that exist. The main purpose of datasets descriptions in the Annex is to identify, sometimes more and sometimes less concretely, the HVDs among those existing documents. Admittedly, the Annex has also the ambition to push the MS towards presenting their documents with the content and structure that the Annex describes, if it is not yet the case. Nevertheless, that content and structure is often taken over from other legislation, in this case binding sectoral rules on statistics. While it would be desirable that this issue is harmonised across the EU, we believe that ESTAT regulations are better tool to achieve that than the implementing act on HVDs.  Datasets without a detailed table should follow the breakdowns laid down in the legal acts referred to in the Annex to the implementing act. |
| What is meant by publishing quality of service criteria for API’s performance, capacity and availability as required by Article 3 of the Implementing Regulation? | API performance and capacity are measured using various metrics such as:   * Concurrent users: Number of users who can interact with the API simultaneously * Response time: The time it takes for an API request to be processed and returned. * Request rate: The number of API requests made per unit of time. * Peak requests: The maximum number of requests that can be made per unit of time. * Error rate: The percentage of API requests that result in errors or failures. * Throughput: The amount of data processed by the API per unit of time. * Latency: The time it takes for data to travel between the client and the API server.   A report by the Commission's Joint Research Centre provides some ideas on the issue of performance (p. 24): <https://publications.jrc.ec.europa.eu/repository/bitstream/JRC118082/jrc118082_api-landscape-standards.pdf>  The Implementing Regulation does not usually prescribe the target figures, neither does it precisely define the concepts of various service criteria. HVD holders, when defining the terms of use of their APIs and the concrete criteria for the quality of service, should bear in mind the needs of re-users. Typically, those should be taken into account when drafting Service Level Agreements (SLAs). We presume that a discussion with re-users can serve as a good basis for the SLAs regarding HVDs. Therefore, we encourage MS to expose through OpenAPI documents all of the metrics (if available) that are listed above.  Another useful reference on SLAs is the OpenAPI SLA initiative: <https://www.openapis.org/blog/2021/03/10/openapi-meets-sla>, including the OpenAPI SLA specification  <https://github.com/isa-group/SLA4OAI-Specification/blob/main/versions/1.0.0-Draft.md>.  JRC’s research on APIs (<https://publications.jrc.ec.europa.eu/repository/handle/JRC129940>) has focused on terms of use as well. |
| [statistics] In the case where a dataset is not published at the national level but is already published by Eurostat, should we additionally publish the data on a national level? Or is it sufficient to rely on Eurostat's publication? | In principle, Eurostat's publication is enough, as long as it fulfils technical conditions imposed by the implementing act (APIs etc.). Offering links to Eurostat’s database can be an acceptable solution. However, please bear in mind that the obligation to make HVDs available is imposed on Member States, and if for any reason Eurostat stops making the HVDs available, MS that had been relying on Eurostat will have to find another solution in order to avoid breaching the implementing act. |
| [statistics] Can a single dataset be included in two different tables? That is, can one part of the data be in one table while the remaining part is in a separate table (both tables would be flagged as HVD tables)? Or is it required for the entire dataset to be contained within a single table? | The Implementing Regulation does not give a clear answer on this. If there are good technical and organizational reasons for that and if it does not complicate the re-use of the actual high-value dataset (e.g. it is not buried inside a large amount of uninteresting data), it may be acceptable. |
| What should be the leading language in the definition field in the API? We would like to use our national language as the leading language in field definitions in our API. In addition, API content is translated into English and it is possible for any English speaking user to retrieve the data. | The Implementing Regulation on HVDs does not prescribe the language. It is expected that the official language of the country of origin of the dataset will be used. Any additional language version (English or another) would be a welcome extra. |
| Formats: section 4.2 states that the datasets shall be made available for re-use in CSV, XML (SDMX), JSON or another publicly documented, Union or internationally recognised open, machine-readable format. It is presumed that there is no need to publish the data in every format indicated above and it is sufficient if the data are retrieved by the API only through the use of JSON. | Yes, it is fine as long as there is at least one format that fulfils the requirements of the Implementing Regulation. |
| For the Geospatial dataset "Administrative Units", the granularity (level of detail) requirement includes ‘maritime units’. In MS X, the administrative areas to be accumulated in the Address Register do not include information on maritime units. | The HVD Implementing Regulation does not oblige Member State to create datasets (or key attributes) that do not exist in that Member State. However, the case may be that the data exist in the given MS, but their existence is not widely known. |
| What does the term ‘historical data’ mean? | Historical versions of high-value datasets should be published to allow for creating time series. These are therefore datasets that reflect a certain completed period and which then allow monitoring the development of the value over a longer period of time. If the new datasets are just corrections and additions, and if older versions of the updated dataset are not even preserved, such an older dataset will not, in our opinion, usually be considered a separate "historical version" of the dataset. In such a case, it is sufficient to publish the currently valid version, which is complete and of the quality prescribed by the Regulation. However, this depends on the context and needs of the data users. For example, in the commercial register, information that was subsequently changed and is no longer valid, but was valid in the past, is also important. |
| Are company documents and accounts containing personal data to be made available for re-use via API in full? | This depends on the situation and practice on the national level. In some Member States the fact that such documents must be made publicly available means that they are automatically reusable, while other MS impose specific conditions on their reuse. This means in practice licensing conditions focusing on conformity with the GDPR (including purpose limitation) or the necessity to follow a certain procedure for signing up for an API.  The Implementing Regulation points to the necessity to take data protection legislation into account - see its Recital 8:  “*Where making high-value datasets available for re-use entails the processing of personal data, such processing should be carried out in accordance with Union law on the protection of individuals with regard to the processing of personal data, in particular Regulation (EU) 2016/679 of the European Parliament and of the Council and including any provisions of national law further specifying the application of the GDPR. Member States should make use of appropriate methods and techniques (such as generalisation, aggregation, suppression, anonymisation, differential privacy or randomisation), thus making as much data as possible available for re-use.”*  This logic has been established in the Open Data Directive, which declares itself without prejudice to Union and national law on the protection of personal data, in particular Regulation (EU) 2016/679 and Directive 2002/58/EC and the corresponding provisions of national law.  More concretely, with respect to the category of data on companies, the Implementing Regulation stipulates that datasets shall be made available for re-use under the conditions of the Creative Commons BY 4.0 licence or any equivalent or less restrictive open licence, with additional conditions concerning the re-use of personal data where relevant. |
| How to make available company documents and accounts prepared in non-machine-readable formats? How to proceed in a situation where the processing of this data requires a disproportionate effort going beyond the simple operation referred to in Art. 5 (3) of the Open Data Directive? | The obligations imposed by the Implementing Regulation with respect to high-value datasets go beyond Article 5(3) of the Open Data Directive, because the Regulation builds on Article 5(8) of the Directive. That said, the Implementing Regulation does take into account the difficulties related to machine readability of documents. Section 5.2(a) of the Annex stipulates that “*machine-readability shall not be imposed to data that are maintained in formats that are not machine readable (e.g. scanned company documents and accounts) or to unstructured/non-machine readable data fields included as part of machine readable documents;*”. Availability via API of data in not machine-readable format will be acceptable in such a case. |
| Company documents and accounts are made available in the ICT system for downloading via a browser in the full version (with personal data).  Und the HVD Regulation, the same IT system would provide access via API to the same documents and accounts, but already anonymized.  What to do in a situation where the application of this regulation results in such a dualism? | It was not the purpose of the Implementing Regulation to introduce the described dualism. However, the differences in approach can be justified, if the techniques used result in different outcomes. For instance, downloading via browser presumably only concerns data on individual companies, not bulk downloads of the whole register. Personal data protection risk is therefore lesser in the former case. We would expect that there is, for instance, some kind of purpose limitation and that such data (in addition to being retrievable on a single company basis) are not openly re-usable. This is why the dualism may be only apparent because while we are talking about the same data, we are not talking about the same extent of accessibility/usability.  Potentially, even API users could benefit from access to non-anonymised data, subject to custom made licensing conditions that would cover personal data protection aspects. |
| Does the Open Data Directive jointly with the HVD Regulation create legal obligations/entitlements to publish data that are not already available based on existing national or EU law? | The Open Data Directive and the HVD Regulation only apply to existing data. They do not require that public sector bodies collect new data. That said, in some cases the fact that some dataset is missing may point to a failure to meet an obligation under relevant sectoral legislation.  Article 3 (1) of the HVD Regulation does not mean an obligation to create new datasets where they do not exist. The datasets to be made available are those, to which the HVD regulation applies (Art. 1(1)), i.e. they are existing and accessible under national rules such as laws on free access to information.  Useful reminder: the OD Directive and the HVD Regulation only concern REUSE, not ACCESS. And it is a prerogative of Member States to define rules for access to data (with several exceptions, where EU law requires accessibility of some information e.g. about companies or about environment). Access is typically guaranteed by national laws on free access to public sector information. Hence, when we speak about ‚open data‘, from the OD Directive perspective the key corresponding term is not ‚accessibility‘ but ‘availability for reuse’. |
| [statistics] The HVD Regulation refers to several statistical sectoral rules that impose transmission deadlines. Is it then reasonable to conclude that HVDs should be made available on MS website not later than when the data should have been transmitted to the EC? | The ‘transmission deadline’ refers to transmission of data from an NSA to the Commission. In that light it seems that ‘*transmission deadline equals publication deadline*’ is a reasonable practical interpretation for the time after the HVD Regulation enters into application.  More generally, the datasets should be published once they are existing and accessible (even theoretically, following a reuser’s request based on their right of access to information), or as soon as practicable after that moment. That is the moment when they enter into the scope of the HVD Regulation and when the reusers can expect/demand them to be published as HVDs. |
|  |  |