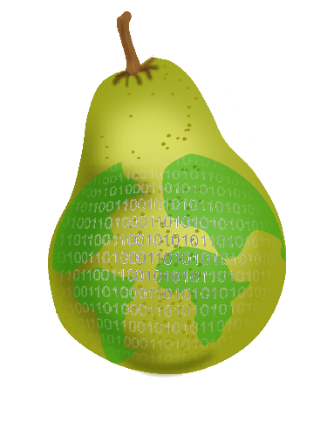
­IBM PAIRS Services

Corpus “basic”

**IBM PAIRS Services:**

**Corpus “basic”**

****

Home Page on the IBM Marketplace:

<https://www.ibm.com/us-en/marketplace/geospatial-big-data-analytics>

Home Page on the IBM developerWorks API Explorer:

<https://developer.ibm.com/api/view/pairs-prod:pairs-api>

Signup URL for the IBM PAIRS Services

<https://www.ibm.com/account/us-en/signup/register.html?a=PAIRS&ctx=C001&cc=us&lc=en&trial=yes&quantity=1&catalogName=Master&partNumber=PAIRS_FREE&siteID=ECOM>

Public GitHub Repository of IPS Client Samples

<https://github.com/webchang/ibm-ips-samples>

IBM Physical Analytics Integrated Repository (PAIRS) User Agreement:

<http://pairs.mybluemix.net/public/PAIRS_Click_License.pdf>

Documentation Modification Date: 2017-11-30

Latest version is available for download at:

<https://github.com/webchang/ibm-ips-samples/blob/master/ips-corpus-basic.pdf>

**Note**

Before using this information and the product it supports, read the information in [Notices](#4.Appendix. Notices|outline)

**Book Edition Notice**

This edition applies to Release 1.0 of IBM PAIRS Services and to all subsequent releases and modifications until otherwise indicated in new editions.

**© Copyright IBM Corporation 2017.**

US Government Users Restricted Rights – Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

[1 Introduction 5](#_Toc499816866)

[1.1 Query Processing Model 6](#_Toc499816867)

[1.2 Spatial Resolution Levels 7](#_Toc499816868)

[2 Corpus “basic” 8](#_Toc499816869)

[2.1 Corpus Datasets and Datalayers 9](#_Toc499816870)

[Dataset 5: MODIS, Aqua, 13 (Global): NASA 250m resolution satellite 9](#_Toc499816871)

[Dataset 6: MODIS, Aqua, 09 (Global): NASA 250m resolution satellite 9](#_Toc499816872)

[Dataset 7: MODIS, Terra, 13 (Global): NASA 250m resolution satellite 10](#_Toc499816873)

[Dataset 8: MODIS, Terra, 09 (Global): NASA 250m resolution satellite 10](#_Toc499816874)

[Dataset 9: PRISM historical climate data (USA) 10](#_Toc499816875)

[Dataset 11: Historical crop planting map (USA) 11](#_Toc499816876)

[Dataset 12: NAM USA weather forecast (USA) 11](#_Toc499816877)

[Dataset 13: California weather condition measurements 11](#_Toc499816878)

[Dataset 14: USGS national elevation data (USA) 12](#_Toc499816879)

[Dataset 15: IBM Evapo-transpiration/irrigation (Global) 12](#_Toc499816880)

[Dataset 16: NOAA Global Forecasting System (Global) 12](#_Toc499816881)

[Dataset 17: IBM blended weather forecast (USA) 13](#_Toc499816882)

[Dataset 24: Daymet historical weather (USA) 13](#_Toc499816883)

[Dataset 25: IBM long term forecast (Global) 13](#_Toc499816884)

[Dataset 34: IBM Research drone 14](#_Toc499816885)

[3 Update History 15](#_Toc499816886)

List of Figures

[Figure 1. Conceptual Query Processing Model of PAIRS. 6](#_Toc499816887)

[Figure 2. Filtering and Joining Datalayers with Different Spatial Grid Resolutions. 6](#_Toc499816888)

List of Tables

[Table 1. Uses Cases for (open) geo-spatiotemporal data. 5](#_Toc499816889)

[Table 2. Global grid spatial resolution levels. 7](file:///C:\rong\proj\IoT\PAIRS\ips\ips-corpus-basic.docx#_Toc499816890)

[Table 3. List of PAIRS “basic” datasets. 8](file:///C:\rong\proj\IoT\PAIRS\ips\ips-corpus-basic.docx#_Toc499816891)

About this document

IBM PAIRS Services (IPS) is a managed API offering that provides geospatial big data curation, repository, and cognitive physical analytics functions via a cloud-centric simple and scalable REST API. PAIRS stands for *Physical Analytics Integrated Data Repository and Services*.

This document describes the “basic” (dataset) corpus supported by the IBM PAIRS Services. Different datasets listed in this document have different geospatiotemporal coverages. Spatial and temporal coverage of a specific datalayer in a specific dataset can be obtained via the IPS REST operation “GET /datalayers/{id}”, where “{id}” is the id of the target datalayer.

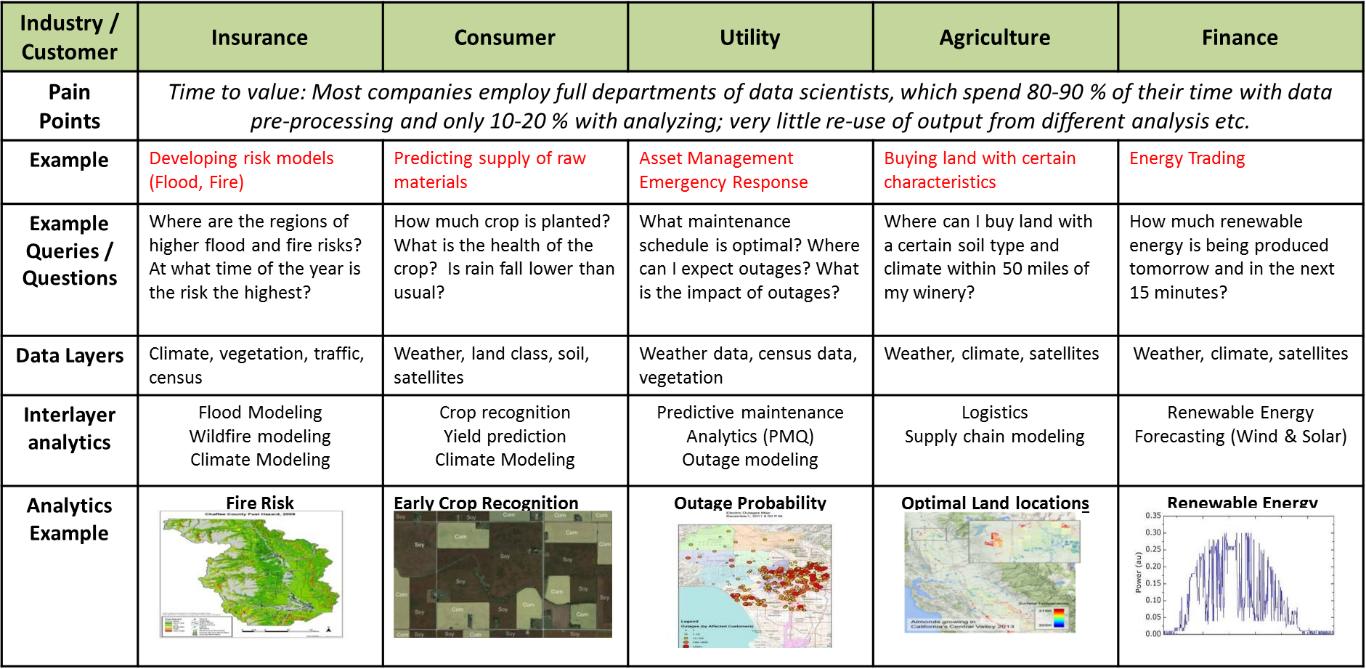
Please provide your questions or suggestions about this document by opening GIT issues at <https://github.com/webchang/ibm-ips-samples>

# Introduction

IBM PAIRS Services is a managed API offering that provides geospatial big data curation, repository, and cognitive physical analytics functions via a cloud-centric simple and scalable REST API. PAIRS stands for *Physical Analytics Integrated Data Repository and Services*. This API offering includes a continuously-updated petabyte repository of curated geo-spatiotemporal data, which lowers big data management and time-to-discovery cost significantly for its customers. It employs cognitive physical modeling and analysis technologies to provision, e.g., high accuracy weather forecasting and agricultural models. Unique industry specific analytics of PAIRS include, among other use cases, global irrigation forecasts, global seasonal weather forecasts, US wide renewable energy generation forecasts, and crop acreage forecasts. Its query performance is scalable in terms of the searched data size. Finally, it supports IBMid and account based self-management of subscriptions, subscribers, and lifecycle of API keys (including API key creation, sharing, and revocation).

Table 1 lists several industrial use cases for PAIRS applications that exploits (open) geo-spatiotemporal data.

Table 1. Uses Cases for (open) geo-spatiotemporal data.



Further details about the IBM PAIRS Services offering are available from the IBM Marketplace at <https://www.ibm.com/us-en/marketplace/geospatial-big-data-analytics> and from the IBM developerWorks API Explorer at <https://developer.ibm.com/api/view/pairs-prod:pairs-api>

## Query Processing Model

Figure 1. Conceptual Query Processing Model of PAIRS.

Every PAIRS query targets at one or more *datalayers*, which are the units of composition of *datasets*. The relation between datalayers and datasets is many-to-one (or child-parent), and data access control is employed only at the dataset level per the IBMid in use. For example, a satellite dataset could comprise of several datalayers, each of which refers to a specific band of satellite images. Moreover, a weather dataset could be composed of several datalayers, each of which represents a set of weather modeling parameters. PAIRS analytics capabilities are delivered through the abstraction of datalayers (e.g., integrated indexing of heterogeneous sets of geo-spatiotemporal data records for a datalayer) and the processing of PAIRS queries. Sample PAIRS datalayers that provide unique industry specific analytics regard global irrigation forecasts, global seasonal weather forecasts, US wide Renewable energy generation forecasts, and crop acreage forecasts.

Figure 1 illustrates the conceptual query processing model of PAIRS. This model is useful for composing a specific PAIRS query against one or more datalayers, which could be managed via several datasets. Per the model, cross-datalayer spatial parameters of the query are processed first to determine the spatial data processing rules for the target datalayers, including the spatial properties of the query result (e.g., resolution level). Target spatial areas can be a set of polygon-shaped geo-spaces. After datalayer-specific processing is done per the additional datalayer-specific temporal processing parameters, mask-based cross-datalayer filtering and joining could be performed per the query specification. Geospatial data output of a query is delivered to the query client as file-based or geoserver-based URLs.

Figure 2 illustrates a use case in which datalayer C is used as a filtering mask for datalayer A. For datalayer C, the mask is constructed by selecting all the grid cells whose value equals 5. After unifying the resolution level, the query result is generated by applying the mask to datalayer A.

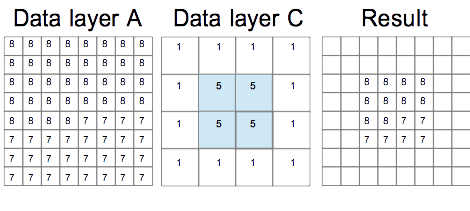


Figure 2. Filtering and Joining Datalayers with Different Spatial Grid Resolutions.

## Spatial Resolution Levels

Table 2. Global grid spatial resolution levels.

Table 2 shows the 26 *resolution levels* that are used in PAIRS for expressing the spatial resolution property of a datalayer. It shows the global grid spatial resolution in degree for longitude (θ) and latitude (φ); in km for longitude; and in km for latitude at the equator, at 20 degrees, at 40 degrees, and at 60, respectively.

# Corpus “basic”

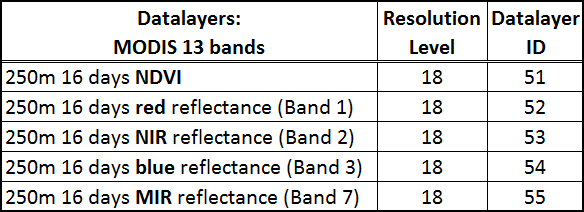
Table 3. List of PAIRS “basic” datasets.

Table 3 lists the datasets that are included in the “basic” corpus of IBM PAIRS Services. They are classified into five categories: (1) analytics, (2) client, (3) satellite images, (4) survey, and (5) weather. The rest of this section briefs the datasets and their respective datalayers by dataset id.

## Corpus Datasets and Datalayers

### Dataset 5: MODIS, Aqua, 13 (Global): NASA 250m resolution satellite

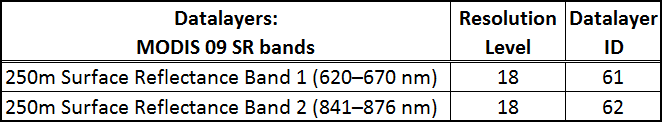
This dataset is derived from USGS MODIS (Moderate-resolution Imaging Spectroradiometer, <https://en.wikipedia.org/wiki/Moderate-resolution_imaging_spectroradiometer>) satellite images sourced from the satellite Aqua, which passes south to north over the

equator in the afternoon.. The table below lists the datalayers in it.

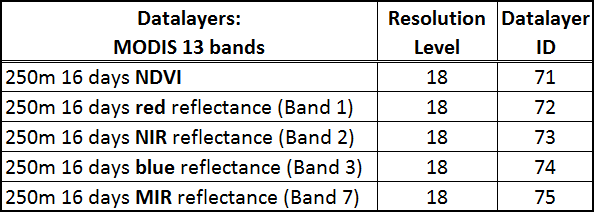
### Dataset 6: MODIS, Aqua, 09 (Global): NASA 250m resolution satellite

This dataset is derived from USGS MODIS (Moderate-resolution Imaging Spectroradiometer) satellite images sourced from the satellite Aqua, which passes south to north over the

equator in the afternoon.. The table below lists the datalayers in it.

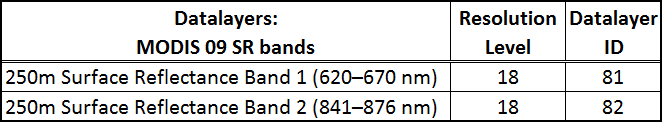


### Dataset 7: MODIS, Terra, 13 (Global): NASA 250m resolution satellite

This dataset is derived from USGS MODIS (Moderate-resolution Imaging Spectroradiometer) satellite images sourced from the satellite Terra, which passes from north to south across the equator in the morning. The table below lists the datalayers in it.

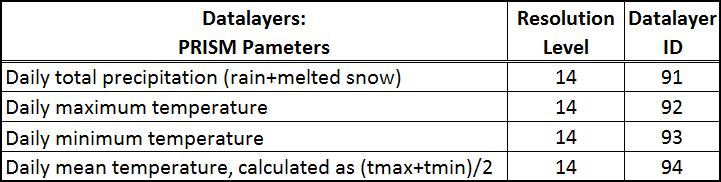
### Dataset 8: MODIS, Terra, 09 (Global): NASA 250m resolution satellite

This dataset is derived from USGS MODIS (Moderate-resolution Imaging Spectroradiometer) satellite images sourced from the satellite Terra, which passes from north to south across the equator in the morning. The table below lists the datalayers in it.



### Dataset 9: PRISM historical climate data (USA)

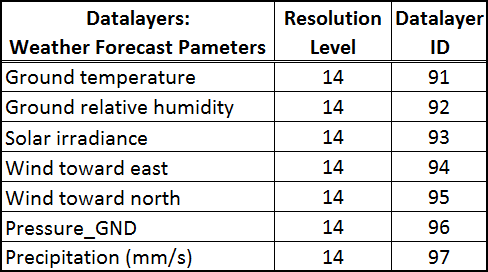
This dataset is derived from the one generated by the PRISM Climate Group, Oregon State University (<http://www.prism.oregonstate.edu/>). It includes historical daily weather condition measurements in USA. The table below lists the datalayers in it.



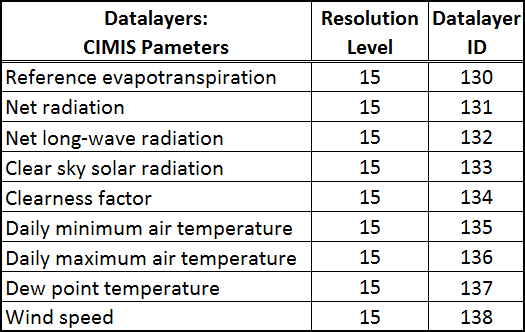
### Dataset 11: Historical crop planting map (USA)

This dataset is derived from CropScape (see <http://nassgeodata.gmu.edu/CropScape/>), generated by United States Department of Agriculture (USDA) National Agriculture Statistics Services (NASS). USDA issues crop information annually in 30m resolution. PAIRS has ingested data from year 2008 to 2015. The table below lists the datalayers in it.

### Dataset 12: NAM USA weather forecast (USA)

This dataset is derived from the North American Mesoscale Forecast System (NAM, <https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/north-american-mesoscale-forecast-system-nam>) of National Oceanic and Atmospheric Administration (NOAA). NAM is a numerical weather prediction system designed for short-term forecasting with finer detail than other forecast models. The model is run four times a day out to 84 hours in advance with 12-km horizontal resolution and three-hour temporal resolution. The table below lists the datalayers in it.

### Dataset 13: California weather condition measurements

This dataset is derived from California Irrigation Management Information System (CIMIS, <http://www.cimis.water.ca.gov/>), a California weather condition measurements dataset, which provides gridded data for the state of California. CIMIS is an integrated network of over 140 automated active weather stations located throughout California providing localized meteorological data useful for crop irrigation scheduling. The table below lists the datalayers in this dataset.

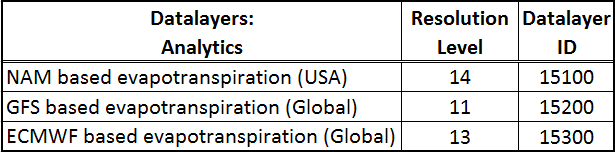
### Dataset 14: USGS national elevation data (USA)

This dataset is derived from a 10-m resolution United States Geological Survey (USGS) National Elevation Dataset (NED), which contains raster-based land elevation data for the conterminous United States, Alaska, Hawaii, and territorial islands, providing basic elevation information for earth science studies and mapping applications. Further details on USGS Visual Identity System Guidance are available at <http://www.usgs.gov/visual-id/> The NED dataset is distributed by the Land Processes Distributed Active Archive Center (LP DAAC, <http://lpdaac.usgs.gov>). The table below lists the datalayers in this dataset.



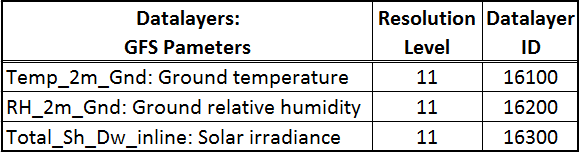
### Dataset 15: IBM Evapo-transpiration/irrigation (Global)

We have multiple one of a kind analytics on PAIRS. Two of them are in the Weather category: SMT (self-learning weather modeling and forecast) and SMT (long term seasonal forecast). The Evapotranspiration model is hosted under Analytics category. When the models are developed based on other datasets on PAIRS and validated, we ingest the derived analytical layers back onto PAIRS as a separate dataset. Currently daily reference evapotranspiration for the continental USA as well as on a global scale (coarser resolution than USA data layer) is available. Reference evapotranspiration is critical in irrigation forecast and decision making. The table below lists the analytics datalayers in this dataset.

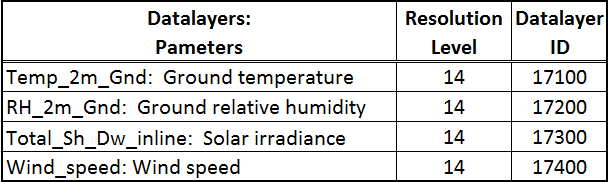


### Dataset 16: NOAA Global Forecasting System (Global)

Global weather forecast dataset is a worldwide forecast model from NOAA with 0.5 degree spatial resolution. 10 days forecast is ingested into PAIRS for weather forecast around the world. All the parameters follow the same conventions as USA weather forecasts except the precipitation is an averaged precipitation rate over 3 hours. The table below lists the datalayers in it.



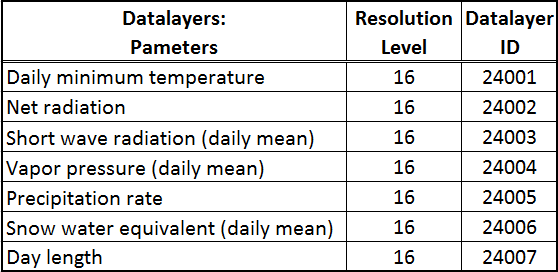
### Dataset 17: IBM blended weather forecast (USA)

An improved weather forecast based on Model blending machine learning algorithm is generated daily for the continental USA. Resolution is the same as USA forecast. The Solar irradiance and wind speed parameters are super important for renewable energy industry. We deliver the forecast to renewable energy utility customers daily. The table below lists the datalayers in it.

### Dataset 24: Daymet historical weather (USA)

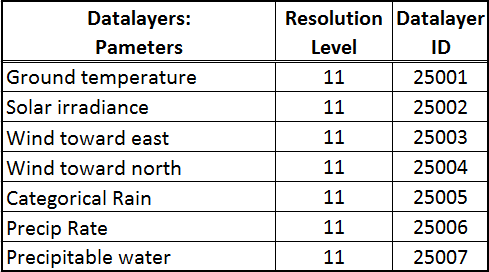
This dataset is derived from Daymet dataset distributed by Oak Ridge National Laboratory, which is under NASA's EarthData license policy (see <https://earthdata.nasa.gov/>). Details on the dataset, including data set citation instructions, are available at

<https://daac.ornl.gov/DAYMET/guides/Daymet_mosaics.html#Daymet_m_citation>

The table below lists the datalayers in it.

### Dataset 25: IBM long term forecast (Global)

Seasonal forecast projecting 6 months ahead is issued by NOAA daily. Based on NOAA’s forecast, we built an improved model using machine learning. The table below lists the analytics datalayers in it.



### Dataset 34: IBM Research drone

This dataset includes three sets of drone images for the IBM T.J. Watson Research Center. The drone data set is based on aerial imagery acquired at 1 inch resolution with drone flying at 300 feet above the ground. Data acquisition is triggered automatically at 5 sec interval and result in imagery with 50% overlap in consecutive images. All images are stitched together to create a continuous image. GPS coordinates of well-defined landmarks are used to georeference the images. Image is acquired as an RGB image, and is uploaded as 3 different layer Reg, Green Blue with range [0,255].



# Update History

[2017-11-28]

* Remove Dataset 26, “ECMWF weather forecast (Global)”

[2017-04-26]

* Initial release