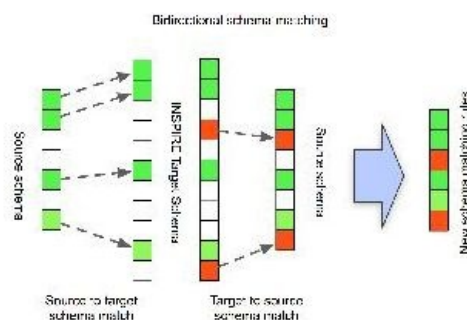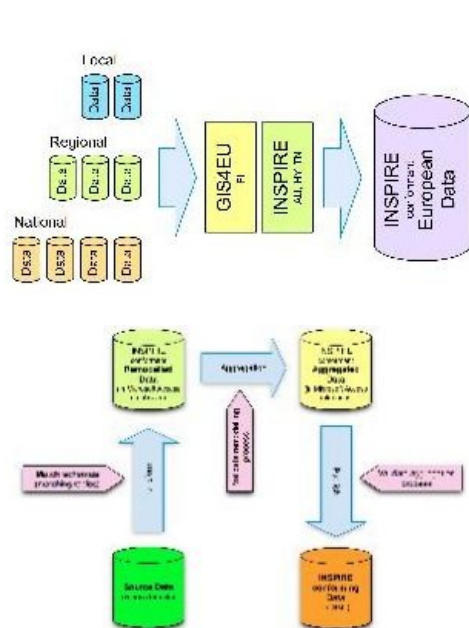# *HARMONISING CROSS-BORDER DATASETS*

GIS4EU

INSPIRE has set a standard towards which GI providers can work and with which GI users can work seamlessly at European scale. GIS4EU has taken this standard and designed a repeatable process that can be implemented to harmonise heterogeneous spatial datasets. By following the GIS4EU process, a set of rules can be determined for almost any spatial dataset that falls within the INSPIRE spatial themes. Applying these rules promotes locally and nationally relevant spatial information to a seamless European SDI. That this can be done without the need to rebuild or modify at source any existing data repositories enhances the economy, promotes reuse, and extends the flexibility of the whole INSPIRE process.

INSPIRE is working toward making a European spatial-data infrastructure (SDI) a reality, an objective that results from a strategy born more than fifteen years ago. Policy initiatives and working groups including GI2000, EUROGI, GENIE, and RISE, developed and furthered the idea of a European SDI, feeding into the INSPIRE initiative to define specifications for a common spatial-data model. INSPIRE became legislation in 2007, necessitating workable process models to implement it.

One model
Spatial datasets can be collected to meet many objectives. The variety of data collection paradigms across Europe has resulted in the creation of a fragmentary archive of differing formats, non-standard definitions and varying levels of completeness. Where use of this data does not exceed the original purpose, problems are likely to be limited. However, the increasing need to collate and implement spatial data at national and international scales exposes the problems inherent in fragmentary and task/location-specific records. To surmount these problems, INSPIRE developed common data models for Administrative Units (AU), Hydrography (HY) and Transportation Networks (TN) on which existing data could be mapped, remodelled and then harmonised (Figure 1). Adopting the INSPIRE methodology for developing data specifications, GIS4EU created a common data model for elevation (EL) data. These common data models provide templates that new spatial data can aspire to. INSPIRE also provides a standardised template for associated metadata: descriptive information about data and services.

Moving Things On
Although the INSPIRE specifications have developed from the results of previous projects, such as RISE, and related policies such as the EU Water Framework Directive, empirical testing of a complete harmonisation process has been limited. The practical process of taking spatial datasets, transforming them into INSPIRE-compliant common data models and then aggregating them into a harmonised geographical theme has remained relatively undocumented. To remedy this, the GIS4EU project, funded through the eContentPlus programme, was set up to provide a robust, documented process model with defined workflows that could be applied across different GI applications to harmonise disparate spatial data. Key to achieving the project's objectives has been the participation of national mapping agencies, local authorities, technical experts, and data users from seven EU member countries.

Source Data Modelling
The GIS4EU process starts with the raw source data. Before anything else, the spatial datasets need to be documented; specifically metadata, including feature type and attribute catalogues. This documentation allows data models to be created and their schemas constructed for mapping against INSPIRE data specifications. The schema derived from the data providers' source data is put into tabular form, called a 'matching table'. In GIS4EU we recognise that the data providers have the deepest knowledge of their own data and are heavily involved in documenting it. This is done through filling in the matching tables with detailed comments against feature types and attributes to describe them as fully as possible. Completed matching tables form the source documents against which the data providers' feature types and attributes are compared to the target, the INSPIRE feature catalogue.

Aspiring to INSPIRE
Cross-matching the source schema to the target catalogue is a bi-directional process (Figure 2). Firstly we look for matches between the source and target feature types and attributes, direct or otherwise. Any candidate feature types in the source data that could be added to the INSPIRE catalogue are also noted. Through this process we may find that information from the source data may be extraneous and

can be ignored, or that the INSPIRE feature catalogue is not comprehensive enough. After the source-to-target matching, the process is reversed and the INSPIRE catalogue is matched against the source schema. While this may appear unnecessary, it isn't. Undertaking this step allows the full range of INSPIRE feature types to be assessed against the source data, whereas the source-led matching targeted only a subset of the INSPIRE catalogue. This process allows a framework to be devised through which all or part of the source-data schema (thus the data which it describes) can be transformed into a common INSPIRE data model ready for aggregation/harmonisation. We refer to this process as 'remodelling', in order to highlight the transformation of the source-data model into the target (INSPIRE) data model. Any feature types or attributes found to be critical for use of the data, but not available in INSPIRE, are documented for provision as feedback to the INSPIRE development teams.

Harmonisation Rules

With our source data in a common model format, we can proceed towards aggregating it into a harmonised geographical dataset. Before we can merge the data, the coherency of the transformed data needs to be confirmed; for example, borders need to be consistent, water courses should respect topography, and transport networks should connect at borders. This is not always simply a case of making sure lines meet and areas are closed; in the case of borders it may require political involvement. Next to be addressed are issues of scale. INSPIRE is scale-independent; it offers a specification set for features and attributes, not one for dealing with different levels of detail. In the real world, however, it is likely that a repository of spatial data will contain data at various levels of detail from high (sub-metre) to low (kilometre) resolution. To provide a consistent data theme where the constituent spatial data resolution differs, higher-resolution datasets need to undergo degradation to match that of the coarsest. The level of detail in terms of features visible at particular scales has also to be defined. Essentially, we are formalising the transformation of multi-scale data into scale-independent data themes. At the same time, we also define the colour and symbology of features to ensure consistency in their portrayal. Finally, the issue of language has to be tackled. For features and attributes in the common data model catalogue, an agreed template was produced for translation between partner languages. After specifying rules for handling cross-border issues, such as different languages and scale, the remodelled datasets are merged. We refer to this process as 'aggregation' (Figure 3).

Validation

As is good practice, the output of the aggregation process is validated to confirm its compliance to INSPIRE and GIS4EU data specifications. This validation step involves checking that the output datasets contain all mandatory feature types, attributes and their values. Additionally, INSPIRE specifies a 'voidable' stereotype for some feature types or attributes. If a voidable feature type or attribute is omitted from data, the reason for doing so must be provided. Applying this validation process to the aggregated data has allowed GIS4EU project partners to address any issues resulting from the previous steps.

Concluding Remarks

Bringing together European partners and spatial datasets, GIS4EU has developed processes, described above, to achieve INSPIRE conformity. We are now at a significant evolutionary point in the implementation and accessibility of GI in Europe; through INSPIRE the opportunities for future applications lie wide open.

Further Reading

Directive 2007/2/EC of the European Parliament and of the Council of 14th March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)GIS4EU, Provision of interoperable datasets to open GI to EU communities.

https://www.gim-international.com/content/article/gis4eu