

Black Magic Meta Data

- get a glimpse behind the scene



Thomas “Voldemort” Vestdam, PhD



Henrik Steen “Saruman” Rasmussen



ELSEVIER

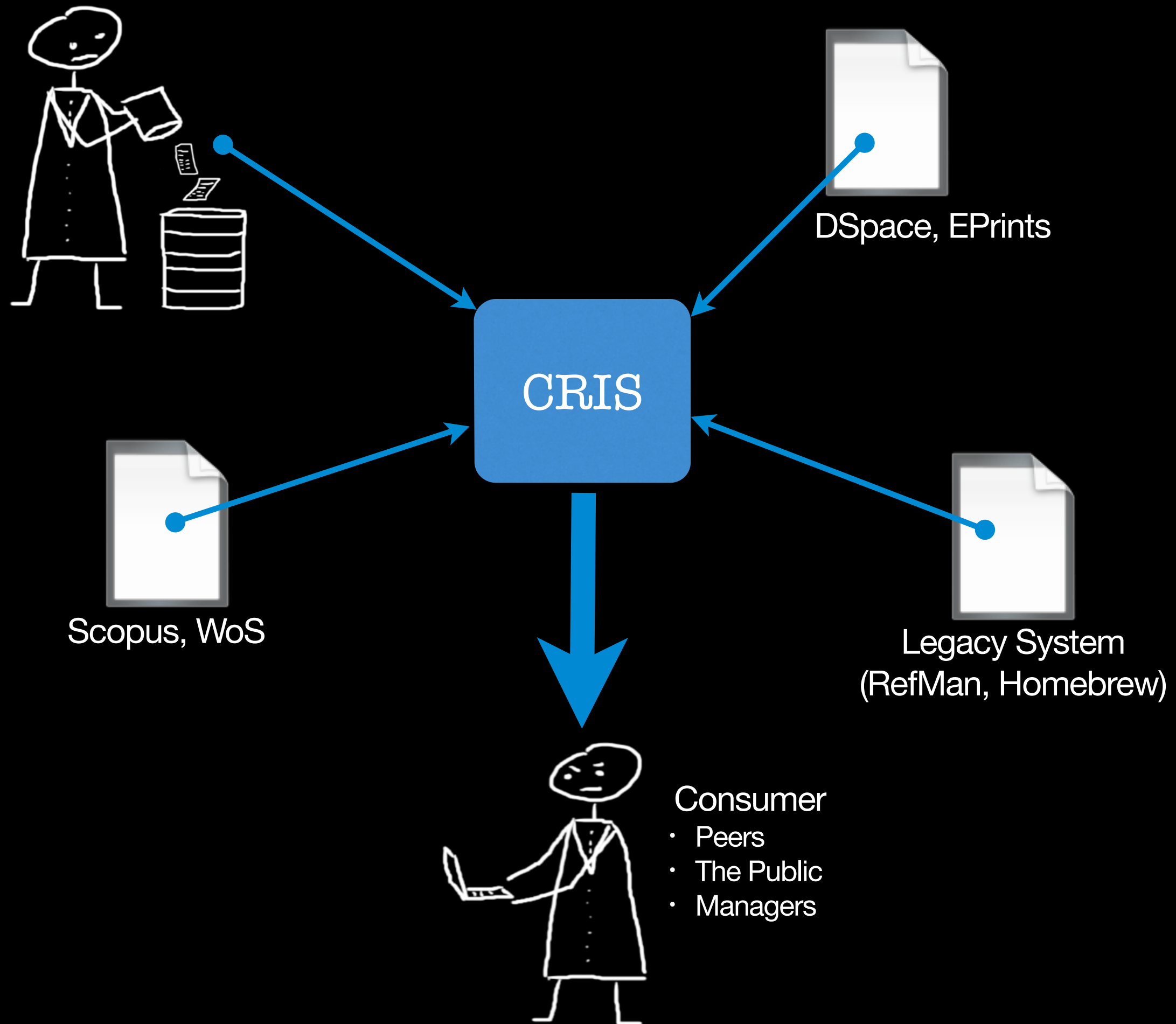
Pure



Marius “Sidious” Doornenbal, PhD

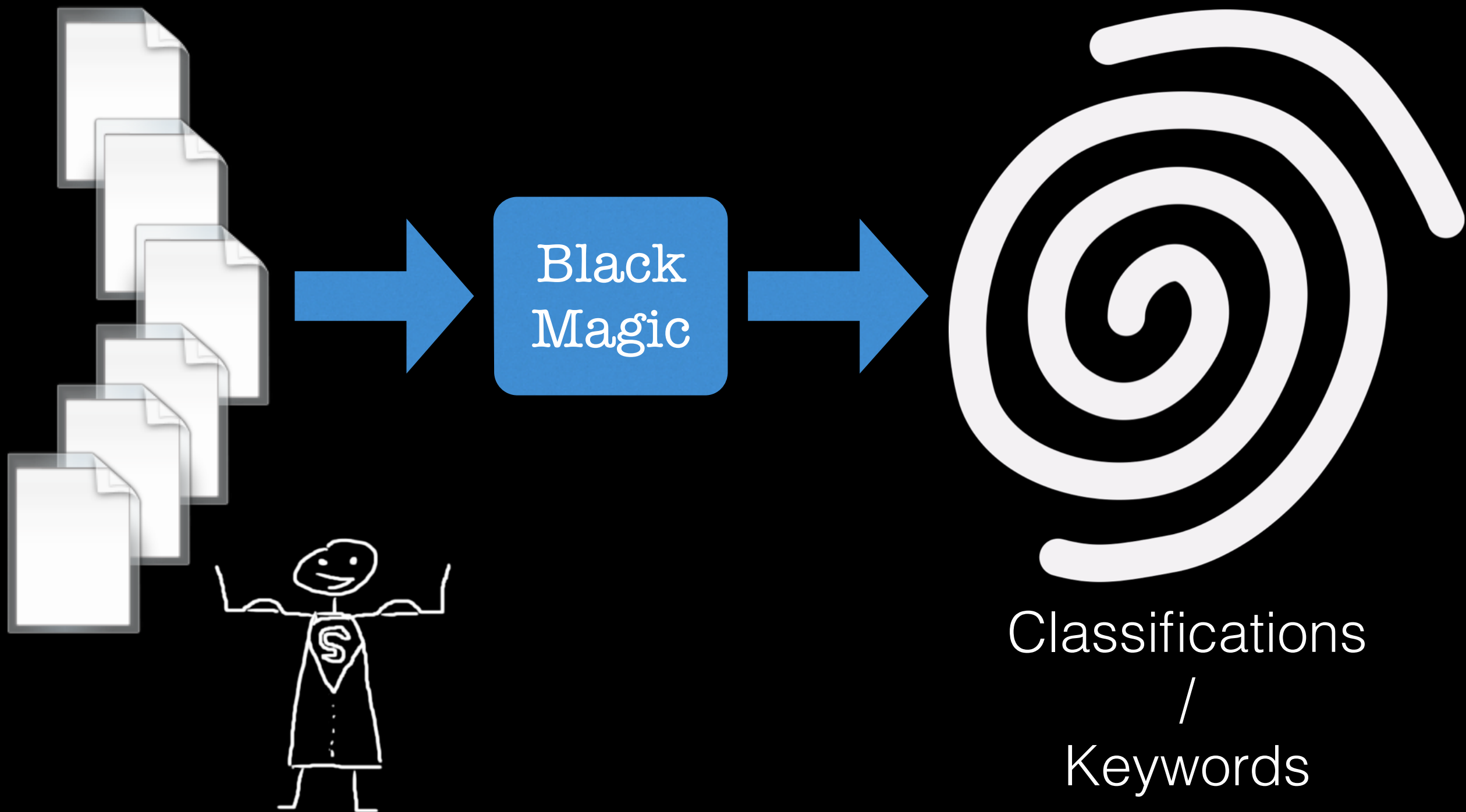
Why?

- If you have accurate, detailed, complete and fine-grained meta-data then you can *relate information, understand information, report on information, ...*
- Producing and maintaining accurate, detailed, complete and fine-grained meta-data is *laborious manual work, difficult, time consuming, ...*



What?

- We need tools to help us generate relevant meta-data that we are not getting from any input sources (for whatever reason)
- We need tools to help us understand our *crappy* meta-data in order to reason about our data



Researchers Portfolio

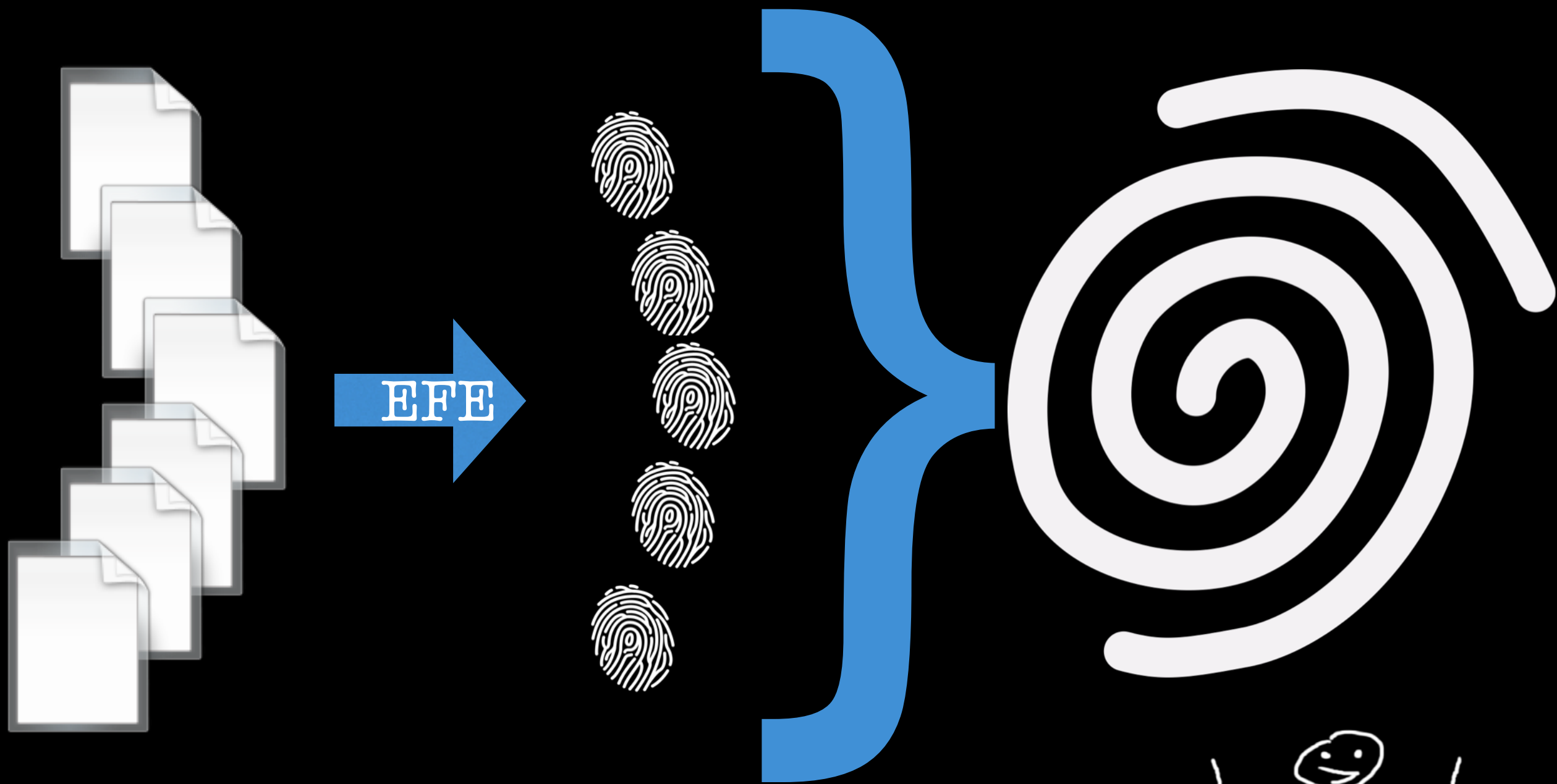
Classifications
/
Keywords

How?

Black
Magic

= Natural Language Processing

- “Understand” text as written by a human
- So, not just “basic” term extraction
 - Tokenization, input analysis and normalization, expansion of abbreviations and coordinations, a number of entity annotations (names, institutions, and citations) and part-of-speech tagging
- We call it “The Elsevier Fingerprint Engine”
 - Yes, this is proprietary, but the principle is not

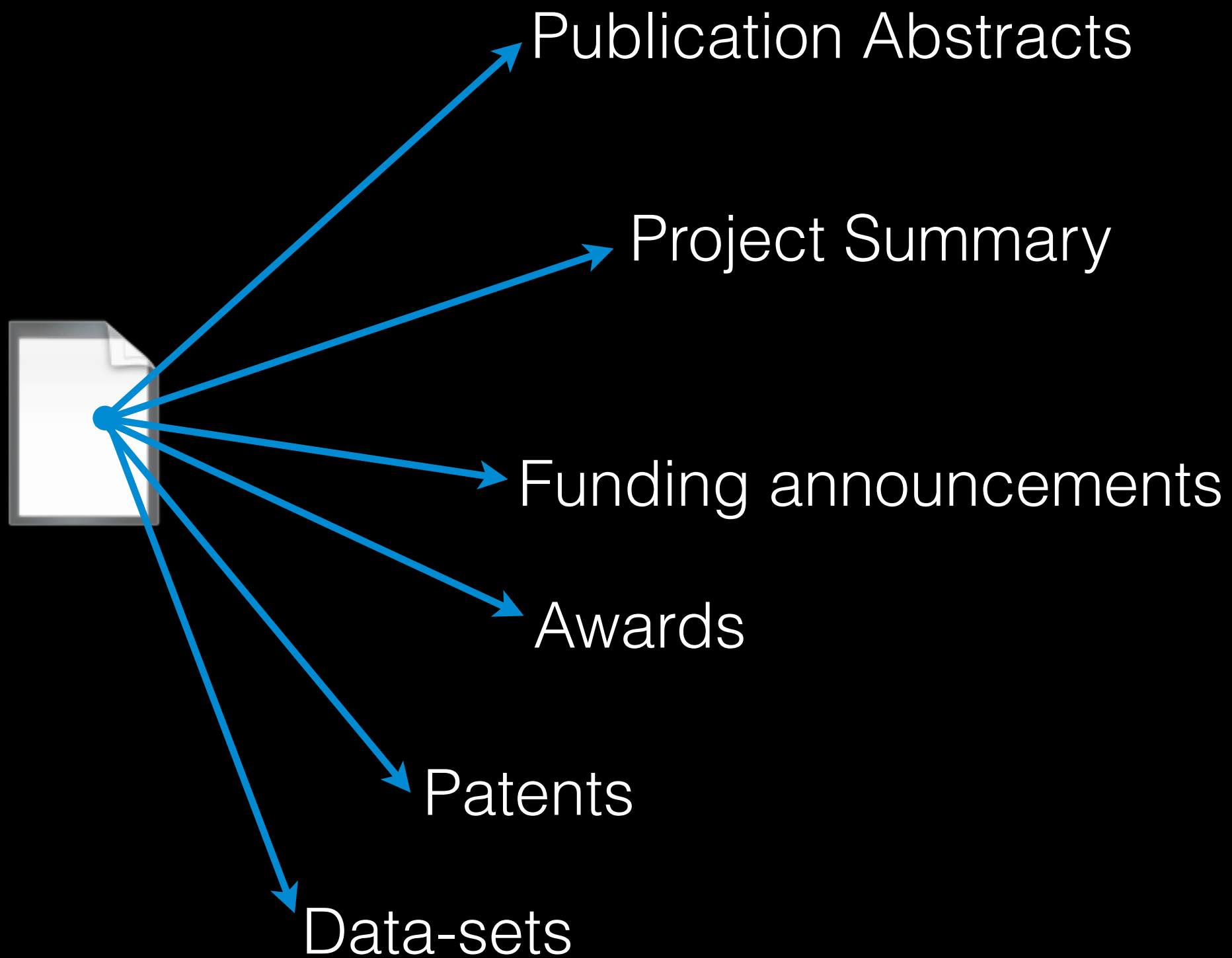


Researchers Portfolio

Requirements

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc,

- Text of a certain “quality” (in English)
- Semantics of the text must be mappable to at least one Thesaurus
- And, text must exist...





Thesaurus

Concept, Weight

⋮
n

- Agriculture & Biology (NAL)
- Arts & Humanities (Humanities Vocabulary)
- Business & Economics (Economics Vocabulary, STW Thesaurus Expanded)
- Chemical Substances & Compounds (Compendex (Ei) Thesaurus, MeSH, Supplementals Thesaurus)
- Earth & Environmental Sciences (Geobase Thesaurus)
- Engineering & Materials Science (Compendex (Ei) Thesaurus)
- Mathematics (Cambridge Math Thesaurus Expanded)
- Medicine & Life Sciences (MeSH Thesaurus)
- Physics & Astronomy (NASA Thesaurus)
- Social Sciences (Gesis Thesaurus Expanded)

Classification Scheme

Classification

Thesaurus

Concept, Weight



- Agriculture & Biology (NAL)
- Arts & Humanities (Humanities Vocabulary)
- Business & Economics (Economics Vocabulary,
- Chemical Substances & Compounds (Compendex (Ei) Thesaurus, MeSH,
- Earth & Environmental Sciences (Geobase Thesaurus)
- Engineering & Materials Science (Compendex (Ei) Thesaurus)
- Mathematics (Cambridge Math Thesaurus Expanded)
- Medicine & Life Sciences (MeSH Thesaurus)
- Physics & Astronomy (NASA Thesaurus)
- Social Sciences (Gesis Thesaurus Expanded)



Researchers, 50%

Text, 25%

Networking, 25%

Natural language processing, 25%

within Business & Economics



George Gettinby

Professor with special portfolio

[University of Europe](#) • [Faculty of Science](#) • [Department of Statistics & Modelling Science](#)

Phone: +45 96 35 61 00 • E-mail: gg.demo@atira.dk

Summary

Research outputs (1996 - 2013)	122
Co-author Network	144
Institutional Network	50

Expert Overview

Fingerprint

Research outputs

Journals

Institutional Network

Co-author Network

Fingerprint ?

Below displays the top Fingerprint concepts, per subject area for this Expert. Fingerprint concepts that appear on this page are based on all the research output produced by this Expert.

Engineering & Materials Science

Thermodynamic properties



Temperature



Thermodynamic stability

Hydrocarbons



Methane



Acetylene



Aromatic hydrocarbons



Hexane



Heptane



Olefins



Paraffins

Nonmetals



Oxygen



Hydrogen



Carbon

Concept: Carbon

Weight: 21%

Domain: Engineering & Materials Science

Semantic group: Chemical Engineering

Reforming reactions



Steam reforming



Catalytic reforming

Catalysts



Catalyst activity



Catalyst supports



Catalyst selectivity



Catalyst deactivation



Catalyst regeneration

Gas fuel manufacture



Hydrogen production

Use-cases

- Find Funding Opportunities



- Networking

- Find collaborators with similar interest



- Find reviewers

- (Carefully) Managing research “directions”

- Automagically annotate with “semantics”



But, this is just an example

- There must be other options for automatically creating or annotating meta-data
 - Based on behaviour (knowledge of the past)
 - Automatic import (and matching)

Summary

- Natural Language Processing is a useful tool within a CRIS, both for managing and presenting meta data
 - Suggest funding opportunities
 - Find “similar” researchers
- But, there must be other ways of applying black magic in CRIS systems
- Fingerprinting rocks
 - Hence, Pure rocks