# DBCells – an open and global multi-scale linked cells

Sergio Souza Costa, Evaldinolia Gilbertoni Moreira, Micael Lopes da Silva, Thamyla Maria de Sousa Lima

it is not about results, it is about an idea

# Outline

- Motivation
- Introduction
- The architecture
- Challenges
- Final remarks

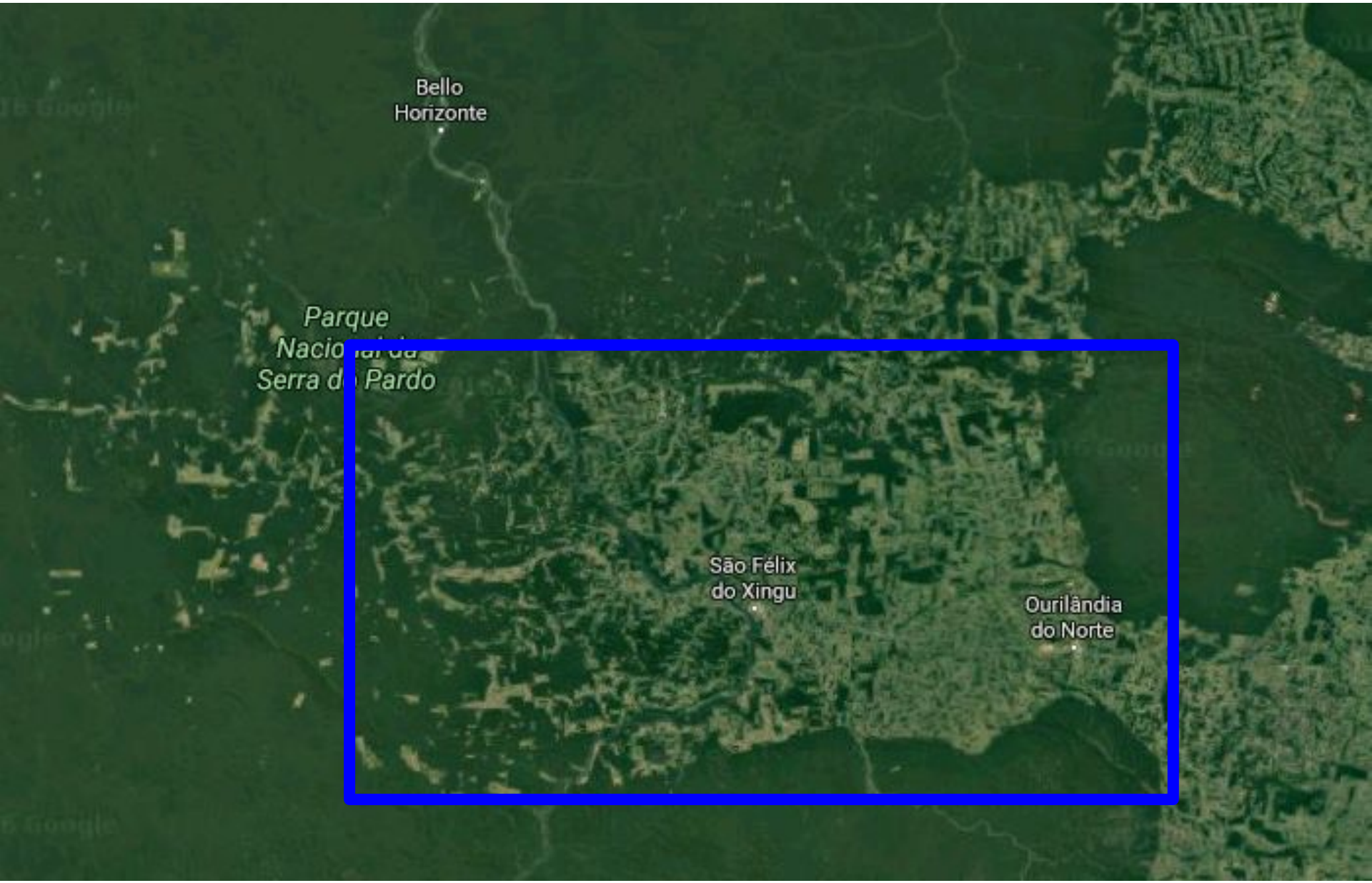
# Motivation

Land change models demand data from environmental, social, technological, and political drivers [Moran et al. 2005, Turner et al. 2007].

# How to share model's data ?

# How to reuse model's data ?

# How to reuse model's data ?



# How to reuse model's data ?

Open linked data allows land change models to become more reusable and shareable

# Introduction: Open Linked data

- Data are the base of the pyramid, and are defined as symbols that represent properties of objects, events and their environment [Ackoff 1989].
- Open data are all those that can be freely used, modified, and shared by anyone for any purpose [The Open Definition 2013].
- The linked data refers to a set of best practices for publishing and interlinking structured data on the Web [Heath and Bizer 2011].

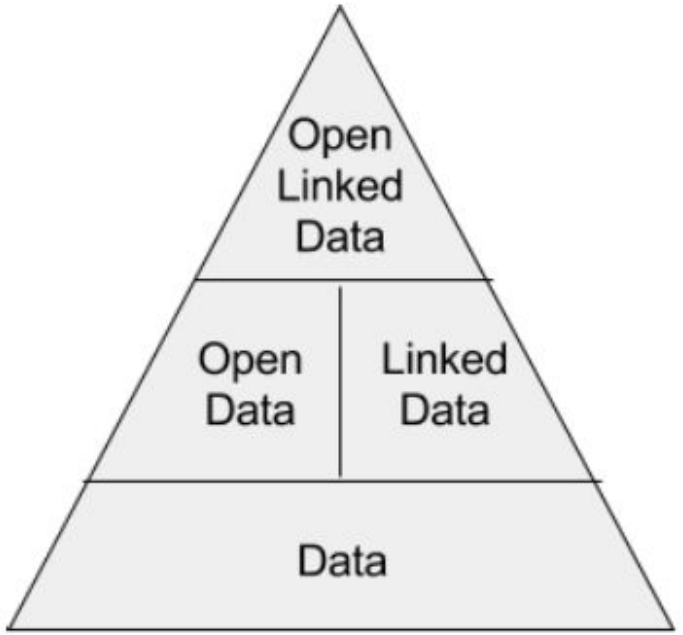



Figure 1. From data to open linked data The

# Introduction: Open Linked data


Figure 2 - Five-stars rating. Source: http://5stardata.info/en/

# Introduction: A global open linked data

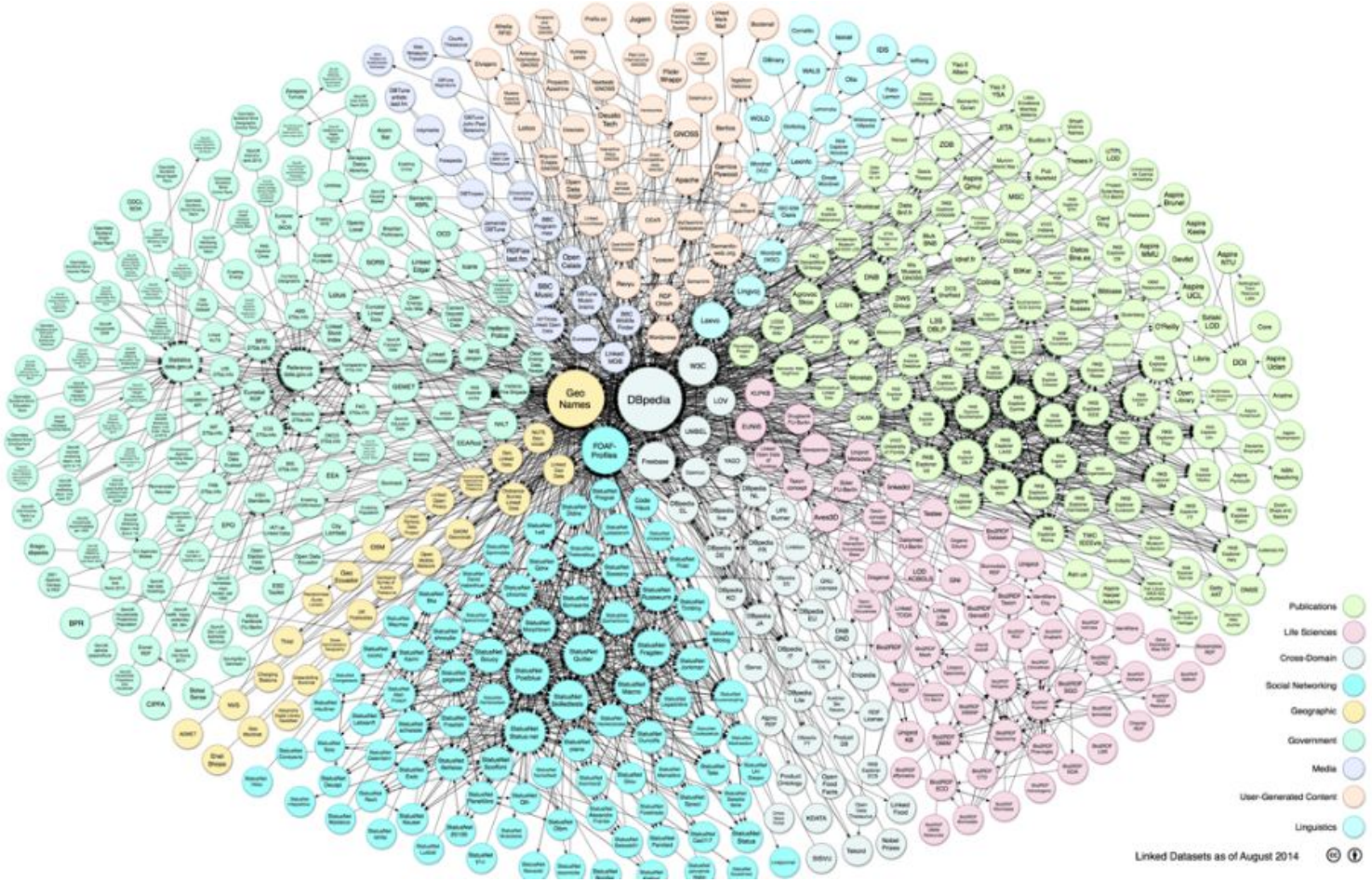



Figure 3 - Linked open data cloud. Source: lod-cloud.net. Last updated: 2014-08-30

# Introduction: Open Science

*“Better science—in terms of transparency, reproducibility, increased efficiency, and ultimately a greater benefit to society—depends on open data” (*Molloy,2011).

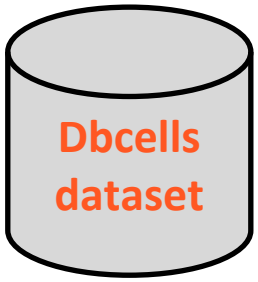
In [Kauppinen and De Espindola 2011], the authors propose the Linked Open Science aiming to be a standardized and generic recipe for executable papers.

A. Linked Data,
B. OpenSource and Web-based Environments
C. Cloud Computing and
D. Creative Commons.

An example of linked open data is the Linked Brazilian Amazon Rainforest Data [Kauppinen et al. 2014].

# The architecture

Publication layer




# Architecture: dbcells dataset

# Cellular space

Our focus is on data from land change models. In general, these models describe phenomena that vary continuously in space and time, as deforestation in the Brazilian Amazon region. Their input and output are represented as geo-fields and discretized in two ways [Kuhn 2012]:

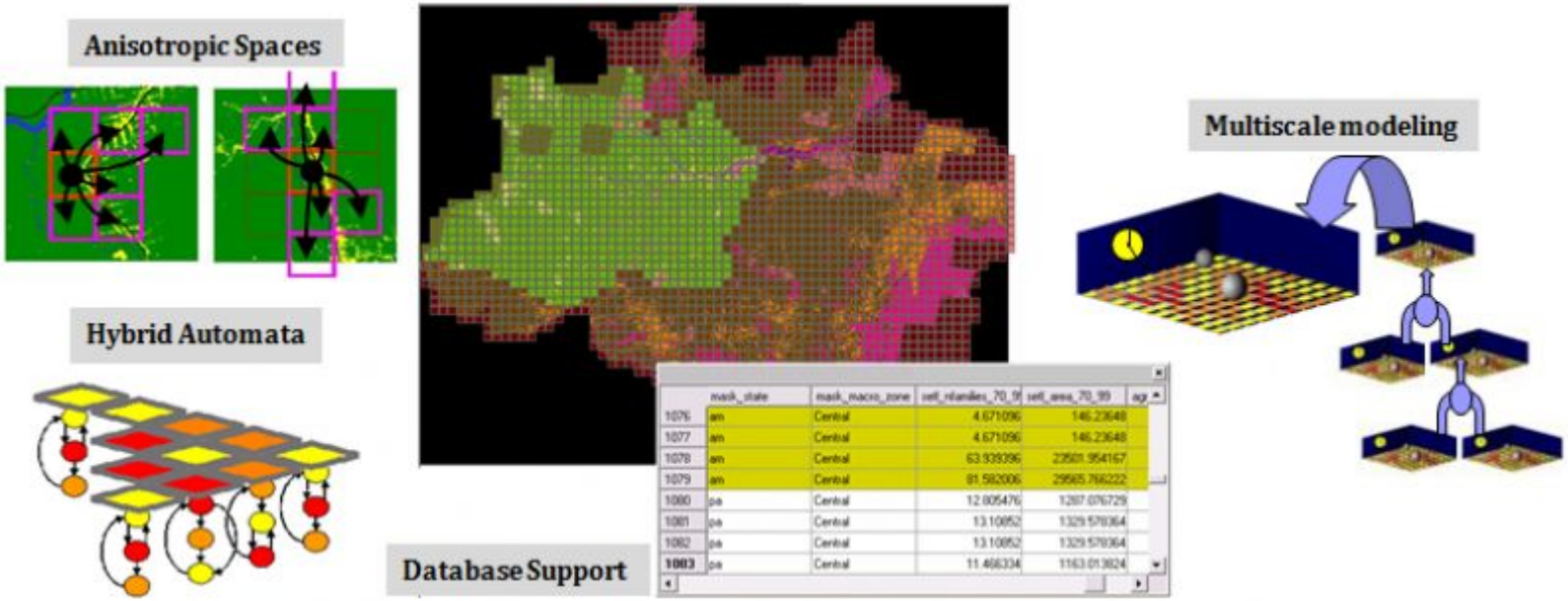
1. through a finite number of cells, within each one the attribute is assumed to remain constant;

**2. through a finite set of sample points with interpolation rules for positions among them.**

In this paper, we are interested in the first way, where the study area is partitioned forming a regular grid of square, triangular, hexagonal, or cubic cells as in **raster based layers** or a **cellular space**.

# Cellular space

Cellular spaces have been used for simulation of urban and environmental models as part of cellular automata models (Batty 2000). In TerraLib ([Câmara et al. 2008]) and TerraME ([de Senna Carneiro et al. 2013]) the cellular space is a native building block.

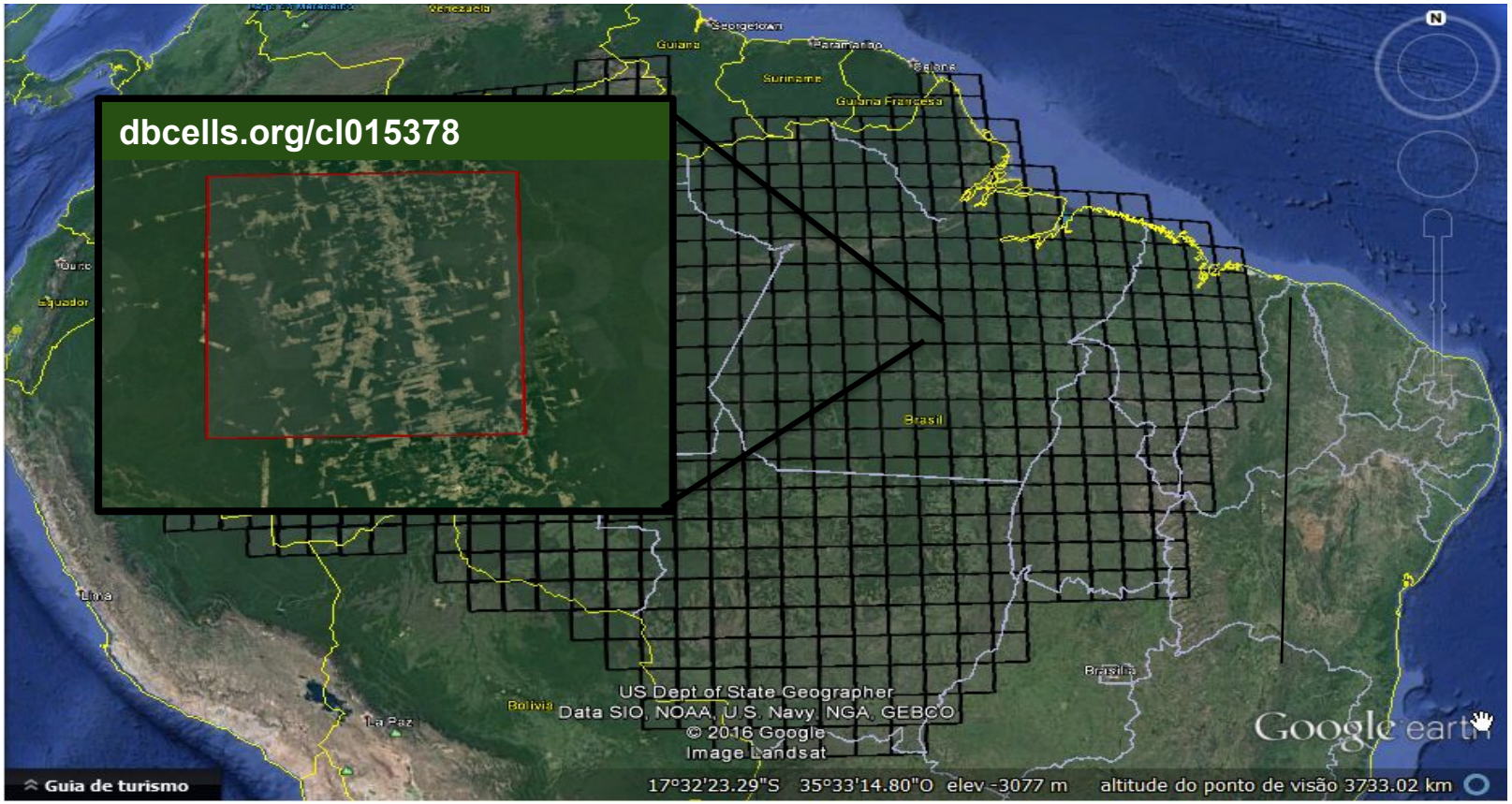



http://www.terrame.org/doku.php

# Divides the space in regular cells

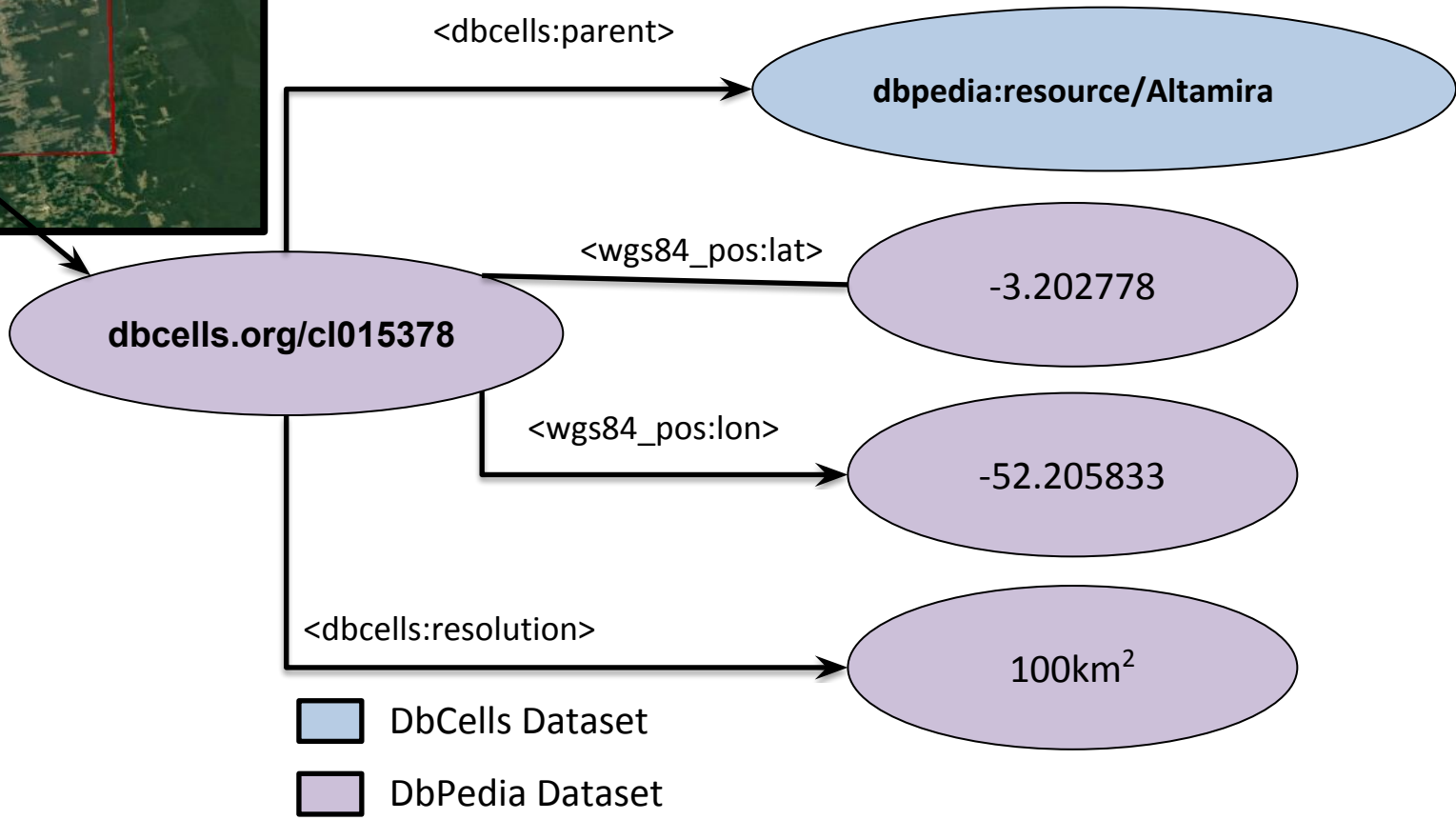
# Associates each cell to an identifier

The Uniform Resource Identifier (URI) is one of the pillars of the web data architecture, which links **data** instead of **pages**.

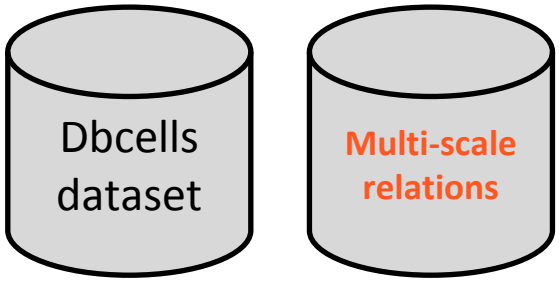


# Represents each cell as an RDF graph

*"The vision for RDF was to provide a minimalist knowledge representation for the Web".*
(Shadbot et. al. 2006)

Publication layer




# Architecture: multi-scale relations

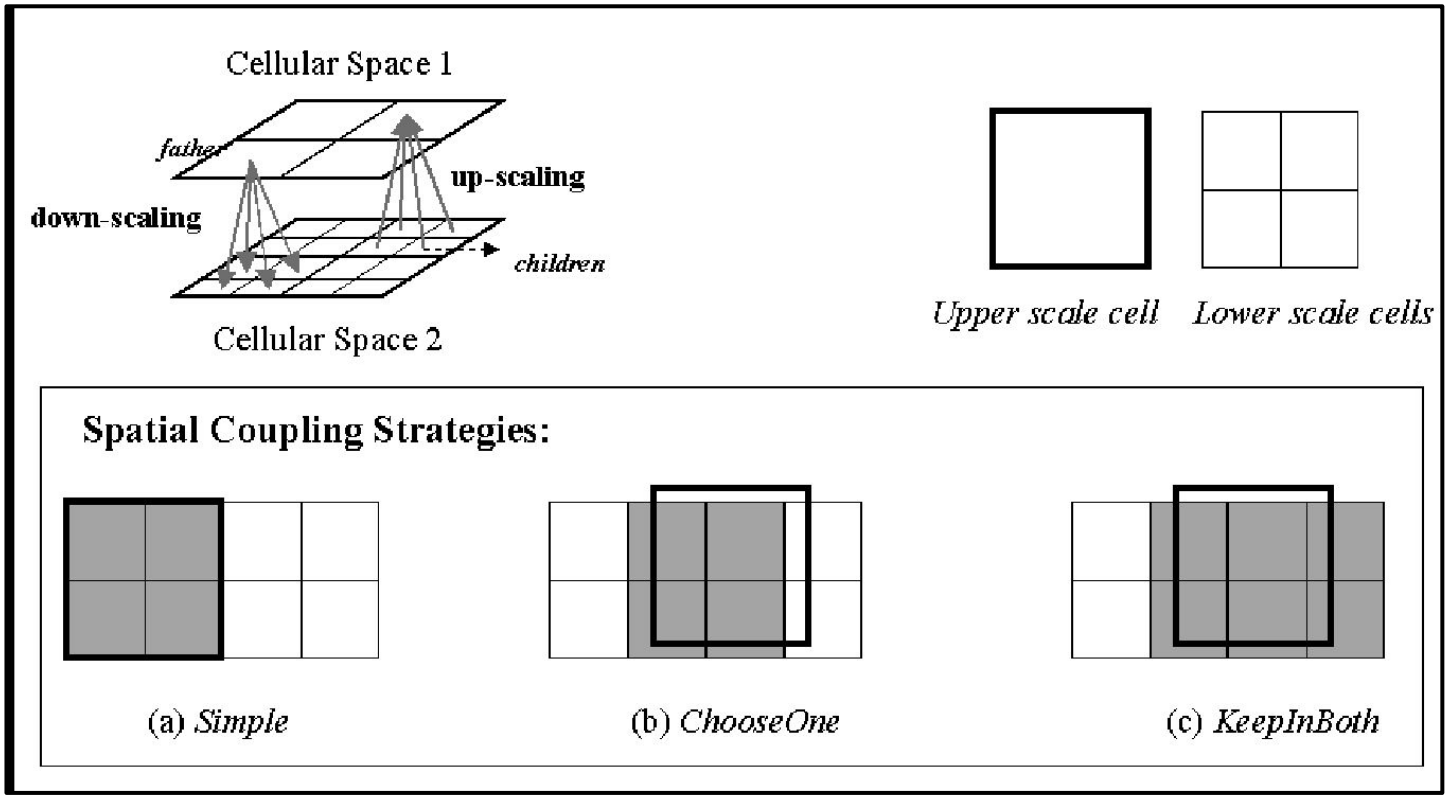
# Multi-scale relations

Our proposal is to describe both, the cells and their relationships, **through RDF graphs.** Graphs express different relations, including:

A. **topological relations**;
B. **network connectivity**, both physical (e.g., transportation infrastructure) and logical (e.g., trade fluxes);
C. **vicinity** in cell spaces and grids;
D. **coupling** between spatial scales.

We propose to describe the relationships in different datasets, allowing a model to select one or more relationships.

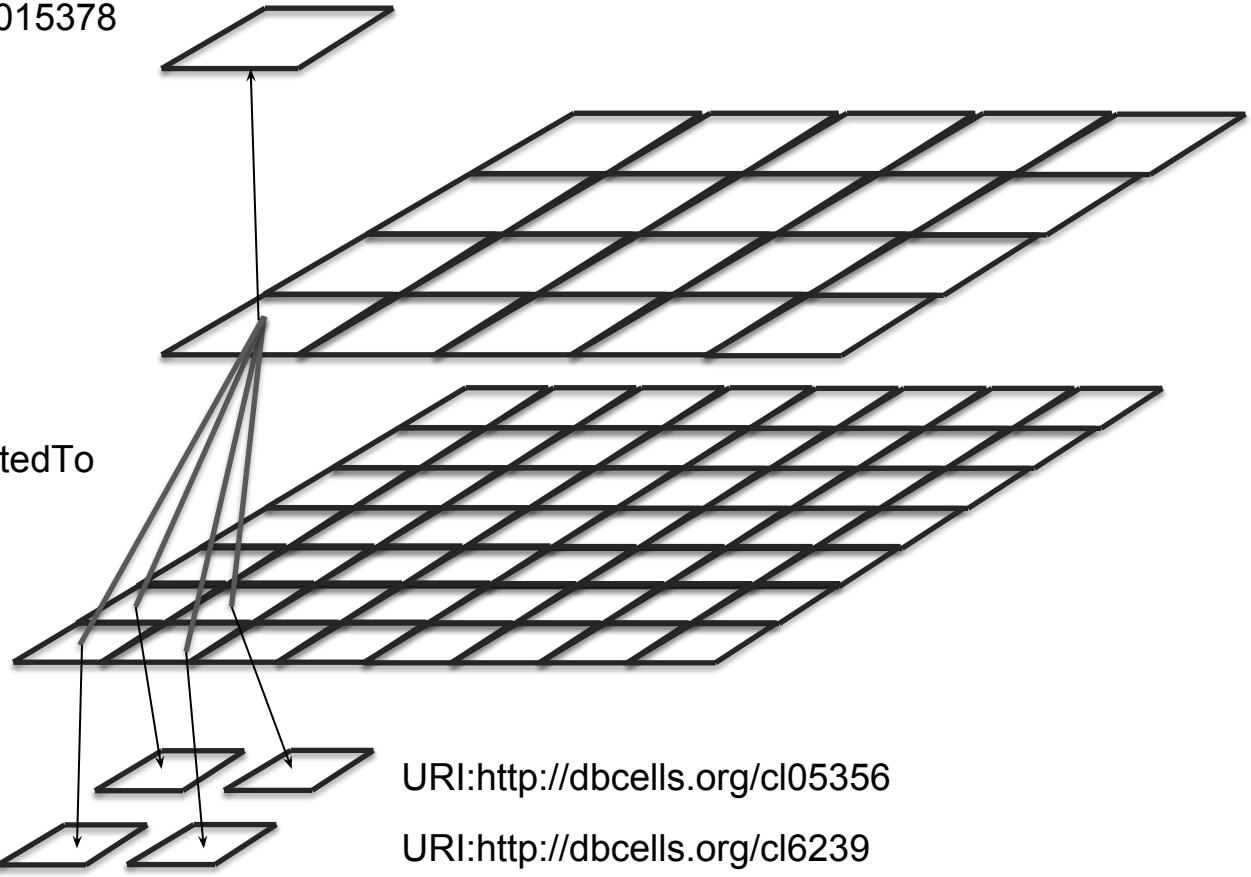
# Multi-scale relations




Representation of strategies for spatial coupling in the case of regular cells. Source: [Moreira et al. 2008]

# Multi-scale relations

URI:http://dbcells.org/cl015378



URI:http://dbcells.org/onto/relatedTo

URI:http://dbcells.org/cl3278

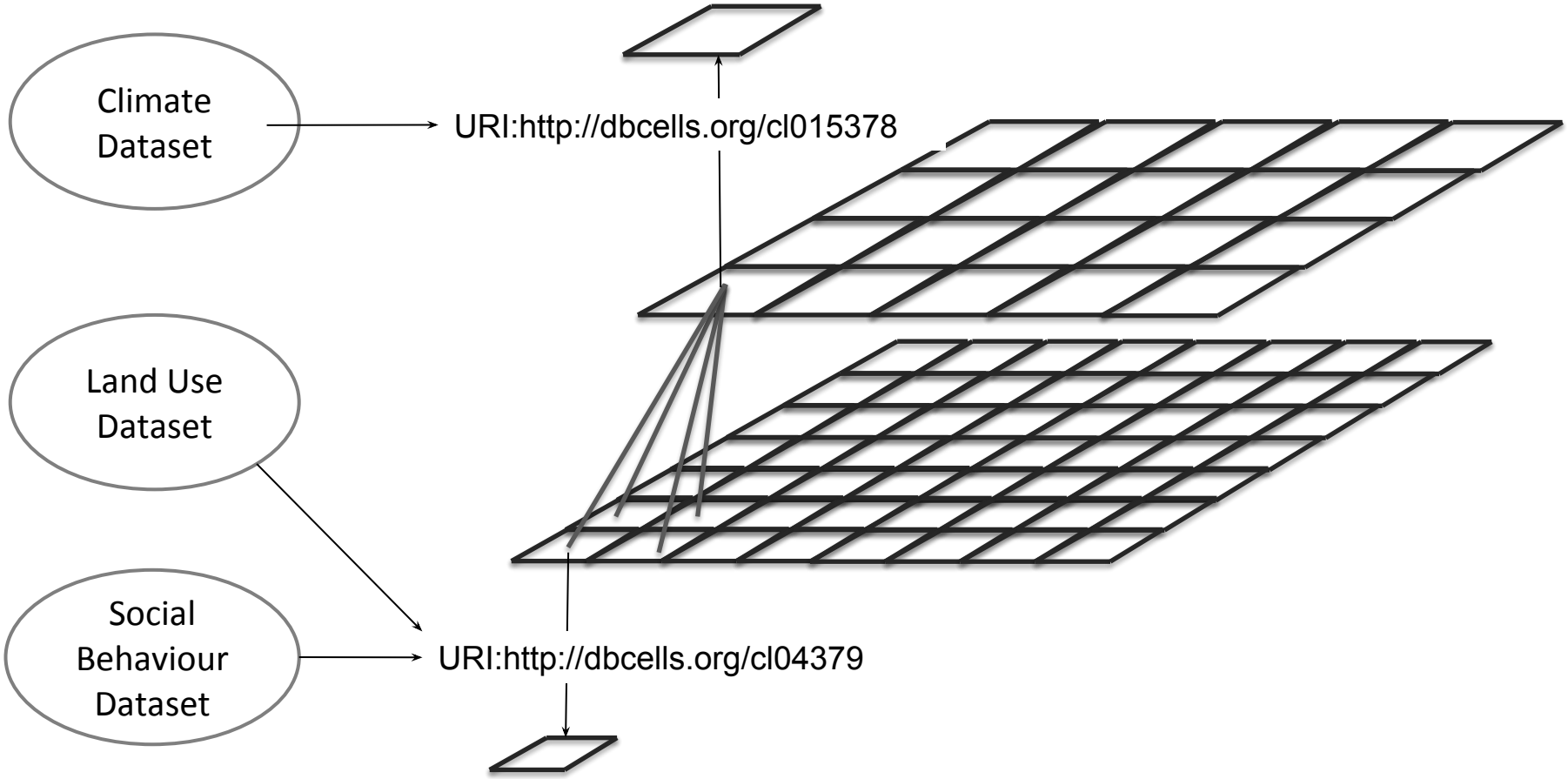
URI:http://dbcells.org/cl5682

URI:http://dbcells.org/cl05356

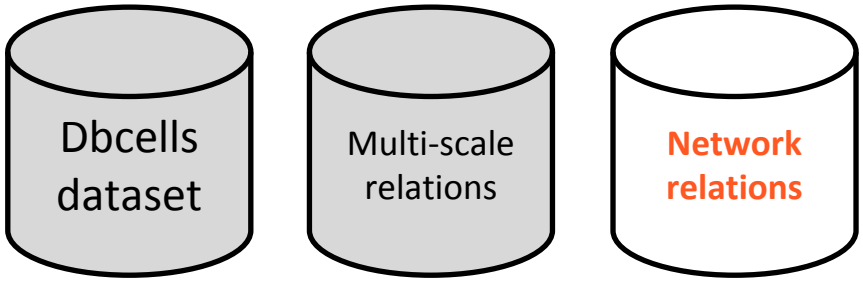
URI:http://dbcells.org/cl6239

# Models at different scales

The relationships between cellular spaces allow the reuse of models at different scales and resolutions. For example, a land use model at a finer scale can use results of a climate model in a coarser scale

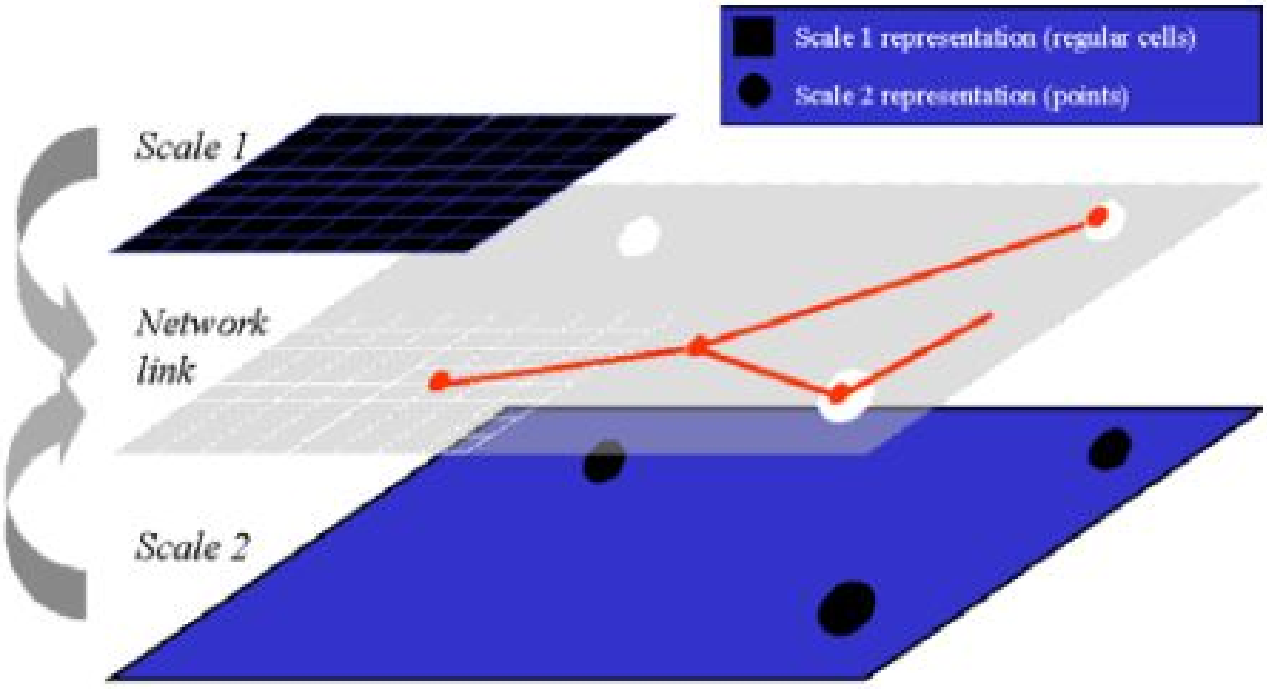


Publication layer



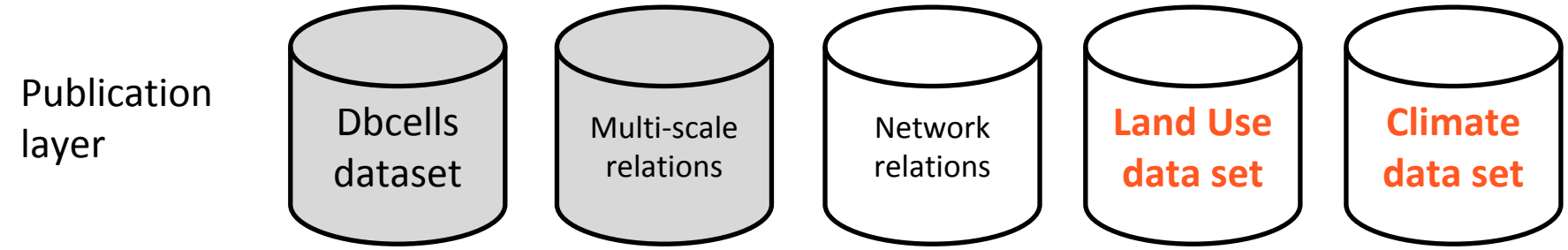

# Architecture: network relations

# Network relations



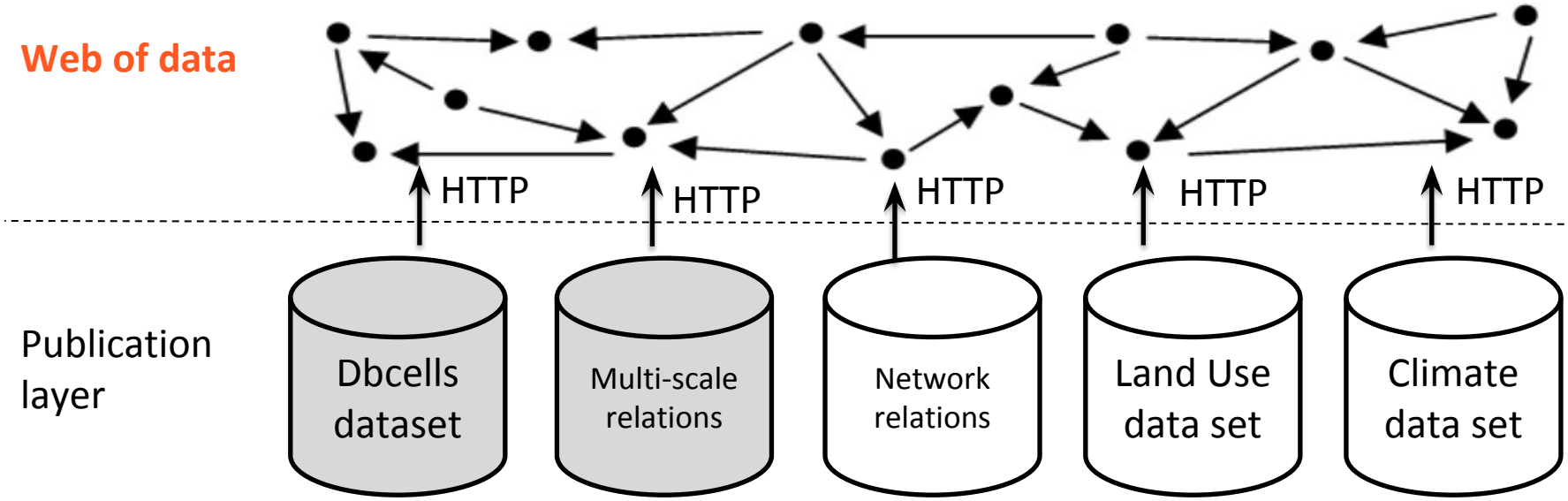

Schematic representation of a network-based spatial relation between cell objects in Scale 1 and point objects in Scale 2.
Source: [Moreira et al. 2008]

The **publication layer** includes a dataset of cellular space, multi-scale relationships and open linked data models.



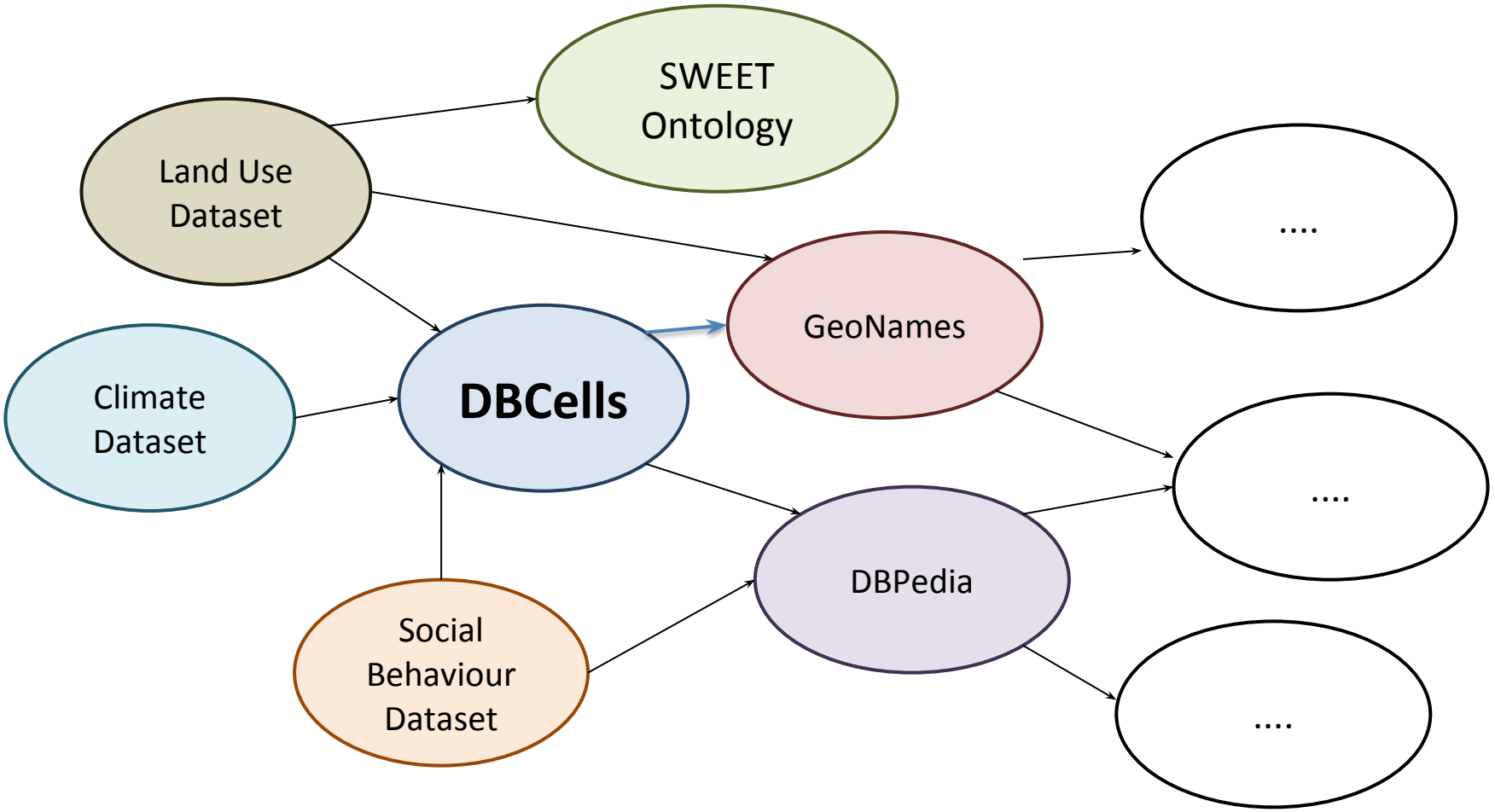

# Architecture: models

The **web of data layer** links the cellular space to existing datasets, like Geonames, DBPedia and SWEET Ontology.

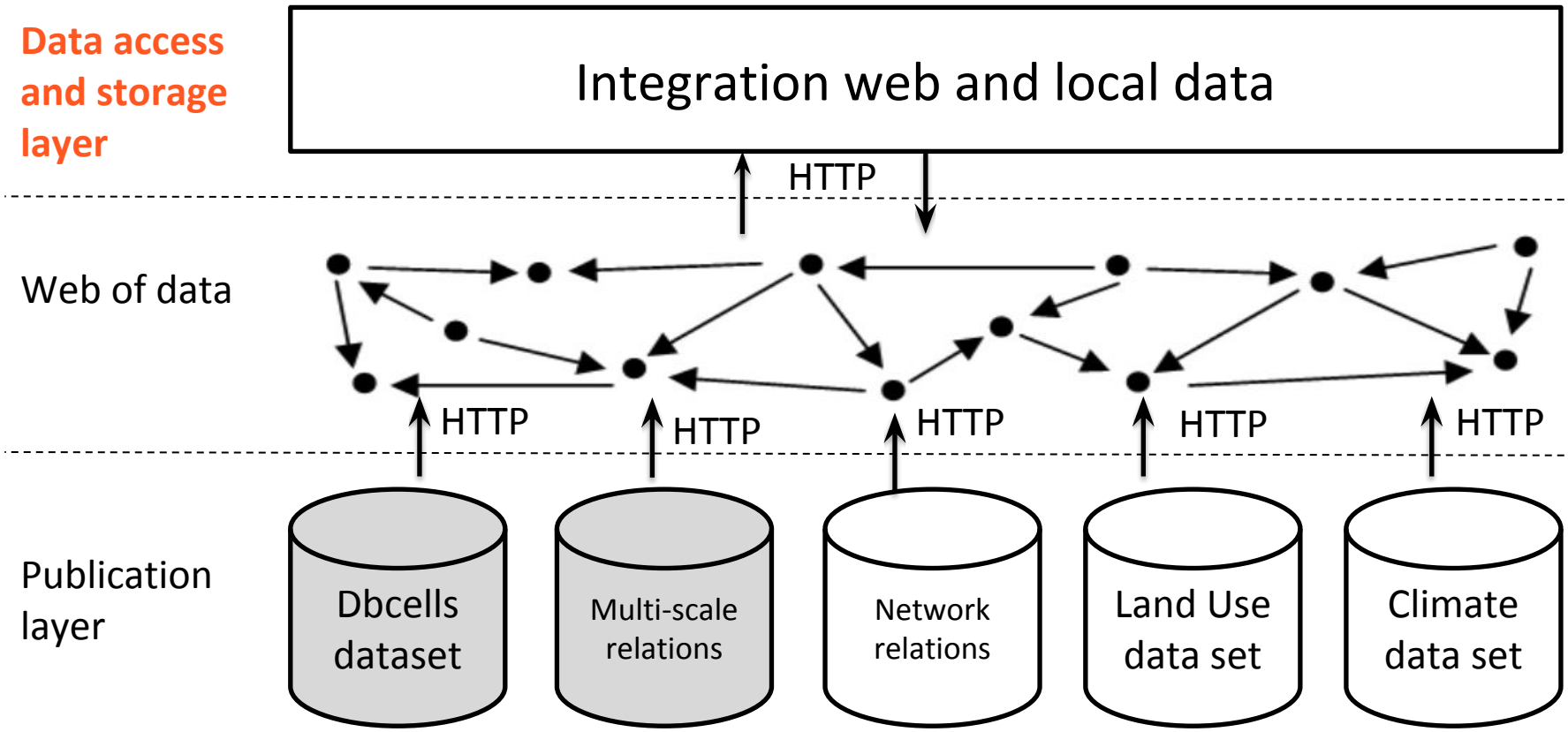



# Architecture: web of data

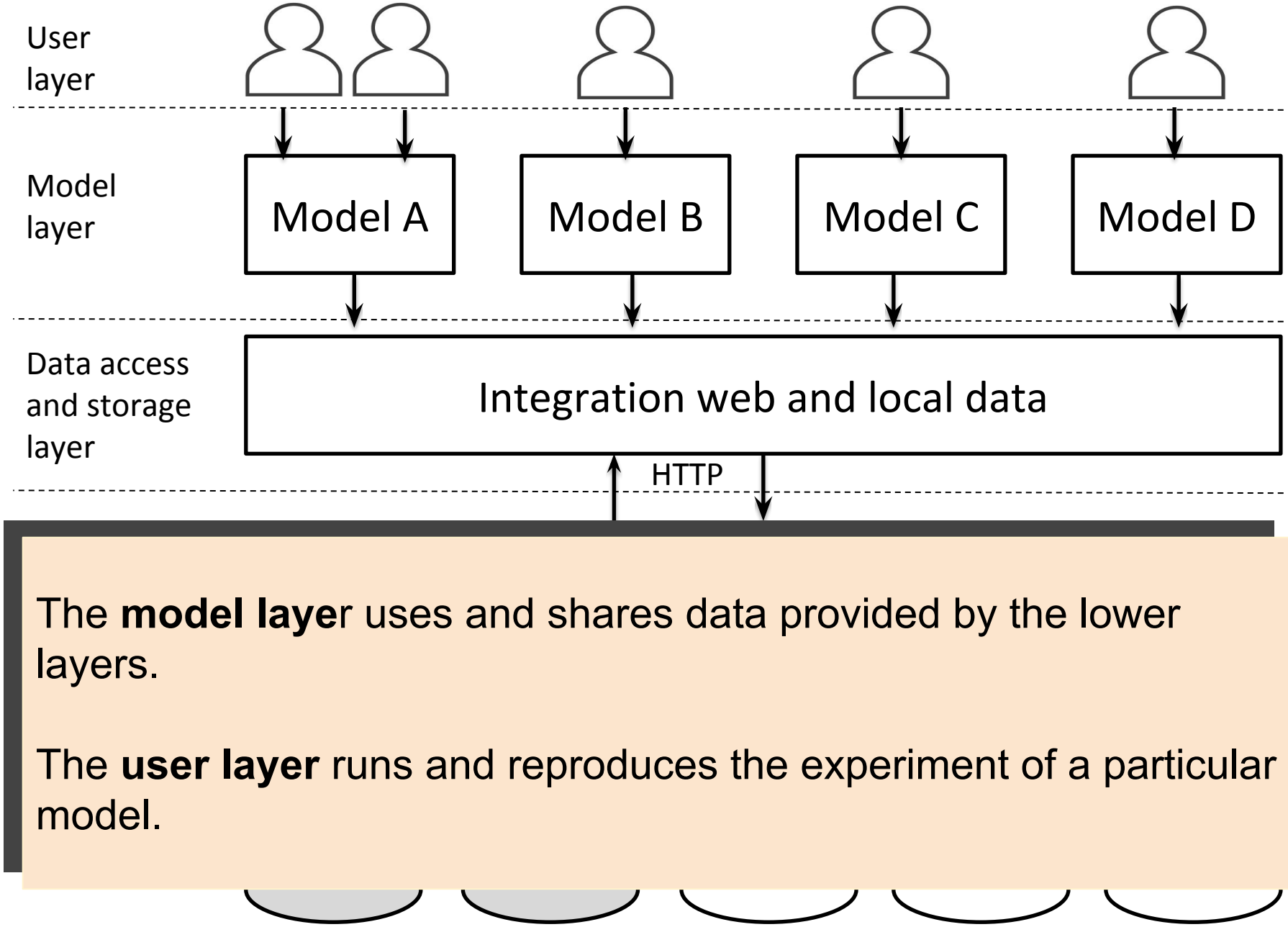
# Web of data layer



The **data access and storage layer** integrate local and web data, providing a transparent access and storage for modeling tools.



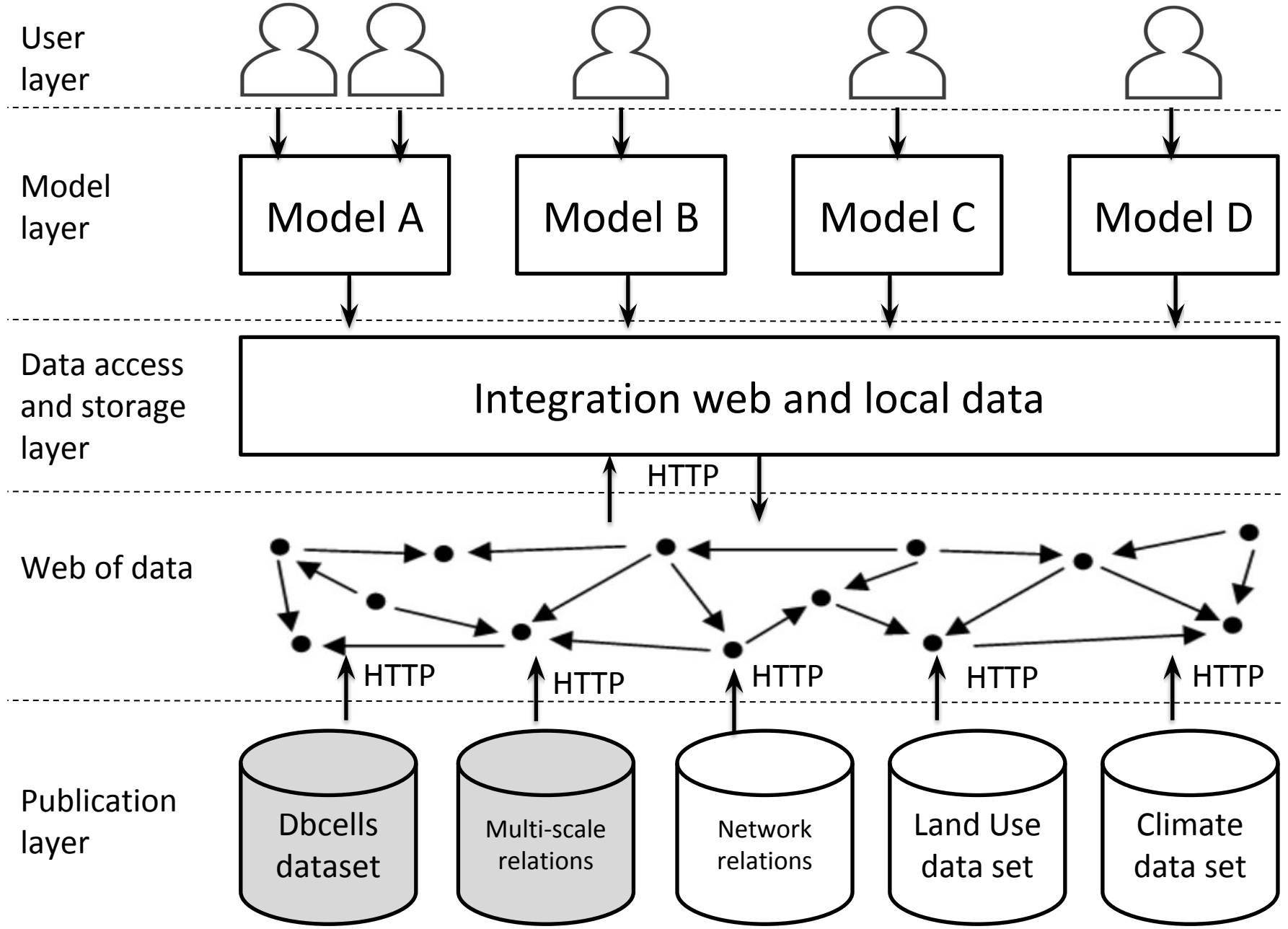

# Architecture: Data access and storage layer




The **model layer** uses and shares data provided by the lower layers.

The **user layer** runs and reproduces the experiment of a particular model.

# Architecture: model and user layers




# Architecture

# Challenges

# Challenges

- Engagement of the scientific community, interest in making their data open and linked.
- Provide data access and storage layer to retrieve and store data on the web.
- Use the already established vocabularies.
- Provide an efficient and distributed computing will be necessary for storage and retrieval of data from a global cellular space at different scales.
- ...

# Final remarks

This paper introduced an innovative architecture – DBCells – that integrates two concepts: cellular spaces and linked data.

The pillar of integration is to treat each cell as a unique and distinct entity that has a universal identifier. To achieve this integration, we propose four steps:

1. divide the space in regular cells,
2. associate each cell to an identifier,
3. represent each cell as an RDF graph available on the web and
4. connect data and results models to these identifiers.

# Final remarks

The main benefits of the new approach are:

1. the reuse,
2. sharing,
3. comparison and
4. reproduction of land change models.

The main challenges are:

1. the participation and interest of the scientific community, and
2. an efficient architecture to store and retrieve large volume of data.

# References


Ackoff, R. L. (1989). From data to wisdom. Journal of applied systems analysis, 16(1). Aguiar,

Moran, E., Ojima, D., Buchmann, N., et al. (2005). Global land project: Science plan and implementation strategy. IGBP Report, 33.

Moreira, E., de Aguiar, A. P. D., Costa, S. S., and Camara, G. (2008). Spatial relations across scales in land change models. In GeoInfo, pages 95–108.

The Open Definition (2013). Open definition. opendefinition.org/od/2.0/pt-br/.

Kauppinen, T. and De Espindola, G. M. (2011). Linked open science-communicating, sharing and evaluating data, methods and results for executable papers. Procedia Com- puter Science, 4:726–731.

Kauppinen, T., De Espindola, G. M., Jones, J., S??nchez, A., Gr??ler, B., and Bartoschek, T. (2014). Linked Brazilian Amazon Rainforest Data. Semantic Web, 5(2):151–155.

Kuhn, W. (2012). Core concepts of spatial information for transdisciplinary research. International Journal of Geographical Information Science, 26(12):2267–2276.