



CROSSMINER

**Developer-Centric Knowledge Mining from
Large Open-Source Software Repositories**

Boris Baldassari, Castalia Solutions

About Crossminer



Helping people better manage their software project by providing in-context recommendations and analytics.

Crossminer:

- **Mines project data** from a wide range of **data sources**
- Builds a **knowledge base**
- And provides **recommendations** to the user through:
 - ◆ A dedicated IDE plugin for in-place advice.
 - ◆ Web dashboards with analytics and recommendations.

About Crossminer



- Crossminer is a EU-funded research project (H2020 program)
- It is the continuation of another EU-funded research project: Ossmeter
- The project involves research teams from all over Europe to address specific concerns and specialties:
 - ◆ Configuration management (puppet, docker..),
 - ◆ NLP (Natural Language Processing)
 - ◆ Deep Code Analysis
 - ◆ Integration with major forges
 - ◆ Software dashboards, recommendations & IDE plugins..
- Sustainability for the project is achieved with the creation of an Eclipse project to foster development and develop community.

CROSSMINER Innovations



Advanced IDEs



Eclipse-based IDE to support development of software systems including a wizard for specifying project components that developers want to reuse in developing new products.

Source Code Mining

Techniques and tools for mining the necessary knowledge to support decision making by software engineers at the software architecture and software design levels.



Text Mining



Text mining to allow developers to analyse a wide range of textual sources including newsgroups, bug trackers, mailing lists, forums, social media and stack overflow, to improve performance and quality of code.

Cross-Project Relationships



Technology Innovations Tools for specifying, discovering and representing different project relationships, based on extracted metrics and on similarity measures underpinning automated creation of clusters.

Configuration Mining



Sophisticated source code analysis tool for system configuration management code and a DevOps dashboard providing inferred insights derived from metrics from analysing system configurations.

Knowledge Extraction



Support for high performance and robust declarative project analysis workflows to plug together analytical components and define dependencies/interactions at a high level of abstraction.



Source code



Q&A systems



Bug Reports



API Documentation

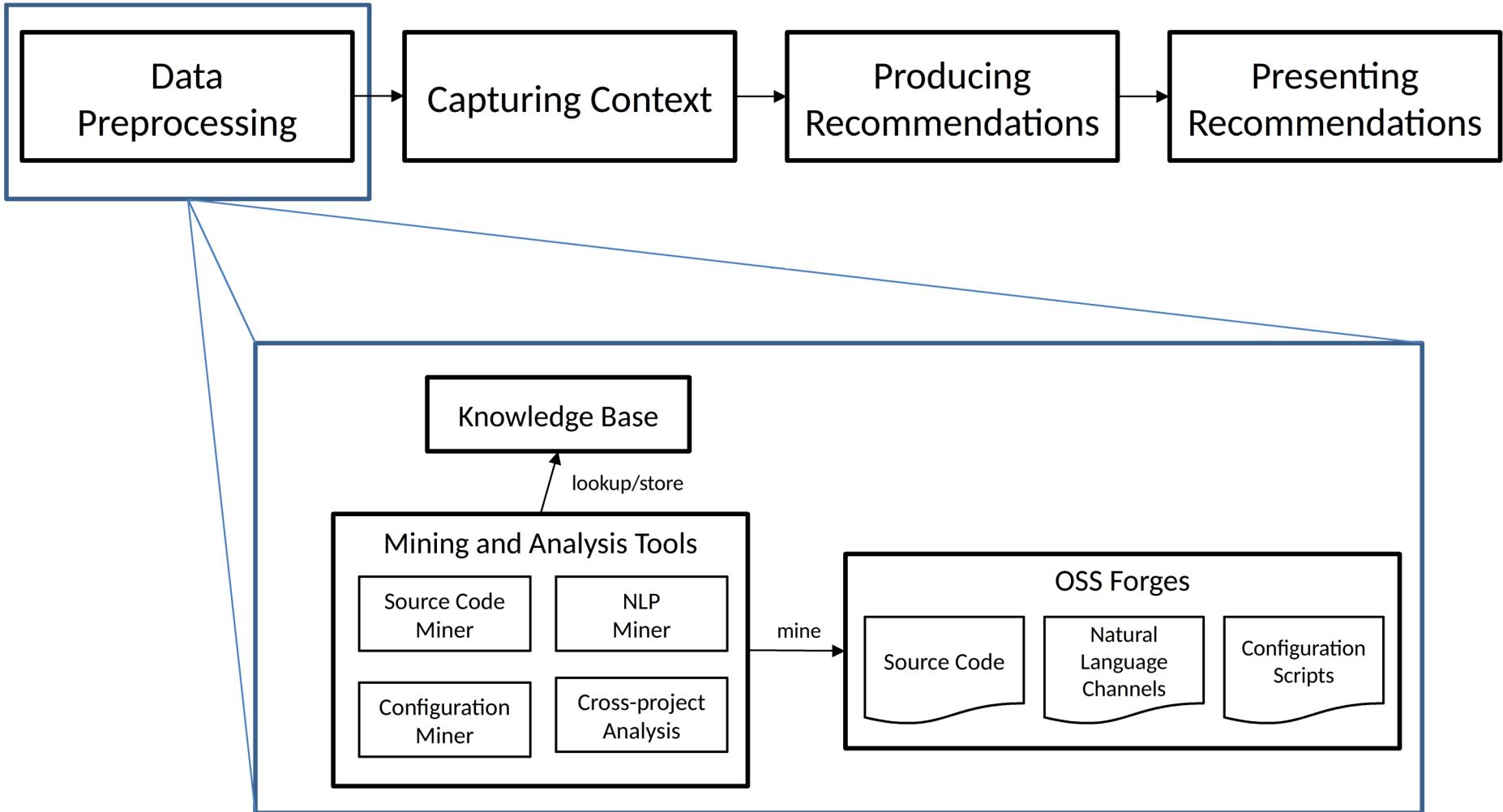


Tutorials

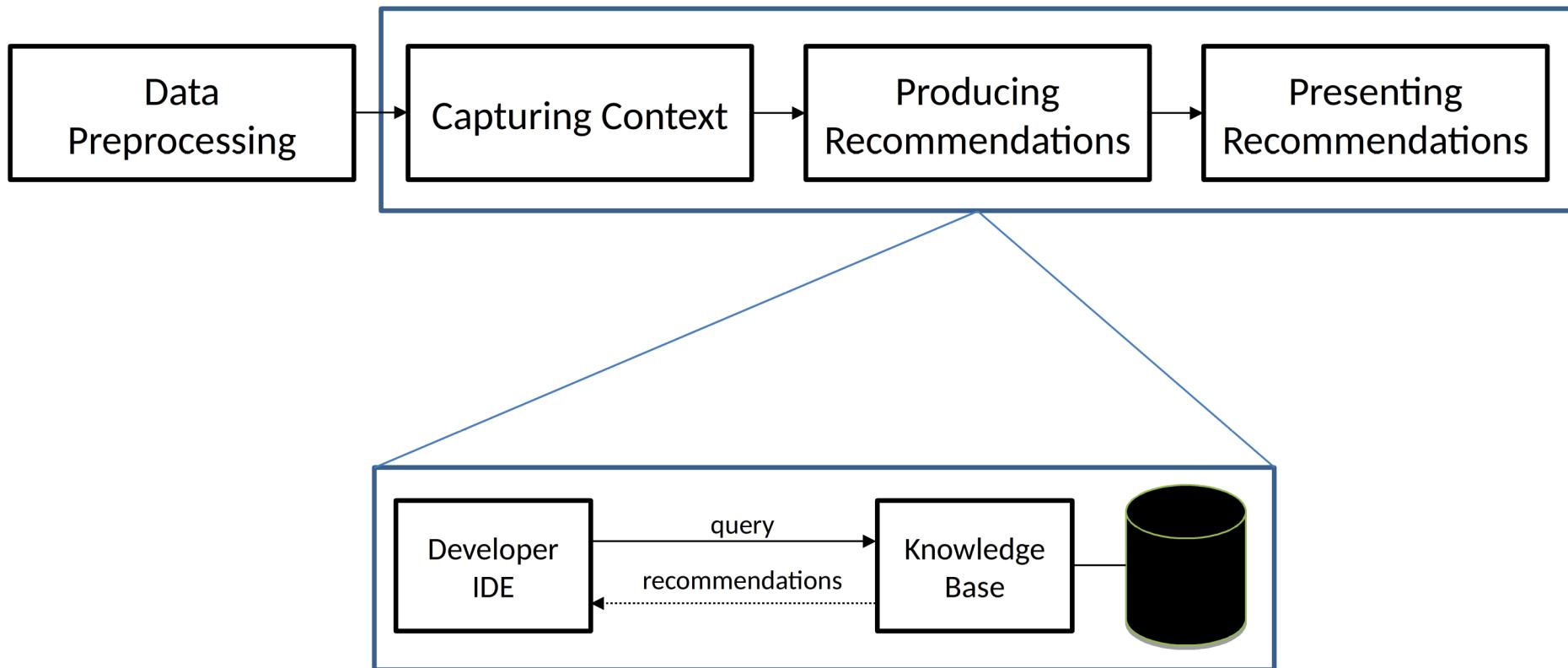


Configuration Management Systems

CROSSMINER: High-level view



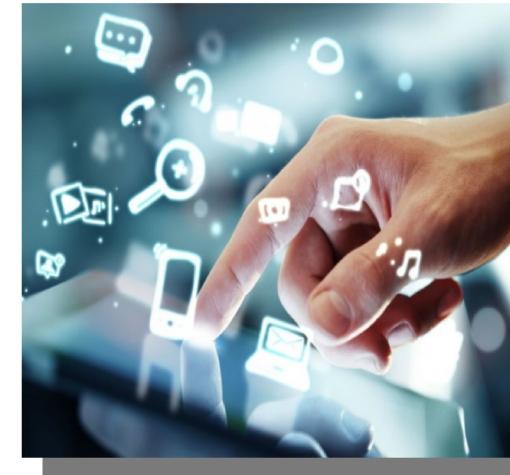
CROSSMINER: High-level view



Real-time recommendations that serve to increase productivity and quality

Industrial Use Cases - examples

- Arduino-based Internet of Things devices
 - ◆ Selecting and monitoring forked libraries aligned to platforms
- Multi-sector IT software services
 - ◆ Analysis and selection of OSS components for bespoke solutions
- Software APIs coupled evolution
 - ◆ Maintaining alignment with evolving third-party APIs and libraries
- Quantitative software analytics
 - ◆ Extending commercial analytics dashboards with additional metrics
- Software quality assurance and trustworthiness
 - ◆ Improving project quality on OSS forges
- Knowledge extraction from Eclipse forge projects
 - ◆ Cross-domain analysis and policy verification for OSS projects





Web: www.crossminer.org

Twitter: @crossminer

GitHub: github.com/crossminer/

Eclipse: projects.eclipse.org/projects/technology.crossmeter



The CROSSMINER project receives funding under the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 732223.