

# Influence Factor: Extending the PROV Model with a Quantitative Measure of Influence

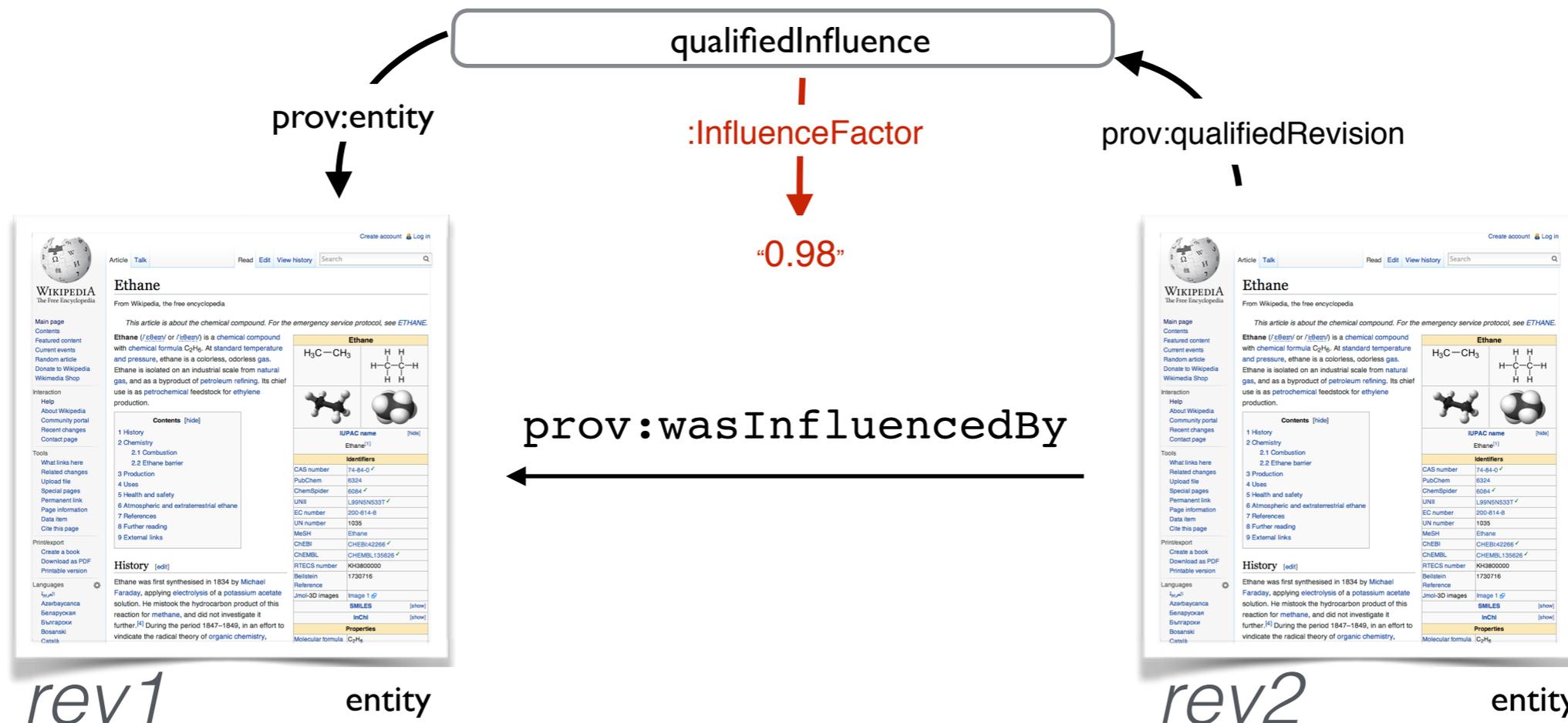
Matthew Gamble  
Prof. Carole Goble

University of Manchester

# In Summary

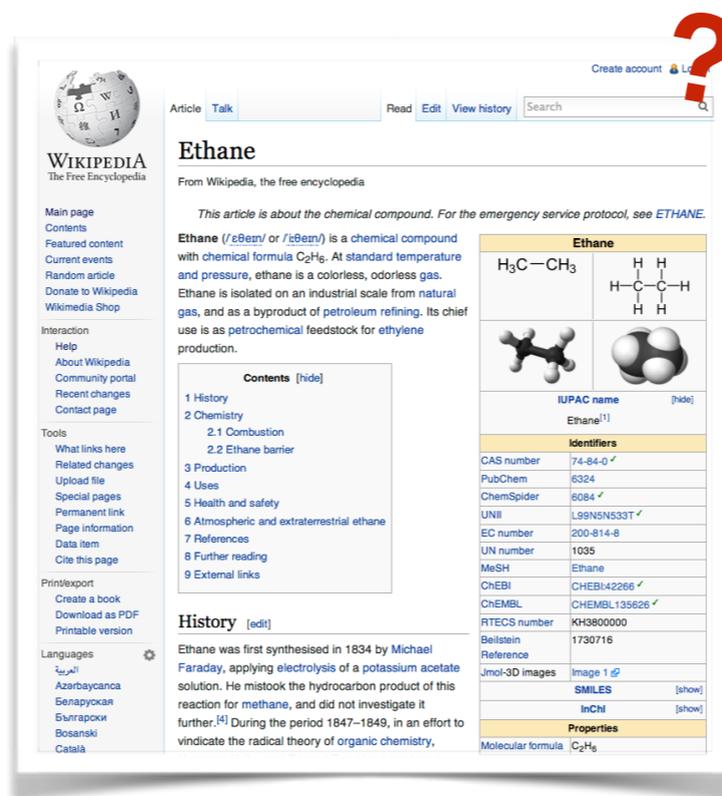
A *quantitative* measure of the influence that one PROV element has had over another.

i.e How Much



# Context

Thesis: Modelling and Computing the Quality of Scientific Information on the Web of Data



The image shows a screenshot of the Wikipedia article for Ethane. The article includes the following content:

**Ethane**  
From Wikipedia, the free encyclopedia

*This article is about the chemical compound. For the emergency service protocol, see ETHANE.*

**Ethane** (/ˈɛbeɪn/ or /ˈiːbeɪn/) is a chemical compound with chemical formula C<sub>2</sub>H<sub>6</sub>. At standard temperature and pressure, ethane is a colorless, odorless gas. Ethane is isolated on an industrial scale from natural gas, and as a byproduct of petroleum refining. Its chief use is as petrochemical feedstock for ethylene production.

**Contents** [hide]

- History
- Chemistry
  - Combustion
  - Ethane barrier
- Production
- Uses
- Health and safety
- Atmospheric and extraterrestrial ethane
- References
- Further reading
- External links

**History** [edit]

Ethane was first synthesised in 1834 by Michael Faraday, applying electrolysis of a potassium acetate solution. He mistook the hydrocarbon product of this reaction for methane, and did not investigate it further.<sup>[4]</sup> During the period 1847–1849, in an effort to vindicate the radical theory of organic chemistry,

**Chemical structures:**  
H<sub>3</sub>C—CH<sub>3</sub>  
  


**IUPAC name:** Ethane<sup>[1]</sup>

**Identifiers**

CAS number	74-84-0 ✓
PubChem	6324
ChemSpider	6084 ✓
UNII	L99N5N533T ✓
EC number	200-814-8
UN number	1035
MeSH	Ethane
ChEBI	CHEBI:42266 ✓
ChEMBL	CHEMBL135626 ✓
RTECS number	KH3800000
Bellstein Reference	1730716
Jmol-3D images	Image 1 
SMILES	<span>[show]</span>
InChI	<span>[show]</span>

**Properties**

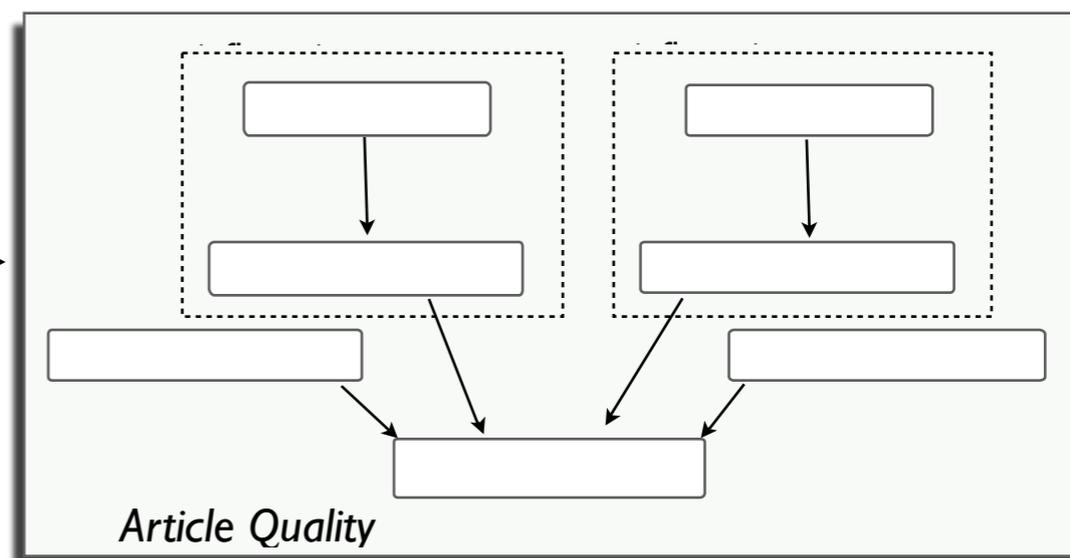
Molecular formula C<sub>2</sub>H<sub>6</sub>

What is the quality of the Wikipedia article for the chemical Ethane?

# Context

Thesis: Modelling and Computing the Quality of Scientific Information on the Web of Data

The screenshot shows the Wikipedia article for Ethane. The article text states: "Ethane (/ˈeɪθən/ or /ˈiːθən/) is a chemical compound with chemical formula C<sub>2</sub>H<sub>6</sub>. At standard temperature and pressure, ethane is a colorless, odorless gas. Ethane is isolated on an industrial scale from natural gas, and as a byproduct of petroleum refining. Its chief use is as petrochemical feedstock for ethylene production." The article includes chemical structures (H<sub>3</sub>C-CH<sub>3</sub> and ball-and-stick models), IUPAC name (Ethane), and a list of identifiers (CAS number, PubChem, ChemSpider, UNII, EC number, UN number, MeSH, ChEBI, ChEMBL, RTECS number, Beilstein Reference, Jmol-3D images). A red question mark is in the top right, and a red 'Metadata' watermark is overlaid on the bottom right.



Probabilistic Graphical Models

What is the quality of the Wikipedia article for the chemical Ethane?

# Provenance

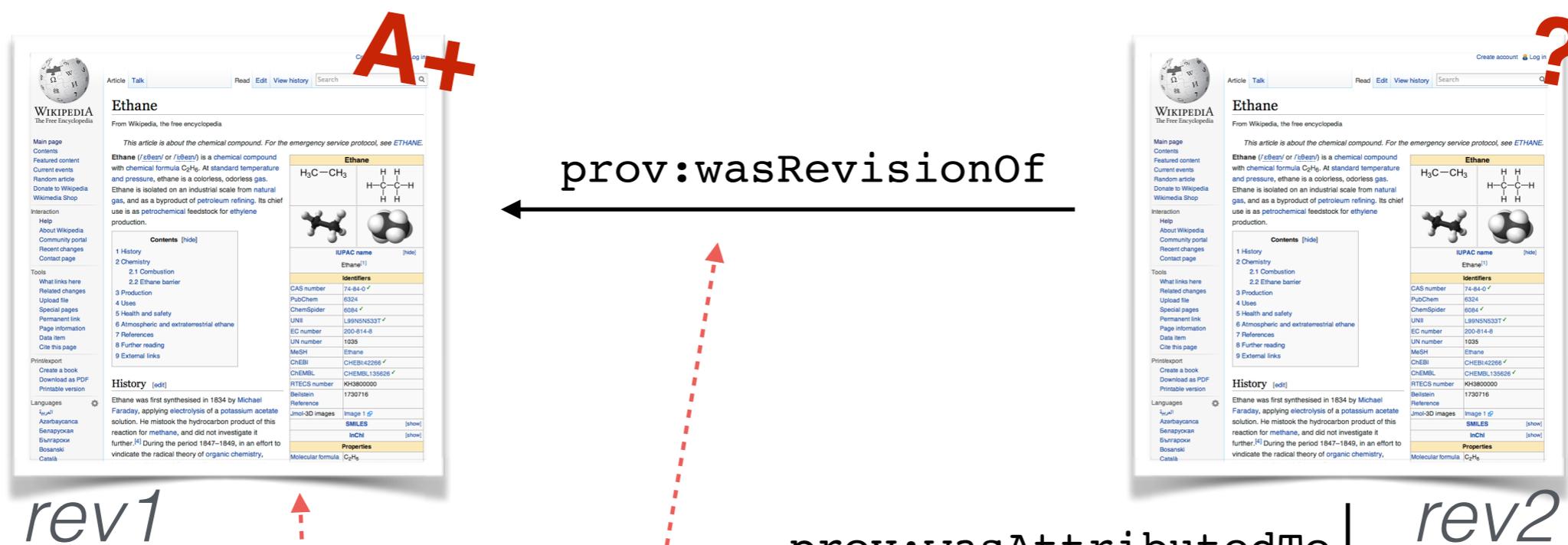
“**Provenance** is information about entities, activities and people involved in producing a piece of data which can be used to form assessments about its **quality, believability, or trustworthiness**”

- *PROV-Overview:  
An Overview of the PROV Family of Documents*



# Exploiting a Common Intuition

Anything that has influenced the production of an entity may have affected its likely quality



Relies on *lineage* and *how* provenance



Administrator

chemical\_lover\_84

# Questions

*Is the how provenance that can expressed in PROV sufficient for our quality assessment?*

*Are publishers of provenance motivated to publish detailed how provenance?*

*Are publishers of provenance motivated to publish detailed how provenance?*

Influencee	Property	Influencer	Datasets								
			Taverna	Wings	Wiki	SRI	OBAMIO	CSIRO	Vis	Chiron	Swift
*	wasInfluencedBy	*	✓	✓	✓	✓	✓	✓	✓	✓	✓
Entity	wasGenerateBy	Activity	✓	✓	✓	✓	✓	✓	✓	✓	✓
Entity	wasDerivedFrom	Entity	✓	×	×	✓	✓	✓	×	✓	✓
Entity	wasAttributedTo	Agent	×	✓	×	×	✓	✓	×	✓	×
Entity	hadPrimarySource	Entity	×	✓	×	×	×	×	×	×	×
Entity	wasQuotedFrom	Entity	×	×	×	×	×	×	×	×	×
Entity	wasRevisionOf	Entity	×	×	✓	×	×	×	×	×	×
Entity	wasInvalidatedBy	Activity	×	×	×	×	×	×	×	×	×
Activity	wasInformedBy	Activity	✓	×	×	✓	✓	✓	×	✓	×
Activity	used	Entity	✓	✓	✓	✓	✓	✓	✓	✓	✓
Activity	wasAssociatedWith	Agent	✓	✓	✓	✓	✓	✓	✓	✓	×
Activity	wasStartedBy	Entity	×	×	×	×	×	×	×	×	×
Activity	wasEndedBy	Entity	×	×	×	×	×	×	×	×	×
Agent	actedOnBehalfOf	Agent	×	×	×	×	×	✓	×	×	✓

All datasets from PROV Bench make use of **between 3 and 7** of the 13 sub-properties of `wasInfluenceBy`

Is the how provenance that can expressed in PROV sufficient for our quality assessment?

**A+**

rev1

← prov:wasInfluencedBy

**?**

rev2

*Is the how provenance that can expressed in PROV sufficient for our quality assessment?*

**A+**

WIKIPEDIA  
The Free Encyclopedia

Article Talk Read Edit View history Search

**Ethane**  
From Wikipedia, the free encyclopedia

This article is about the chemical compound. For the emergency service protocol, see ETHANE.

**Ethane** (/ˈeθeɪn/ or /ɪˈein/) is a chemical compound with chemical formula C<sub>2</sub>H<sub>6</sub>. At standard temperature and pressure, ethane is a colorless, odorless gas. Ethane is isolated on an industrial scale from natural gas, and as a byproduct of petroleum refining. Its chief use is as petrochemical feedstock for ethylene production.

**Contents** [hide]

- History
- Chemistry
  - Combustion
  - Ethane barrier
- Production
- Uses
- Health and safety
- Atmospheric and extraterrestrial ethane
- References
- Further reading
- External links

**History** [edit]

Ethane was first synthesised in 1834 by Michael Faraday, applying electrolysis of a potassium acetate solution. He mistook the hydrocarbon product of this reaction for methane, and did not investigate it further.<sup>[1]</sup> During the period 1847–1849, in an effort to vindicate the radical theory of organic chemistry,

**Identifiers**

CAS number	74-84-0 ✓
PubChem	6324
ChemSpider	6084 ✓
UNII	L39NNS33T ✓
EC number	200-814-6
UN number	1035
MeSH	Ethane
CHEBI	CHEBI:42266 ✓
CHEMBL	CHEMBL135626 ✓
RTECS number	KH380000
Beilstein Reference	1730716
Jmol-3D images	Image 1 <span>g</span>
	<b>SMILES</b> <span>[show]</span>
	<b>INChI</b> <span>[show]</span>
	<b>Properties</b>
Molecular formula	C <sub>2</sub> H <sub>6</sub>

*rev1*

`prov:wasRevisionOf`

**?**

WIKIPEDIA  
The Free Encyclopedia

Article Talk Read Edit View history Search

**Ethane**  
From Wikipedia, the free encyclopedia

This article is about the chemical compound. For the emergency service protocol, see ETHANE.

**Ethane** (/ˈeθeɪn/ or /ɪˈein/) is a chemical compound with chemical formula C<sub>2</sub>H<sub>6</sub>. At standard temperature and pressure, ethane is a colorless, odorless gas. Ethane is isolated on an industrial scale from natural gas, and as a byproduct of petroleum refining. Its chief use is as petrochemical feedstock for ethylene production.

**Contents** [hide]

- History
- Chemistry
  - Combustion
  - Ethane barrier
- Production
- Uses
- Health and safety
- Atmospheric and extraterrestrial ethane
- References
- Further reading
- External links

**History** [edit]

Ethane was first synthesised in 1834 by Michael Faraday, applying electrolysis of a potassium acetate solution. He mistook the hydrocarbon product of this reaction for methane, and did not investigate it further.<sup>[1]</sup> During the period 1847–1849, in an effort to vindicate the radical theory of organic chemistry,

**Identifiers**

CAS number	74-84-0 ✓
PubChem	6324
ChemSpider	6084 ✓
UNII	L39NNS33T ✓
EC number	200-814-6
UN number	1035
MeSH	Ethane
CHEBI	CHEBI:42266 ✓
CHEMBL	CHEMBL135626 ✓
RTECS number	KH380000
Beilstein Reference	1730716
Jmol-3D images	Image 1 <span>g</span>
	<b>SMILES</b> <span>[show]</span>
	<b>INChI</b> <span>[show]</span>
	<b>Properties</b>
Molecular formula	C <sub>2</sub> H <sub>6</sub>

*rev2*

*Is the how provenance that can expressed in PROV sufficient for our quality assessment?*

**rev1**

`prov:wasRevisionOf`

**rev2**

To what extent is rev2 influenced by rev1?

# How Much?

This screenshot shows the Wikipedia article for Ethane. The article text is in blue, and the navigation sidebar on the left is also in blue. A large red 'A+' grade stamp is overlaid in the top right corner of the article area.

rev1

Small change.

This screenshot shows the Wikipedia article for Ethane. The article text is in blue, but the navigation sidebar on the left is in red. A large red question mark stamp is overlaid in the top right corner of the article area.

rev2

This screenshot shows the Wikipedia article for Ethane. The article text is in blue, and the navigation sidebar on the left is also in blue. A large red 'A+' grade stamp is overlaid in the top right corner of the article area.

rev1

Significant Change

This screenshot shows the Wikipedia article for Ethane. The article text is in red, and the navigation sidebar on the left is also in red. A large red question mark stamp is overlaid in the top right corner of the article area.

rev2



# How Much?

This screenshot shows the Wikipedia article for Ethane at revision 1. The page includes a title bar, a search bar, and a main content area with text, chemical structures, and a table of identifiers. A large red 'A+' is overlaid on the top right of the page.

*rev1*

Small change.

This screenshot shows the Wikipedia article for Ethane at revision 2. The page is nearly identical to revision 1, but with a large red '?' overlaid on the top right, indicating a question about the change.

*rev2*

**influence of previous revision:** (words common / total words in previous)

**influence of the author :** 1 - (words common/total words in previous).

## How Much?

WIKIPEDIA  
The Free Encyclopedia

Article Talk Read Edit View history Search

## Ethane

From Wikipedia, the free encyclopedia

This article is about the chemical compound. For the emergency service protocol, see ETHANE.

Ethane (/ˈeɪθən/ or /ˈiðɪn/) is a chemical compound with chemical formula C<sub>2</sub>H<sub>6</sub>. At standard temperature and pressure, ethane is a colorless, odorless gas. Ethane is isolated on an industrial scale from natural gas, and as a byproduct of petroleum refining. Its chief use is as petrochemical feedstock for ethylene production.

**Contents** [hide]

- History
- Chemistry
  - Combustion
  - Ethane barrier
  - Production
  - Uses
  - Health and safety
  - Atmospheric and extraterrestrial ethane
  - References
  - Further reading
  - External links

**History** [edit]

Ethane was first synthesised in 1834 by Michael Faraday, applying electrolysis of a potassium acetate solution. He mistook the hydrocarbon product of this reaction for methane, and did not investigate it further.<sup>[H]</sup> During the period 1847–1849, in an effort to vindicate the radical theory of organic chemistry,

**Identifiers**

CAS number	<span>74-84-0</span> <span> </span> <span><span>✓</span></span>
PubChem	6324
ChemSpider	6084 <span> </span> <span><span>✓</span></span>
UNII	L89NS5337 <span> </span> <span><span>✓</span></span>
EC number	200-614-6
UN number	1035
MDL	Ethane
CHEBI	CHEBI:42266 <span> </span> <span><span>✓</span></span>
CHEMBL	CHEMBL135628 <span> </span> <span><span>✓</span></span>
RTECS number	KH880000
Beilstein	1730716
References	
Jmol-3D images	Image 1 of 1 <span> </span> <span><span>✓</span></span>
SMILES	<span><span></span></span> <span> </span> <span><span>✓</span></span>
InChI	<span><span></span></span> <span> </span> <span><span>✓</span></span>

**Properties**

Molecular formula	C <sub>2</sub> H <sub>6</sub>
-------------------	-------------------------------

rev1

Small change.

WIKIPEDIA  
The Free Encyclopedia

Article Talk Read Edit View history Search

## Ethane

From Wikipedia, the free encyclopedia

This article is about the chemical compound. For the emergency service protocol, see ETHANE.

Ethane (/ˈeɪθən/ or /ˈiðɪn/) is a chemical compound with chemical formula C<sub>2</sub>H<sub>6</sub>. At standard temperature and pressure, ethane is a colorless, odorless gas. Ethane is isolated on an industrial scale from natural gas, and as a byproduct of petroleum refining. Its chief use is as petrochemical feedstock for ethylene production.

**Contents** [hide]

- History
- Chemistry
  - Combustion
  - Ethane barrier
  - Production
  - Uses
  - Health and safety
  - Atmospheric and extraterrestrial ethane
  - References
  - Further reading
  - External links

**History** [edit]

Ethane was first synthesised in 1834 by Michael Faraday, applying electrolysis of a potassium acetate solution. He mistook the hydrocarbon product of this reaction for methane, and did not investigate it further.<sup>[H]</sup> During the period 1847–1849, in an effort to vindicate the radical theory of organic chemistry,

**Identifiers**

CAS number	<span>74-84-0</span> <span> </span> <span><span>✓</span></span>
PubChem	6324
ChemSpider	6084 <span> </span> <span><span>✓</span></span>
UNII	L89NS5337 <span> </span> <span><span>✓</span></span>
EC number	200-614-6
UN number	1035
MDL	Ethane
CHEBI	CHEBI:42266 <span> </span> <span><span>✓</span></span>
CHEMBL	CHEMBL135628 <span> </span> <span><span>✓</span></span>
RTECS number	KH880000
Beilstein	1730716
References	
Jmol-3D images	Image 1 of 1 <span> </span> <span><span>✓</span></span>
SMILES	<span><span></span></span> <span> </span> <span><span>✓</span></span>
InChI	<span><span></span></span> <span> </span> <span><span>✓</span></span>

**Properties**

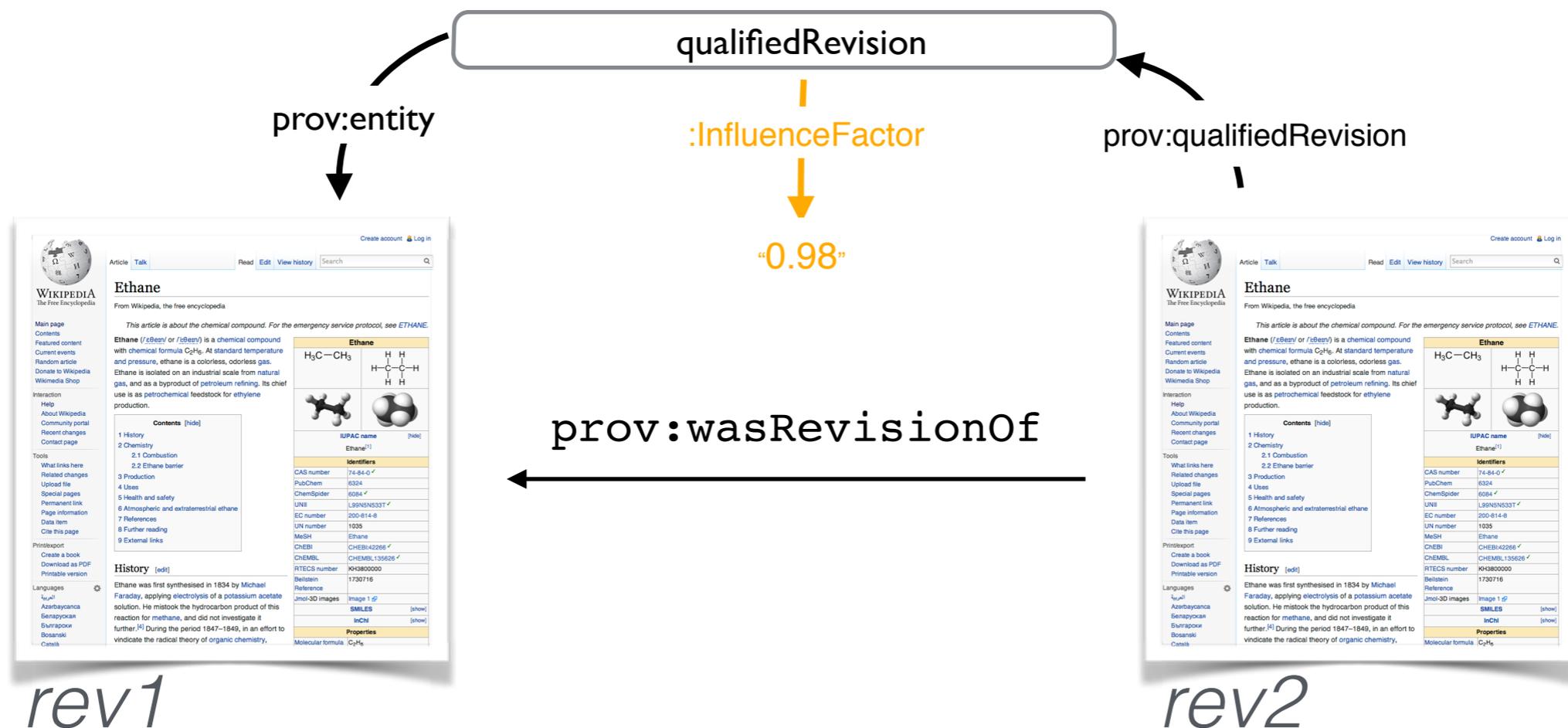
Molecular formula	C <sub>2</sub> H <sub>6</sub>
-------------------	-------------------------------

rev2

Other possibilities?

# Influence Factor

A *quantitative* measure of the influence that one *PROV* entity has had over another.



# Influence Factor

A *quantitative* measure of the influence that one PROV entity has had over another.

`evident:influenceFactor`



`evident:discreteInfluenceFactor`



`evident:continuousInfluenceFactor`



`evident:normalInfluenceFactor`



`rdfs:subProperty`

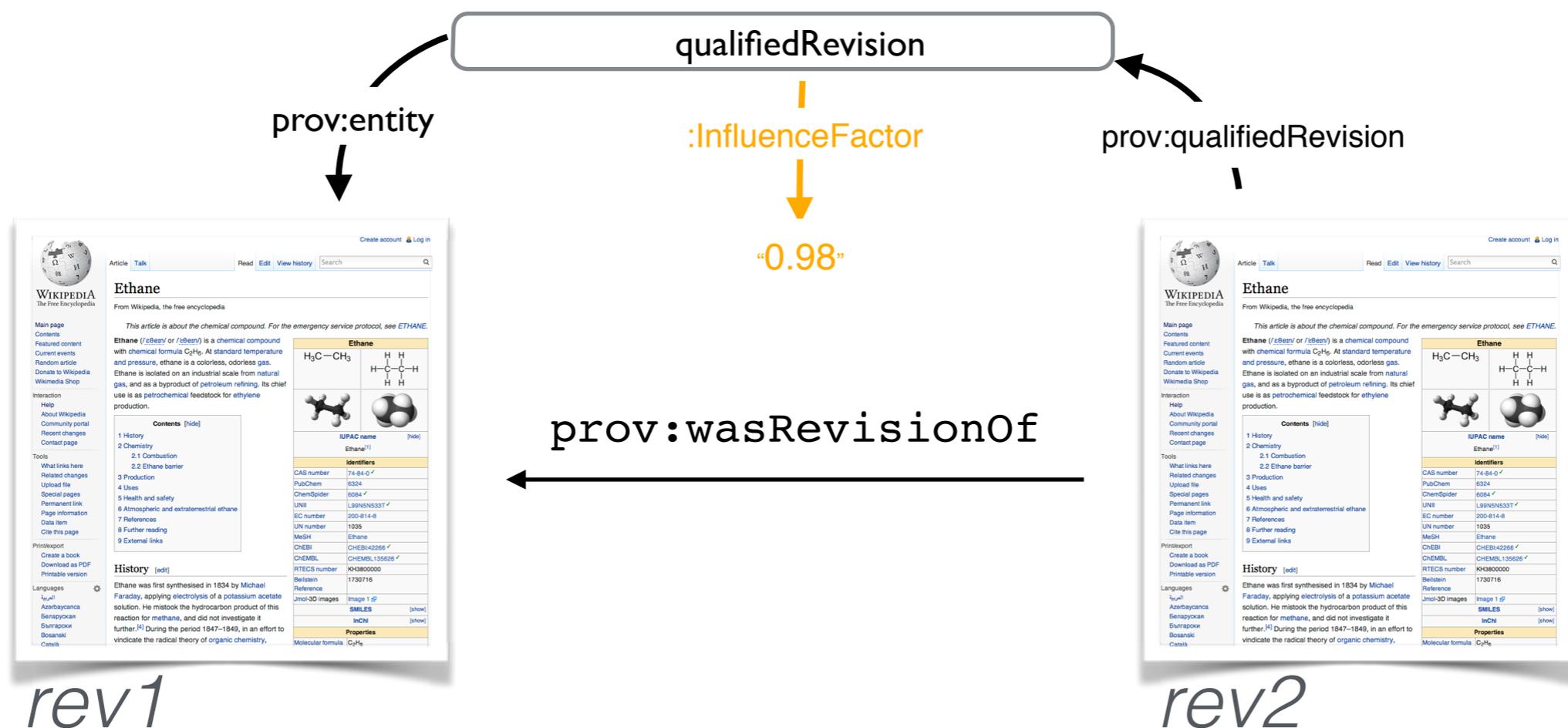
<http://purl.org/net/evident>

# The New Data

## ProvBench - wikipedia-provenance

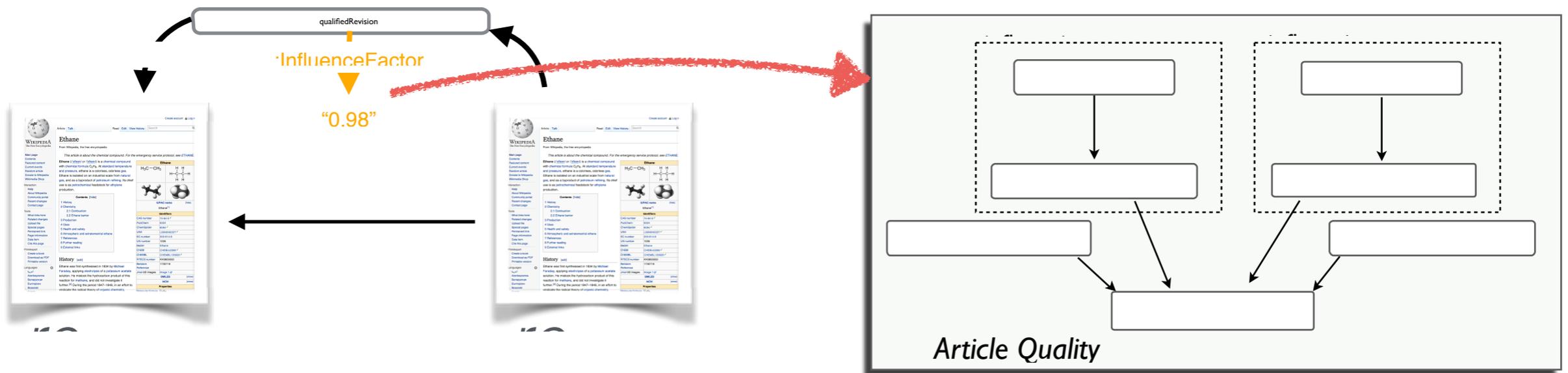
<https://github.com/matthewgamble/wikipedia-provenance>

*Now with added RDF and influenceFactor!*



# Influence Factor

Used the enriched data to support quality assessment with probabilistic models.



Influence factor informs priors in our probabilistic model  
- like a weighting/dampening factor.

# Summary & Questions

Provenance publishers appear to be motivated to provide detailed how provenance beyond **:wasInfluencedBy**

A quantitative measure of influence supported our ability to automatically evaluate quality.

Is the information generic enough to be at the interchange level?

What is the scope of influence factor & where does it break?