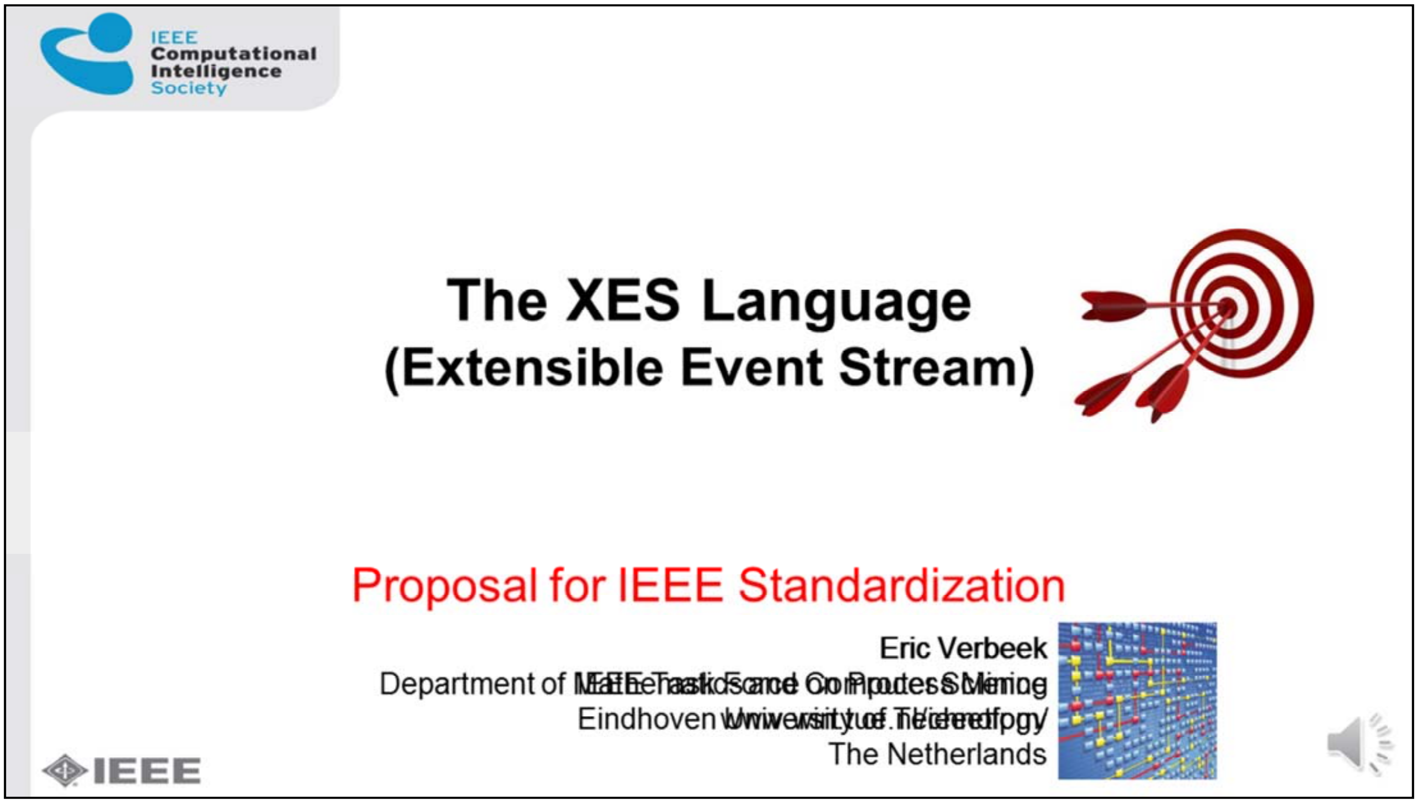


IEEE Computational Intelligence Society

The XES Language (Extensible Event Stream)

Proposal for IEEE Standardization

Eric Verbeek
Department of Mathematics and Computer Science
IEEE Task Force on Process Mining
Eindhoven University of Technology
www.win.tue.nl/ieeetfpm/
The Netherlands

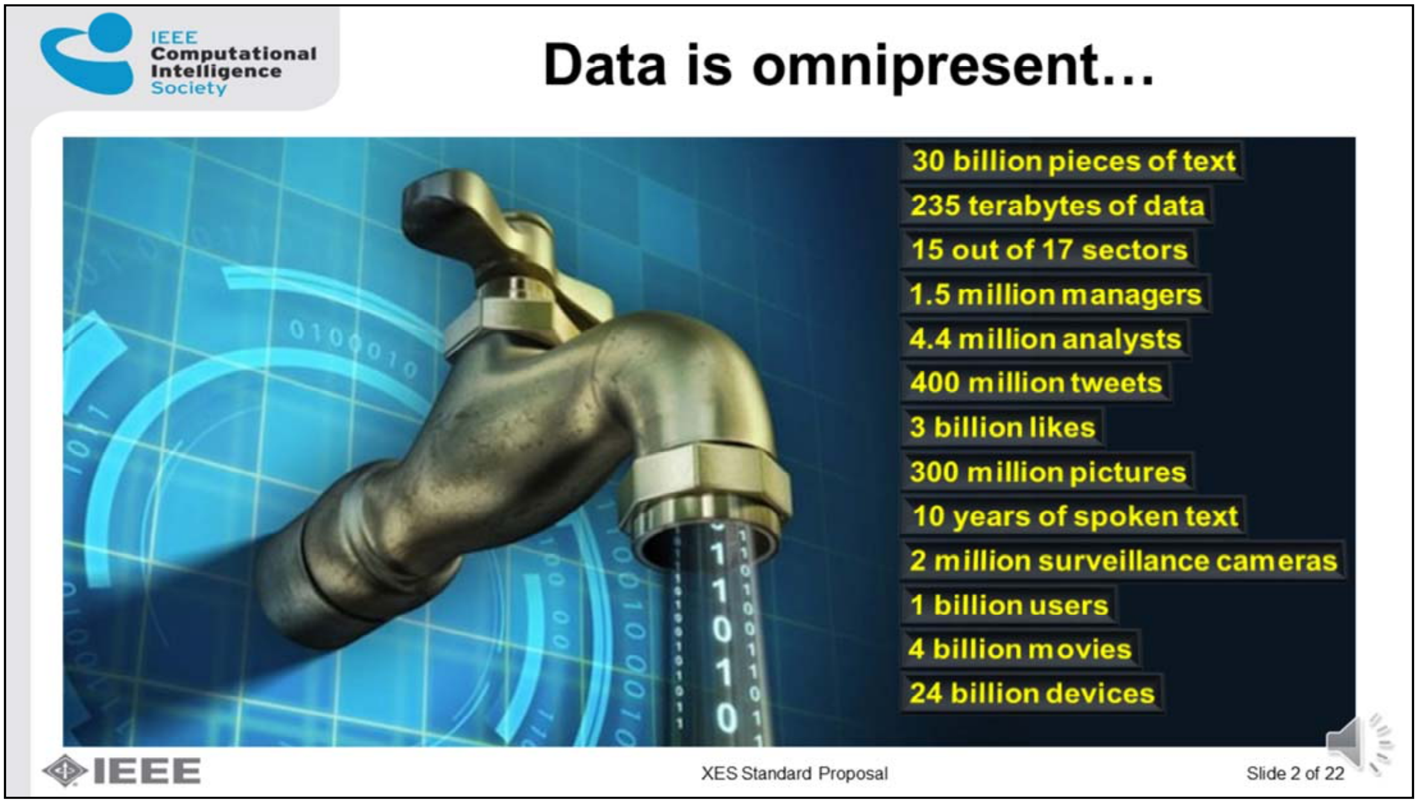
IEEE

Welcome to this webinar on the proposal for IEEE Standardization on the XES language, where XES stands for Extensible Event Stream.

My name is Eric Verbeek, and I work as a Scientific Engineer at the Department of Mathematics and Computer Science at the Eindhoven University of Technology in the Netherlands.

I will be presenting this webinar today, but please note that I do so on behalf of the [TAB]IEEE Task Force on Process Mining.

The [TAB]goal of this webinar is to introduce the XES language, which is a language to transfer event data from one location to another location.



Data is omnipresent…

- 30 billion pieces of text
- 235 terabytes of data
- 15 out of 17 sectors
- 1.5 million managers
- 4.4 million analysts
- 400 million tweets
- 3 billion likes
- 300 million pictures
- 10 years of spoken text
- 2 million surveillance cameras
- 1 billion users
- 4 billion movies
- 24 billion devices

XES Standard Proposal

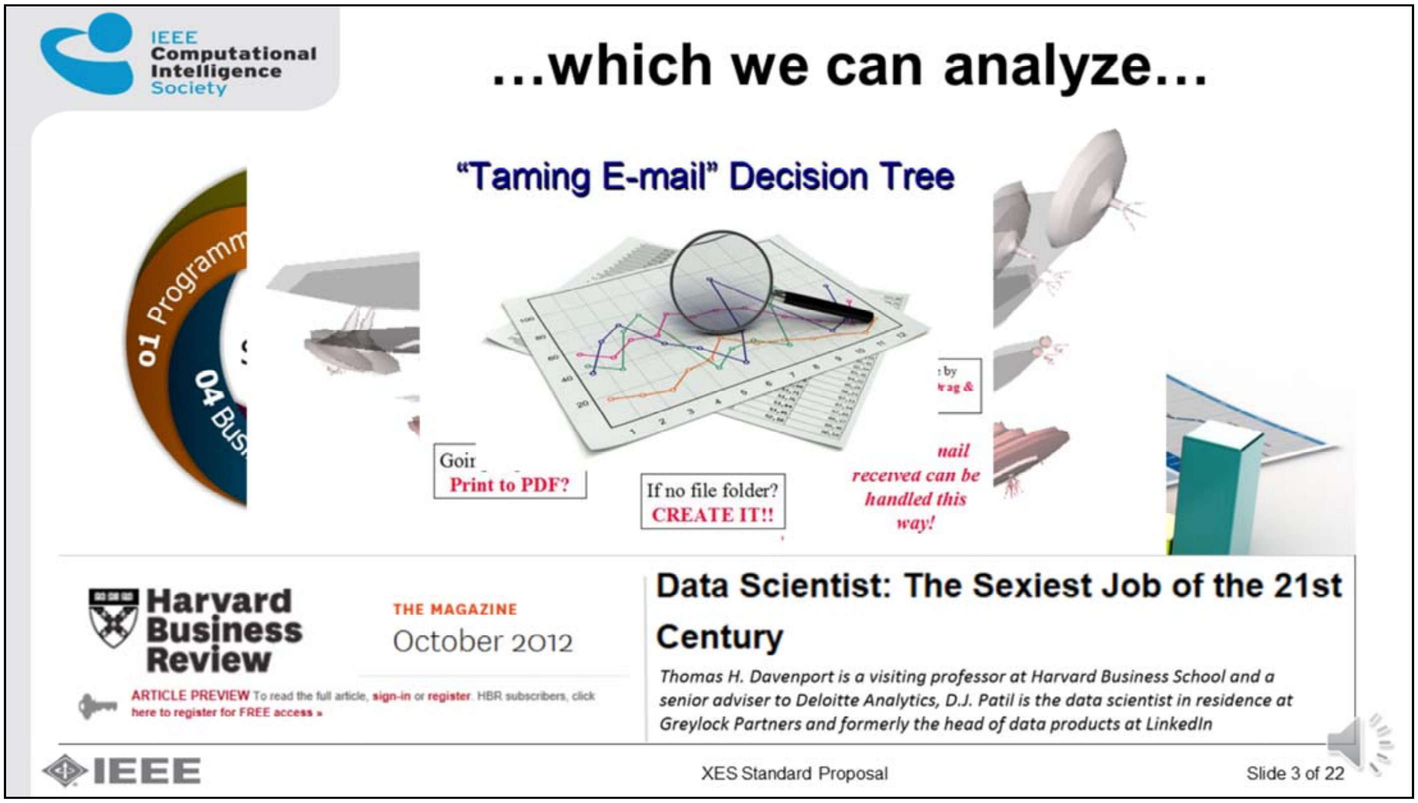
Slide 2 of 22

The real need for such a language originates from the fact that, these days, data is almost omnipresent, and coming at us in a rapid pace.

For example, in two-thousand-eleven, McKinsey estimated that [TAB]thirty billion pieces of content were shared on Facebook, every month, that [TAB]two-hundred-and-thirty-five terabytes of data have been collected by the US Library of Congress by April of that year, that [TAB]fifteen out of seventeen sectors in the US have more data stored per company than the US Library of Congress, and that we need [TAB]one-and-a-half million more data-savvy managers to take full advantage of big data in the US alone.

In two-thousand-twelve, Gartner stated that in two-thousand-fifteen there would be a need of [TAB]four-point-four million analysts worldwide; of which only twenty-five percent can be met.

Figures from a Dutch newspaper from two-thousand-thirteen show a similar story: People produce per day [TAB]four-hundred million tweets, send over [TAB]three billion likes, and upload [TAB]three-hundred million pictures. Furthermore, Google Voice processes [TAB]ten years of spoken text daily, the UK has [TAB]two million surveillance cameras, and Facebook has [TAB]one billion users, who watch [TAB]four billion movies daily. Finally, in two-thousand-twenty there will be [TAB]twenty-four billion internet connected devices.



IEEE Computational Intelligence Society

…which we can analyze…

Harvard Business Review

THE MAGAZINE

October 2012

ARTICLE PREVIEW To read the full article, sign-in or register. HBR subscribers, click here to register for FREE access »

Data Scientist: The Sexiest Job of the 21st Century

Thomas H. Davenport is a visiting professor at Harvard Business School and a senior adviser to Deloitte Analytics, D.J. Patil is the data scientist in residence at Greylock Partners and formerly the head of data products at LinkedIn

IEEE XES Standard Proposal Slide 3 of 22

As a result of this omnipresence of data, a new breed of scientist has emerged: The [TAB]Data Scientist, which has become of very promising job [TAB].

Among other things, a data scientist must be very innovative and distinctive in his approach in applying various techniques intelligently to extract data and get useful insights in solving business problems and challenges, have hands-on experience in, for example, data mining techniques such as [TAB]graph analysis, [TAB]pattern detection, [TAB]decision trees, and [TAB]statistical analysis, be able to analyze data from a variety of sources, and have the ability to locate and construe rich data sources.[TAB]

IEEE Computational Intelligence Society

…if we can get it.

IEEE

XES Standard Proposal

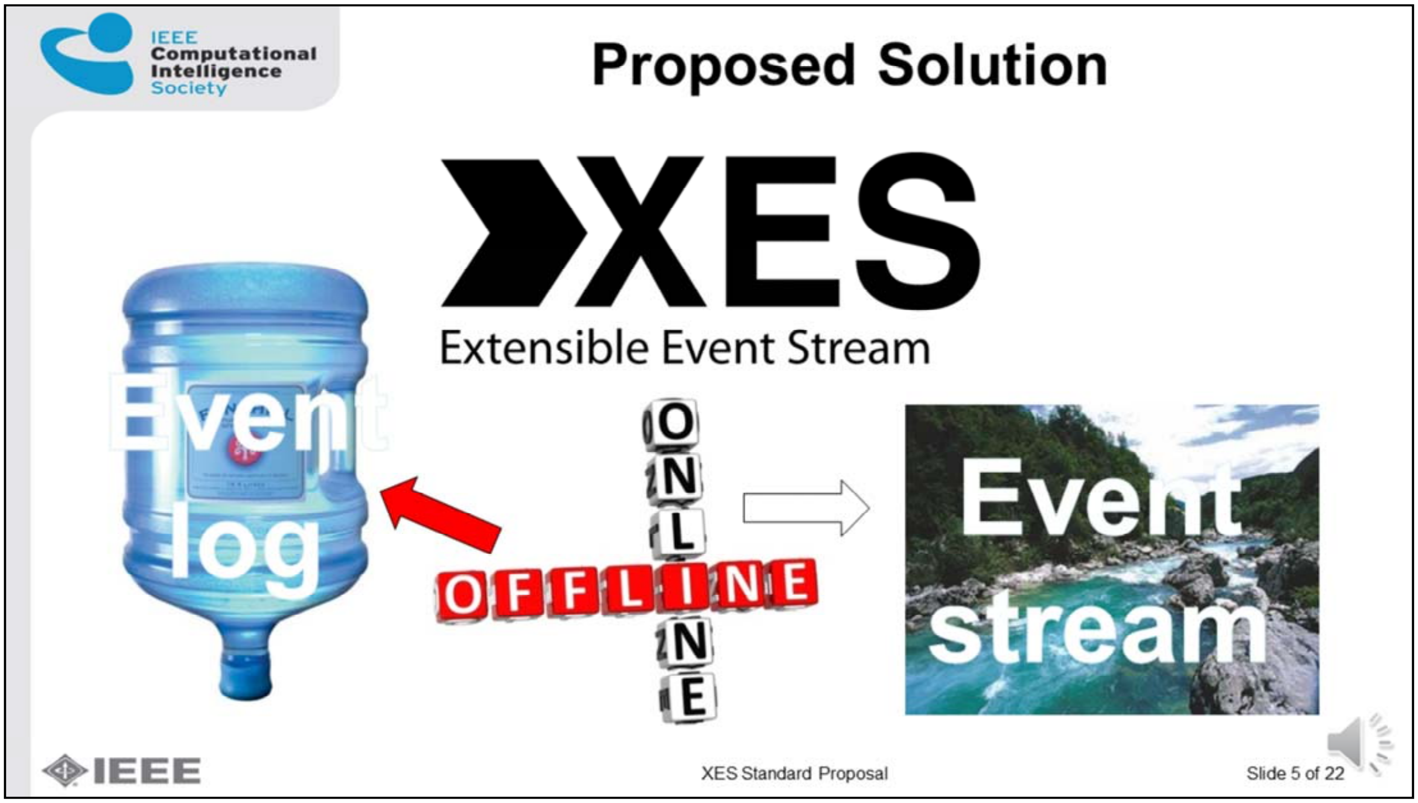
Slide 4 of 22

However, we need to transfer the data from the location where it is being generated to the location where it can be analyzed, that is, to the location where the data scientists can do their thing.

Also, once transferred, the data might have to be stored for later use at the analysis location.

Clearly, this transfer and storage needs to be done in a standardized way, that is, in a way that is clear, and does not obfuscate the data in any way.

Thus, the [TAB]syntax of the standard should be clear and well-understood, its [TAB]semantics should also be clear and well-understood, and, finally, it should be [TAB]extensible, as at the moment of speaking, we cannot foresee which event data will be generated and analyzed in the future, like map makers in the eighteenth century could not have foreseen the need for a map with metro lines.



IEEE Computational Intelligence Society

Proposed Solution

XES

Extensible Event Stream

Event
log
OFFLINE
ONLINE
Event
stream

IEEE

XES Standard Proposal

Slide 5 of 22

For this transfer and storage, the IEEE Task Force on Process Mining would like to propose a [TAB]new standard, called the Extensible Event Stream standard, or XES in short.

This XES standard should allow for both an [TAB] off-line setting and an on-line setting.

In the off-line setting, [TAB]the required data is transferred and stored at the analysis site, after which it can be thoroughly analyzed by the data scientists.

In this setting, the data is fixed, and does not change after it has been stored.

As a result, the data may be stored in an efficient way after transfer is done.

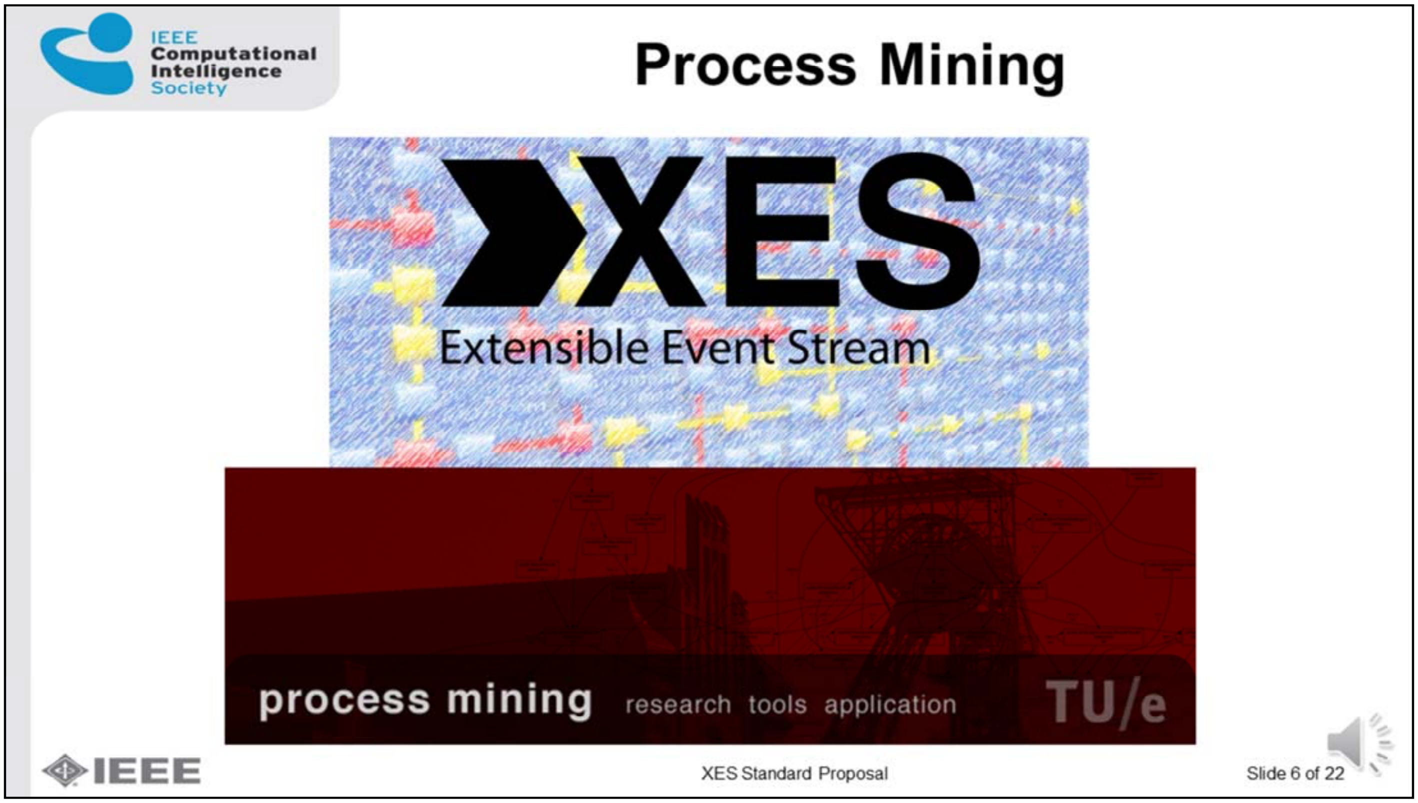
We refer to this kind of data transfer and storage as an [TAB]event log.

In the on-line setting, [TAB]the required data is transferred in a piece-meal fashion to the analysis site, and the required data typically keeps on coming.

In this setting, the data being analyzed is not fixed, and keeps changing.

As a result, the data can typically not be stored as a whole.

We refer to this kind of data transfer as an [TAB]event stream.



IEEE Computational Intelligence Society

Process Mining

XES
Extensible Event Stream
process mining research tools application
TU/e

IEEE

XES Standard Proposal

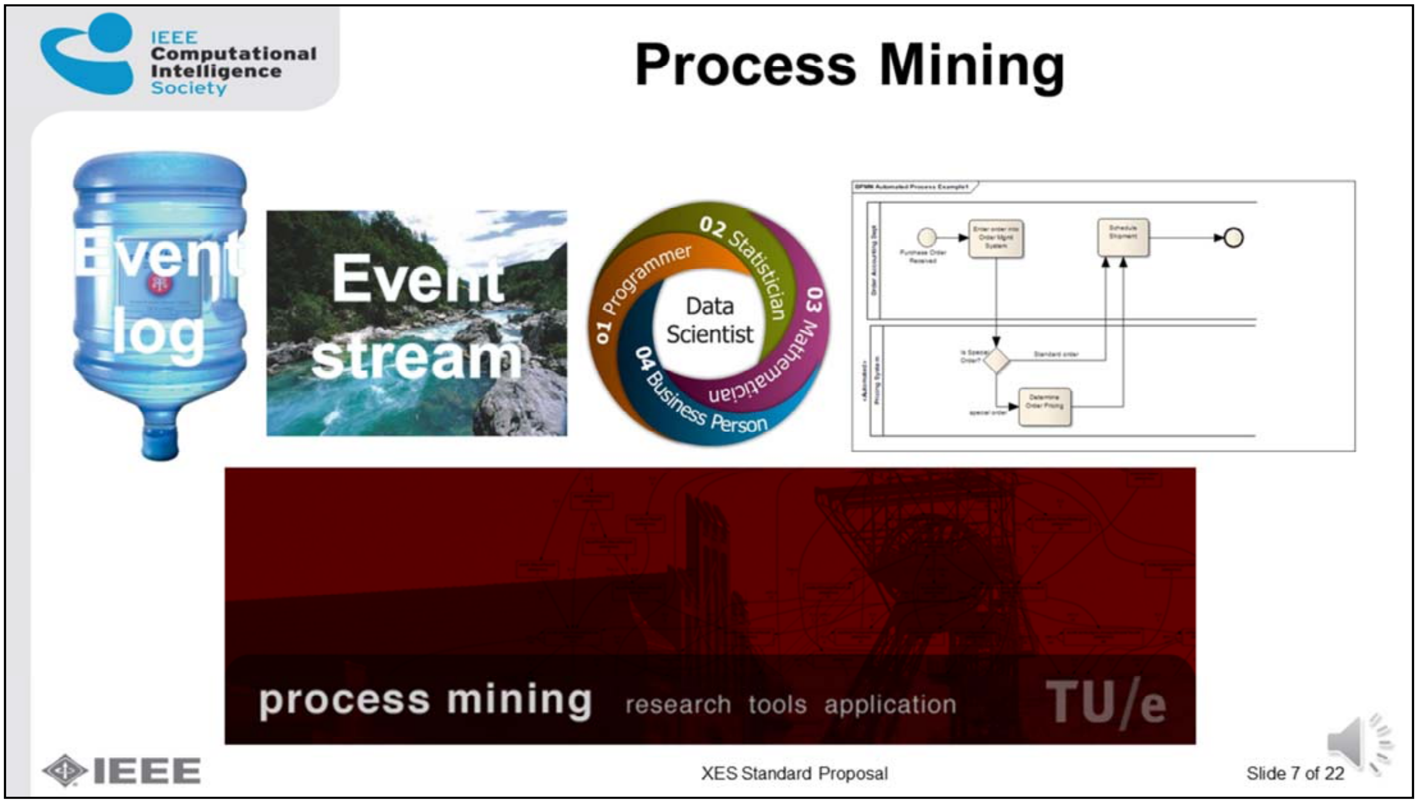
Slide 6 of 22

During the remainder of this webinar, we would like to inform you on the XES standard we propose, how it should look like, what it should be, which requirements it should fulfil, etc.

Being the [TAB]IEEE Task Force on Process Mining, we would like to do so in the context of [TAB]process mining.

Therefore, we will first provide a brief introduction on the process mining field, but please be aware of the fact that the use of the XES standard is not limited to the process mining field.

Any field where event driven data is being generated and analyzed can benefit from this standard.



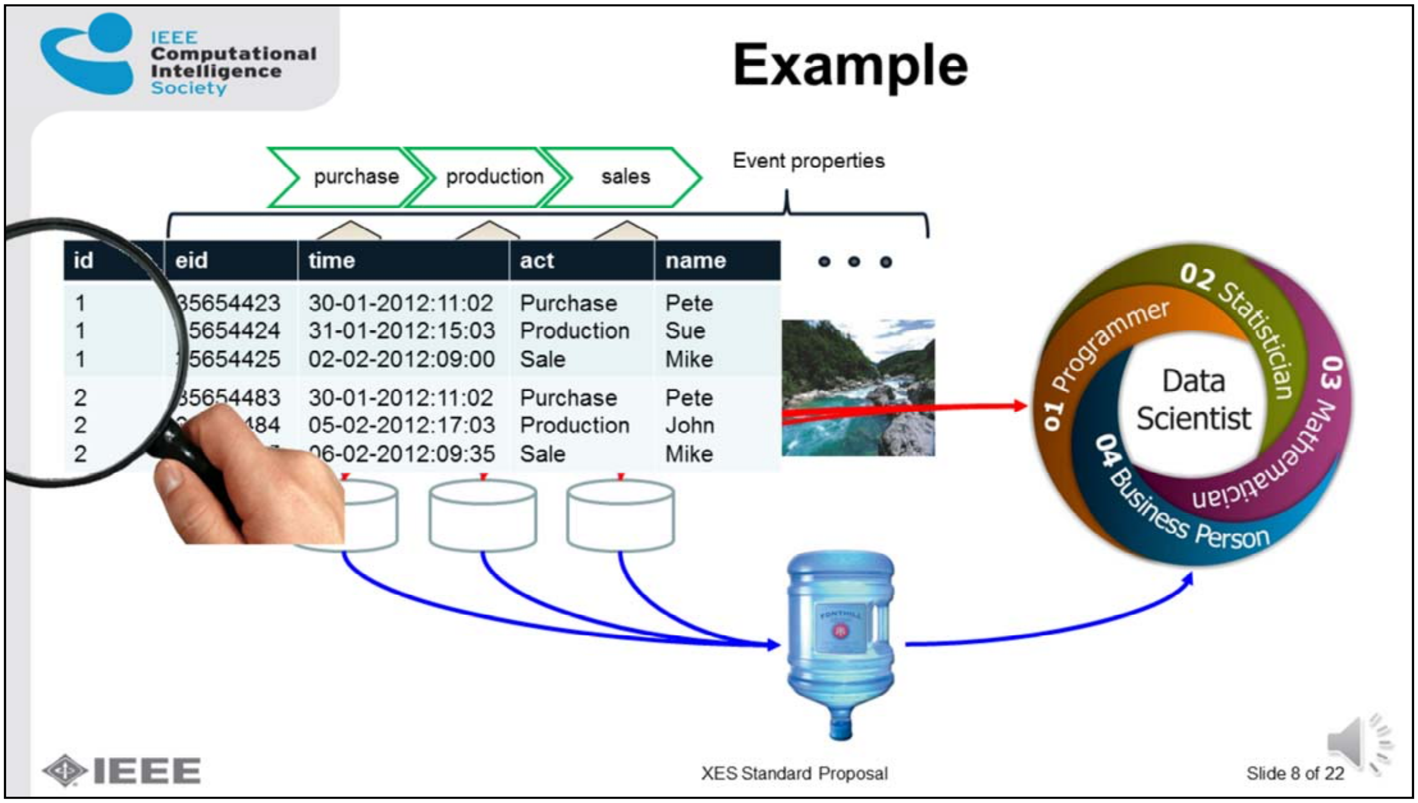
Process Mining

XES Standard Proposal

Slide 7 of 22

Using process mining, a [TAB]data scientist can establish a link between an [TAB]actual process and its data on the one hand and the [TAB]process model on the other hand.

As such, process mining offers a data scientist the possibility to use the recorded data, that is, actual and factual information, to provide a better view on the actual process, that is, deviations can be analyzed and the quality of the process model can be improved.



Example

id	eid	time	act	name
1	35654423	30-01-2012:11:02	Purchase	Pete
1	5654424	31-01-2012:15:03	Production	Sue
1	5654425	02-02-2012:09:00	Sale	Mike
2	35654483	30-01-2012:11:02	Purchase	Pete
2	[illegible]484	05-02-2012:17:03	Production	John
2	[illegible]	06-02-2012:09:35	Sale	Mike

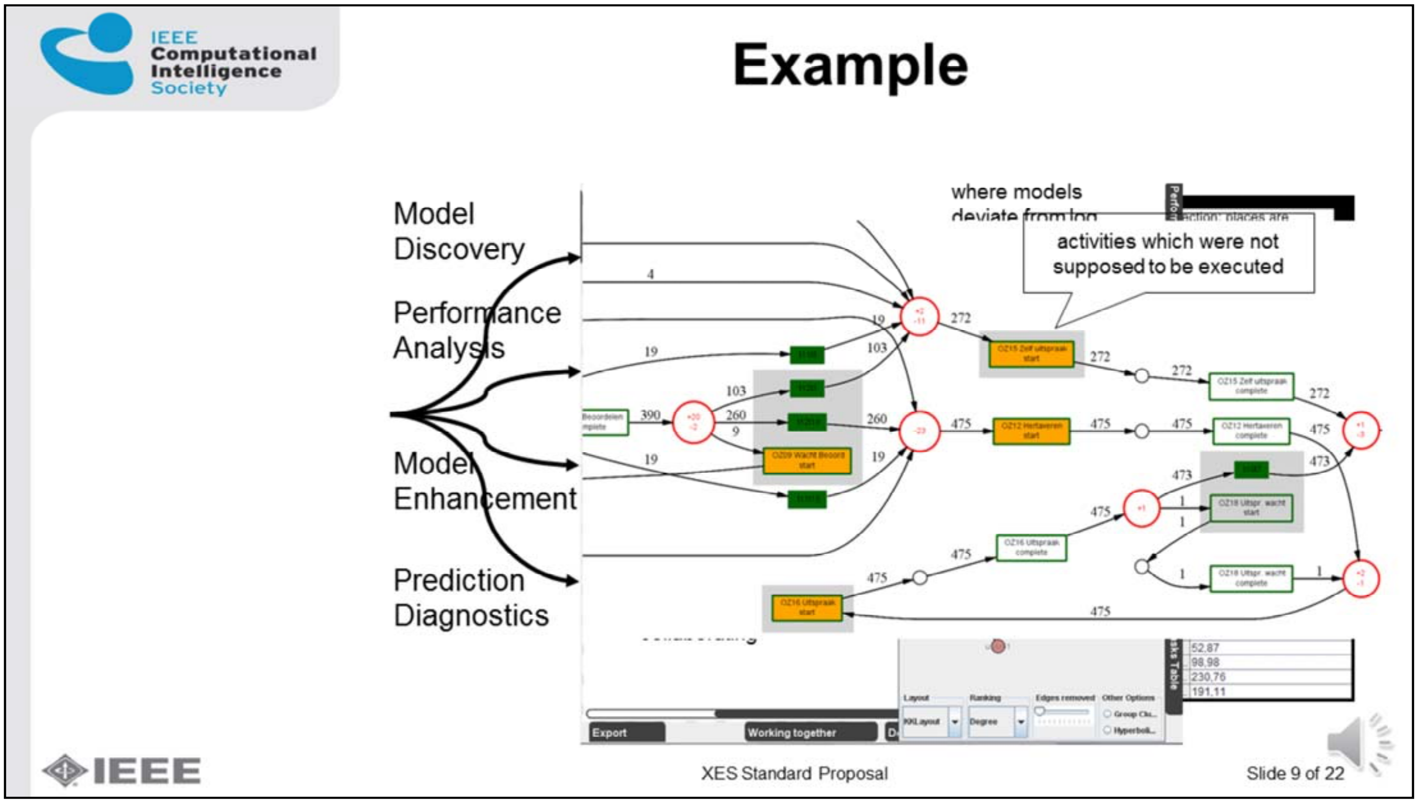
IEEE

XES Standard Proposal

Slide 8 of 22

Consider, for example, a [TAB]purchase-production-sales system. In this system, each part uses its own supporting [TAB]software system. Each of these systems uses its own [TAB]database, in which everything is stored and recorded. Somehow, we need to transfer the relevant recorded data to the data [TAB]scientist for analysis. We can do the transfer either in an [TAB]off-line setting, using a fixed event log, or in an [TAB] on-line setting, using an event stream.

Either way, typical [TAB]data for an event will contain information on the [TAB]corresponding case, the [TAB]id of the event, the [TAB]time the event was generated, the [TAB]activity the event refers to, the resource whose action triggered the event, and [TAB]possibly many more.



IEEE Computational Intelligence Society

Example

Model Discovery

Performance Analysis

Model Enhancement

Prediction Diagnostics

where models deviate from log

activities which were not supposed to be executed

IEEE

XES Standard Proposal

Slide 9 of 22

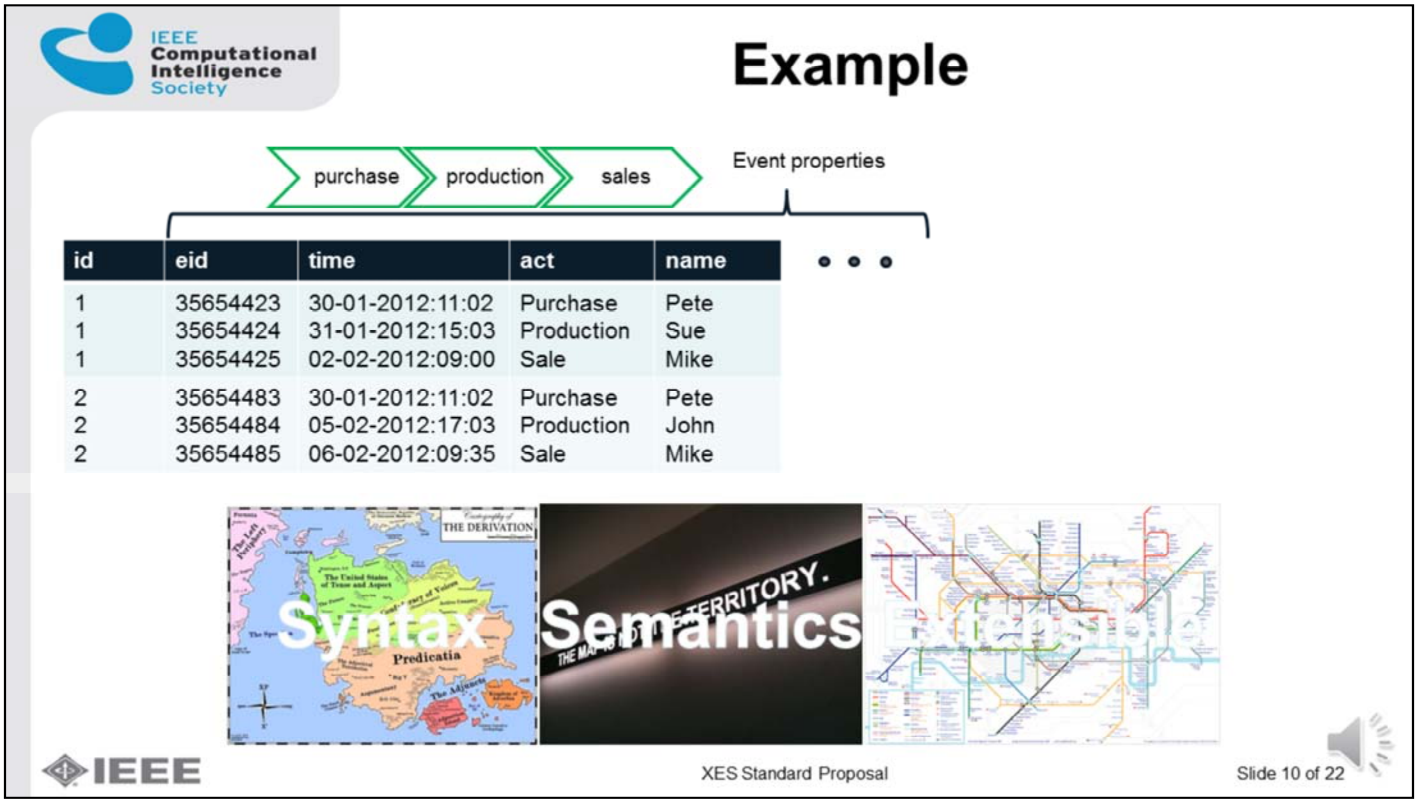
[TAB] Based on event data from such a purchase system, a data scientist could do several things.

As a first example, the data scientist can [TAB]discover a [TAB]process model from the event data, without using any a-priori information. If the event data also contains information about resources, a data scientist can also discover resource-related models, like a [TAB]social network showing how people work together in the organization.[TAB]

Second, on either an existing or a discovered model, the data scientist can [TAB]analyze the performance of the system. For example, [TAB]bottlenecks can be detected, information on flow time can be provided, and the average time it takes to move from one activity to another can be reported.[TAB]

Third, the data scientist can [TAB]enhance either an existing or a discovered model. For example, [TAB]information on where models deviate from an event log can be projected onto the model, frequent execution paths could be visualized, information on execution and waiting times could be provided for activities, and people who often hand over work to each other could be shown.[TAB]

Finally, the data scientist can [TAB]provide prediction diagnostics on the current process. For example, he can report [TAB]which activities occurred in situations where they were not supposed to occur according to the model.[TAB]



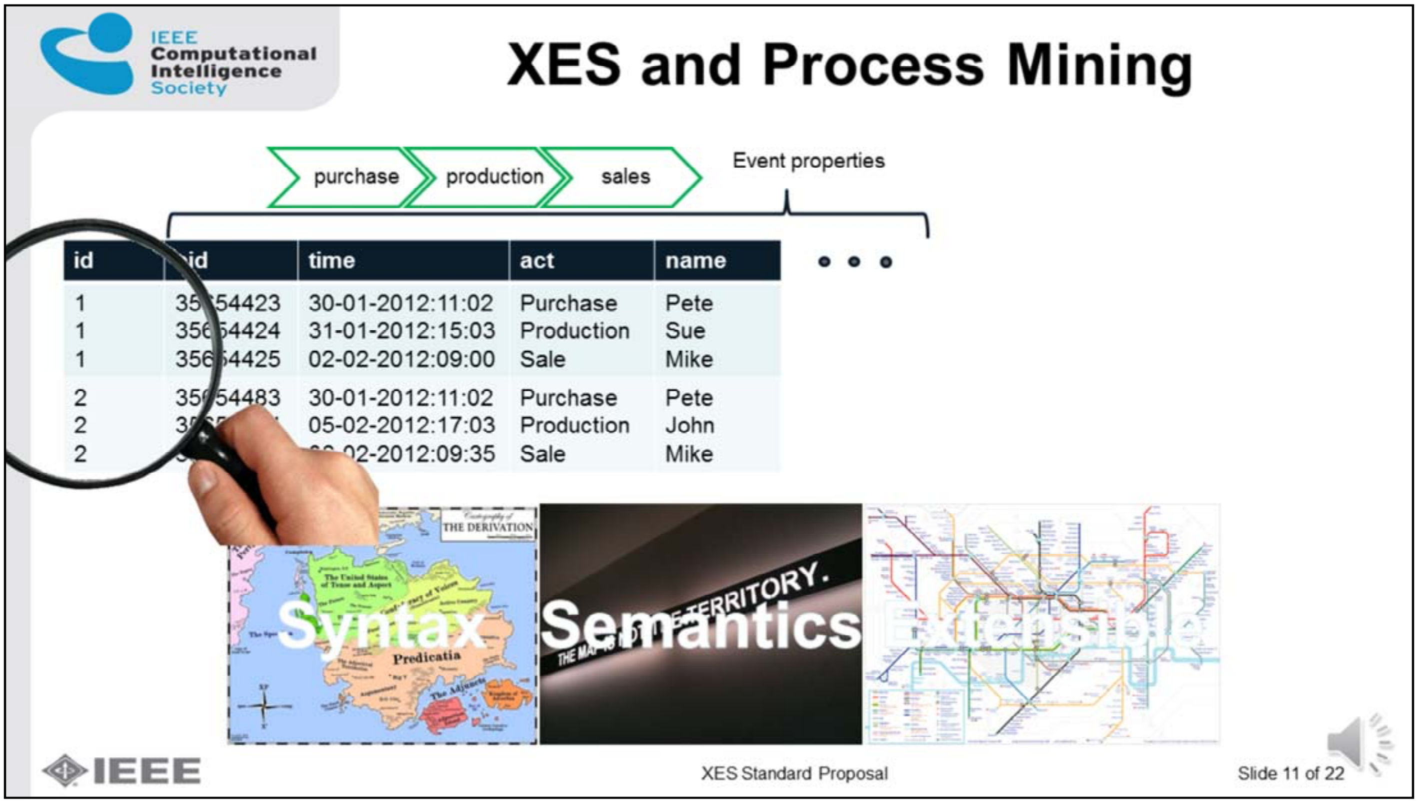
Example

id	eid	time	act	name
1	35654423	30-01-2012:11:02	Purchase	Pete
1	35654424	31-01-2012:15:03	Production	Sue
1	35654425	02-02-2012:09:00	Sale	Mike
2	35654483	30-01-2012:11:02	Purchase	Pete
2	35654484	05-02-2012:17:03	Production	John
2	35654485	06-02-2012:09:35	Sale	Mike

XES Standard Proposal

Slide 10 of 22

The data scientist can do al of this, provided that the event data gets to him in a way that he can handle it. To achieve this, we need to provide [TAB]syntax for the event data, we have to assign [TAB]proper semantics to the data, and we have to cater for [TAB] possible extensions in the future, that is, we have to be able to add new data to the event log or event stream with ease.



XES and Process Mining

id	id	time	act	name
1	35[illegible]54423	30-01-2012:11:02	Purchase	Pete
1	356[illegible]4424	31-01-2012:15:03	Production	Sue
1	356[illegible]4425	02-02-2012:09:00	Sale	Mike
2	35[illegible]54483	30-01-2012:11:02	Purchase	Pete
2	3[illegible]	05-02-2012:17:03	Production	John
2	[illegible]	[illegible]-02-2012:09:35	Sale	Mike

• • •

Syntax Semantics Extensible

XES Standard Proposal

Slide 11 of 22

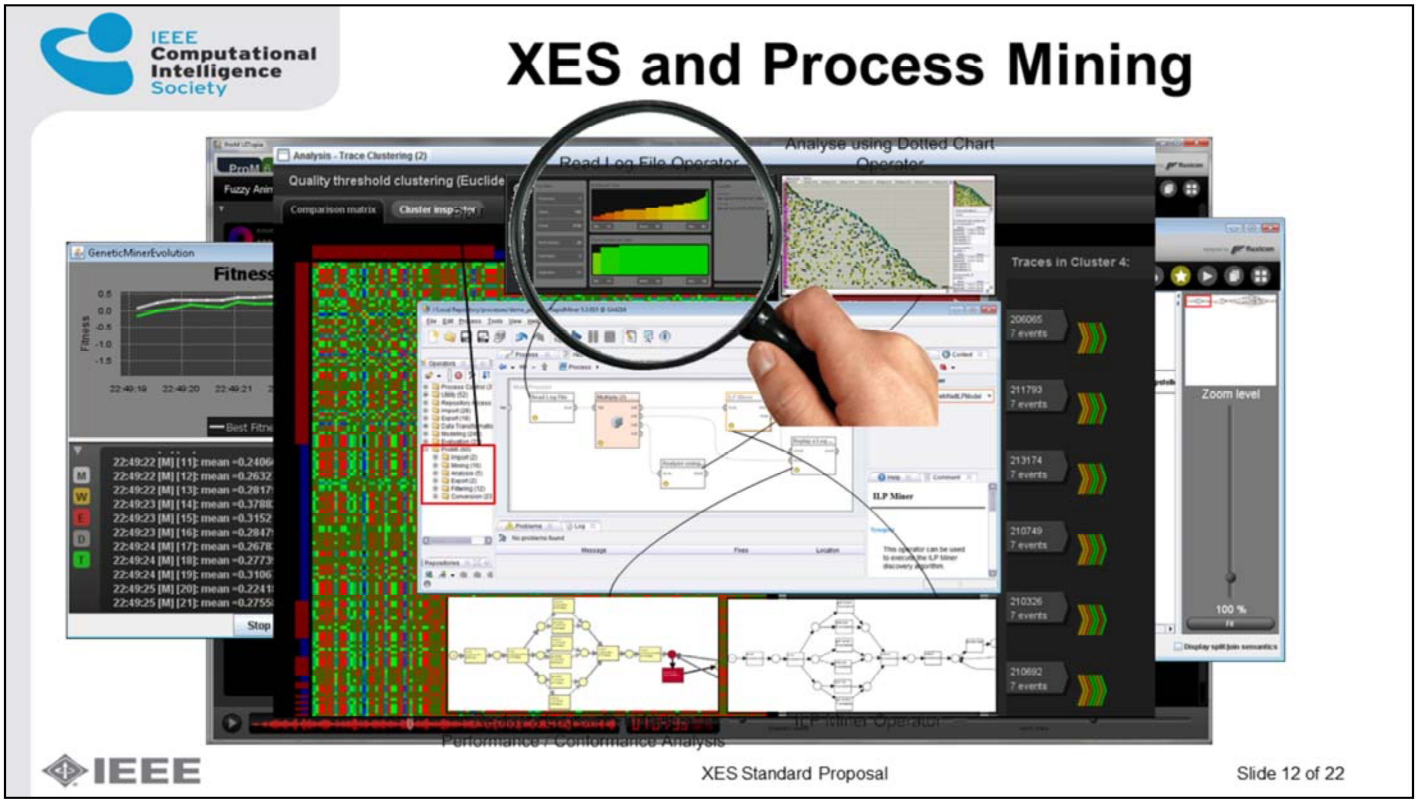
To summarize, all three types of process mining (discovery, conformance, and extension) require event data as input.

As a result, any process mining technique should be able to read an event log or subscribe to an event stream.

For this, it is vital that the event log and the event stream are standardized in a way that leaves no room for misinterpretation.

It should be clear [TAB]to which case an event corresponds, to [TAB]which activity it refers, at [TAB]which time it was executed, by [TAB]whom it was performed, etc.[TAB]

Furthermore, as we cannot predict which kind of data will become important in next years, it should be possible to include new data into the log or stream with ease.[TAB]



IEEE Computational Intelligence Society

XES and Process Mining

Read Log File Operator
Analyse using Dotted Chart Operator
ILP Miner Operator
Performance / Conformance Analysis
GeneticMinerEvolution
Fitness
Analysis - Trace Clustering (2)
Quality threshold clustering (Euclide
Traces in Cluster 4:
Zoom level
100 %

IEEE

XES Standard Proposal

Slide 12 of 22

From the previous, it is clear that the data scientist will be using a lot of [TAB]computational intelligence ‘under the hood’ when doing [TAB]process mining.

For example, process models can be discovered using [TAB]fuzzy algorithms, event logs can be [TAB]replayed on such fuzzy models, [TAB]genetic algorithms can be used, and the traces in the event log can be [TAB] clustered prior to discovering a model.

On top of this, many of these cutting edge algorithms are now available in tools like [TAB]RapidMiner, which allows the RapidMiner use to [TAB]import event logs into RapidMiner, [TAB]analyse the log using a dotted chart, [TAB]discover a process model from it using an ILP-based miner, and [TAB]replay it on a process model to check performance and conformance.[TAB]

As such, computational intelligence is a natural habitat for the process mining field. As a result, the XES standards proposal also finds its place in computational intelligence.

IEEE Computational Intelligence Society

XES and Process Mining

IEEE

XES Standard Proposal

Slide 13 of 22

As we need a standard like [TAB]XES in the process mining field, we need a standard like XES in the area of computational intelligence.

The XES standard will be an XML-based standard for event logs and event streams.

Its purpose is to provide a generally-acknowledged format for the interchange of event data between tools and application domains.

A possible purpose is for process mining, that is, the analysis of operational processes based on their event data.

However, XES will not be limited to process mining, and will [TAB] be designed to also be suitable for general data mining, text mining, and statistical analysis.

IEEE Computational Intelligence Society

XES Goals

XES

Extensible Event Stream

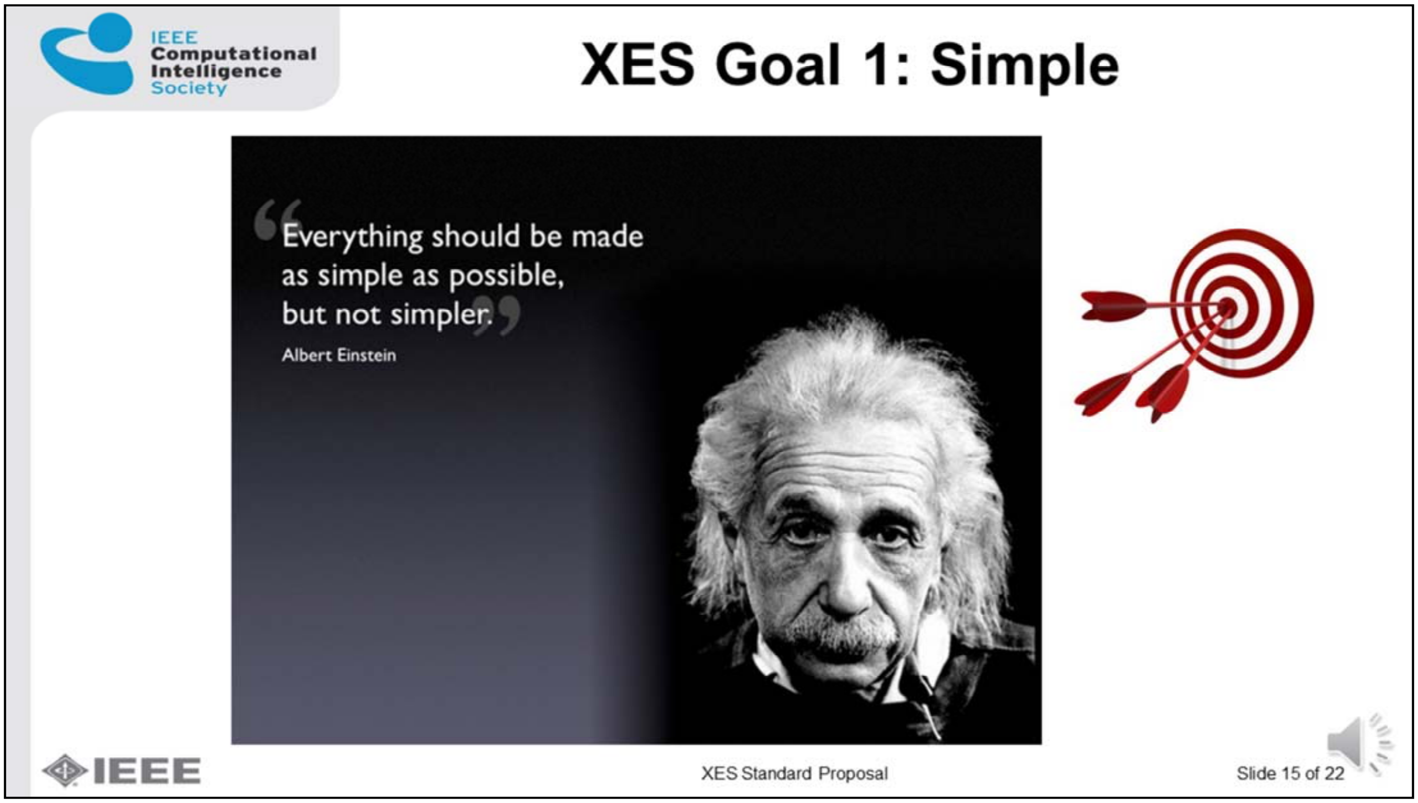
Simple
Syntax
Semantics
Extensible
Flexible
Expressive
THE DERIVATION
The United States of Tense and Aspect
THE MAP IS NOT THE TERRITORY

IEEE

XES Standard Proposal

Slide 14 of 22

When designing the XES standard, the following goals will be used as guiding principles: It should be [TAB]simple, [TAB]flexible, [TAB]extensible, and [TAB]expressive.



XES Goal 1: Simple

"Everything should be made as simple as possible, but not simpler."

Albert Einstein

XES Standard Proposal

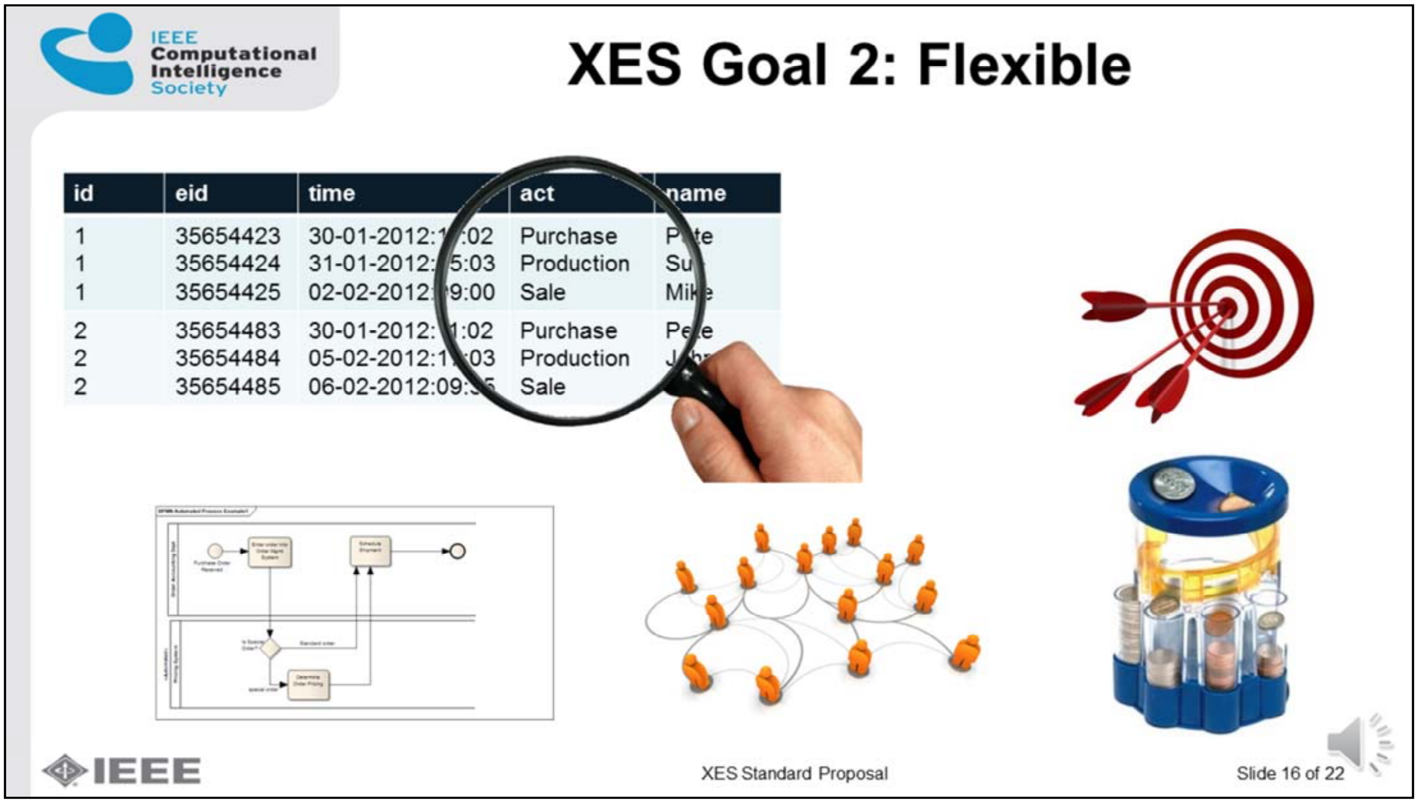
Slide 15 of 22

The first goal is simplicity.

XES will use the [TAB]simplest possible way to represent information.

XES logs and streams should be easy to parse and to generate, and they should be equally well human-readable.

In designing this standard, care will be taken to take a pragmatic route wherever that benefits an ease of implementation.



IEEE Computational Intelligence Society

XES Goal 2: Flexible

id	eid	time	act	name
1	35654423	30-01-2012:1[illegible]:02	Purchase	Pete
1	35654424	31-01-2012:[illegible]5:03	Production	Su[illegible]
1	35654425	02-02-2012:[illegible]9:00	Sale	Mike
2	35654483	30-01-2012:[illegible]1:02	Purchase	Pete
2	35654484	05-02-2012:1[illegible]:03	Production	J[illegible]
2	35654485	06-02-2012:09:[illegible]5	Sale	

IEEE

XES Standard Proposal

Slide 16 of 22

The second goal is flexibility.

In XES, there will be no predefined attributes with any well-understood meaning.

In particular, there will be no fixed identity for an event.

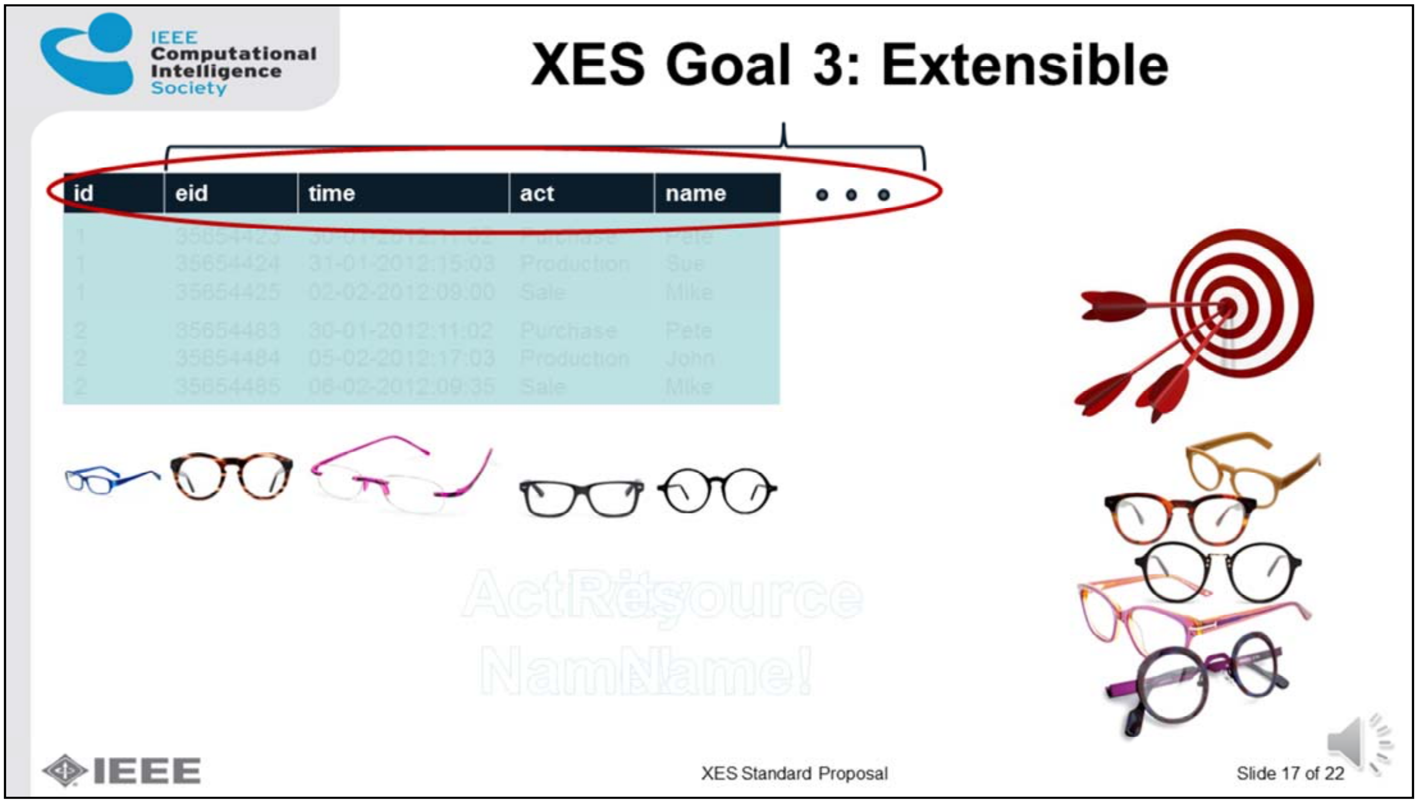
Instead, *event classifiers* will be a mandatory feature of the XES standard.

Like [TAB] coin sorters classify coins by their size, event classifiers will classify events by their attributes.

In contrast to the coin sorters, however, classifiers will be configurable through the [TAB]set of attributes it uses for the classification.

For example, if the activity is stored in [TAB]the act attribute, then a classifier configured by that attribute will provide you the activity for an event, and we will be discovering a [TAB]process model. However, by changing the classifier to use, for example [TAB] the name attribute, we would immediately be discovering a [TAB]social network.[TAB]

Example standard classifiers include an [TAB] activity classifier and, especially in case of an event stream, a [TAB] trace classifier.



XES Goal 3: Extensible

id	eid	time	act	name	• • •
1	35654423	30-01-2012:11:02	Purchase	Pete	
1	35654424	31-01-2012:15:03	Production	Sue	
1	35654425	02-02-2012:09:00	Sale	Mike	
2	35654483	30-01-2012:11:02	Purchase	Pete	
2	35654484	05-02-2012:17:03	Production	John	
2	35654485	06-02-2012:09:35	Sale	Mike	

ActResource
NamName!

XES Standard Proposal

Slide 17 of 22

The third goal is extensibility.

The XES standard will not define a [TAB]specific set of attributes.

In the example, we have used labels like act and name, and until now we have assumed that act refers to the activity while name refers to the resource.

However, it would also be possible that act refers to the actor, or resource, while name refers to the activity.

Clearly, the label itself is not sufficient to decide what exactly the semantics is of the attribute.

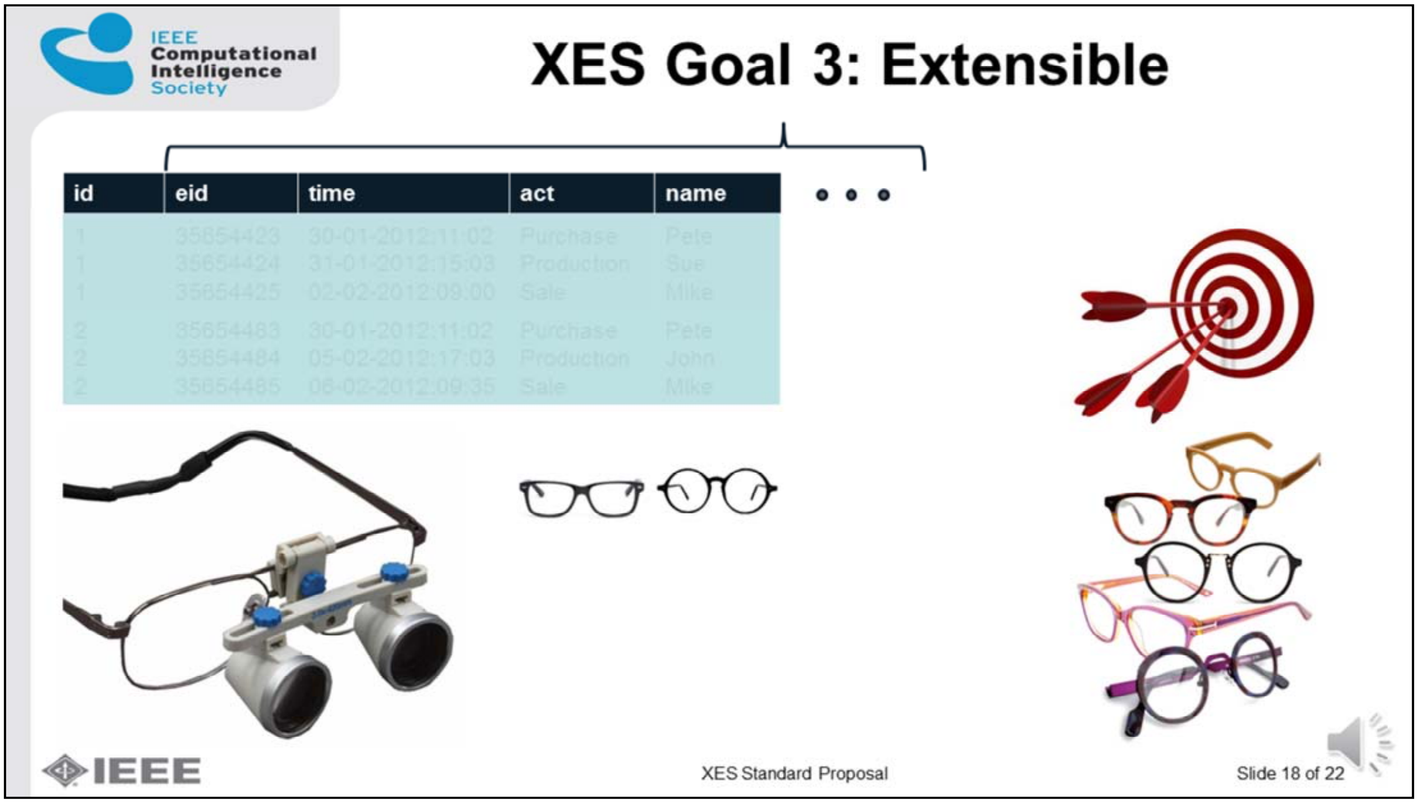
As such, the semantics of these attributes must necessarily be ambiguous, [TAB]hampering the interpretation of that data.

Instead of having a fixed set of attributes, we introduce the notion of an extension, which can be compared to a [TAB]pair of glasses.

An extension provides focus for a restricted set of attributes.

For example, we could have a [TAB]pair of glasses that fixes the [TAB]act attribute to indeed correspond to the activity, we could have a second [TAB] pair of glasses that does the same for the [TAB] resource attribute, and we could have [TAB] additional pairs of glasses for [TAB] other attributes.

As a result, in XES, an attribute will only have well-defined semantics in combination with a correct pair of glasses.



IEEE Computational Intelligence Society

XES Goal 3: Extensible

id	eid	time	act	name
1	35654423	30-01-2012:11:02	Purchase	Pete
1	35654424	31-01-2012:15:03	Production	Sue
1	35654425	02-02-2012:09:00	Sale	Mike
2	35654483	30-01-2012:11:02	Purchase	Pete
2	35654484	05-02-2012:17:03	Production	John
2	35654485	06-02-2012:09:35	Sale	Mike

IEEE

XES Standard Proposal

Slide 18 of 22

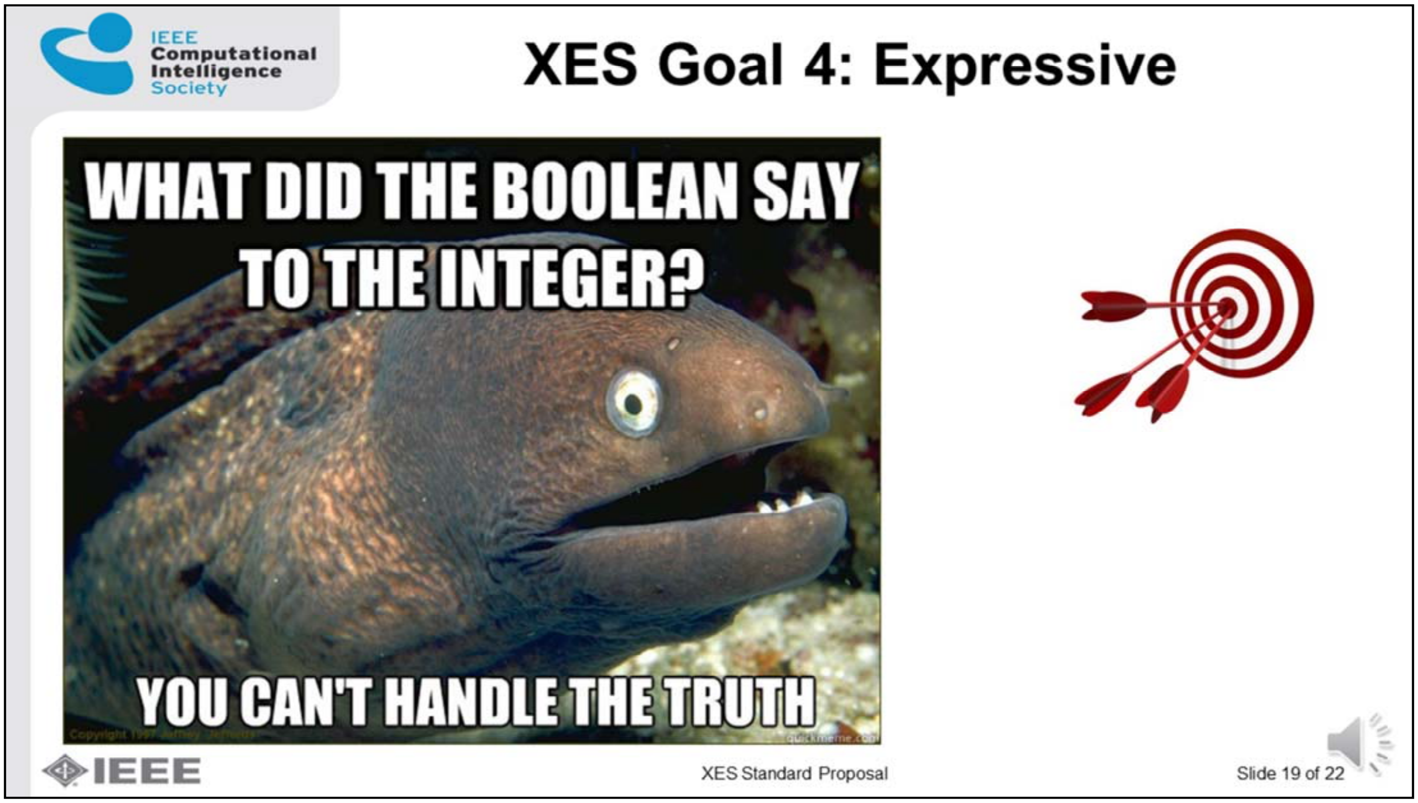
Extensions have many possible uses.

One important use is to introduce a set of commonly understood attributes which are vital for a specific perspective or dimension of event data analysis, and which may even not have been foreseen at the time of designing the XES standard.

A default set of standard extensions will be introduced in XES, which includes a concept extension, which defines [TAB]generally understood names, and an organizational extension, which identifies the [TAB] resource having caused the event, and his position in the organizational structure.

Other uses include the definition of generally-understood attributes for a specific application domain, like [TAB] medical attributes for hospital processes, or for supporting special features or requirements

of a specific analysis application.



IEEE Computational Intelligence Society

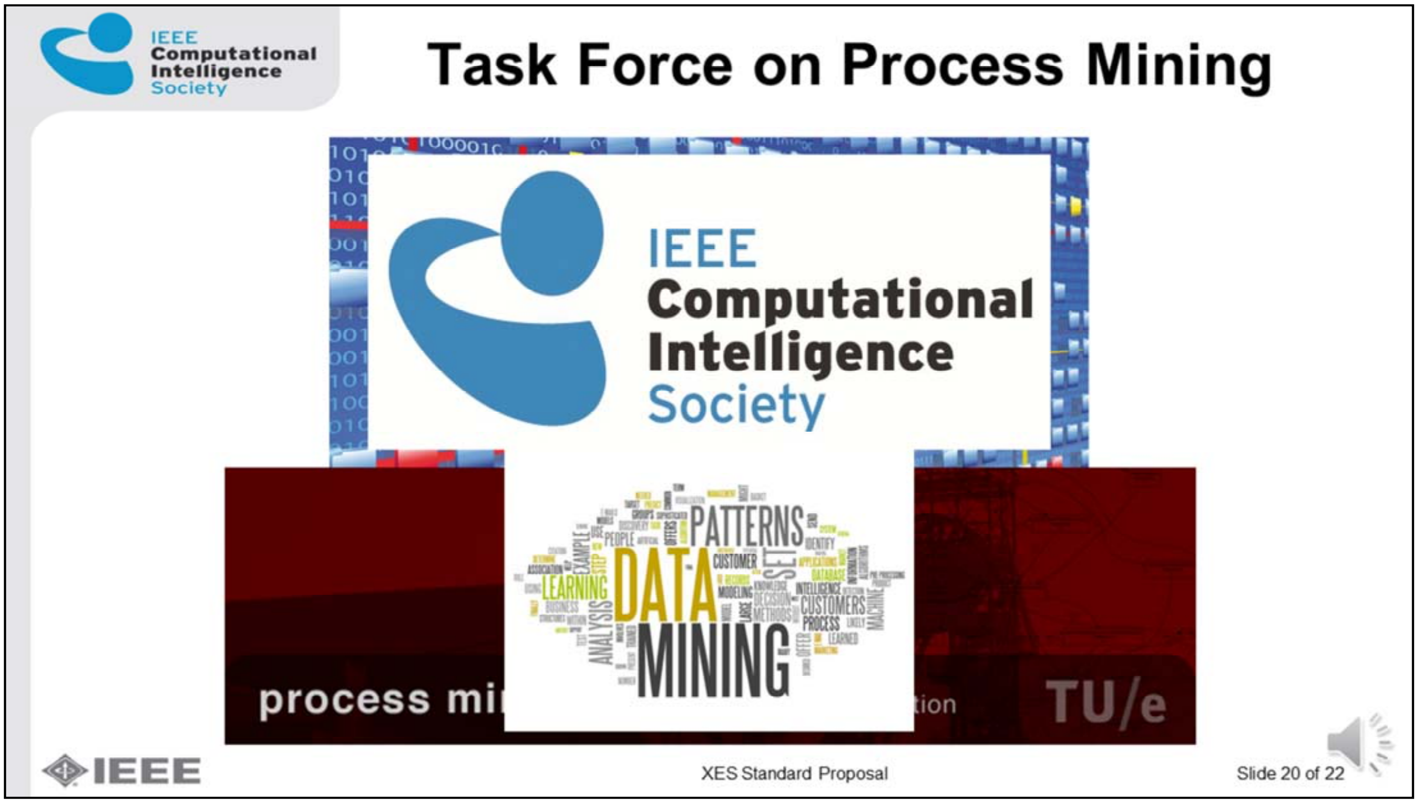
XES Goal 4: Expressive

WHAT DID THE BOOLEAN SAY
TO THE INTEGER?
YOU CAN'T HANDLE THE TRUTH

IEEE XES Standard Proposal Slide 19 of 22

The fourth, and last, goal is expressiveness.

All information in an event log will be stored in *attributes* of specific types. [TAB]String attributes will hold literal information which is generally untyped and of arbitrary length. [TAB]Date attributes will hold information about a specific point in time. [TAB]Int attributes will hold a discrete integer number, whereas float attributes will hold a continuous floating-point number. [TAB]Boolean attributes will hold a boolean value which can either be true or false.



IEEE Computational Intelligence Society

Task Force on Process Mining

IEEE
Computational
Intelligence
Society
PATTERNS
DATA
MINING
LEARNING
ANALYSIS
CUSTOMER
CUSTOMERS
PROCESS
MACHINE
process mi
tion
TU/e

IEEE

XES Standard Proposal

Slide 20 of 22

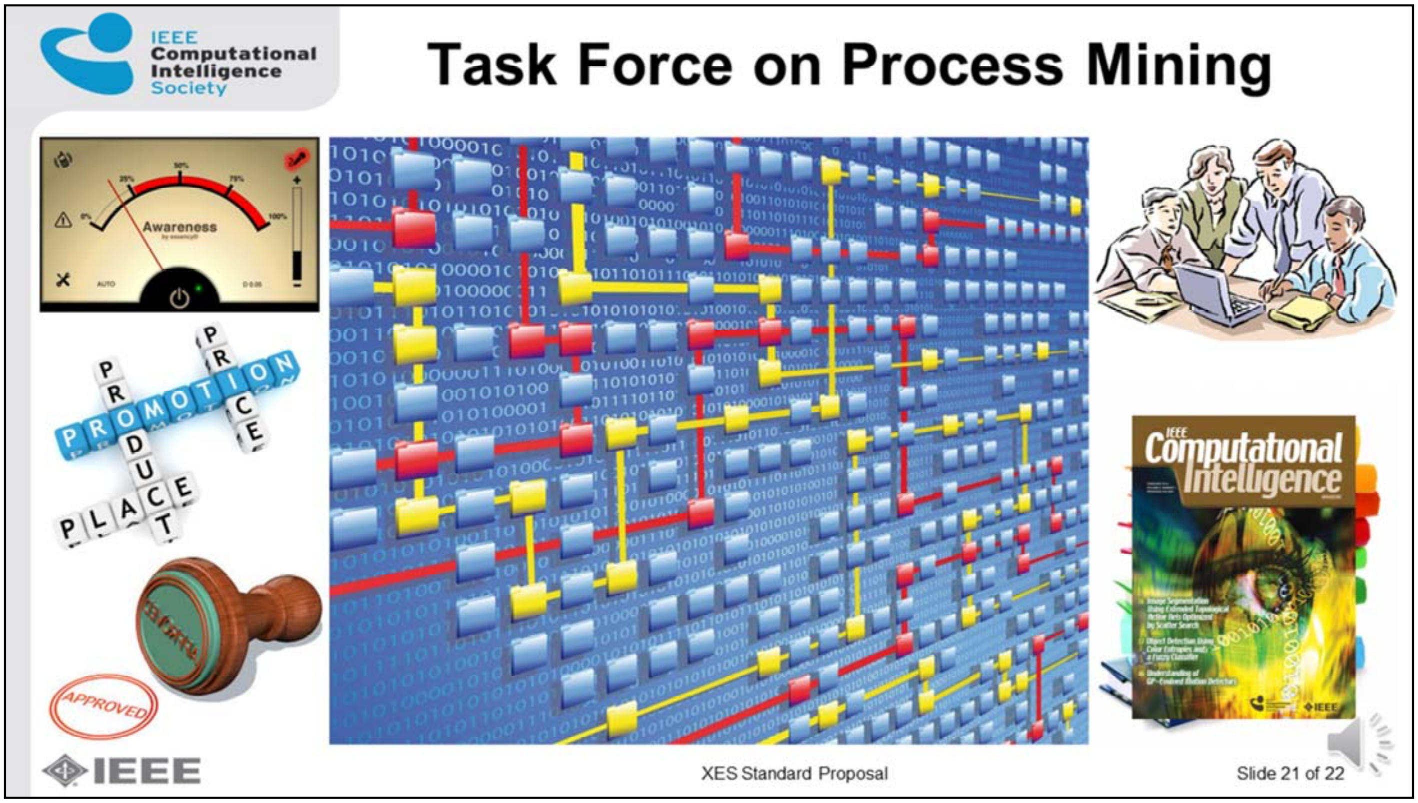
We now have motivated the need for a XES standard and its basic requirements.

In the remaining part, we would like to inform you about the IEEE Task Force on Process Mining, as this explains why we want the IEEE Computational Intelligence Society to sponsor the XES standard proposal.

As more and more people, both in industry and academia, consider [TAB]process mining as one of the most important innovations in the field of business process management, the IEEE has established a Task Force on Process Mining.

This Task Force is established in the context of the [TAB]Data Mining Technical Committee of the [TAB]IEEE Computational Intelligence Society.

The goal of this Task Force is to promote the research, development, education and understanding of process mining.



IEEE Computational Intelligence Society

Task Force on Process Mining

IEEE

XES Standard Proposal

Slide 21 of 22

More concretely, the goal is to:

- [TAB]make end-users, developers, consultants, and researchers aware of the state-of-the-art in process mining,

- [TAB]promote the use of process mining techniques and tools and stimulating new applications,

- [TAB]play a role in standardization efforts for logging event data,

- [TAB] organize conferences, tutorials, special sessions, workshops, and panels, possibly with technical co-sponsorship of the IEEE Computational Intelligence Society, and

- [TAB]publish in the form of articles, books, and special issues of journals, like for example the [TAB]IEEE Computational Intelligence Magazine.

Task Force on Process Mining

XES Standard Proposal

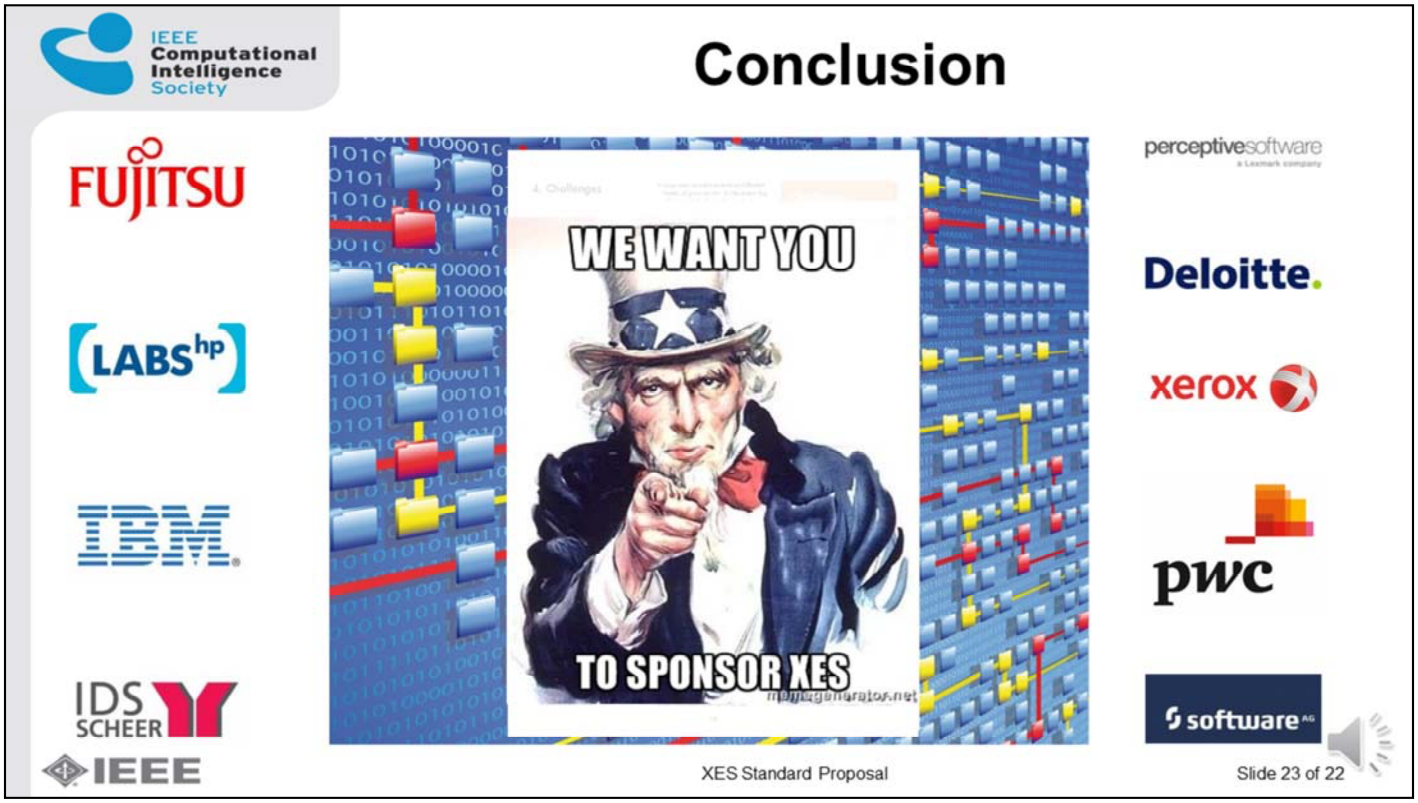
Slide 22 of 22

In the context of this task force, a group of more than seventy-five people involving more than fifty organizations created the Process Mining Manifesto, which has been translated into twelve [TAB]other [TAB]languages.

By defining a set of [TAB]guiding principles and listing [TAB]important challenges, this manifesto hopes to serve as a guide for [TAB]software developers, [TAB]scientists, [TAB]consultants, [TAB]business managers, and [TAB]end-users.

As one of its stated goals is to play a role in the standardization efforts for logging data, the task force proposes the [TAB]XES standard.

As a fist step in this standardization process, we ask the IEEE Computational Intelligence Society to sponsor this standard proposal.



IEEE Computational Intelligence Society

Conclusion

XES Standard Proposal

Slide 23 of 22

The XES standard proposal is a product of this IEEE Task Force on Process Mining, which is an active task force in the IEEE CIS domain.

This task force has adopted the XES standard proposal as the default interchange format for event data, as it sees that there is a very general and clear need for such a standard.

Given that the task force contains many industry partners, like [TAB]Fujitsu, [TAB]HP, [TAB]IBM, [TAB]IDS Scheer, [TAB]Perceptive Software, [TAB]Deloitte, [TAB]Xerox, [TAB]Pricewaterhouse Coopers, and [TAB]Software AG, there is also a large interest from industry for the standard proposal.

Finally, the standard proposal has full support from the [TAB]DMTC, or Data Mining Technical Committee.

To help us advance, we ask [TAB]you to sponsor the XES standard proposal.