

BIG DATA EUROPE

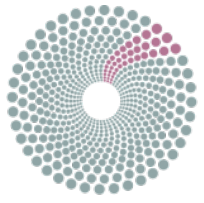
Empowering Communities
with Data Technologies



BIG DATA AGGREGATOR

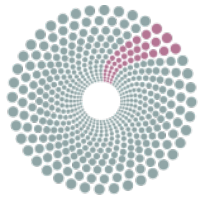


STASINOS KONSTANTOPOULOS
NCSR "DEMOKRITOS", GREECE



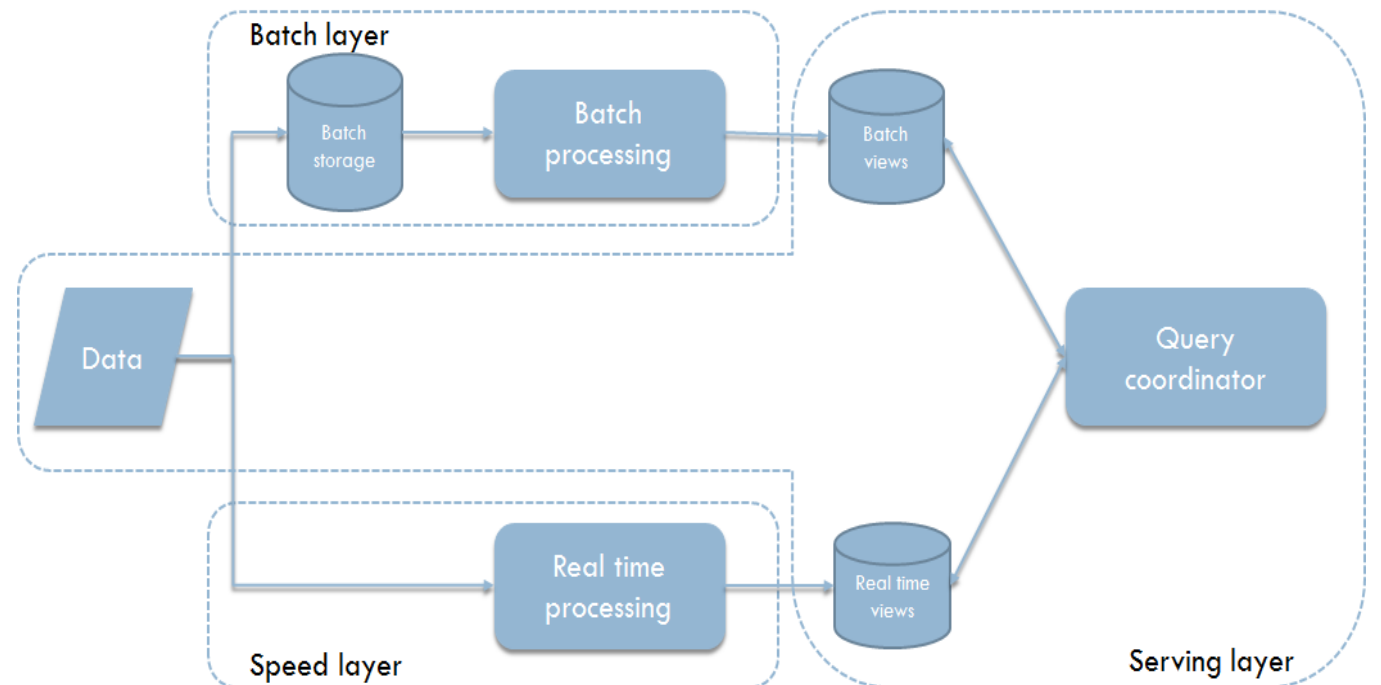
The Big Data Aggregator

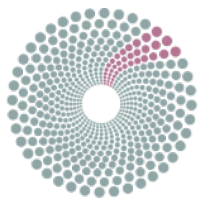
- ⊙ The Big Data Aggregator:
 - A general-purpose architecture for processing Big Data
 - An implementation of the core architecture
 - ❖ Integrating existing mature components
 - An ecosystem of tools around the core system
 - ❖ Driven by our use cases across all Horizon 2020 challenges



Conceptual basis

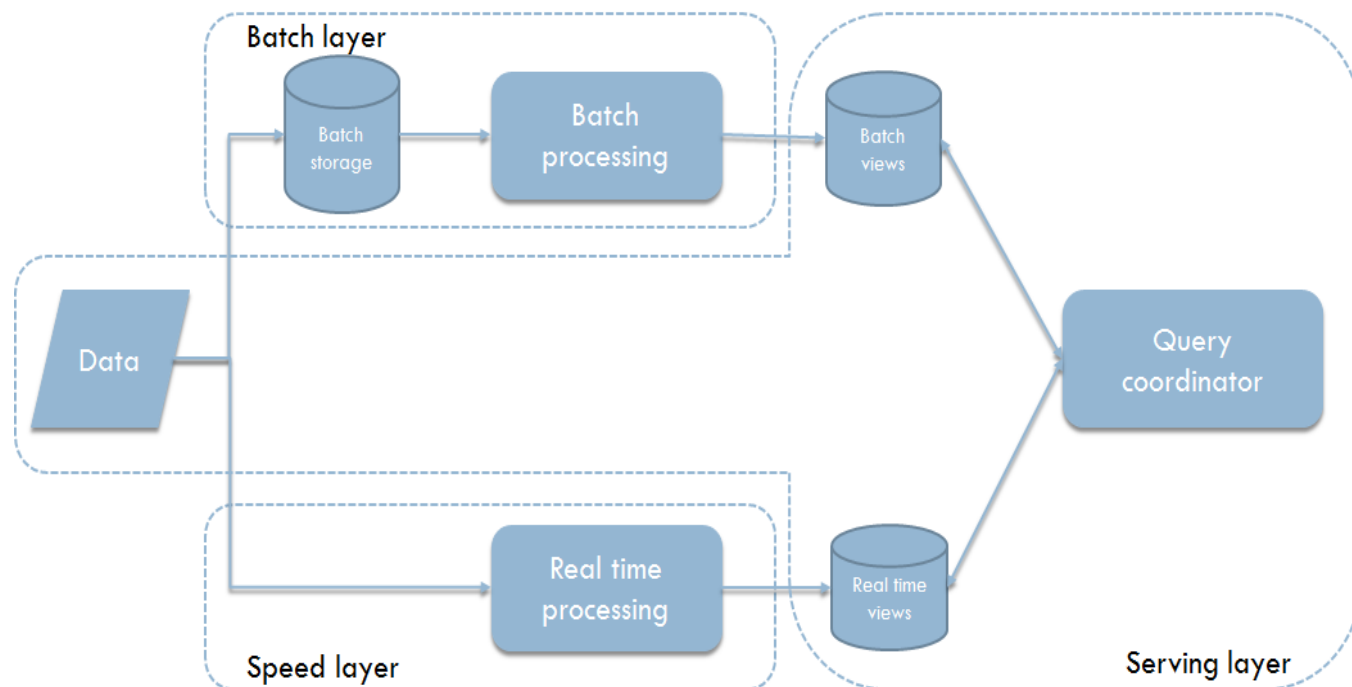
- ⊙ Big Data Aggregator architecture builds upon the *Lambda Architecture*
 - Generic, scalable and fault-tolerant data processing architecture
- ⊙ Batch layer
 - Time-consuming computations
 - Physically available data
 - Typically in large chunks

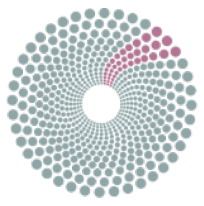




Conceptual basis

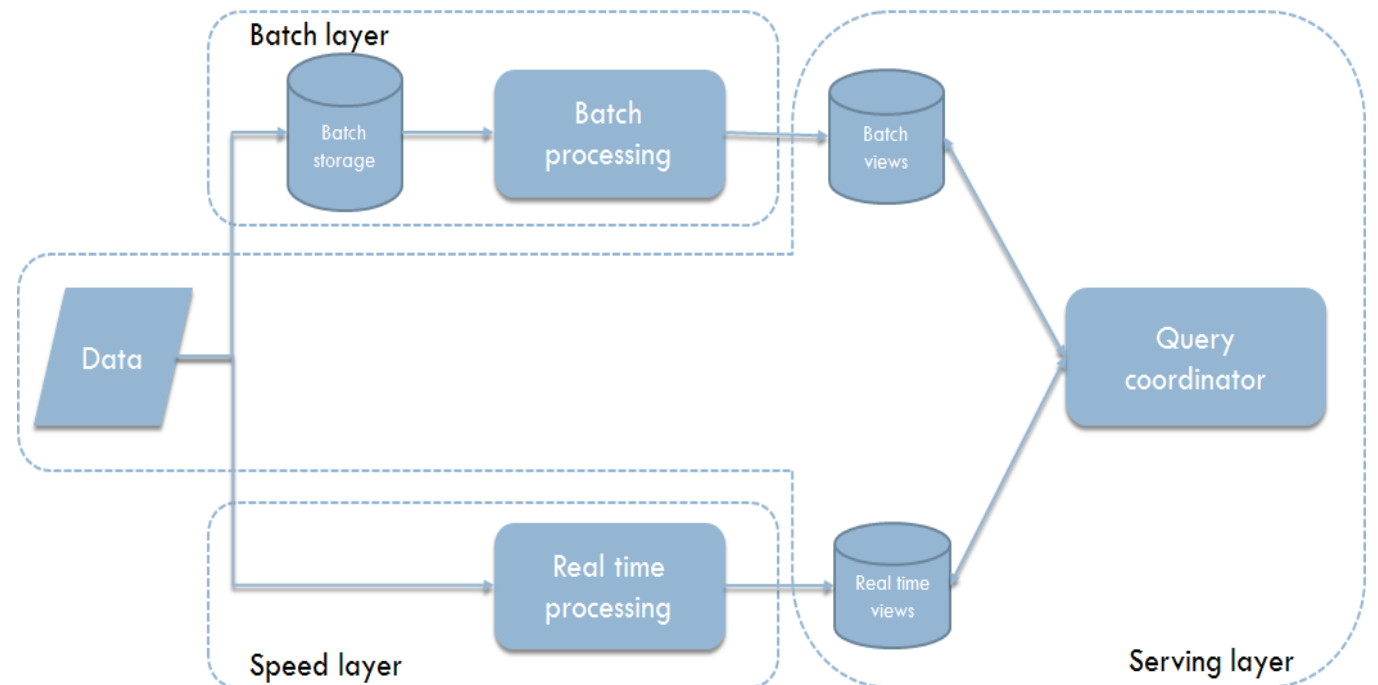
- ⊙ Big Data Aggregator architecture builds upon the *Lambda Architecture*
 - Generic, scalable and fault-tolerant data processing architecture
- ⊙ Speed layer
 - Computations expected to provide results in real time
 - Smaller amounts of data
 - Often streams

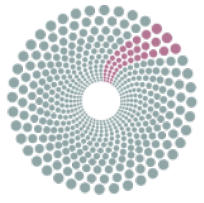




Conceptual basis

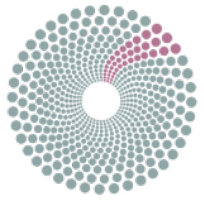
- ⊙ Big Data Aggregator architecture builds upon the *Lambda Architecture*
 - Generic, scalable and fault-tolerant data processing architecture
- ⊙ Data serving layer
 - Data input and data consumption
 - Results offered for consumption as *views*, predefined queries required by the application
 - Views combine batch and speed layer results to offer a unified view to the application



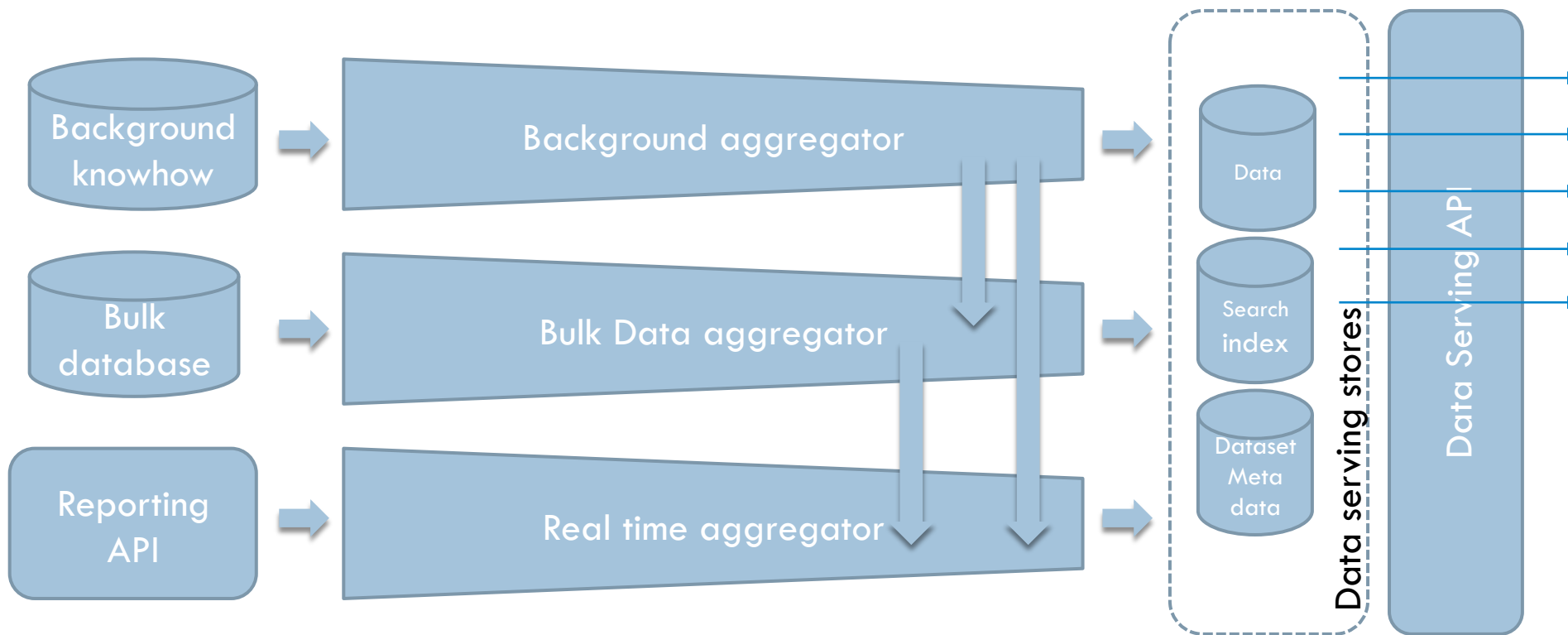


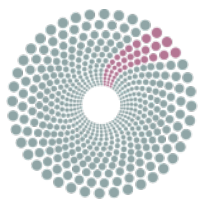
Big Data Aggregator Conceptual Architecture

- ⊙ *A Lambda Architecture for the Semantic Web*
 - Generic, scalable and fault-tolerant data processing architecture
 - In the presence of semantic knowledge about the data
 - Maintaining metadata about provenance
 - ❖ Especially when pooling together multiple data sources
 - ❖ Including non-trivial data integration where substantial transformations are carried out

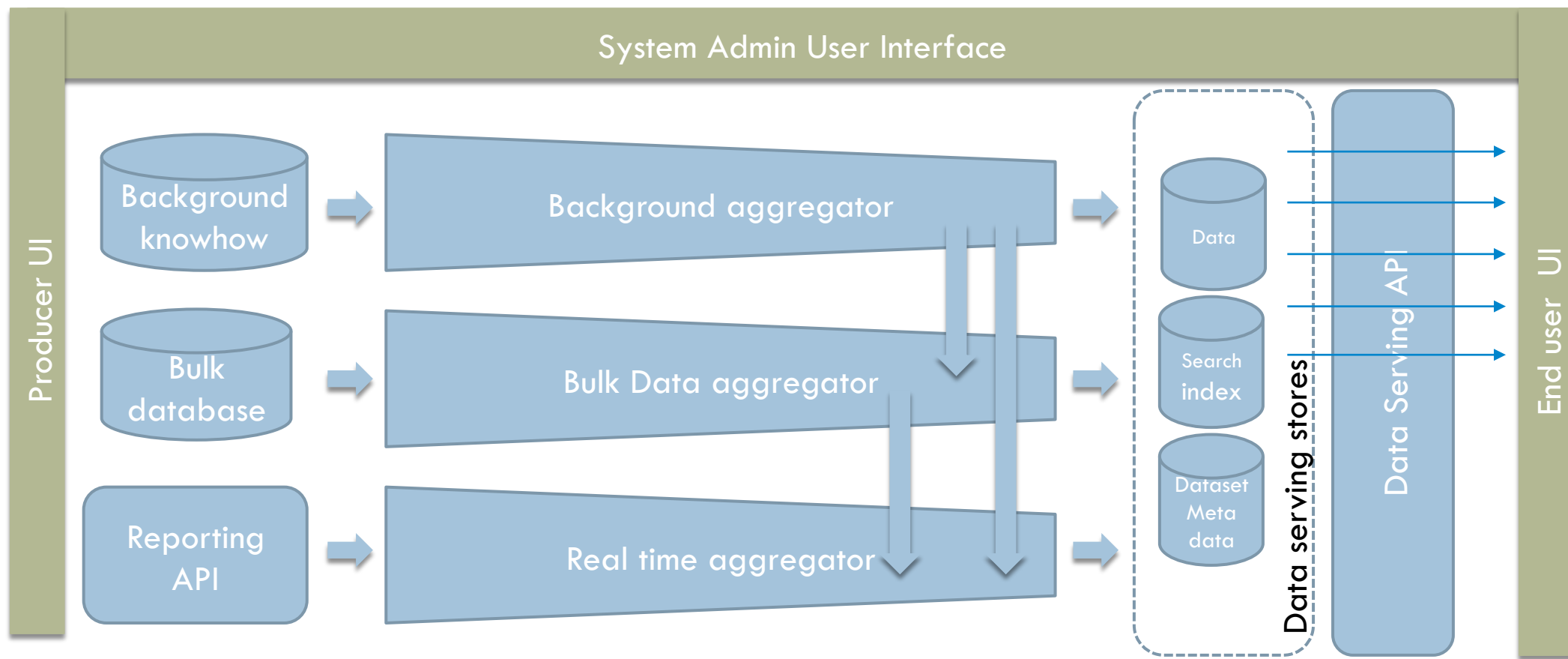


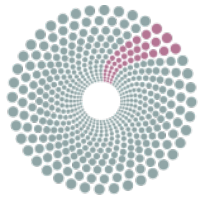
Big Data Aggregator Conceptual Architecture





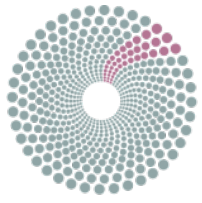
Big Data Aggregator Conceptual Architecture





Semantic Web aspects: Background

- ⊙ Background knowledge
 - Integrating different pieces of background
 - Making it available to data processing
- ⊙ Vertical links:
 - E.g., stream processor receives aggregated background



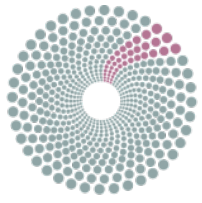
Semantic Web aspects: Provenance

⊙ Provenance and other metadata

- Metadata about data sources providing to this computation
- Metadata “travels” down the processing pipeline without getting disassociated from the data it describes
- Metadata is available as a data serving view

⊙ Granularity

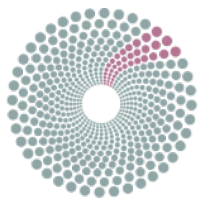
- Per result tuple can become Big Data by itself
- Per resultset can be less useful
 - ❖ Invalidates enormous processing for the slightest now-invalid input
- Something inbetween or user configuration?
 - ❖ To be discussed



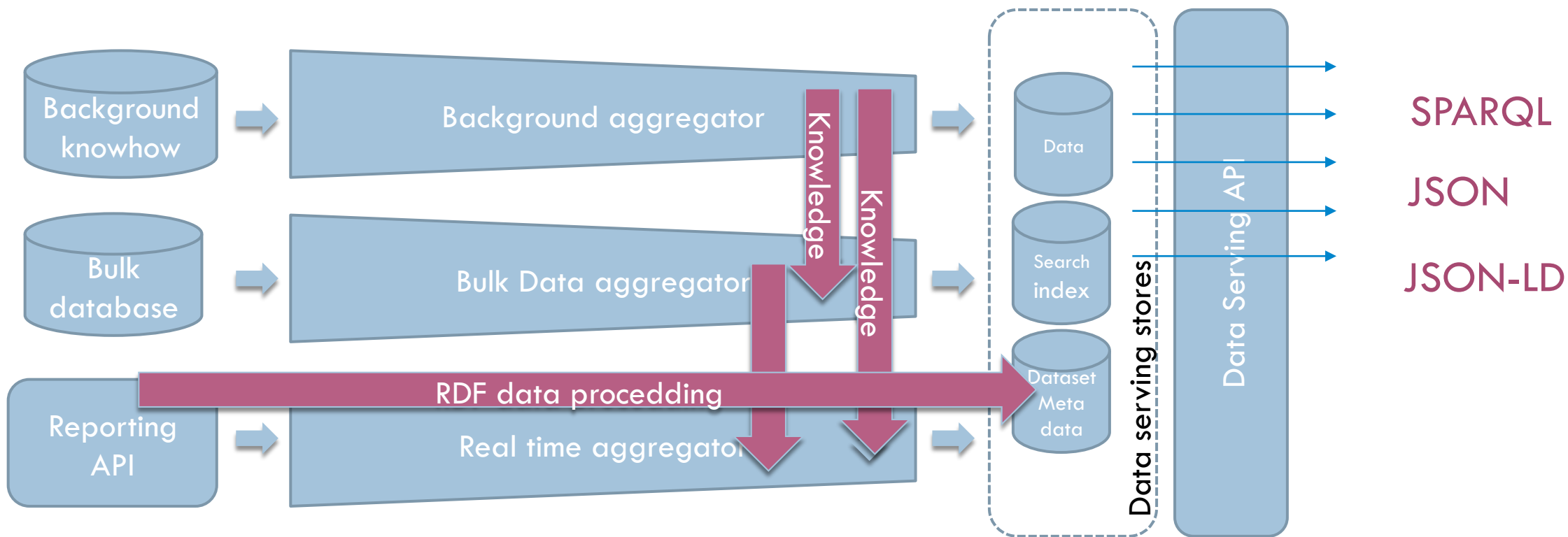
Semantic Web aspects: Data

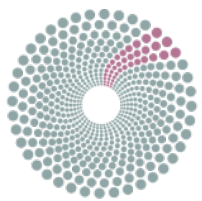
⊙ Semantic Web data

- All component interfaces exchange RDF
- Data serving API supports LD/SW formats
 - ❖ JSON, SPARQL & co
- Besides ingesting RDF data and LD, run-time accessing SPARQL endpoints
 - ❖ At least for the purposes of dynamically ingesting

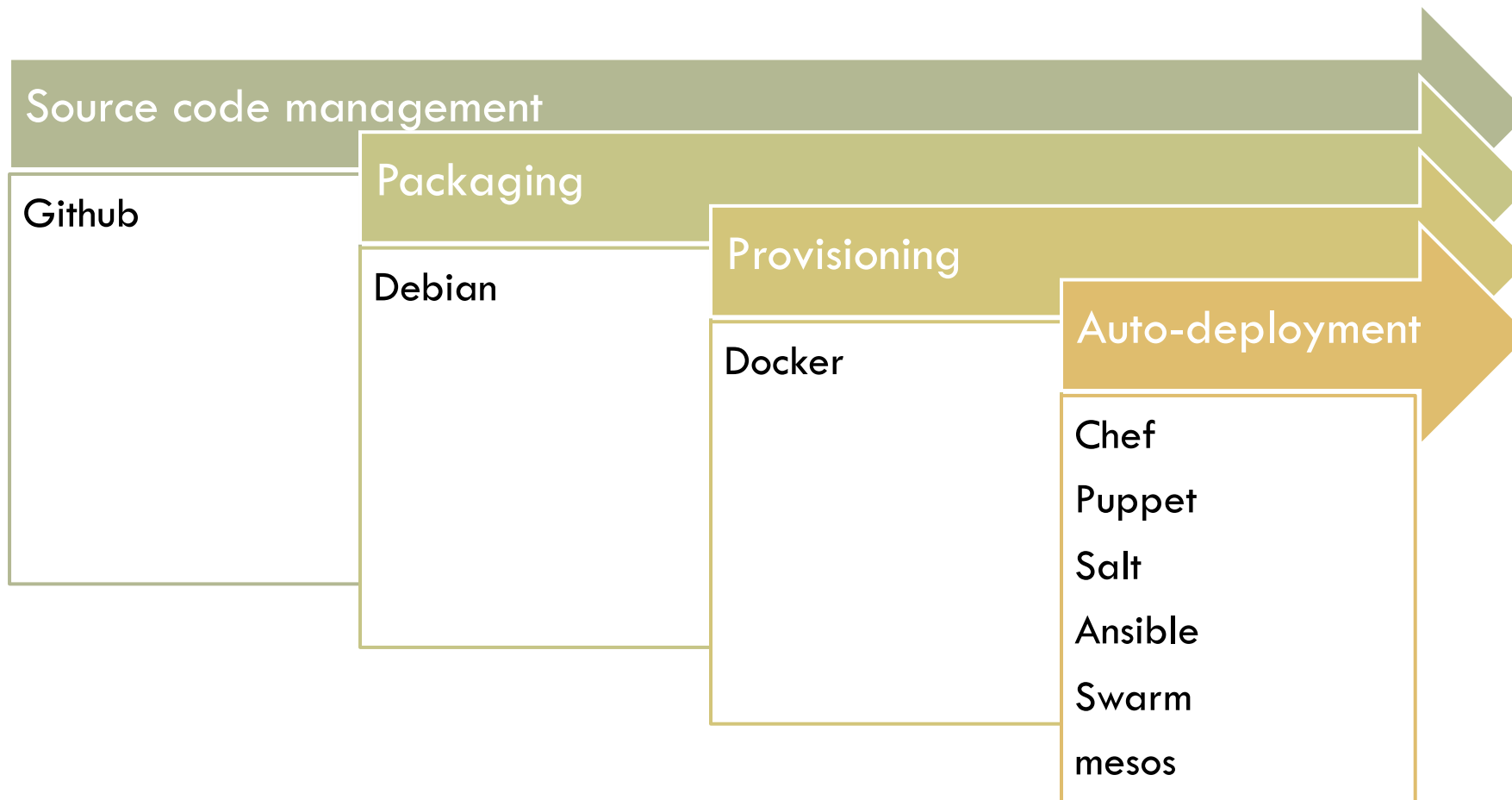


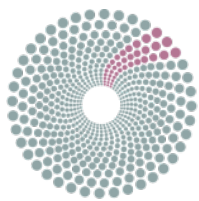
Big Data Aggregator Conceptual Architecture





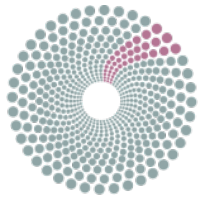
From source code to deployed instance





Starting Point

- ⊙ **Horton Works**, <http://hortonworks.com>
 - Integrated suite of Big Data processing tools
- ⊙ **LOD2 stack**, <http://www.lod2.eu>
 - Tools for the Data Web
 - Ontologies
 - Automatically interlinking and fusing Web data
 - Provenance, privacy, security, quality
 - Searching, browsing, and authoring of Linked Data.
- ⊙ **SemaGrow**, <http://www.semagrow.eu>
 - Federated SPARQL querying
 - Data integration
 - Optimized query execution
 - ❖ Including over uncooperative endpoints
 - Provenance metadata
- ⊙ **GeoKnow**, <http://www.geoknow.eu>
 - Geospatial and LD integration
 - Data provenance
 - Adaptive geospatial exploration, authoring and curation



Thank you for your attention

🎯 Questions?