This protocol shows a simple data processing pipeline. Press the green Run Protocol button to execute it.

View the Quick Tour in the help center or look in the Examples folder on the Protocols tab for more examples.

Pipeline Pilot – A Fancy Galaxy

with four cylinders

horsepower

Viewer

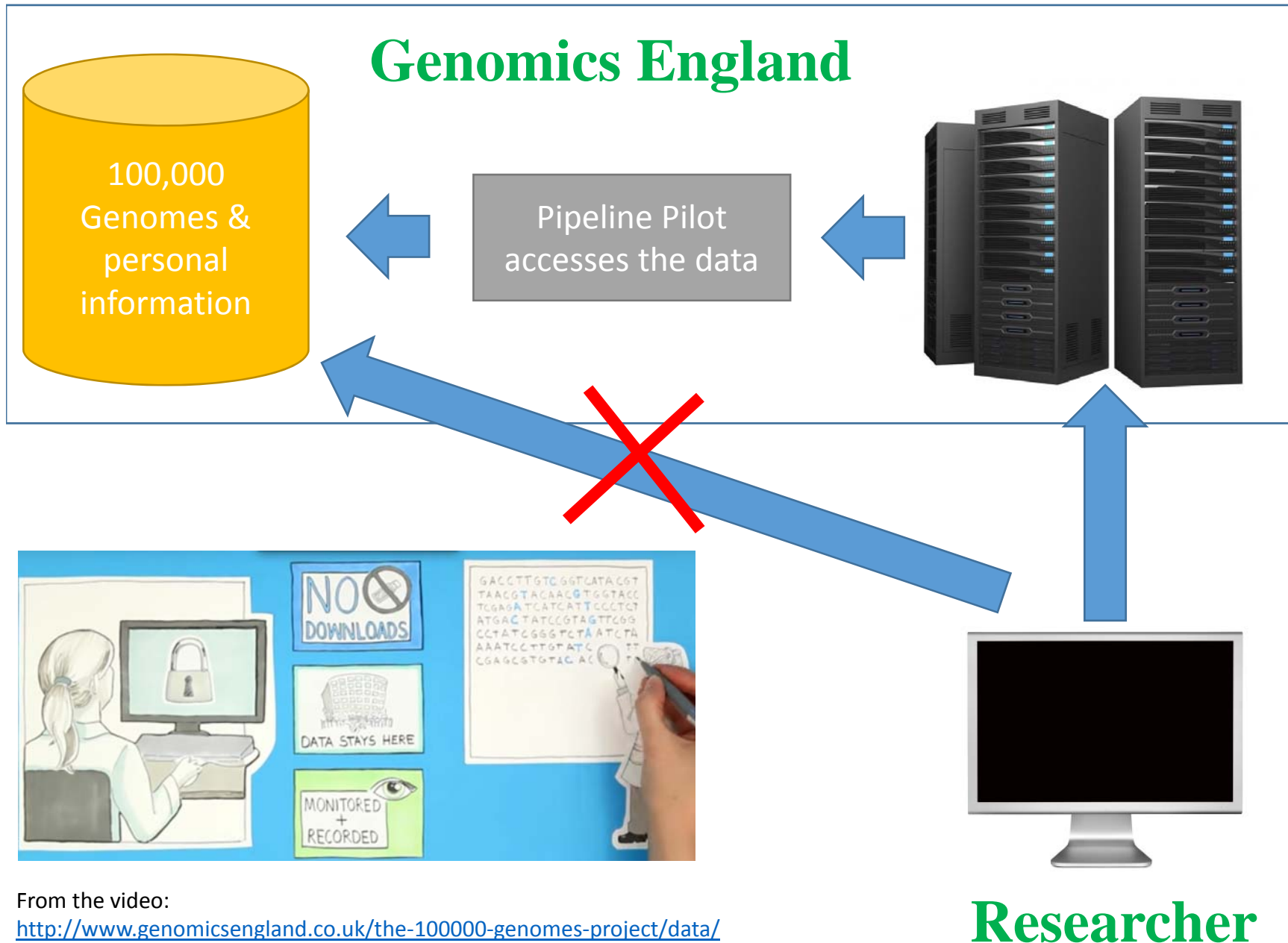| The Delimited Text Reader reads data from simple text files | The Keep Properties component removes data fields that are not in our defined list | A Custom Filter lets you write simple logical expressions for splitting a data stream | The Sort Data component reorders the data records based on the value of a given property | At the end of a pipeline you often use a viewer component to display results |

- ❖ What is Pipeline Pilot?
- ❖ Why learning about Pipeline Pilot?
- ❖ How does Pipeline Pilot work?
- ❖ Why using Pipeline Pilot?
- ❖ When should we use Pipeline Pilot?
- ❖ Learn more about Pipeline Pilot

1

Why learning about Pipeline Pilot?

Genomics England

100,000 Genomes & personal information

Pipeline Pilot accesses the data

From the video:
http://www.genomicsengland.co.uk/the-100000-genomes-project/data/
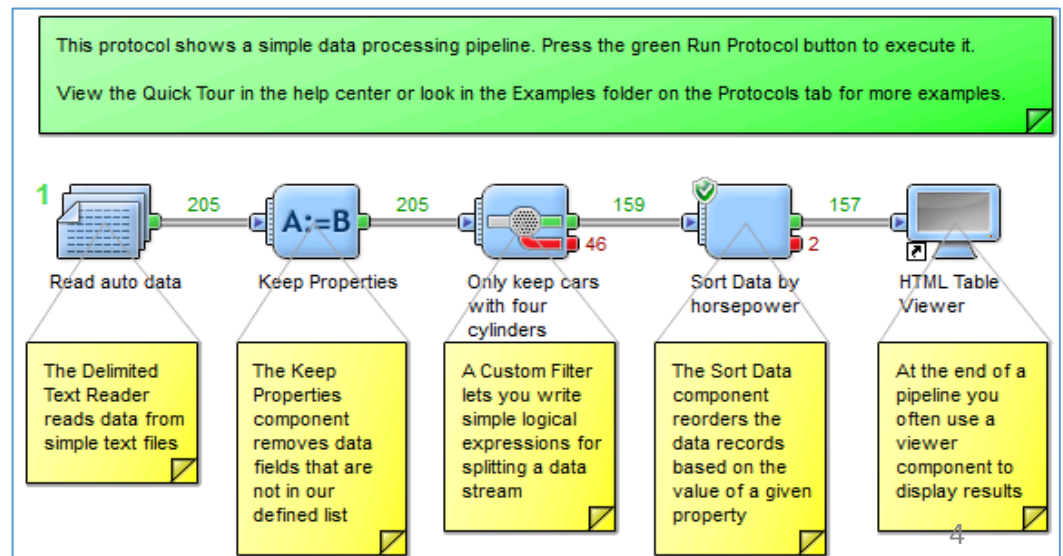
Researcher

Pipeline Pilot overview (3 min video)

https://www.youtube.com/watch?v=nZ2YMNv-uTk&feature=youtube_gdata

What is Pipeline Pilot?

- Pipeline Pilot is a commercially developed workflow environment (like Galaxy) that will be incorporated into Genomic England.

- PP enables scientists to create, test and publish scientific services that automate the process of accessing, analyzing and reporting scientific data.

- **Anyone** can create scientific protocols that provide access to research data locked in silos (like Genomic England). These protocols automate the scientific analysis of the data and enable researchers to rapidly explore, visualize and report results.

- PP allows sharing workflows

This protocol shows a simple data processing pipeline. Press the green Run Protocol button to execute it.

View the Quick Tour in the help center or look in the Examples folder on the Protocols tab for more examples.

| Read auto data	Keep Properties	Only keep cars with four cylinders	Sort Data by horsepower	HTML Table Viewer
The Delimited Text Reader reads data from simple text files	The Keep Properties component removes data fields that are not in our defined list	A Custom Filter lets you write simple logical expressions for splitting a data stream	The Sort Data component reorders the data records based on the value of a given property	At the end of a pipeline you often use a viewer component to display results

Why using Pipeline Pilot?

- To access Genomics England's Data

- Complete R&D projects faster by leveraging existing research and automating routine data gathering and analysis processes.

- Encapsulate and deploy best practices to ensure compliance and enable collaboration across your research and development organizations.

- Build and deploy your solution based upon standard technologies.

- Faster results through real-time reporting dashboards.

- Reduce support costs by integrating incompatible software and databases.

How does Pipeline Pilot work?

What is a Component?

- A component is the building block used to create workflows
- Each component performs a task like reading, writing or manipulating data
- Components can have one input and up to two output ports
- Highlighting a component displays its parameter panel which control its behavior

Input — Sample Component — Pass / Fail

Generic Sequence Reader
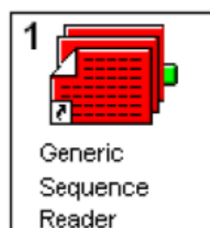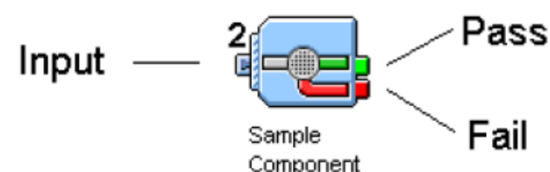
Parameters		
Source		...
File Format	Infer From File Extension	▼
Maximum		
SourceTag	None	▼
Keep Properties		...
Read Features	False	▼

MSC ONLY·THE·BEST

What is a Pipeline?

- A series of *components* connected through pipes through which data flows.

- Each component acts on the data and passes it on to subsequent components.

Delimited Text Reader   Create New Property   HTML Table Viewer

Data Flow in a Pipeline
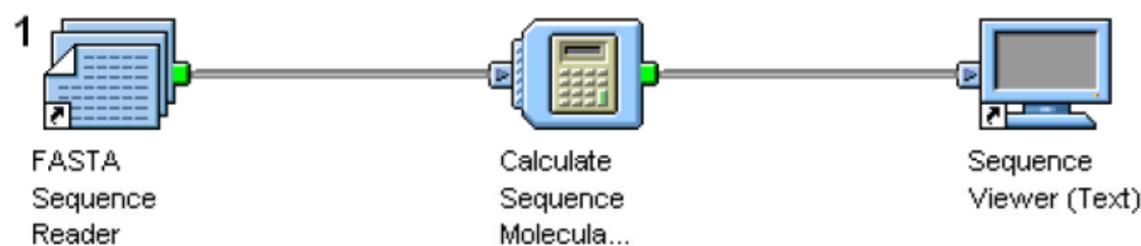
File Reader

Reads data records from flat file

Molecular Weight

Calculates MW and adds to record

Notepad Viewer

Displays records as rows in NotePad

8

What is a Protocol?

- A protocol consists of one or multiple pipelines that are run sequentially.

- Components are executed left to right, top to bottom.

- Pipeline 2 will not start processing data until pipeline 1 has completely finished processing all the records.

1   699   699   241   241
458

Read breast cancer data   Count Data into @AllCases   Filter for malignant cases   Count Data into @MalignantCases   Don't Pass Data

2

Display Message
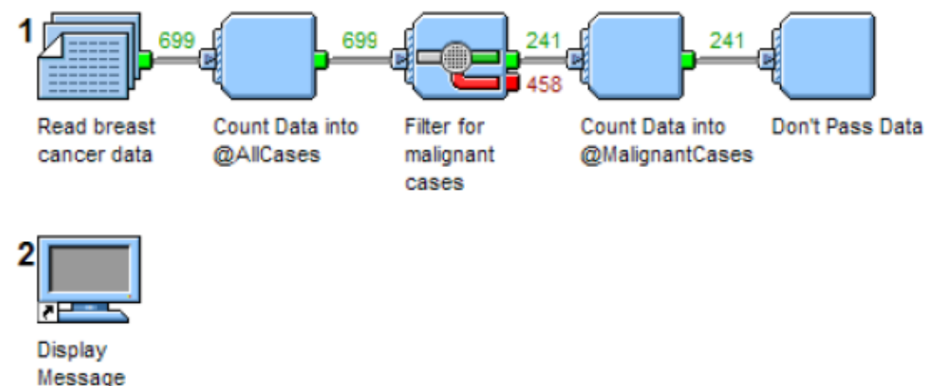
Pipeline Pilot Main Window

Learn more about Pipeline Pilot

Tutorials

http://lsl.sinica.edu.tw/Services/Class/files/4162009.pdf

http://www.genengnews.com/gen-articles/getting-more-value-from-bioinformatics/1768/

Videos

http://meta-guide.com/videography/best-pipeline-pilot-videos

BIOVIA Pipeline Pilot Overview. You can also register for webinars

http://accelrys.com/products/collaborative-science/biovia-pipeline-pilot/

http://accelrys.com/services/training/pipeline-pilot/fundamentals.html

Link to Genomics England Pipeline Pilot:

https://emb.extge.co.uk/ovd/

About Genomics England

http://www.genomicsengland.co.uk/the-100000-genomes-project/data/