Query-Driven Visualization
Bridging the gaps between Processing, Archiving and Analyzing
H. Buddelmeijer, buddel@astro.rug.nl
Kapteyn Astronomical Institute, University of Groningen

Traditional 'Pushing' Approach


With **data pushing**, there are two possibilities:
A) All data is pushed through processing to storage automatically, even if never used
   - large amount of **unnecessary processing and storage**
   - it is **difficult to adapt** for user-specific needs
B) Data is pushed through manually
   - requires **manual work** that could be automated
   - requires **access** for the user to the processing
   - makes it **difficult to share** data between scientists
The scientist has to make the **trade-off between speed and reusability**!

Query-Driven 'Pulling' Approach


With **data pulling** the process **starts at the end**:
1) The visualization requests the data that is needed for the analysis at hand.
   - The desired **end result is known** at every step.
2) Data is stored with **full data lineage**: it is know how all the data was derived.
   - **Automatic reuse** of existing data.
   - **Implicit sharing** of data.
3) New data is only **processed if necessary** for the analysis at hand.
   - The result is **derived as fast as possible**.
4) (Ultimately, even raw data could be retrieved or even measured on request.)

Furthermore, data pushing is still possible, therefore
- the benefits of the traditional approach remain, while
- the problems disappear thus
- **both speed and reusability** are achieved!

Interested in query-driven visualization? Scan the QR code to visit:
http://www.esciencecenter.nl/projects/project-portfolio/e-visualization-of-big-data/