

DATA FLOW DIAGRAMS

IN GEOGRAPHIC INFORMATION SYSTEMS: A SURVEY

Ing. Zdena Dobesova, Ph.D.

Palacký University, Olomouc, **Czech Republic**

ABSTRACT

Data flow diagrams (DFD) are one of the oldest graphical representations in informatics. The data flow diagrams can be considered as visual programming language (VPL). Nowadays, geographic information systems (GIS) have some graphical environment that assists in the designing of flow chart model. The article brings an overview of notations and formal semantics for data flow diagrams in GIS. The survey contains software ArcGIS for Desktop 10.2, ERDAS IMAGINE 2013, IDRISI Selva and AutoCAD Map 3D 2014. Beside the commercial software, the open source QGIS 2.2 and its component Processing Modeler is mentioned.

The article concerns more to the graphical notation than the amount of functionality. The graphical notation is important from the point of perception and cognition. The main task of data flow diagrams is to support the batch processing data. The next function is a “graphical report” of steps of processing. From the point of these two tasks, the graphical notation is important. For all that purposes, it is necessary to use cognitively effective visual notations. Cognitively effective means optimized for processing by the human mind. An optimized notation helps to user very quickly understand to the diagram [1].

Keywords: geoinformatics; notation; visual programming language; diagram; cognition; perception.

INTRODUCTION

Data flow diagrams express the steps of processing in a graphical way. GIS products are supplemented by these components more and more. The article brings overview of components for assembling and executing flow charts with geospatial processes. Unfortunately, the history of data flows diagrams in GIS is more than 20 years old.

Model Maker for ERDAS IMAGINE, a graphic flow chart model building editor, was introduced in 1993 [2]. http://en.wikipedia.org/wiki/Erdas_Imagine_-_cite_note-ASPRS_p568-8In 2004 Esri company was inspired by the idea of data flow diagram environment and released an ArcGIS solution named Model Builder. Another older DFD component is Macro Modeler that was added to the GIS software IDRISI 32 Release 2 in 2001 [3]. The producer of IDRISI is Clark Labs at Clark University in the USA.

Later, the Autodesk company implemented the component Workflow Designer in version AutoCAD Map 3D 2010. The component is improved mainly in the amount list of spatial function from version to version. The graphical notation remains the same till latest version AutoCAD Map 3D 2014. The newest data flow component is the Processing Modeler in QGIS that was implemented in version QGIS 2.0 Dufor in 2013.

The latest version is QGIS Valmiera from February 2014. The graphical notation has been slightly changed from the first version.

Graphical notation in DFD can be described and assessed by different methods. Alan Blackwell designed “The Cognitive Dimensions Questionnaire” with open questions [4]. Theory of “Physics of Notation” by Daniel Moody is an alternative and more suitable theory [5]. There are nine principles that are focused on the physical (perceptual) properties of notations rather than their logical (semantic) properties. Detail evaluation according to “Physics of Notation” is described in the article about ModelBuilder for ArcGIS for desktop [1]. Semantic properties of symbol and example models of ModelBuilder are described in several articles [6, 7, 8, 9].

ERDAS IMAGINE

The oldies VPL component is Model Maker in ERDAS IMAGINE. The Model Maker’s graphical user interface was changed to Spatial Model Editor in the release of ERDAS IMAGINE 2013 in December 2012. The architecture was rewritten from a push model to a pull model. This re-architecture allows many complex models to be processed directly to a viewer in near-real time. A Python capability was added at this time [10]. Both component, Model Maker and Spatial Model Editor, are available simultaneously in ERDAS IMAGINE 2013.

Model Maker uses only black and white graphical symbols. The shapes express different types of data in a good way. From the point of Moody’s principle of Semantic Transparency, the appearance (shapes) might suggest their meaning. The shape of raster data is near to the set of the bands, the symbol of the table and the matrix is a rectangle with one or more columns and rows. Functions are expressed by circle symbol. Connectors are the solid straight arrows (Fig.1). The orientation of a flow can be variable. User can construct a top-down oriented model, a left-right oriented model or mixture of both. The automatic aligning is not available. The labels are used for description. The position of labels is under the graphic elements, and labels are often crossed by connecting lines. Crossing reduces the readability of diagram.

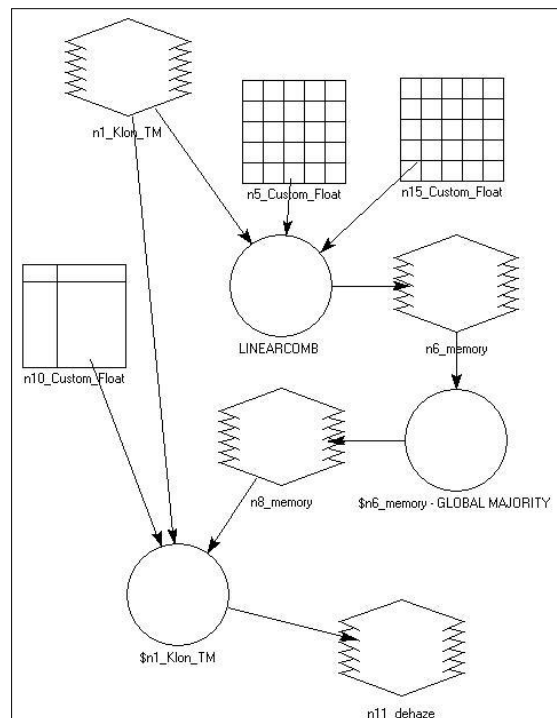


Fig. 1 Example model in Model Maker for ERDAS IMAGINE

Spatial Model Editor has totally different notation. All symbols are rounded rectangles with a color fill (Fig.2). The pink color is for data (input/output raster, scalar, table, vector,...).The light green fill is for operators and operations. Each rectangle has ports,

small arrows that are connecting points for lines. The color fill of arrow depends on setting: red fill – an unset value, grey color – the set value, yellow is the output/derived value. Each rounded rectangle has an inner color icon that expresses the type of data or type of operation (e.g. icon π means scalar). The icons are very meaningful (expressive) and helpful in the cognition of diagram. Each symbol has inner label that is added automatically – e.g. the name of the operation. The label can be renamed by the user to better express the value or operation. For instance, the default label is “Scalar Value” can be renamed to “Scalar Value=5 (water)”. The labelling is considered as Dual Coding in Moody’s theory of “Physics of Notation”. The notation of Spatial Model Editor perfectly fulfills this idea that state: the graphical symbol has to be supplemented by text labels. The size of rectangles is different and depends on the number of ports (they can be added) or the length of the inner text labels.

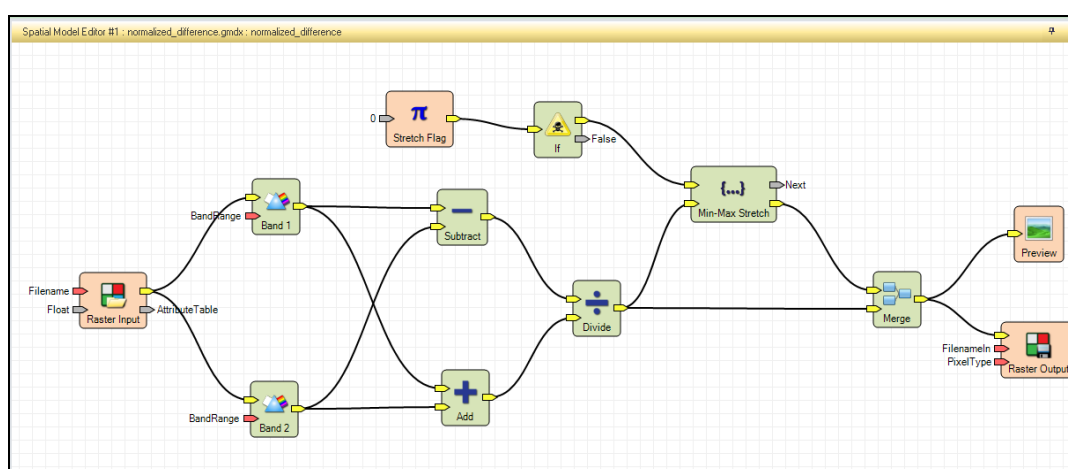


Fig. 2 Example model in Spatial Model Editor for EDRAS IMAGINE 2013

Very uncommon is the shape of connectors. They are the black curved lines that automatically changed the shape when the positions of rectangles have shifted. The Auto Layout function is available to align the graphical elements.

There are several differences between Model Maker and Spatial Model Editor, but one of them is not so evident. One function draw as one circle in Model Maker is depicted by several graphic elements in Spatial Model Editor. This change is space consuming in a diagram, however, the diagram is better readable.

Macro Modeler in IDRISI Selva

Clark Labs released the newest IDRISI Version 17 (known as IDRISI Selva) in January 2012. The component Macro Modeler is graphical modeling tool, exposes all of IDRISI’s GIS modules as objects that can be linked, dynamically and with feedback, with map layers in an algorithmic chain [11].

Graphical vocabulary in IDRISI has simple sharp rectangles for data and rhombus for operation. The rhombus has a pink color fill. The color fill of rectangles differs the type of data: violet is raster, green is vector, yellow is attribute (Fig. 3). Moreover, there are rectangles for dynamic group (red label) and group file. The size is the same for all symbols. The text labels differ in thickness. The label of operation is in bold face, others have normal face (Fig. 3). From the point of the principle of Perceptual Discriminability

of Moody’s theory [5], the level of discriminability is very low. Only color fill differs the symbols. No inner icons or text label with other information about data is in symbols.

The connecting lines are straight or right-angled dark blue lines with the arrow in the end. The extraordinary function is the construction of the process with feedback in that modeler. The color of the feedback line is red. The orientation of a flow is variable in editor. It depends on user choice. There is the possibility insert a submodel.

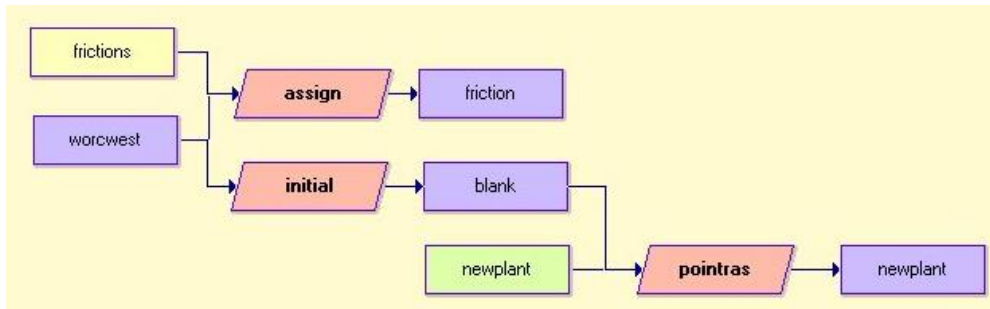


Fig. 3 Example model in Macro Modeler for IDRISI Selva

ModelBuilder in ArcGIS for Desktop

The vocabulary of graphic elements contains various shapes and color fills of elements in ModelBuilder. The ovals are input/output/derived data with blue/green fill. The rectangles for tools are yellow with inner icon of hammer (Fig. 4). The special tools are Iterators. A hexagonal orange symbol is used as an iterator. Iterators are tools that help to repeat the same operation or loop through a set of input data or values [12]. Detail description and assessment according to Moody’s theory is in [1]. The level of Perceptual Discriminability and Semiotic Clarity are good in ModelBuilder. Auto layout button is available.

There are more options to change the appearance of symbols. The color fill and shape can be changed optionally (any color, star, diamond symbol, etc.). The text label of an element can be renamed to the user label that expresses more information.

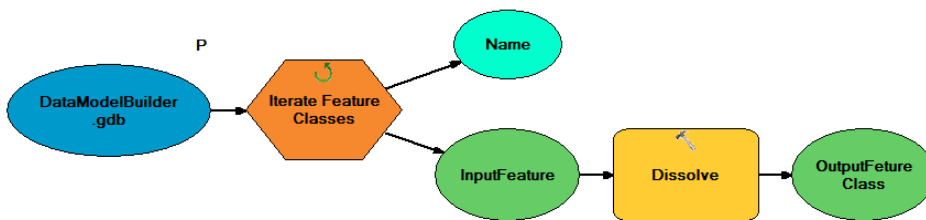


Fig. 4 Parametric model with iterator to dissolve all feature classes in dataset in ModelBuilder

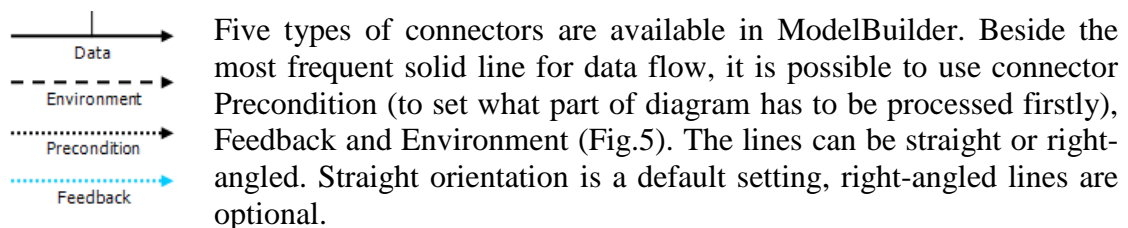


Fig. 5 Types of graphical connectors in ModelBuilder

Workflow Designer in AutoCAD Map 3D



Workflow Designer is one of newer editor of flow chart in GIS software. The symbols are only grey rectangles with several lines of labels. Data and processes have no distinguished graphic representation. The first text line, that is in bold, is a name of the operation. Description of each activity is supplemented by small icons on the right upper corner of its box (Fig. 6). Information about input and output data is included in a box with activities as a text. The names of input and output data are specified sometimes as parameter [6, 8].

It is also possible to skip a box with activities by disabling it. It is not necessary to remove the activity from dataflow, disabling is enough. The disabled activity changes its color to green. Boxes with Perform Overlay analysis are provided with an explanation picture of the type of overlay on the bottom. The picture is very illustrative and helps in the process of cognition. The dual coding by the icon and text fulfill the principle of Dual Coding.

Connectors are very short black lines that are added automatically to diagram. There is the fixed top-down orientation in diagram, nevertheless, there is a possibility to zoom in/out the diagram to see detail parts.

Fig. 6 Diagram in Workflow Designer in AutoCAD Map 3D

Programming languages usually differentiate two types of graphical elements (icons). One element for data and separate element for process. These simple elements depict primitive objects in the languages, whereas the composite graphical elements depict complex arrangements of simple object [13]. The graphical entities in the Workflow Designer are composite entities.

Processing Modeler in QGIS

There is a lack of visual programming languages in open source software in the area of GIS. Only one exists in software QGIS. The name of the component is Processing Modeler. The opportunity to combine tools graphically was the new ability in version 2.0. The newest version is QGIS 2.2 Valmiera. There is possible to build up complex analysis from a series of small single purpose modules. These models can be saved and then use them as building blocks in even more complex models [14]. The algorithm and commands can be takes from several libraries: GDAL/OGR, QGIS, Orfeo, GRASS, GeoServer/PostGIS tools and SAGA. Model-only tools are also available.

The graphical entities are three rectangles with the same dimensions. The rectangles differ by a color fill. The filling for input data is blue. The rectangle for the process has a white fill, and the rectangle for output data has light blue (turquoise) color (Fig.7).

The symbols differ by inner icon. The input data has the small plus symbol on the left side, the output data has a small arrow on the left side too. The inner symbol (on the left side) for tool depends on the origin library. Saga library is used in Fig. 7. The examples

of set of tools from different libraries with various inner icons are on Fig. 8. The input rectangle and tool rectangle have two icons on the left side: symbol x and pencil symbol. By pencil symbol, it is possible to set the parameters of tool. Alternative setting is by double click on the rectangle.

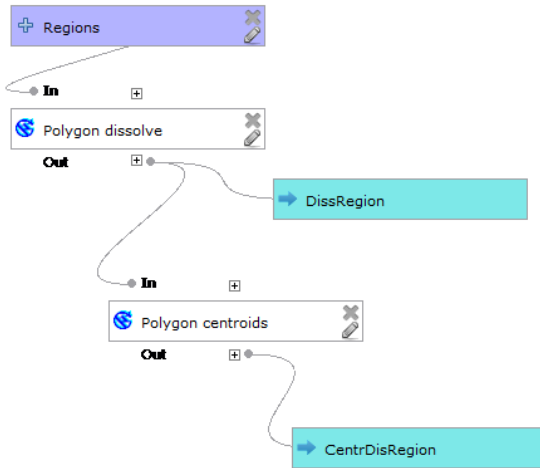


Fig. 7 Data flow diagram in QGIS Processing Modeler

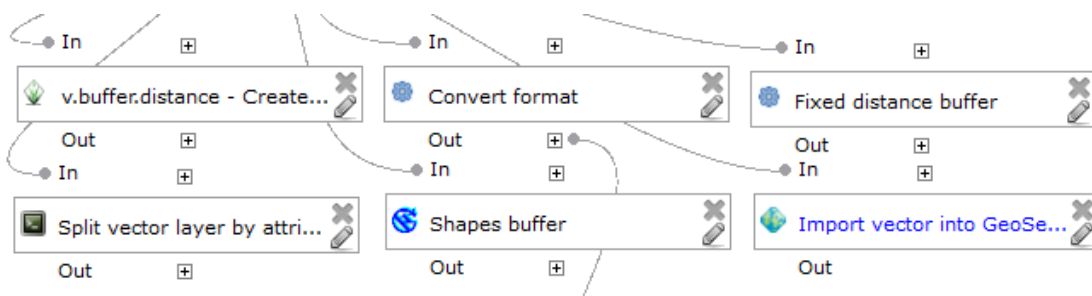


Fig. 8 Inner icon of tools (GRASS, ORG, QGIS, Script, Saga, Geoserver) in rectangles

The connectors are added automatically. Interesting is the shape, it is a curved line. The lines are automatically reshaped when rectangles are moved. The connection points are the points above and under the rectangles. Unpacking of a plus sign displays parameters. Given a model, it is possible to create automatically Python code that performs the same task as the model itself.

CONCLUSION

The article brings six examples of data flow diagram that are accessible in GIS. Three of them ERDAS Model Maker, IDRISI Macro Modeler and ArcGIS ModelBuilder were released nearly in the same time. The symbology and the concept of functionality are very close. They can be assumed as the first generation of dataflow diagrams in GIS. Nevertheless, ArcGIS ModelBuilder version 10 was improved by new symbols of Iterators and functionality.

The different and modern notations have ERDAS Spatial Model Editor and AutoCAD Map Workflow Designer. Graphic elements are various. Very young is notation in

QGIS Process Modeler. In fact, it has modern concept that can be expected to be improved in following versions. Basic functions as auto align, zoom in/out will be useful.

In formalizing the visual languages for DFDs, only the Workflow Designer has composite graphical entities, unlike others mentioned in the article. Others have simple graphical entities to depict primitive objects in the language (data and process).

Final overview of basic description of notations is in Table 1.

Table 1 Overview of graphic notation in GIS

GIS product	VPL component	Number of elements	Color	Shape of connectors	Orientation	Auto align
ERDAS IMAGINE	Model Maker	5	B/W	Straight lines	Variant	No
ERDAS IMAGINE	Spatial Model Editor	2	Yes	Curved lines	Variant (Left-right default)	Yes
IDRISI	Macro Modeler	6	Yes	Straight or right-angled lines	Variant	No
ArcGIS for Desktop	Model-Builder	6	Yes	Straight lines (default) right-angled (optionally)	Variant (Left-right default)	Yes
AutoCAD Map 3D	Workflow Designer	1 start+stop symbol	Yes	Straight lines (automatically added)	Top-Down	Yes automatic
QGIS	Processing Modeler	3	Yes	Curved lines	Variant	No

We are now at the stage of complex testing of DFD from the point of principles of “Physics of Notation”. In addition, the tests by eye-tracking equipment were realized. The partial results agree with the statement of the article [15]. The perceptual discriminability and dual notation have a stronger effect on process diagram comprehension than semantic transparency and aesthetics. Perceptual discriminability is achieved by shape, color, inner icons supplemented by text labels. The semantic transparency states that visual appearance suggests their meaning. It only appears at Model Maker notation that is obsolete and pushed up by new Spatial Model Editor in ERDAS IMAGINE.

ACKNOWLEDGEMENT

Work was supported by the project CZ.1.07/2.3.00/20.0166.

REFERENCES

1. Dobesova, Z. Using the “Physics” of Notation to Analyse ModelBuilder Diagrams. SGEM 2013 13th International Multidisciplinary Scientific GeoConference, Albena, Proceedings Volume I STEF92 Technology Ltd. Sofia, Bulgaria, p. 595–602, ISBN 978-954-91818-9-0,
2. ERDAS IMAGINE Ad, PE&RS May 1993. American Society for Photogrammetry and Remote Sensing (ASPRS), Bethesda, Maryland, p568.
3. Wheate R. D. Software Review - Idrisi32 Release 2. Retrieved 24. 4. 2014 Available: <http://www.asprs.org/a/resources/software/review-08-2002/>
4. Green, T.R.G. and M. Petre, Usability analysis of visual programming environments: A 'cognitive dimensions' framework. *Journal of Visual Languages and Computing*, 1996. 7(2): p. 131-174.
5. Moody, D. L., The "Physics" of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering. *IEEE Transactions on Software Engineering*, 2009. 35(6): p. 756-779.
6. Dobesova, Z. Visual programming language in geographic information systems. In *Recent Researches in Applied Informatics, 2nd International Conference on Applied Informatics and Computing Theory*, AICT 11. 2011. Prague: NAUN/IEEE, WSEAS Press.
7. Dobesova, Z. and Dobes, P. Differences in visual programming for GIS . *Applied Mechanics and Materials Vols. 519-520*, Computer and Information Technology Trans Tech Publications, Switzerland, 2014, p. 353-356. doi:10.4028/www.scientific.net/AMM.519-520.353,
8. Dobesova, Z. and P. Dobes. Comparison of visual languages in Geographic Information Systems. In *IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. 2012.
9. Dobesova, Z. Strengths and weaknesses in data flow diagrams in GIS. *International Conference on Computer Sciences and Applications*, IEEE, Computer Society, Wuhan, China, Dec. 14th-15th, 2013, 803-807 p. ISBN 978-0-7695-5125-8, DOI 10.1109/CSA.2013.192
10. ERDAS IMAGINE® 2013 Features Next-Generation Spatial Modeler. Available: <http://www.intergraph.com/assets/pressreleases/2012/10-23-2012b.aspx>. Retrieved 24. 4. 2014.
11. IDRISI Brochure. Retrieved 24.4.2104. Available: <http://www.clarklabs.org/applications/upload/IDRISI-Selva-GIS-Image-Processing-Brochure.pdf>
12. ArcGIS Help 10.2. A quick tour of using iterators. Retrieved 30.4.2014 Available: <http://resources.arcgis.com/en/help/main/10.2/index.html#//002w0000001w000000>
13. Kulkarni, G., Sathyaraj, R. A Survey on Graphical Programming Systems. *CompuSoft* 3(4), April 2014, 709-713 p., ISSN 2320-0790
14. QGIS: What’s new in QGIS 2.0, Retrieved 29.4.2014. Available: https://www.qgis.org/en/docs/user_manual/preamble/whats_new.html
15. Figl, K., Recker, J., Mendling, J. A Study on the effects of routing symbol design on process model comprehension. *Decision Support Systems* (54) 2, 2013, pp. 1104-1118.