“As-is” vs. INSPIRE-compliant data sets

Daniela Hogrebe, Ine de Visser,
Michael Lutz

MIG-T meeting #60 Prague
15-16 October 2019

Joint Research Centre
The European Commission’s science and knowledge service

European Commission

Issue

- Data Sets in scope of INSPIRE directive (art. 4) shall be made available in compliance with the IRs on data interoperability
- The INSPIRE data models do not always cover all object types or properties available in the source data sets → Some parts of source data sets cannot be mapped to the INSPIRE data models
- INSPIRE data models can be extended, in order to enable the mapping of the whole source data set into the (extended) INSPIRE data models, or MS can continue to make available in the INSPIRE infrastructure the full source data sets „as-is“
- In most cases, the transformation is not done on the fly, so a seperate representation of the data set (that is compliant with the IRs) is created

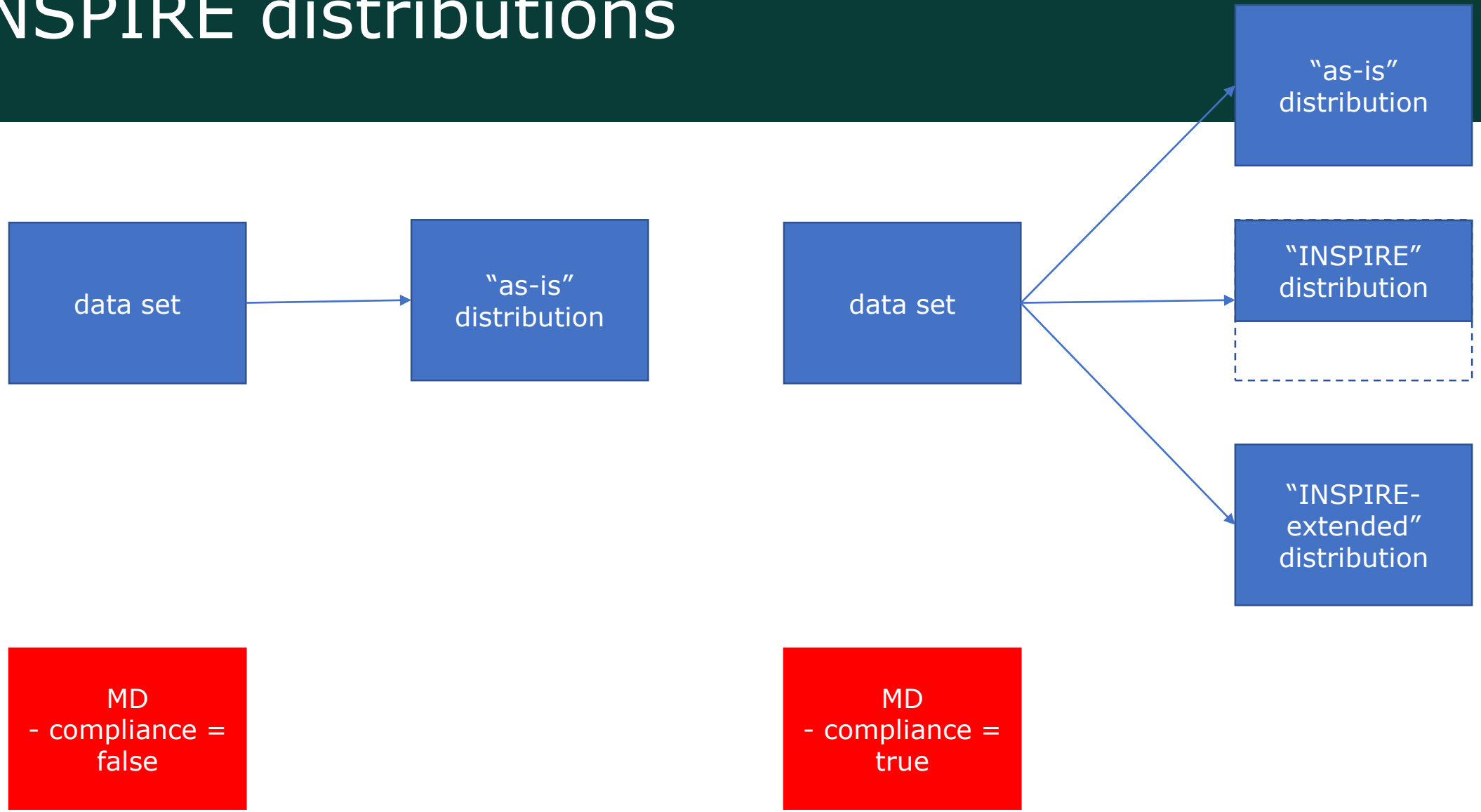
Issue

- This representation may have different content (less if the model is not extended) and specifications (CRS, format ….) than the „as-is“ representation of the source dataset
- If the harmonised representation is considered to be a separate data set from the source data set, metadata will be created for both of them
- In this case, the indicator on the percentage of compliance of data sets (DSi2) will never reach 100% (since there will always be “as-is” data sets in the infrastructure that are not compliant)
- In this case, MSs might be tempted to remove the “as-is” data sets from the infrastructure in order to improved their DSi2 indicator

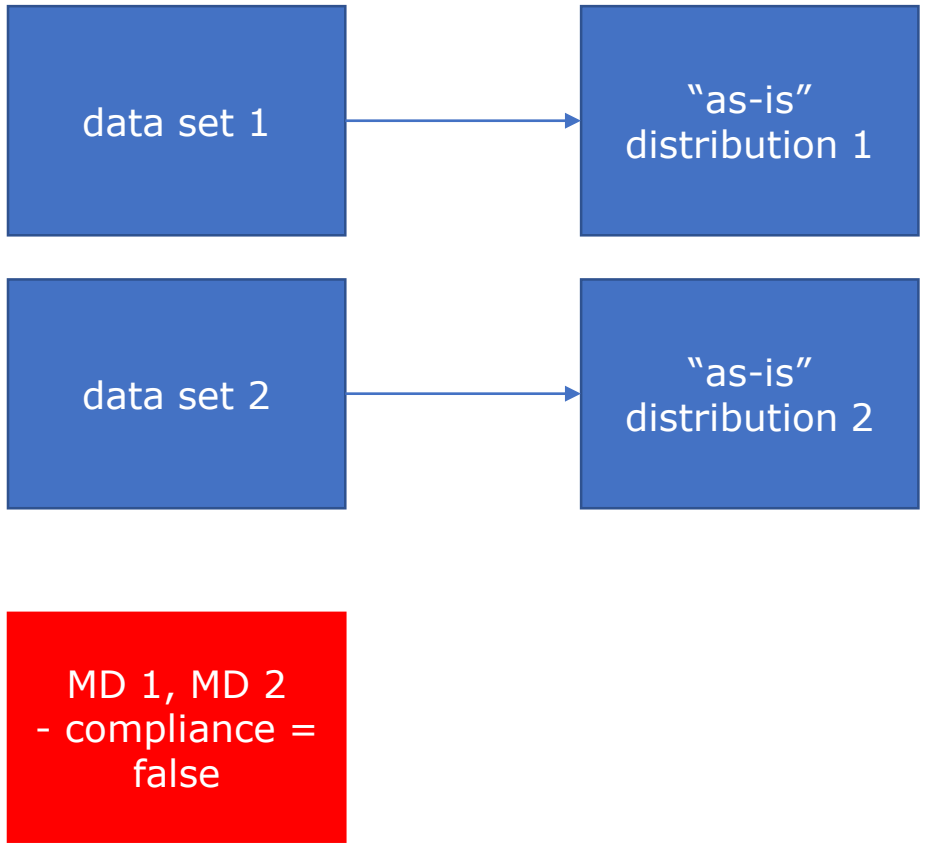
Proposed solution – Principles

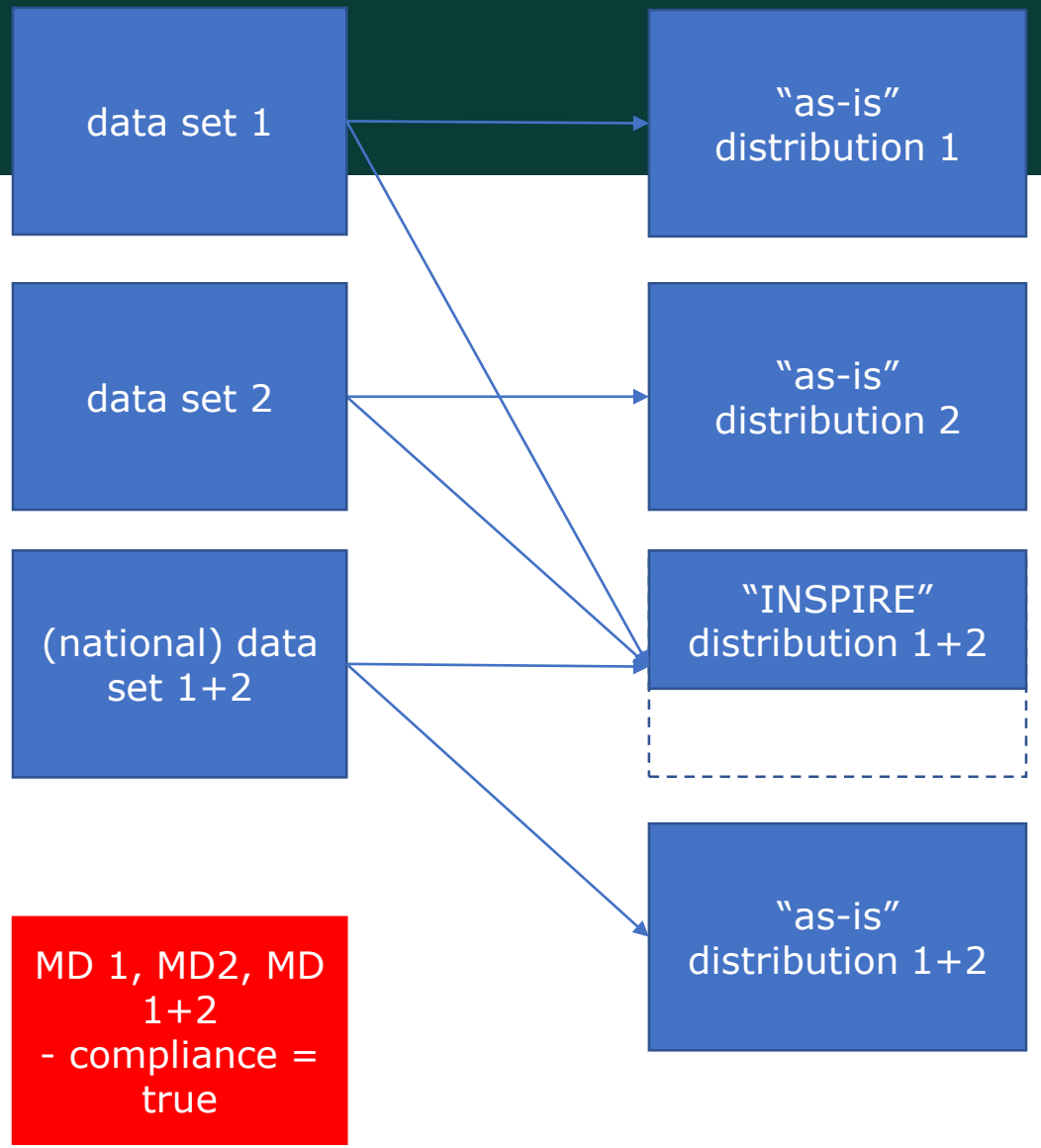
- MS should not exclude "as-is" data from the INSPIRE infrastructure just to improve the indicators
- A data set should be considered "compliant with the IR on interoperability" if there is at least one distribution of it that meets the IR requirements.
- A distribution is "a specific representation of a dataset. A dataset might be available in multiple serializations that may differ in various ways, including natural language, media-type or format, schematic organization, temporal and spatial resolution, level of detail or profiles (which might specify any or all of the above)." (DCAT)
  - In some cases all distributions of a dataset will be fully informationally equivalent, in the sense that lossless transformations between the representations are possible. However, in other cases the distributions might have different levels of fidelity to the underlying data.

INSPIRE distributions



Combining data sets





Open issues

- How to implement “INSPIRE distributions” in ISO 19115?
  - One option could be the resource locator element
  - combine this with the work on Simplification of data-service linking?