Can Data Warehouses Evolve to Become Information Warehouses?

By Charles Hoffman, CPA
February 27, 2017

A global standard replacement for data warehouses that behaves more like a knowledge based system and which business users can interact with using pivot tables and spreadsheets

Data, Information, Knowledge, Wisdom

- **Data**: The basic compound for Intelligence is data -- measures and representations of the world around us, presented as external signals and picked up by various sensory instruments and organs. *Simplified: raw facts and numbers*.
- **Information**: Information is produced by assigning relevant meaning to data. *Simplified: information is data in context*.
- **Knowledge**: Knowledge is the subjective interpretation of information and approach to act upon the information in the mind of the perceiver. *Simplified: knowledge is the interpretation of information*.
- **Wisdom (or Intelligence)**: Intelligence or wisdom embodies awareness, insight, moral judgments, and principles to construct new knowledge and improve upon existing understanding. *Simplified: wisdom is the creation of new knowledge*.

DIKW Pyramid, Wikipedia, retrieved February 24, 2016; https://en.wikipedia.org/wiki/DIKW_Pyramid

Gene Bellinger, Durval Castro, Anthony Mills; *Data, Information, Knowledge, and Wisdom*; Retrieved February 24, 2016, http://www.systems-thinking.org/dikw/dikw.htm

Differentiating Data and Information

- There is a subtle but important to understand difference between data and information.
- Data are the facts or details from which information is derived. Individual pieces of data are rarely useful alone. For data to become information, data needs to be put into some context.
- The same fact can be used in many different contexts.

Diffen, *Data vs Information*, http://www.diffen.com/difference/Data_vs_Information

Data vs. Information

	Data	Information
Meaning	Data is raw, unorganized facts that need to be processed. Data can be something simple and seemingly random and useless until it is organized.	When data is processed, organized, structured or presented in a given context so as to make it useful, it is called information.
Example	Each student's test score is one piece of data.	The average score of a class or of the entire school is information that can be derived from the given data.
Etymology	"Data" comes from a singular Latin word, datum, which originally meant "something given." Its early usage dates back to the 1600s. Over time "data" has become the plural of datum.	"Information" is an older word that dates back to the 1300s and has Old French and Middle English origins. It has always referred to "the act of informing, " usually in regard to education, instruction, or other knowledge communication.

Diffen, *Data vs Information*, http://www.diffen.com/difference/Data_vs_Information

Data vs. Information - Differences in Meaning

Data are simply facts or figures — bits of information, but not information itself. When data are processed, interpreted, organized, structured or presented so as to make them meaningful or useful, they are called information. Information provides context for data.

For example, a list of dates — data — is meaningless without the information that makes the dates relevant (dates of holiday).

"Data" and "information" are intricately tied together, whether one is recognizing them as two separate words or using them interchangeably, as is common today. Whether they are used interchangeably depends somewhat on the usage of "data" — its context and grammar.

Examples of Data and Information

- The history of temperature readings all over the world for the past 100 years is data. If this data is organized and analyzed to find that global temperature is rising, then that is information.
- The number of visitors to a website by country is an example of data. Finding out that traffic from the U.S. is increasing while that from Australia is decreasing is meaningful information.
- Often data is required to back up a claim or conclusion (information) derived or deduced from it. For example, before a drug is approved by the FDA, the manufacturer must conduct clinical trials and present a lot of data to demonstrate that the drug is safe.

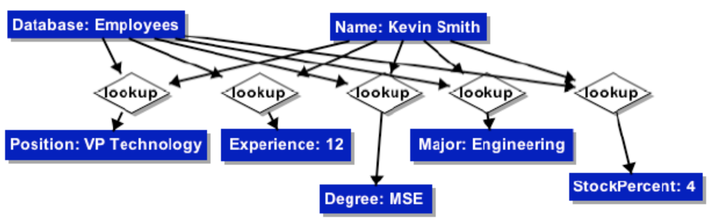
Data warehouse – Shortcomings of OLAP

- There is no global standard for OLAP
- Cube rigidity, hard for business professionals to adjust and maintain
- Limited computation support, mainly roll ups
- Limited business rule support and inability to exchange business rules between implementations
- Inability to transfer cubes between systems, each system is a "silo" which cannot communicate with other silos
- Inability to articulate metadata which can be shared between OLAP systems
- Focus on numeric-type information and inconsistent support for text data types
- OLAP systems tend to be internally focused within an organization and do not work well externally to organzation, for example across a supply chain
- OLAP tends to be read only
- Metadata is “lean” and therefore information is not “rich”

Richness of Information
Data Warehouse Contrast to Information Warehouse

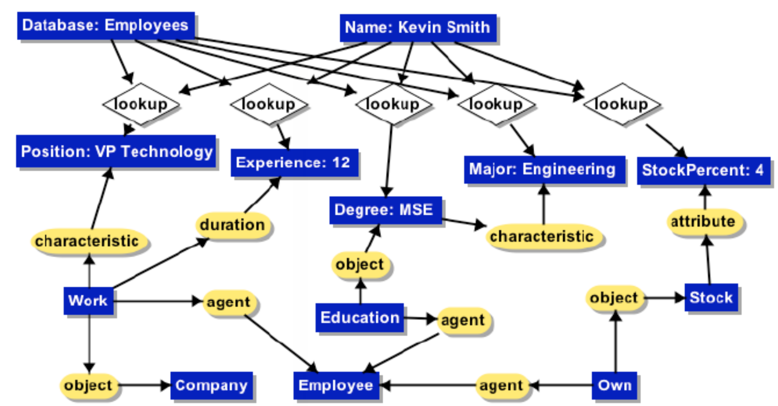
Database Values as Data

Name	Position	Yrs Experience	Degree	Major	Percent Stock
Karen Jones	VP Marketing	18	MBA	Marketing	3
Kevin Smith	VP Technology	12	MSE	Engineering	4
Keith Williams	VP Finance	15	BS	Accounting	3



Database Values as Information (data in context)

Name	Position	Yrs Experience	Degree	Major	Percent Stock
Karen Jones	VP Marketing	18	MBA	Marketing	3
Kevin Smith	VP Technology	12	MSE	Engineering	4
Keith Williams	VP Finance	15	BS	Accounting	3



Multidimensional Model

- Information is inherently multidimensional
- There is not global standard multidimensional model
- XBRL, a global standard, has a standard multidimensional model
- Model
  - Fact
  - Characteristic
  - Fact Table
  - Relation
    - Is a (or Type of, Class of)
    - Whole part
    - Computational business rule
  - Grain

Introduction to the Multidimensional Model for Professional Accountants, http://xbrl.squarespace.com/journal/2016/3/18/introduction-to-the-multidimensional-model-for-professional.html

Data warehouse – Other Shortcomings

- Lack of business rules to check quality of data in system; result is that errors enter the system
- If business rules do exist, they tend to be hard-coded into applications; result that rule maintenance is hard and adding new rules is hard resulting in inflexible systems
- Where rules can be expressed, the rules are not very rich; result is that information is missing from the system

Information Warehouse Description

- Information warehouses are general purpose, task-specific, solution-specific, knowledge based systems for *business reporting*
- Basic architecture is star schema (fact table) or snowflake schema (fact table) just like a *data warehouse*
- Information warehouse fact table is semantically equivalent to an XBRL fact table
- Knowledgebase is XBRL taxonomy schema, XBRL linkbases, and XBRL Formulas
- Fact database can be seen as a set of XBRL instances, for example the XBRL-based public company financial filings to the US SEC EDGAR system is a big fact database
- **Information warehouse is a general-purpose expert system built using standard, interchangeable components**

EXPERT SYSTEMS BUILDING TOOLS: DEFINITIONS: http://www.wtec.org/loyola/kb/c3_s2.htm

Capabilities of Information Warehouse

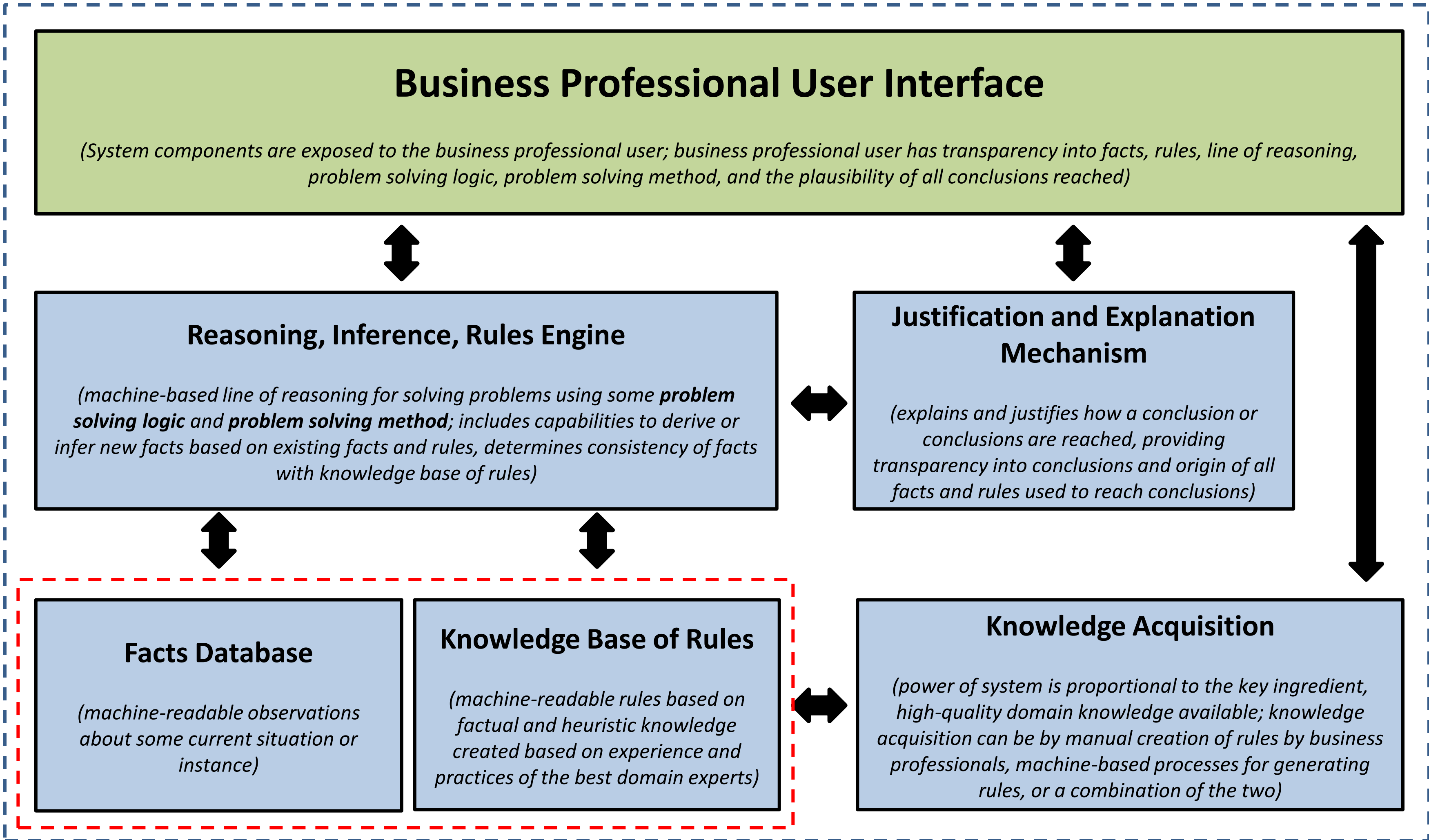
- Global standard multidimensional model based on the XBRL Dimensions model
- Global standard metadata based on XBRL taxonomies
- Global standard fact database based on XBRL instances
- Allows, but does not require, an OLAP engine
- Flexible, rather than rigid, hypercubes (cubes)
- Global standard problem solving logic, based on ISO/IEC Common Logic
- Global standard business rules/reasoning/inference engine (i.e. via Common Logic, RuleLog)
- Information warehouses interoperate with other information warehouses and legacy data warehouses
- Metadata is interchangeable between information warehouses using XBRL or OWL
- Supports both numeric information and nonnumeric information including text and prose
- Inherently web-enabled, usable internal within one organization, between organizations, or across an entire supply chain
- Information warehouse is read/write as opposed to read only
- Standard business user interface for using information, dynamic reports, similar to pivot tables and spreadsheets

Knowledge Based System

- Knowledge allows you to take effective action.
- Knowledge based systems provide capability and know how.
- Knowledge management is about creating a managed system that routinely and systematically institutionalizes knowledge and ensures that people have the knowledge they need to make correct decisions.
- Result of poor knowledge management:
  - Mistakes repeated
  - Successful practices not replicated
  - Slow rate of learning
  - Knowledge lost when staff retires

Nick Milton of Knoco.com discusses data information and knowledge , https://www.youtube.com/watch?v=sdzUfHwNCVQ

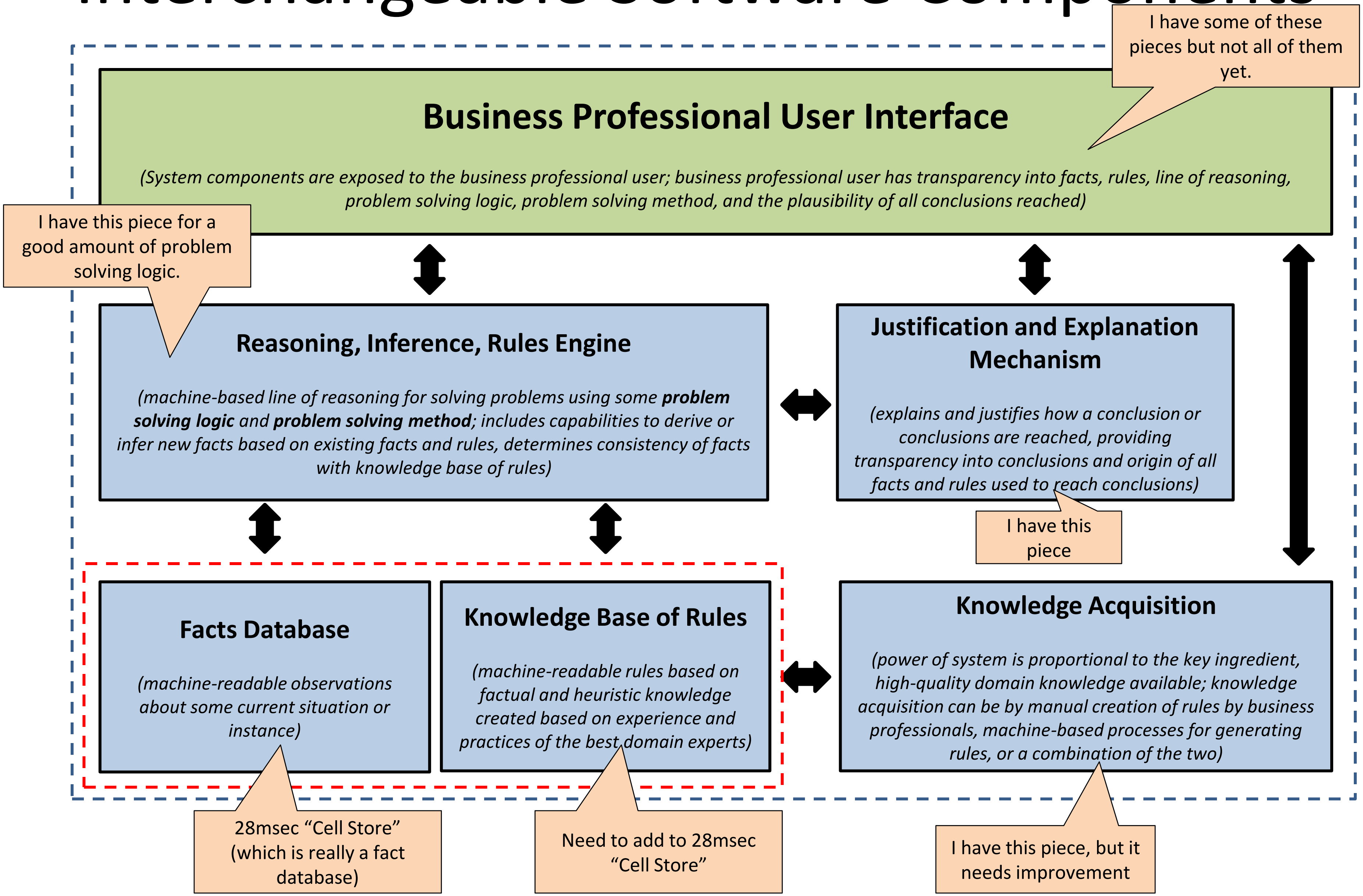
Components of a Knowledge Based System



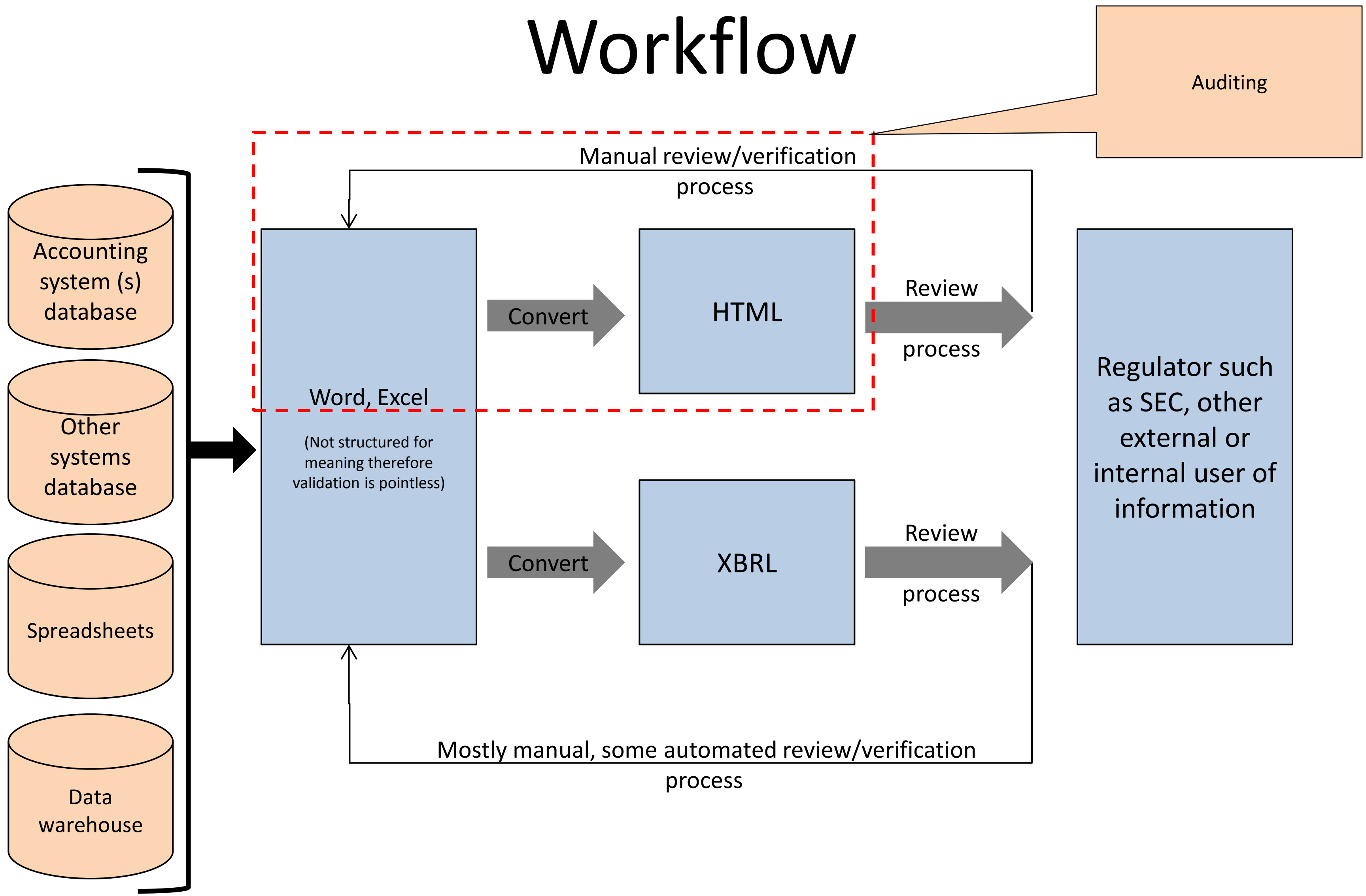
EXPERT SYSTEMS BUILDING TOOLS: DEFINITIONS: http://www.wtec.org/loyola/kb/c3_s2.htm

Comprehensive Introduction to Expert Systems, http://xbrlsite.azurewebsites.net/2017/IntelligentDigitalFinancialReporting/Part01_Chapter02.6_ComprehensiveIntroductionToExpertSystems.pdf#Page=6

Interchangeable Software Components



Current Old School Reporting Workflow



Using Deming's World Class Manufacturing Techniques on Information

- Old way to build a car
  - Build car
  - Find defects (adds costs)
  - Fix defects (adds costs)
- New way to build a car
  - Fix processes (adds costs)
  - Remove causes of defects (reduces costs)
  - Build defect free cars

Deming: A Theory of Systems for Educators and Managers, http://xbrl.squarespace.com/journal/2015/7/24/deming-a-theory-of-a-system-for-educators-and-managers.html

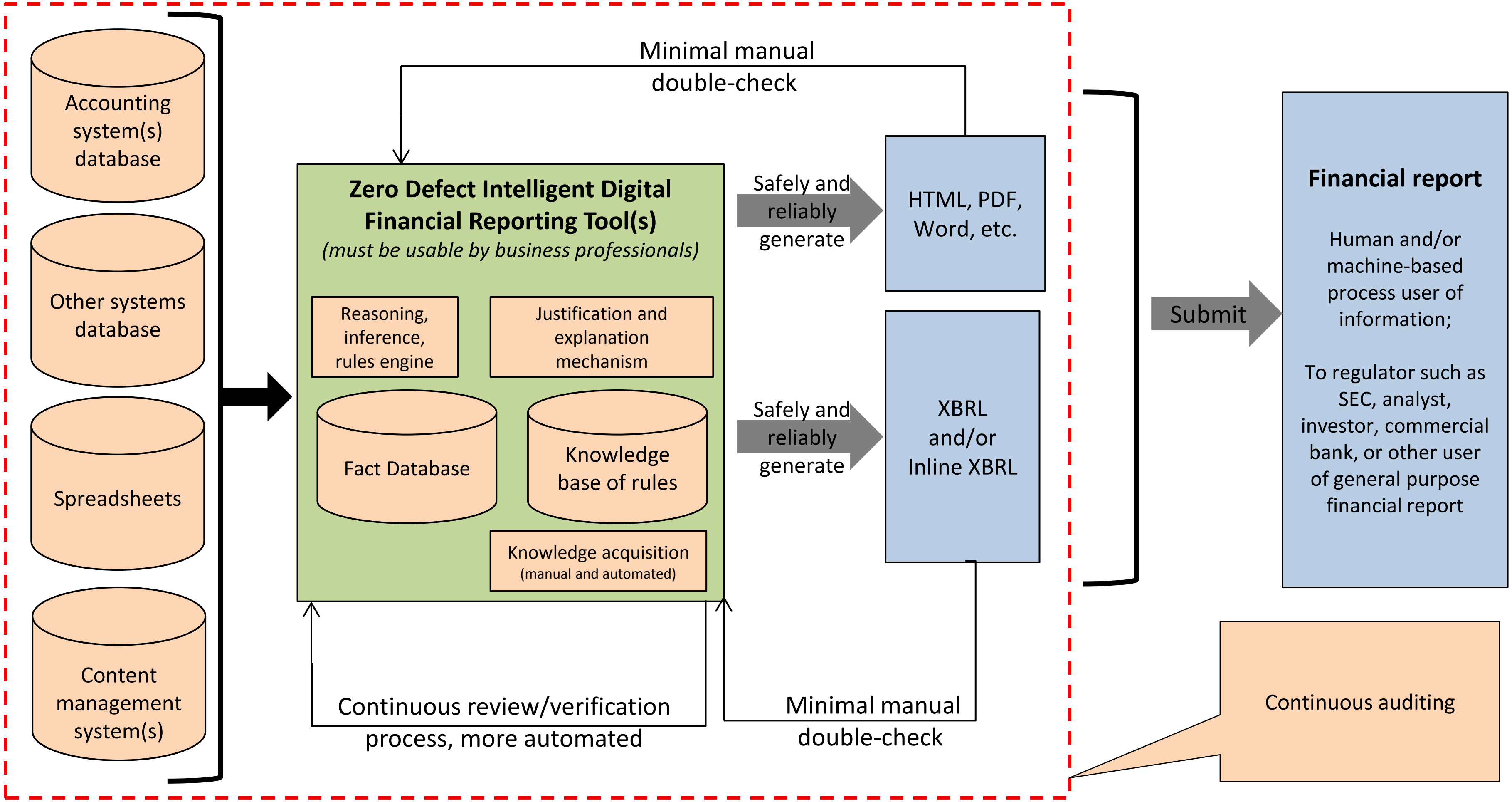
Applying World Class Manufacturing Techniques to Information

- Old way
  - Put data into database which contains errors
  - Make management mistakes
  - Find defects (i.e. errors in information)
  - Fix defects
- New way
  - Fix processes, use automation to detect defects
  - Remove causes of defects
  - Use defect free information

Deming: A Theory of Systems for Educators and Managers, http://xbrl.squarespace.com/journal/2015/7/24/deming-a-theory-of-a-system-for-educators-and-managers.html

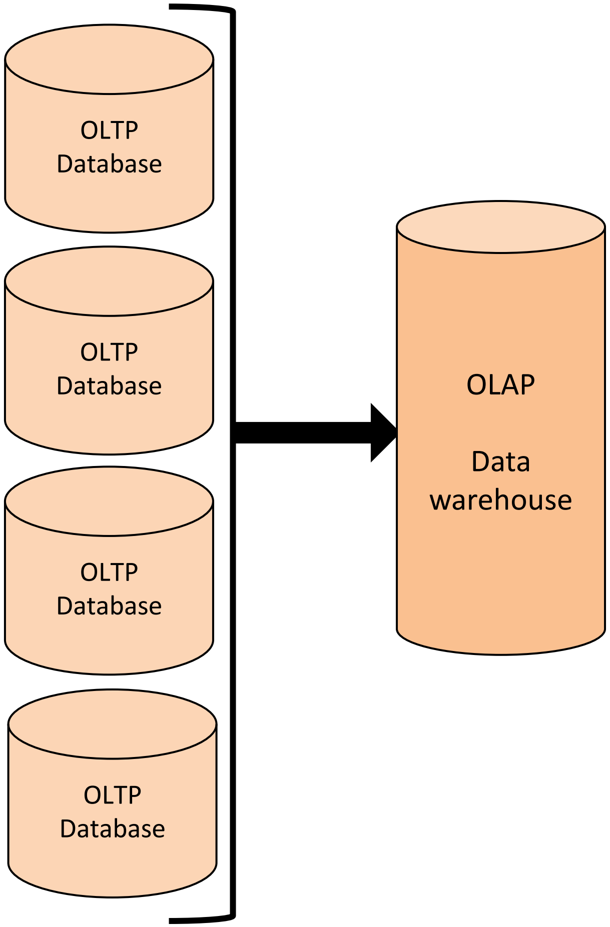
Modern Reporting Workflow

(Integrated, knowledge based system, one version of the truth)

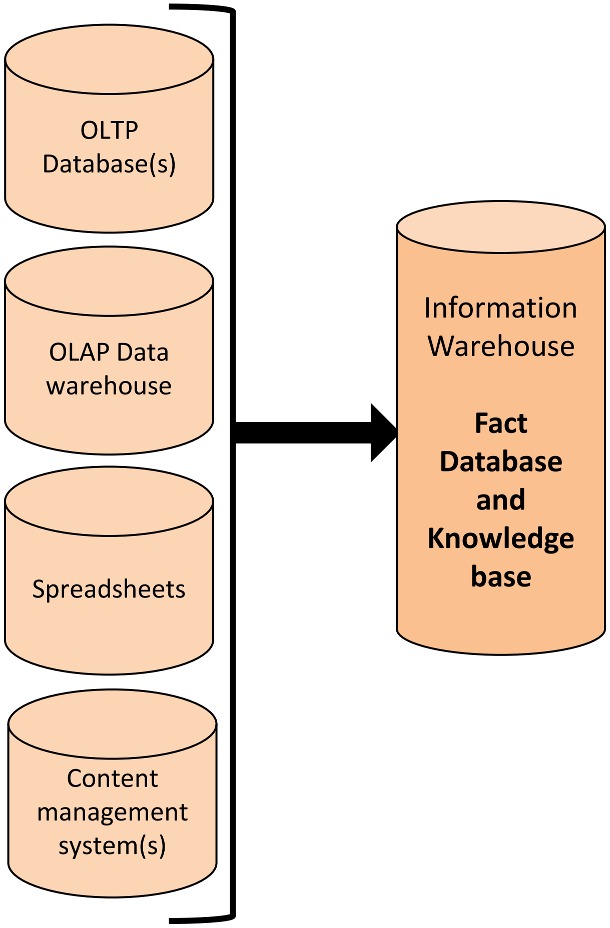


https://en.wikipedia.org/wiki/Continuous_auditing

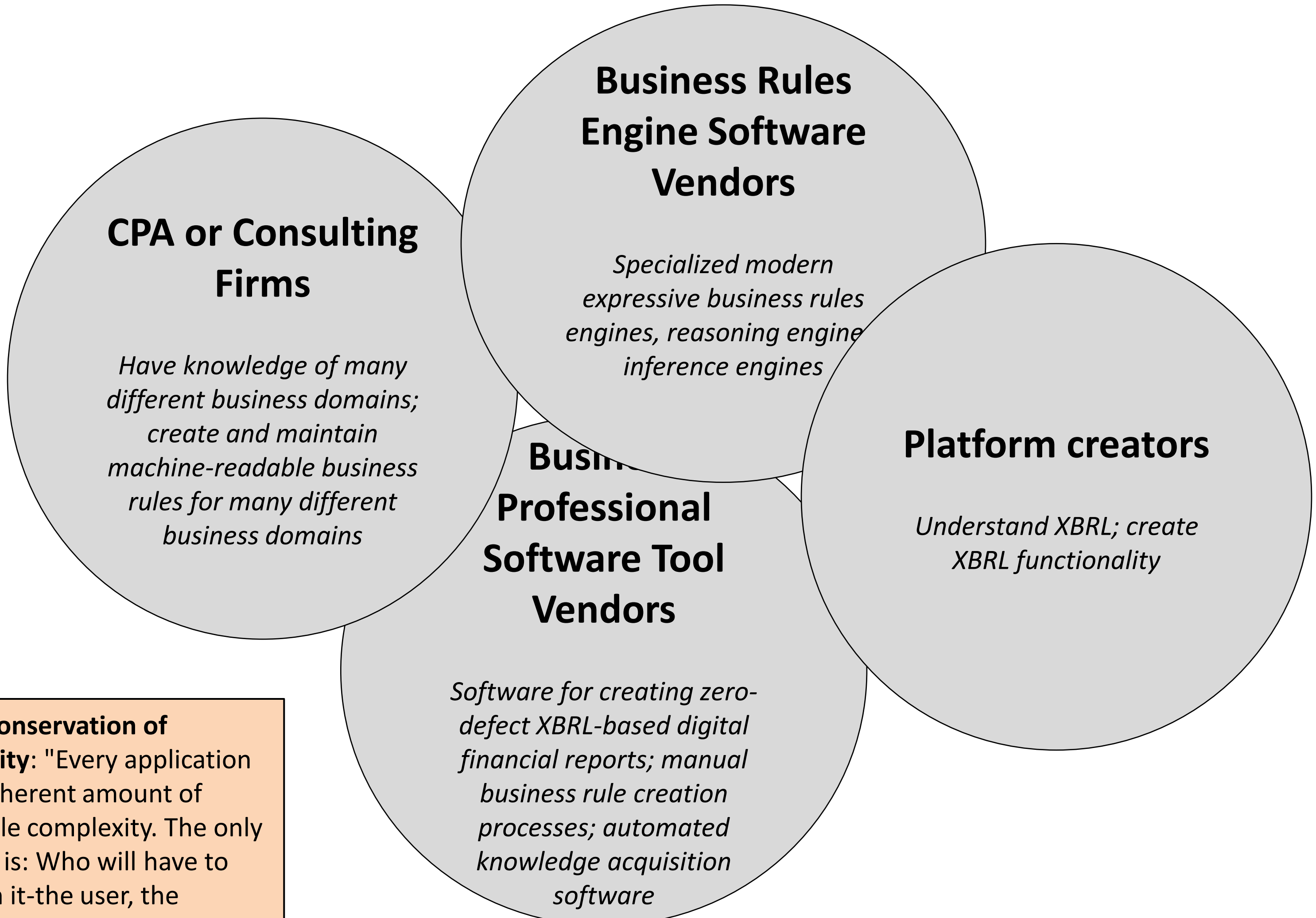
Data Warehouse



Information Warehouse



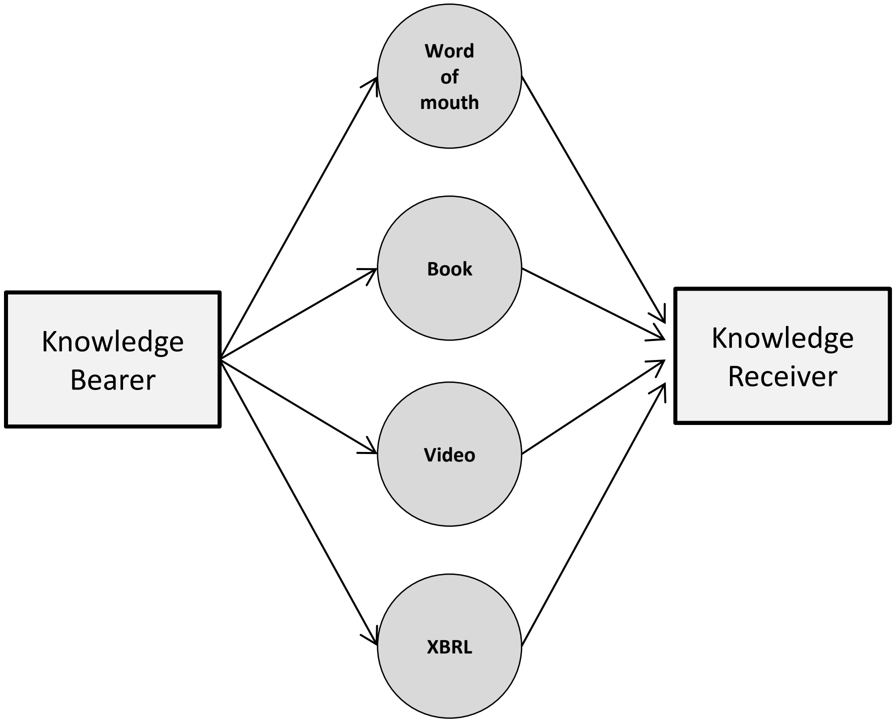
Coordination to Make it Work Effectively



Law of Conservation of Complexity: "Every application has an inherent amount of irreducible complexity. The only question is: Who will have to deal with it-the user, the application developer, or the platform developer?"

Law of Conservation of Complexity: http://xbrl.squarespace.com/journal/2015/5/24/understanding-the-law-of-conservation-of-complexity.html

Recognizing that XBRL is a Knowledge Media



Understanding that XBRL is a Knowledge Media: http://xbrl.squarespace.com/journal/2017/1/16/understanding-that-xbrl-is-a-knowledge-media.html

Effective Knowledge Media

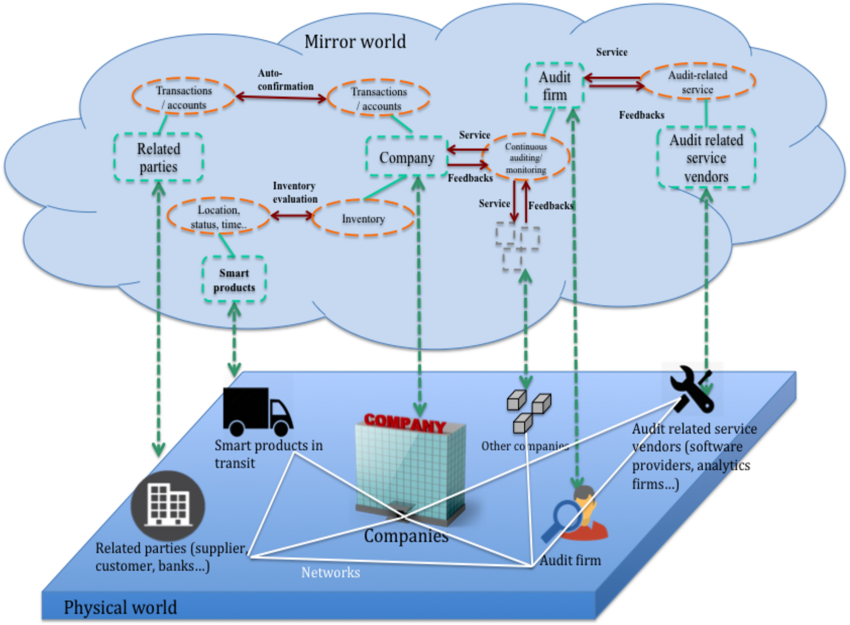
In order to make use of a knowledge media effectively, the following three conditions must be satisfied:

1. **Easy for knowledge bearer to represent information**: The effort and difficulty required for the knowledge bearer to successfully formulate the knowledge in the medium must be as low as possible.
2. **Clear, consistent meaning**: The meaning conveyed by the knowledge bearer to the knowledge receiver must be clear and easily followed by human beings and be consistent between different software applications. The result cannot be a "black box" or a guessing game and users of the information should not be able to derive different knowledge simply by using a different software application.
3. **High-quality information representation**: The form in which the knowledge is represented to the receiver must be as good as possible. The quality must be high whether the knowledge receiver is a human-being or an automated machine-based process. Sigma level 6 is a good benchmark, 99.99966% accuracy.

Big Picture: Enabling Industry 4.0

Industry 4.0 emphasizes six major principles in its design and implementation:

- interoperability,
- virtualization,
- decentralization,
- real-time capability,
- service orientation, and
- modularity.



Understanding XBRL's Role in the Fourth Industrial Revolution:: http://xbrl.squarespace.com/journal/2017/2/26/understanding-xbrls-role-in-the-fourth-industrial-revolution.html

High Quality Information is Key

The only way to achieve a meaningful exchange of information without disputes is with the prior existence of and agreement as to a standard set of technical syntax rules, business semantics rules, and workflow rules.

Meaningful exchange relates to exchange without disputes as to precise meaning, it means unambiguous interpretation, it means resolving conflicts and inconsistencies.

Consider this scenario: Two public companies, A and B, each have some knowledge about their financial position and financial condition. They must communicate their knowledge to an investor who is making investment decisions which will make use of the combined information so as to draw some conclusions. All three parties are using a common set of basic logical principles (facts known to be true, deductive reasoning, inductive reasoning, etc.) and common financial reporting standards (i.e. US GAAP, IFRS, etc.), so they should be able to communicate this information fully, so that any inferences which, say, the investor draws from public company A's input should also be derivable by public company A using basic logical principles and common financial reporting standards, and vice versa; and similarly for the investor and public company B.

Business Processes will be Monitored Against Pre-determined Rules

- ISO TR 9007:1987 ("Helsinki principles"):
  - Any meaningful exchange of utterances depends upon the prior existence of an agreed set of semantic and syntactic rules
  - The recipients of the utterances must use only these rules to interpret the received utterances, if it is to mean the same as that which was meant by the utterer
- "Business rules prevent anarchy"

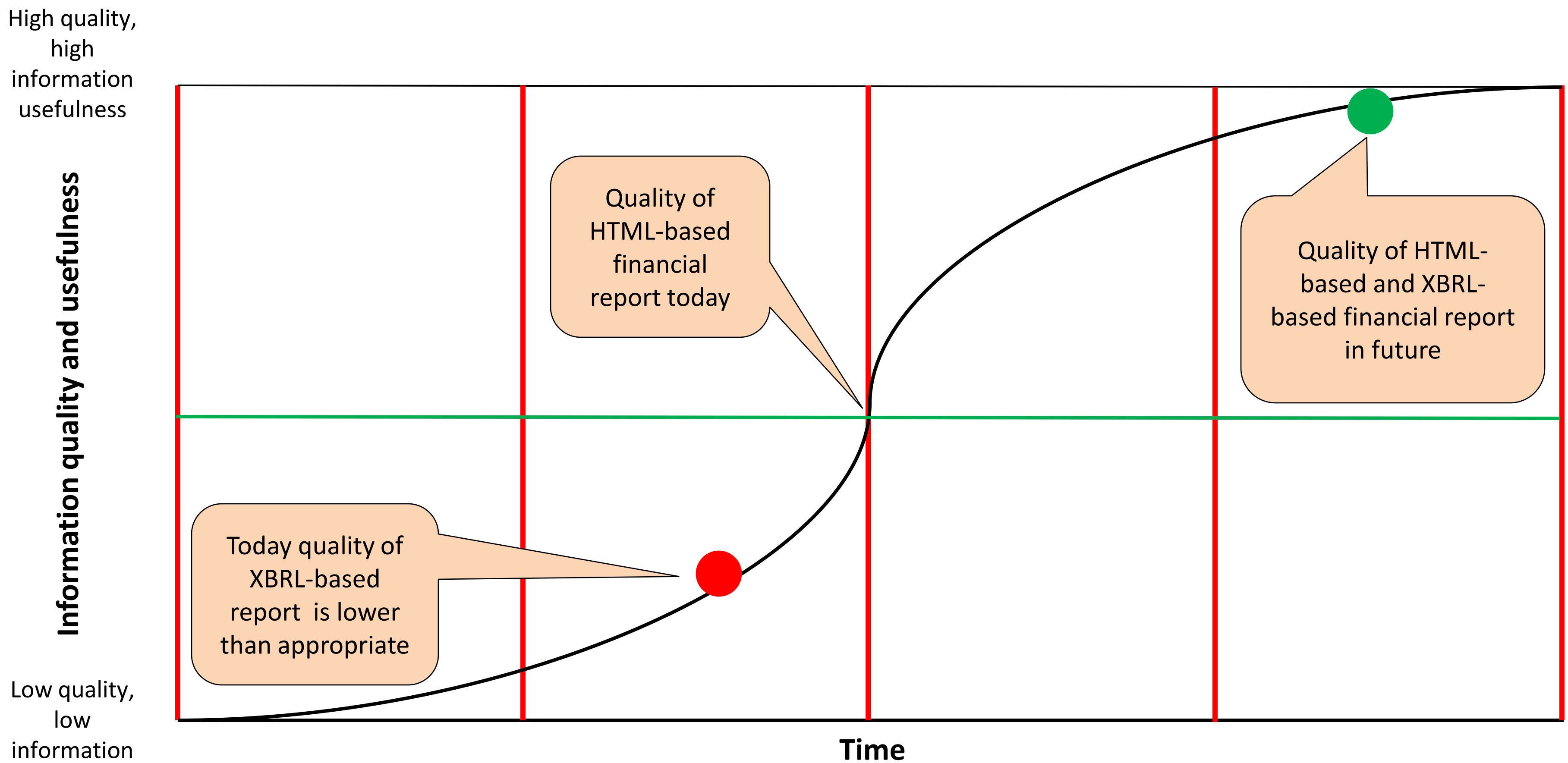
Reimagining the Spreadsheet

- **Readable by both humans and machines**: A spreadsheet should be readable by both humans and machines. Information provided within a spreadsheet should be more a representation of information than presentation oriented. The representation can be presented in sheets, rows, columns, and cells but this is done leveraging information metadata and commonly understood patterns. 100% pixel perfect renderings are specifically not a requirement.
- **Global standard format**: The format of the spreadsheet should be a global standard, not controlled by one software vendor.
- **Agreed upon level of semantics**: The creators and consumers can agree on the level of semantic clarity they will make use of for a spreadsheet. The spectrum can range from no semantics at all (which is similar to today's spreadsheet) or a high level of semantics expressed by a highly controlled representation model.
- **Separation of representation and presentation**: The "representation" and the "presentation of the representation" should not be intermingled.
- **Business rules separable from spreadsheet**: Business rules should be separated from the information when desired, integrated with the spreadsheet when necessary. Business rules which are external to the spreadsheet can be used to "watch over" the things and relations within the spreadsheet. The business rules can be made available publicly via a URL, privately via a private URL, etc.
- **Managed global standard**: The better spreadsheet should be a global standard under the control of someone like OMG, XBRL International, ISO, Apache OpenOffice, or some other such organization.
- **Provide a formal shape but be domain neutral filler**: One formal shape should be agreed to, for example the multidimensional model, but the pieces which fit into that shape or "fill" the shape are domain neutral, controlled by the business domain.
- **Format should allow for versioning, collaboration, etc**: The syntax format should allow for ease of versioning, constructing systems which are collaborative in nature (multi-user).
- **Straightforwardly usable over the Internet**: The format should be compliant with internet standards.
- **Support a wide variety of common business use cases**: A wide variety of common business use cases would be served, but it is not a goal to solve every business problem which exists.
- **Highly limited options**: The number of optional features is to be kept to the absolute minimum, ideally zero. Multiple approaches to solving a problem are not necessary when one will do.
- **Formal and concise design**: The design must be formal, concise, well designed and well engineered.

What is Changed?

- **Document format changed**: *(enabled by XBRL)*
  - **OLD**: Unstructured information; **not measurable**
  - **NEW**: Structured information; **measurable**
- **Review process can change**: *(enabled by structured information)*
  - **OLD**: Almost 100% manual review; because not measurable therefore not automatable
  - **NEW**: More automated review, less manual review because; (a) machine-readable rules and (b) structured information is measurable

Intelligent XBRL-based Financial Reporting Maturity Levels



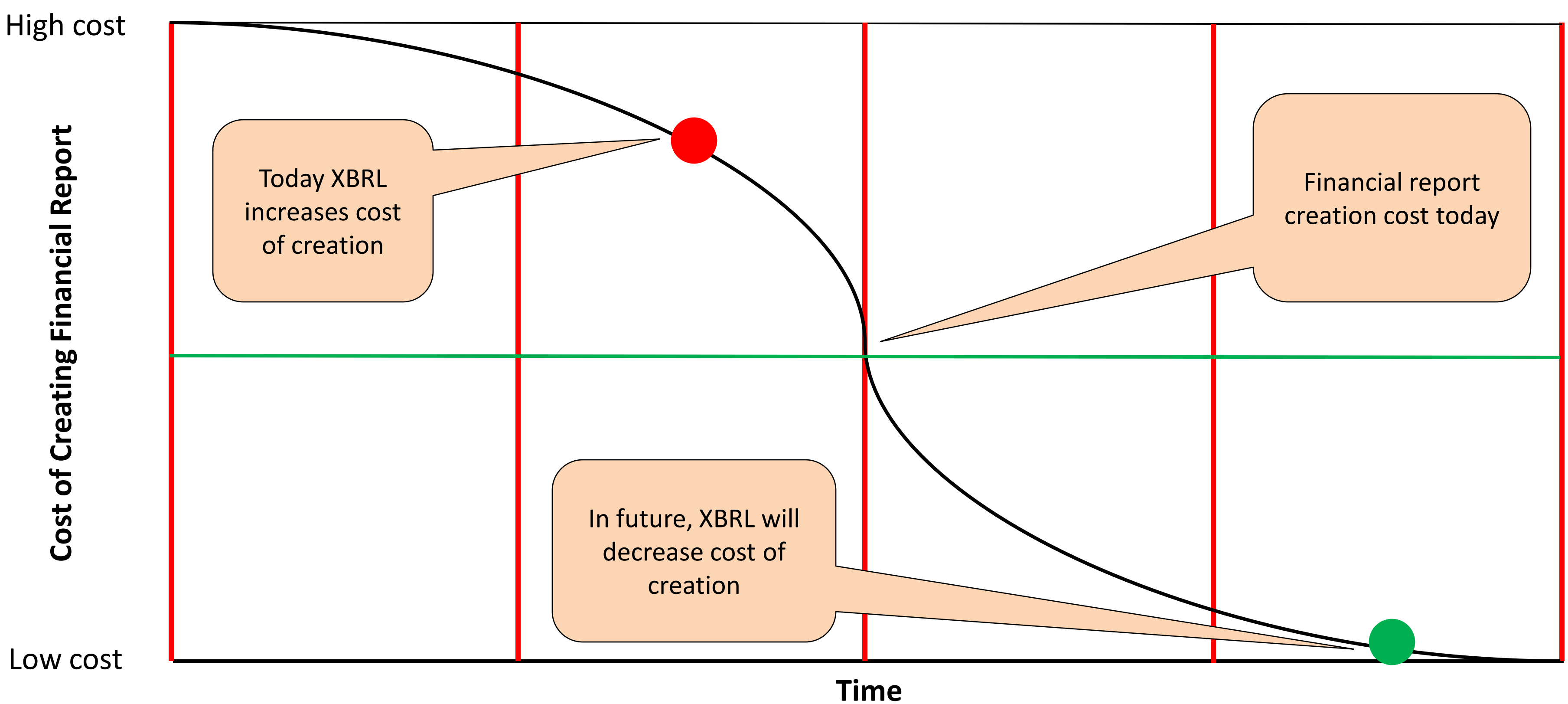
LEVEL 0 – Creation of Global Standard XBRL, first creation and use of US GAAP XBRL Taxonomy

LEVEL 1 – Creation of first generation of software tools, refinement of US GAAP XBRL Taxonomy

LEVEL 2 – Creation of next generation of software tools, further refinements of US GAAP XBRL Taxonomy, adding additional financial reporting rules and other metadata

LEVEL 3 – Further refinement of software, further refinement of US GAAP XBRL Taxonomy, refinement of additional financial reporting metadata, increased experience

Intelligent XBRL-based Financial Reporting Maturity Levels



LEVEL 0 – Creation of Global Standard XBRL, first creation and use of US GAAP XBRL Taxonomy

LEVEL 1 – Creation of first generation of software tools, refinement of US GAAP XBRL Taxonomy

LEVEL 2 – Creation of next generation of software tools, further refinements of US GAAP XBRL Taxonomy, adding additional financial reporting rules and other metadata

LEVEL 3 – Further refinement of software, further refinement of US GAAP XBRL Taxonomy, refinement of additional financial reporting metadata, increased experience

Contrasting Data Warehouse and Information Warehouse

- **Data warehouse**
  - Proprietary
  - Internally facing
  - Numbers only
  - OLAP required
  - Information technology professionals maintain
  - Lean information
  - Read only
  - More errors
- **Information warehouse**
  - Standard
  - Internal or external
  - Numbers and text
  - OLAP optional
  - Business professionals maintain
  - Rich information
  - Read/write
  - Less errors