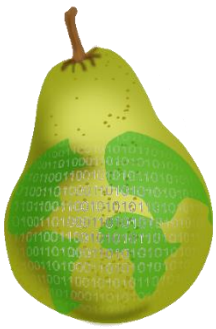
IBM PAIRS Services: Corpus “basic”



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: 2017-11-28
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

List of Figures

List of Tables

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” corpus supported by the IBM PAIRS Services.

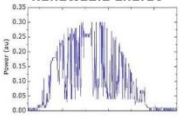
Please email pairs@us.ibm.com your questions or suggestions about this document.

1 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-spatio-temporal 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.

Industry / Customer	Insurance	Consumer	Utility	Agriculture	Finance
Pain Points	*Time to value: Most companies employ full departments of data scientists, which spend 80-90 % of their time with data pre-processing and only 10-20 % with analyzing; very little re-use of output from different analysis etc.*				
Example	Developing risk models (Flood, Fire)	Predicting supply of raw materials	Asset Management Emergency Response	Buying land with certain characteristics	Energy Trading
Example Queries / Questions	Where are the regions of higher flood and fire risks? At what time of the year is the risk the highest?	How much crop is planted? What is the health of the crop? Is rain fall lower than usual?	What maintenance schedule is optimal? Where can I expect outages? What is the impact of outages?	Where can I buy land with a certain soil type and climate within 50 miles of my winery?	How much renewable energy is being produced tomorrow and in the next 15 minutes?
Data Layers	Climate, vegetation, traffic, census	Weather, land class, soil, satellites	Weather data, census data, vegetation	Weather, climate, satellites	Weather, climate, satellites
Interlayer analytics	Flood Modeling Wildfire modeling Climate Modeling	Crop recognition Yield prediction Climate Modeling	Predictive maintenance Analytics (PMQ) Outage modeling	Logistics Supply chain modeling	Renewable Energy Forecasting (Wind & Solar)
Analytics Example	Fire Risk 	Early Crop Recognition 	Outage Probability 	Optimal Land locations 	Renewable Energy 

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

1.1 Query Processing Model

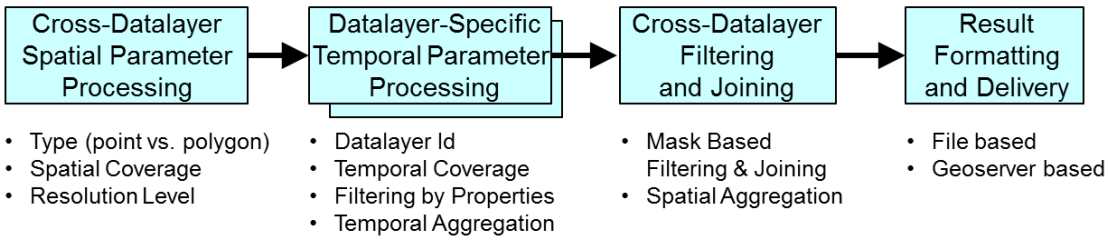


Figure 1. Conceptual Query Processing Model of PAIRS.

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.

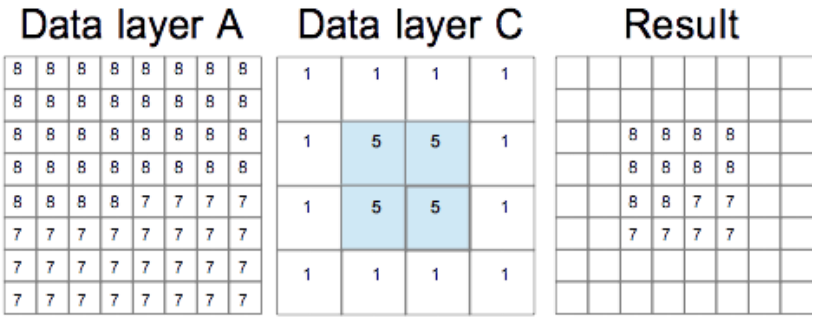


Figure 2. Filtering and Joining Datalayers with Different Spatial Grid Resolutions.

1.2 Spatial Resolution Levels

Table 2. Global grid spatial resolution levels.

Resolution Level	dDegree [degree]	dyCell [km]	dxCell@0 [km]	dxCell@20 [km]	dxCell@40 [km]	dxCell@60 [km]
26	0.000008	0.00089	0.00089	0.00084	0.00067	0.00045
25	0.000016	0.00178	0.00178	0.00168	0.00134	0.0009
24	0.000032	0.00356	0.00356	0.00336	0.00268	0.0018
23	0.000064	0.00712	0.00712	0.00672	0.00536	0.0036
22	0.000128	0.01424	0.01424	0.01344	0.01072	0.0072
21	0.000256	0.02848	0.02848	0.02688	0.02144	0.0144
20	0.000512	0.05696	0.05696	0.05376	0.04288	0.0288
19	0.001024	0.11392	0.11392	0.10752	0.08576	0.0576
18	0.002048	0.22784	0.22784	0.21504	0.17152	0.1152
17	0.004096	0.45568	0.45568	0.43008	0.34304	0.2304
16	0.008192	0.91136	0.91136	0.86016	0.68608	0.4608
15	0.016384	1.82272	1.82272	1.72032	1.37216	0.9216
14	0.032768	3.64544	3.64544	3.44064	2.74432	1.8432
13	0.065536	7.29088	7.29088	6.88128	5.48864	3.6864
12	0.131072	14.58176	14.58176	13.76256	10.97728	7.3728
11	0.262144	29.16352	29.16352	27.52512	21.95456	14.7456
10	0.524288	58.32704	58.32704	55.05024	43.90912	29.4912
9	1.048576	116.6541	116.65408	110.10048	87.81824	58.9824
8	2.097152	233.3082	233.30816	220.20096	175.63648	117.9648
7	4.194304	466.6163	466.61632	440.40192	351.27296	235.9296
6	8.388608	933.2326	933.23264	880.80384	702.54592	471.8592
5	16.7772	1866.465	1866.46528	1761.60768	1405.09184	943.7184
4	33.5544	3732.931	3732.93056	3523.21536	2810.18368	1887.4368
3	67.10886	7465.861	7465.86112	7046.43072	5620.36736	3774.8736
2	134.2177	14931.72	14931.7222	14092.8614	11240.7347	7549.7472
1	268.4355	29863.44	29863.4445	28185.7229	22481.4694	15099.494

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.

2 Corpus “basic”

Table 3. List of PAIRS “basic” datasets.

ID	Display Name	Resolution Layers	Category
5	MODIS, Aqua, 13 (Global): NASA 250m resolution satellite	18	Satellite
6	MODIS, Aqua, 09 (Global): NASA 250m resolution satellite	18	Satellite
7	MODIS, Terra, 13 (Global): NASA 250m resolution satellite	18	Satellite
8	MODIS, Terra, 09 (Global): NASA 250m resolution satellite	18	Satellite
9	PRISM historical climate data (USA)	14	Weather
11	Historical crop planting map (USA)	21	Survey
12	NAM USA weather forecast (USA)	14	Weather
13	California weather condition measurements	15	Weather
14	USGS national elevation data (USA)	23	Survey
15	IBM Evapo-transpiration/irrigation (Global, USA)	11, 13, 14	Analytics
16	NOAA Global Forecasting System (Global)	11	Weather
17	IBM cognitive weather forecast (USA)	14	Weather
24	Daymet historical weather (USA)	16	Weather
25	IBM long term forecast (Global)	11	Weather
34	IBM Research drone	29	Client

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.

2.1 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. The table below lists the datalayers in it.

Datalayers: MODIS 13 bands	Resolution Level	Datalayer ID
250m 16 days **NDVI**	18	51
250m 16 days **red** reflectance (Band 1)	18	52
250m 16 days **NIR** reflectance (Band 2)	18	53
250m 16 days **blue** reflectance (Band 3)	18	54
250m 16 days **MIR** reflectance (Band 7)	18	55

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. The table below lists the datalayers in it.

Datalayers: MODIS 09 SR bands	Resolution Level	Datalayer ID
250m Surface Reflectance Band 1 (620–670 nm)	18	61
250m Surface Reflectance Band 2 (841–876 nm)	18	62

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. The table below lists the datalayers in it.

Datalayers: MODIS 13 bands	Resolution Level	Datalayer ID
250m 16 days **NDVI**	18	71
250m 16 days **red** reflectance (Band 1)	18	72
250m 16 days **NIR** reflectance (Band 2)	18	73
250m 16 days **blue** reflectance (Band 3)	18	74
250m 16 days **MIR** reflectance (Band 7)	18	75

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. The table below lists the datalayers in it.

Datalayers: MODIS 09 SR bands	Resolution Level	Datalayer ID
250m Surface Reflectance Band 1 (620–670 nm)	18	81
250m Surface Reflectance Band 2 (841–876 nm)	18	82

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.

Datalayers: PRISM Pameters	Resolution Level	Datalayer ID
Daily total precipitation (rain+melted snow)	14	91
Daily maximum temperature	14	92
Daily minimum temperature	14	93
Daily mean temperature, calculated as (tmax+tmin)/2	14	94

Dataset 11: Historical crop planting map (USA)

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

Datalayers: USDA Crop Information	Resolution Level	Datalayer ID
CROP	21	111

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). The table below lists the datalayers in it.

Datalayers: Weather Forecast Pameters	Resolution Level	Datalayer ID
Ground temperature	14	91
Ground relative humidity	14	92
Solar irradiance	14	93
Wind toward east	14	94
Wind toward north	14	95
Pressure_GND	14	96
Precipitation (mm/s)	14	97

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. The table below lists the datalayers in this dataset.

Datalayers: CIMIS Pameters	Resolution Level	Datalayer ID
Reference evapotranspiration	15	130
Net radiation	15	131
Net long-wave radiation	15	132
Clear sky solar radiation	15	133
Clearness factor	15	134
Daily minimum air temperature	15	135
Daily maximum air temperature	15	136
Dew point temperature	15	137
Wind speed	15	138

Dataset 14: USGS national elevation data (USA)

This dataset is derived from a 10-m resolution NED dataset for elevation for the USA, which is derived from the USGS (see http://www.usgs.gov/visual-id/ for further details on USGS Visual Identity System Guidance). 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.

Datalayers: Elevation	Resolution Level	Datalayer ID
Elevation	23	140

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.

Datalayers: Analytics	Resolution Level	Datalayer ID
NAM based evapotranspiration (USA)	14	15100
GFS based evapotranspiration (Global)	11	15200
ECMWF based evapotranspiration (Global)	13	15300

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.

Datalayers: GFS Pameters	Resolution Level	Datalayer ID
Temp_2m_Gnd: Ground temperature	11	16100
RH_2m_Gnd: Ground relative humidity	11	16200
Total_Sh_Dw_inline: Solar irradiance	11	16300

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.

Datalayers: Pameters	Resolution Level	Datalayer ID
Temp_2m_Gnd: Ground temperature	14	17100
RH_2m_Gnd: Ground relative humidity	14	17200
Total_Sh_Dw_inline: Solar irradiance	14	17300
Wind_speed: Wind speed	14	17400

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.

Datalayers: Pameters	Resolution Level	Datalayer ID
Daily minimum temperature	16	24001
Net radiation	16	24002
Short wave radiation (daily mean)	16	24003
Vapor pressure (daily mean)	16	24004
Precipitation rate	16	24005
Snow water equivalent (daily mean)	16	24006
Day length	16	24007

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.

Datalayers: Pameters	Resolution Level	Datalayer ID
Ground temperature	11	25001
Solar irradiance	11	25002
Wind toward east	11	25003
Wind toward north	11	25004
Categorical Rain	11	25005
Precip Rate	11	25006
Precipitable water	11	25007

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].

Datalayers: Pameters	Resolution Level	Datalayer ID
Red	29	34001
Green	29	34002
Blue	29	34003

3 Update History

[2017-11-28]

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

[2017-04-26]

- Initial release