## Components of a Data Portal

- **Research Communities**
- **Public/Media**
- **Data Portal**
  - **Web Crawler**
  - **WWW**
  - **Web Pages**
  - **Information Extractor**
  - **Climate Change Data**
  - **Geographical Data**
  - **Research Data**
  - **Integrated Server**
  - **Data Portal**
- **Sensor Info. Extractor**
- **Data Portal Reliability**
  - **Abort on 3, Undo Task**
  - **Resume Transaction**
  - **Log Completed, but Failure Occurred, Restart System and Undo Transactions**
  - **Transaction Redo**
  - **Transaction Undo**
  - **System Failure**
  - **Serial Redo**
  - **Undo Action**
  - **Return to Serial Redo**

### Types of a Data Portal

**Academic** - heavy scientific data, focused on current data allowing the audience to reach a conclusion.

**Commercial** - contains basic information for a general audience, including current events/political information.

### Components of a Data Portal

- **Web Crawler** – Searches the Web for results related to the data portal topic (i.e., climate change).
- **Information Extractor** – Extracts information from websites into raw data for the portal to read and lay out to the portlet displays. Also extracts information from sensors and collects real-time data.
- **Query Engine** – Allows queries on the engine and data fields.

### Approaches to Extracting Information

- **“Shotgun” Approach** – Web crawler continuously searches all relevant results related to the given topic and extracts information from all data.
- **Top-Down Incremental Approach** – Web crawler searches the top websites and its links. The extractor extracts data in order of its relevancy based on a combination of PageRank and TF-IDF algorithms.

### Data Portal Reliability

- A data portal is reliable if it is serializable, recoverable, and resilient, much like all very large database and nested database systems.
- **ARIES** allows an approach for better recovery of database systems. Experiments have also been done for the scheduling of logs through TMR scheduling.

---

**NSF Nevada EPSCoR**

**Survey of environmental data portals: features and characteristics**

David Walker – University of Nevada, Las Vegas

---

**Graduate Research Assistant: David Walker**

David Walker is a student in the University of Nevada, Las Vegas Electrical Engineering program. He is a life long Nevadan and has earned a Bachelor's in Computer Engineering (2007), Computer Science (2007), and a Master's in Electrical Engineering (2009). Currently, he is working towards a PhD in Electrical Engineering at the same university. He is a member of the Cyberinfrastructure team for the NSF-EPSCoR sponsored project “Nevada Infrastructure for Climate Change, Education, and Science.” David is currently working under his advisor Dr. Shahram Latifi in this project. His research is based on the optimization and reliability of climate change data portals.